



Chemistry
Education Research
and Practice

**Why do we assess students? Investigating General
Chemistry Instructors' Conceptions of Assessment Purposes
and Their Relationships to Assessment Practices**

Journal:	<i>Chemistry Education Research and Practice</i>
Manuscript ID	RP-ART-05-2024-000147.R1
Article Type:	Paper
Date Submitted by the Author:	19-Jun-2024
Complete List of Authors:	Shi, Lu; University of Virginia Wang, Ying; University of Virginia, Chemistry Mitchell-Jones, Jherian; University of Virginia Stains, Marilyne; University of Virginia, Department of Chemistry

SCHOLARONE™
Manuscripts

Why do we assess students? Investigating General Chemistry Instructors' Conceptions of Assessment Purposes and Their Relationships to Assessment Practices

Lu Shi‡, Ying Wang‡, Jherian K. Mitchell-Jones, Marilyne Stains*

Department of Chemistry, University of Virginia, Charlottesville, VA 22904-4319

*Corresponding author's email: mstains@virginia.edu

‡Authors contributed equally to this work and the names are listed in alphabetical order.

ABSTRACT

Assessment plays a critical role in instruction and curriculum. Existing literature on instructors' assessment practices and related factors has been intensively focused on primary and secondary education. This study extended the contexts of previous literature to post-secondary chemistry education by exploring general chemistry instructors' conceptions of assessment purposes and their assessment practices. Semi-structured interviews were conducted with 19 general chemistry instructors from 14 institutions across the East Coast region of the United States of America. The results demonstrate that instructors predominately perceive the purpose of Assessment of Learning (i.e., evaluation of student performance) with only few of them mentioning purposes of Assessment for Learning (i.e., assessment provides actionable feedback for both the instructors and the students) and Assessment as Learning (i.e., assessment promotes self-regulation). The use of various assessment practices is related to the number of assessment purposes instructors recognize. In addition, the study demonstrates that instructors perceive their assessment practices to be influenced by academic culture and departmental norms. This nuanced understanding can guide practical and research efforts to improve chemistry instructors' engagement in assessment reforms.

INTRODUCTION

Assessment is traditionally viewed as a means to track student progression toward qualifications and assess students' achievement in reaching educational goals (Brown *et al.*, 2013). With evolving perspectives on learning, many education researchers have advocated for changes in this perception and use of assessment (Stiggins, 1994; Black and Wiliam, 1998; Shepard, 2000). For instance, Black and Wiliam (1998) reviewed 250 studies and suggested that effective classroom assessment can enhance student learning, particularly when paired with quality feedback. Shepard (2000) proposed a conceptual framework aligned with social-constructivist perspectives, emphasizing that classroom assessment should be an ongoing process integrated with instruction, addressing both learning processes and outcomes with

clear expectations, challenging tasks, and student involvement. This conceptual framework on assessment supports instructors in adapting instruction and students in improving their learning.

The idea of using assessment to facilitate learning has garnered significant attention in science education, as evidenced by the National Academy report "Knowing What Students Know: The Science and Design of Educational Assessment," which defines assessment as a process of reasoning from evidence collected on various aspects of student learning (Glaser *et al.*, 2001). This perspective underscores the need for assessment tools to target diverse learning aspects and emphasizes instructors' roles in interpreting and acting on student data. However, shifting the focus of assessment from evaluating students to informing teaching and learning requires more than just adopting new strategies (Darling-Hammond and Snyder, 2000). Shepard (2000) argues that instructors must also change their views on purposes of assessment (i.e., why they assess students). Indeed, empirical evidence in education literature demonstrates a link between instructors' conceptions of assessment purposes and their assessment practices (Sadler and Reimann, 2017). Moreover, instructors may encounter psychological and social barriers when trying to change assessment practices to support learning (Dwyer, 2006), with many contextual factors influencing these practices at varying levels (e.g., student responses and social pressures)(Fulmer *et al.*, 2015; Yan *et al.*, 2021).

Calls to reform assessment practices in postsecondary chemistry courses have been ongoing for over a decade (Holme *et al.*, 2010; Stowe and Cooper, 2017; Stowe *et al.*, 2021) and have led to the development of new assessment tools (Lavery *et al.*, 2016; Schultz *et al.*, 2017; Barbera *et al.*, 2023). However, there has been limited research on chemistry instructors' knowledge and thinking about assessment (Emenike *et al.*, 2011; Emenike *et al.*, 2013; Raker *et al.*, 2013; Raker and Holme, 2014; Feola *et al.*, 2023) and their assessment practices (Gibbons *et al.*, 2018; Gibbons *et al.*, 2022), leading to challenges in the adoption of these new tools (Gibbons *et al.*, 2022). Without these knowledge bases, it is difficult to develop assessment-focused reform efforts that effectively address and build on the needs and ways of thinking of chemistry instructors. The goal of this study is to inform future assessment reform efforts by providing insights into chemistry instructors' conceptions of assessment purposes, their assessment practices, and factors influencing these practices. In particular, this study seeks to address the following research questions (RQ):

- 1) What are the assessment practices that general chemistry instructors employ in their classrooms?

- 2) What are the external factors and pressures that general chemistry instructors perceive to influence their assessment practices?
- 3) What are general chemistry instructors’ conceptions of assessment purposes?
- 4) To what extent do general chemistry instructors’ conceptions of assessment purposes and perceived external influences relate to their assessment practices?

CONCEPTUAL FRAMEWORK

The current study is situated within the Teacher-Centered Systemic Reform (TCSR) model (Figure 1; Woodbury and Gess-Newsome, 2002; Gess-Newsome *et al.*, 2003). The model suggests that in order to achieve desirable outcomes on instructional practices, including assessment practices, reform efforts need to consider one or more of the following domains: 1) teacher thinking factors (e.g., knowledge and beliefs about teaching and learning); 2) contextual factors (e.g., cultural, institutional and classroom contexts); and 3) teacher personal factors (e.g., demographic profiles and pedagogical training). The assessment literature related to each aspect of the framework is presented in the next sections.

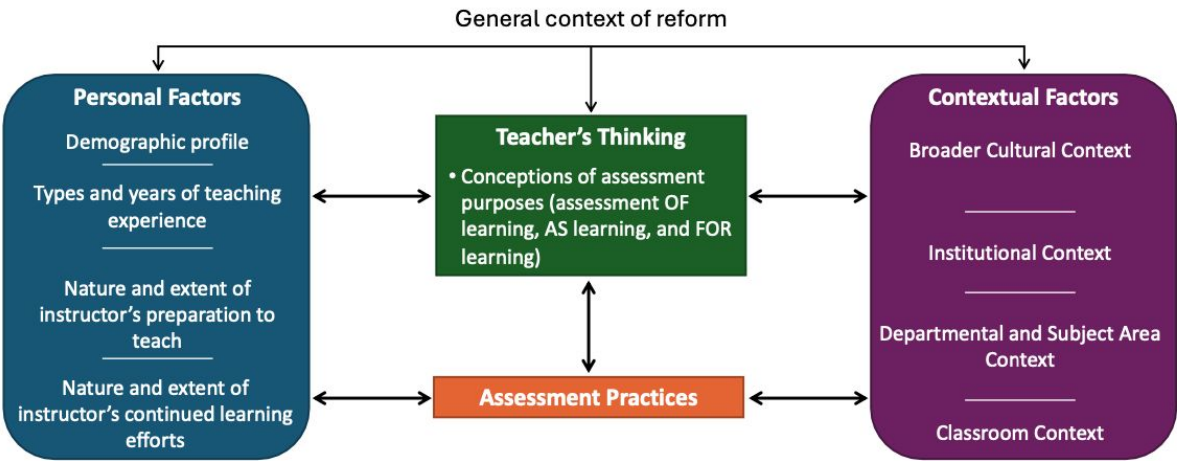


Fig. 1 Teacher-Centered Systemic Reform (TCSR) model applied to the goal of this study

Assessment practices

Assessment practices in the current study refer to the assessment strategies used by chemistry instructors in their courses (Fulmer *et al.*, 2015). A couple of studies have explored the assessment practices of chemistry instructors through the collection of surveys distributed nationally. Wang *et al.* (2024) found

that general chemistry instructors have a lot of autonomy in developing their examinations. Gibbons and colleagues (2022) characterized postsecondary chemistry instructors' enacted assessment practices and found that more than 90 % of these instructors use exams, highlighting varying assessment practices were used across different levels of chemistry courses and courses with different class sizes. While this latter work provides valuable insights regarding assessment practices used by chemistry instructors across different contexts, it primarily focuses on assessment tools generated based on the authors' experiences and those that have extensive research bases in the literature. As indicated by the authors, there is a more nuanced picture of assessment tools beyond the list in their survey. Thus, it is important to have this nuanced understanding of postsecondary chemistry instructors' assessment practices.

Teacher's thinking about assessment

Several aspects of teacher thinking about assessment have been explored in the STEM literature at both the secondary and postsecondary levels, including instructors' beliefs about assessment (Park *et al.*, 2024), conceptions of assessment (Fletcher *et al.*, 2012), and knowledge of assessment (Mertler and Campbell, 2005; Raker *et al.*, 2013; Raker and Holme, 2014). Few chemistry-specific studies in this domain of the TCSR framework have been conducted. These studies have focused on chemistry instructors' familiarity with assessment terminology such as summative and formative assessment (Emenike *et al.*, 2011; Raker *et al.*, 2013; Raker and Holme, 2014). Confusion surrounding the terms utilized in the literature to explore aspects of teacher thinking (e.g., beliefs versus conceptions versus knowledge) has been noted by Fulmer and colleagues (2015) who emphasized the need to clarify definitions and establish common terminology. In this paper, we adopt the term "instructors' conceptions of assessment purposes" to refer to instructors' "general views and beliefs about what assessment is, and its purposes in school and society", aligning with the definition in Fulmer and colleagues' work (2015).

Earl and Katz's (2006) proposed a conceptual framework that provides insights into understanding the purposes of classroom-based assessment. The framework identifies three main purposes of assessment: Assessment of Learning (AoL), Assessment for Learning (AfL), and Assessment as Learning (AaL).

Assessment of Learning (AoL). AoL is summative in nature and typically occurs after learning has taken place to provide evidence of students' performance (Zeng *et al.*, 2018). The core of AoL is to evaluate whether students have achieved the goals of the curriculum. Additionally, AoL may be used to compare students' performance relative to each other, often through ranking or grading. AoL has traditionally been

associated with summative assessment tools such as mid-terms and final exams. It has historically been a core purpose of educational evaluation (Brown *et al.*, 2013). However, researchers argue that the emphasis on AoL alone is not effective in promoting students' learning because it is enacted after the learning is completed (Black *et al.*, 2004). This means that while AoL provides valuable information about student achievement at a specific point in time, it does not inherently contribute to ongoing learning and improvement during the learning process. To address this limitation, the concept of AfL has been proposed.

Assessment for Learning (AfL). AfL refers to any assessment for which the priority in its design and practice is to promote student learning (Black *et al.*, 2003; Black *et al.*, 2004; Stiggins, 2005; Earl and Katz, 2006). It builds upon the traditional idea of formative assessment (e.g., minute papers and homework), which is focused on providing feedback to instructors. However, AfL assessment tools are designed to provide valuable information for both instructors AND students. AfL aims to empower instructors and students to adapt teaching and learning activities based on assessment data. AfL assessment tools provide insights to instructors about students' progress on their learning journey so that they can adjust their instructional strategies, target specific areas of need, and allocate resources effectively to support student success. Moreover, AfL encourages ongoing feedback and collaboration between instructors and students, fostering a dynamic learning environment focused on continuous improvement and achievement.

Assessment as Learning (AaL). It has been argued that the ultimate goal of assessment should be to support self-regulated learners (Earl and Katz, 2006). To achieve this goal, the approach of AaL has been proposed. AaL is defined as a process of developing and supporting student's metacognition (i.e., the ability to reflect on one's own thinking), encouraging students to become critical connectors between assessment and learning (Earl and Katz, 2006). Similar to AfL, AaL has evolved from formative assessment principles. AaL emphasizes the active participation of students in their own assessment and views the assessment process as an integral part of the learning process. This approach primarily involves peer assessment and self-assessment with the goal of helping students to become autonomous learners (Earl and Katz, 2006). For example, by engaging in self-assessment and reflective practices, students develop a deeper understanding of their strengths and areas for improvement, which contributes to their ability to take ownership of their learning.

All three assessment purposes have their unique value and should be integrated to make assessment a powerful process for enhancing student learning (Earl and Katz, 2006; Zeng *et al.*, 2018). As noted in Earl

and Katz's work (2006, p14), "*It is important for educators to understand the three assessment purposes, recognize the need to balance among them, know which one they are using and why, and use them all wisely.*"

Research on instructors' conceptions of assessment purposes has been extensive in primary and secondary education, both domestically and internationally (e.g., Barnes *et al.*, 2017; Brown *et al.*, 2019). A significant body of literature emerged following the development of the Teacher Conceptions of Assessment inventory (Brown, 2006). This instrument consists of statements measuring four main purposes of educational assessment, including "assessment makes schools accountable", "assessment makes students accountable", "assessment improves education", and "assessment is irrelevant". The validity of this instrument was established through research with primary instructors in New Zealand and Queensland. Subsequently, researchers have employed this instrument or refined versions to different groups of instructors (e.g., in-service and pre-service) across various school contexts (e.g., primary school and secondary school) and in different countries (e.g., United States, Spain, and China)(Brown *et al.*, 2011; Brown and Remesal, 2012; Barnes *et al.*, 2017; Brown *et al.*, 2019). Findings from these studies reveal significant variations in instructors' conceptions of assessment purposes across different cultural contexts, underscoring the importance of exploring this topic beyond primary and secondary education. To date, this topic remains relatively understudied in higher education contexts. The few studies that exist include one large-scale survey study characterizing both instructors' and students' conceptions of assessment purposes (Fletcher *et al.*, 2012), a longitudinal case study exploring the evolution of instructors' assessment purposes (Offerdahl and Tomanek, 2011), semi-interviews characterizing instructors' assessment purposes (Reimann and Sadler, 2016) and the relationships between instructors' assessment purposes and their assessment practices (Postareff *et al.*, 2012). The contexts of these studies vary across all disciplines with only one specifically focused on chemistry instructors (Offerdahl and Tomanek, 2011). Therefore, more research efforts are needed to provide a comprehensive understanding of chemistry instructors' conceptions of assessment purposes and their relationship to their assessment practices.

Contextual factors

Contextual factors encompass environmental factors at various levels that are outside instructors' control yet influence their instructional practices, including classroom, departmental, institutional, and broad cultural contexts (Figure 1; Gess-Newsome *et al.*, 2003). Specifically, classroom context factors include class size, classroom resources, individual characteristics of students, and social factors within the

classroom such as student-instructor interactions. Departmental and institutional contexts include factors that are not part of the classroom but may directly influence the classroom dynamics (e.g., policy and support from department and university). Examples of broad cultural contexts include national educational policies and disciplinary norms related to education and assessment.

Contextual factors that influence instructors' assessment thinking and practices have received little attention in the literature. While a few review articles explore this topic, they are mainly focused on primary and secondary education (Fulmer *et al.*, 2015; Heitink *et al.*, 2016; Yan *et al.*, 2021). These studies have identified various contextual factors that may influence assessment practices, including immediate course context (e.g., subject area and student characteristics), school and community factors (e.g., school culture and climate, internal school support), and macro-social factors (e.g., cultural norms and national or international policies). For example, instructors working in a culture that emphasizes assessment for accountability and achievement standards are likely to value and use AoL (Deneen *et al.*, 2019). Additionally, school leaders who emphasize the importance of AfL and provide continuous support are more likely to encourage the implementation of AfL-aligned assessment tools (Moss *et al.*, 2013).

In discussions about assessment reforms within higher education, the concept of institutional assessment culture has emerged as crucial. Institutional assessment culture refers to "*the deeply embedded values and beliefs collectively held by members of an institution who influence assessment practices on their campuses*" (Fuller *et al.*, 2016, 20). Several theoretical and empirical studies have highlighted the significance of institutional assessment culture on the success of assessment change initiatives (Medland, 2014; Fuller *et al.*, 2015; Fuller *et al.*, 2016; Skidmore *et al.*, 2018; Simper *et al.*, 2022). For example, Fuller *et al.* (2015) found that instructors commonly are able to describe general assessment norms in their discipline and these norms differ across disciplines. However, although they generally are motivated to change assessment, their actions for change appeared to be influenced by institutional contexts.

Related research in chemistry education focuses primarily on the variations across different course contexts (e.g., lower-level vs. upper-level courses). For instance, a couple of national survey studies suggest that postsecondary chemistry instructors' assessment practices vary significantly across courses at different levels and with different class sizes (Gibbons *et al.*, 2018; Gibbons *et al.*, 2022). Overall, little is known regarding how contextual factors at other levels (e.g., institutional level and departmental level)

influence chemistry instructors' assessment practices. Therefore, more research is needed to address this knowledge gap.

Personal factors

Personal factors include instructors' demographic profiles, professional development experiences, teaching experiences, preparation for the course, and experiences with innovative pedagogy as a student. Empirical evidence supports an association between these factors and instructors' teaching practices (Apkarian *et al.*, 2021; Yik *et al.*, 2022; Kraft *et al.*, 2024). Gibbons and colleagues' national survey work (Gibbons *et al.*, 2018; Gibbons *et al.*, 2022) also captured how postsecondary chemistry instructors' assessment practices are related to three personal factors (i.e., number of years teaching, participation in workshops, and funding received to improve courses). This work suggested that attending teaching-focused workshops is related to a higher-level use of research-based assessment tools. Given the limitations imposed by our small sample size, which restricts the reliability of comparisons across different personal factors, we did not include personal factors in our data analysis. However, we acknowledge that personal factors are essential to the TCSR model that serve as the theoretical foundation for the current work.

METHODOLOGY

A convenience sample (Robinson, 2014) was obtained for this study by contacting general chemistry instructors *via* email at institutions that were geographically close to the University of Virginia. A total of 143 instructors were contacted across 51 institutions spanning 2 states. Nineteen general chemistry instructors at 14 different four-year institutions voluntarily participated in our study. The descriptive demographics and course contexts for the participants are presented in Tables 1 and 2. Approval to conduct this study was obtained by the institutional review board at the University of Virginia (protocol number 001406).

Table 1. Descriptive demographics for participants

Demographic Variables		Number of participants
Gender	Women	11
	Men	8
Academic rank	Lecturer/Visiting professor	8

	Assistant Professor	3
	Associate Professor	4
	Professor	4
Teaching experience	Less than 5 years	5
	5.1 - 10 years	5
	10.1 – 15 years	4
	15 more years	5

Table 2. Participant course contexts

Context Variables		Number of participants
Class Size	Less than 50	12
	more than 100	7
Classroom Layout	Fixed	13
	Flexible	5
	Both	1
Institution Type*	R1	6
	ML	4
	MM	3
	MS	1
	PUI	5

* Based on Carnegie Classifications: **R1**: Doctoral Universities: Very High Research Activity; **ML**: Master's Colleges & Universities: Larger Programs; **MM**: Master's Colleges & Universities: Medium Programs; **MS**: Master's Colleges & Universities: Small Programs; **PUI**: Baccalaureate Colleges: Arts & Sciences Focus

Data collection

Semi-structured interviews with these 19 instructors were conducted online through Zoom during Spring and Summer 2021 in the midst of the COVID-19 pandemic. The interview protocol, designed by L.S. and M.S., consists of a set of questions targeting several main topics related to instructors’ characteristics, course context, enacted assessment practices, conceptions of the purposes of these practices, and perceived external influences on these practices (The protocol can be found in the appendix). In order to understand assessment practices under normal teaching circumstances, instructors were asked to provide responses based on the course they taught during the academic year of 2019-2020 (before the COVID-19 pandemic). Course artifacts, including syllabi, exams, homework, as well as any assessment materials, were collected from the interviewees and reviewed before each interview in order to familiarize the

interviewer with the particulars of the interviewee's practices. Interviewees were prompted to confirm the use of each specific assessment practice found in their artifacts and provided opportunities to offer information regarding any assessment practices that were not reflected in their artifacts. The interviewer (L.S.) audio-recorded the dialogue for transcription purposes. Interview lengths ranged from 47 to 93 minutes.

Data analysis

The recorded interviews were transcribed verbatim through the transcription service Temi. L.S. then reviewed and cleaned the transcripts to remove any identifiable information. To further protect the instructors' identity, a document number was assigned to each instructor prior to data analysis (e.g., instructor #1).

The data analysis includes multiple stages involving several researchers. Firstly, a master codebook for all research questions was developed by a team of three researchers (L.S., J.K.M.J. and M.S.). They first independently open-coded three transcripts using NVivo, with the aim to identify instructors' demographic information, course context, assessment practices, purposes of assessment, and external influences. The researchers then met to revise the codes, discuss the unique cases, and reach an agreement on the initial master codebook.

To answer RQ 1, two researchers (L.S. and J.K.M.J.) independently coded seven more transcripts (around 30% of the total number of transcripts) through an iterative process, during which any disagreement on the codebook was discussed and used to update the codebook. The final version of the codebook was applied to analyze the seven transcripts which resulted in a strong inter-rater reliability ($\kappa = 0.77$; Sim and Wright, 2005). L.S. coded the rest of the transcripts independently ($n=12$). Y.W. and M.S. then reviewed and updated the codebook slightly. In particular, the codes "poll questions", "in-class worksheet" and "in-class questions" were grouped under "in-class assessment" and assessment tools coded under "homework" were split into two new codes, "pre-class assessment" or "post-class assessment." The final version of the codebook can be found in Appendix 2.

To answer RQ 2, Y.W. reviewed all 19 transcripts and developed a codebook pertaining to external pressures that instructors perceived on implementing assessment practices. Six chemistry education researchers (including M.S. and J.K.M.J.) formed into pairs, with each pair collectively coding five to six of

the transcripts simultaneously under the facilitation of Y.W. During the interactive coding process, any confusion about the codes was discussed by the whole research group, and transcripts were re-coded by each member of the research team after updating the codebook. Once the research team had completed coding, Y.W. coded all the transcripts independently, compared coding results with the ones generated by the research team, and discussed any discrepancies with them until a full consensus was reached. The final codebook for RQ 2 can be found in Appendix 3.

To answer RQ 3, two researchers (L.S. and J.K.M.J.) used a consensus coding strategy. Specifically, both researchers independently coded one transcript using the initial master codebook. After debriefing with M.S., both researchers continued coding another four transcripts, discussed the codes and developed a final codebook (Appendix 4). Subsequently, using the final codebook, both researchers independently coded all the transcripts, discussed, and resolved any discrepancies until a full consensus was reached.

After the completion of coding, a thematic analysis approach was employed to analyze the links between instructors' assessment practices and their conceptions of assessment purposes and their perceived external pressures (RQ 4) (Saldaña, 2021).

Trustworthiness

The trustworthiness was addressed by fulfilling four major criteria: 1) *credibility*, defined as the extent to which the presented findings reflect the original views of the participants, 2) *transferability*, defined as the extent to which the findings can be transferred to other contexts with other participants, 3) *dependability*, defined as the stability of findings over time, and 4) *confirmability*, defined as the extent to which the findings can be confirmed by other researchers (Lincoln and Guba, 1985; Anney, 2014). The transferability was addressed through an effort to recruit general chemistry instructors from diverse types of institutions and provide a thick description of these participants, study context, and interpretation of the data. This transparency enables readers to evaluate the extent to which the findings may be applicable to other contexts and different populations. The credibility, dependability, and confirmability were supported in multiple ways: peer debriefing, prolonged engagement, and rigorous coding process. Specifically, the authors of this paper consist of members who have intensive engagement in supporting faculty members to promote chemistry education reforms. This study also involves multiple peer debriefing meetings with chemistry education researchers who were not involved in the project but provided constructive comments on the development of interview protocols, data interpretation, and

research findings. Additionally, the development and refinement of the codebooks involved multiple different researchers across time, with a consensus or high-level interrater reliability reached in each round of iterative coding, indicating the stability of the findings across time and across different raters.

RESULTS

General chemistry instructors reported mostly using assessments that align with AoL and AfL purposes

Instructors were prompted about their assessment practices in their general chemistry courses in the interview with support from their course artifacts. Across our sample, general chemistry instructors reported using nine different assessment practices with an average of around five different types of assessment per course (Fig. 2).

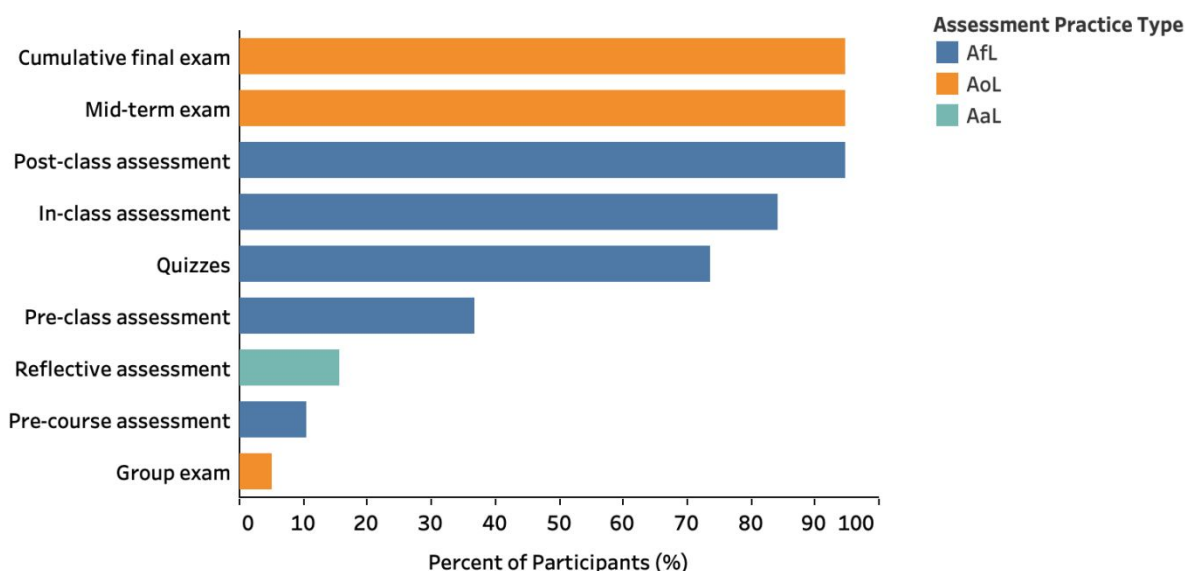


Fig. 2 Assessment practices used by general chemistry instructors in this study

The assessment practices of our sample can be categorized into three groups based on the opportunity they offer to fulfill Earl and Katz's three purposes for assessment (2006): AoL, AfL, and AaL (Fig. 2). Exams, including cumulative final exams, mid-terms, and group exams, align with the purpose of AoL as they primarily evaluate overall student understanding after the learning opportunities have been provided. Eighteen (95%) of the study participants reported using AoL-aligned assessment tools with an average of around two different types per participant.

AfL-aligned assessment tools comprise assessment practices that provide opportunities for instructors and students to receive feedback on teaching and learning throughout the course. This includes pre-class, in-class, and post-class assessments, as well as quizzes and pre-course assessments. All 19 instructors reported implementing at least one type of AfL-aligned assessment tool when teaching general chemistry, with an average of around three different types per participant. Post-class assessments are the most frequently used (n=17; 95% of participants), often embedded within commercial online homework/learning systems, such as ALEKS (Fang *et al.*, 2019) and Mastering Chemistry (Eichler and Peeples, 2013). These systems provide built-in, immediate feedback to students upon completing questions and may adjust questions' difficulty based on student progress. In-class assessments rank second, incorporated by sixteen participants (84%). Among these, thirteen instructors use the assessment within group work sessions during class. In these sessions, students are assigned chemistry problems or worksheets to work on collaboratively, or instructors orally ask questions while facilitating group activities. Additionally, eleven instructors (58%) use clicker questions and poll questions, which students engage with either individually or collaboratively. Quizzes are also commonly employed (n=14; 74% of participants) on a varied frequency basis, ranging from daily to monthly. Most of the quizzes are completed outside of the classroom. Pre-class assessments are utilized by 37% of the instructors (n=7), often through pre-class homework functions in commercial learning systems. These assessments include a series of questions associated with readings or tutorials and provide opportunities to prepare students for their classes. A smaller portion (n=2; 10% of participants) uses pre-course assessment, including diagnostic tests covering prerequisite topics and questions associated with tutorials to prepare students for the course. For example, instructor #3 talked about giving a pre-course assessment testing materials from the first semester of general chemistry at the beginning of the second semester of general chemistry:

"We give them an assessment as soon as they come in; it's all about the general chemistry I. Do you remember very fundamental ideas, you know, I think electron configuration, structure of matter. Maybe like solve a geometry problem, or write chemical equations, or a combustion reaction, so like, do you remember any chemistry?"

Reflective assessment tools align with an AaL purpose since they typically involve prompts or activities encouraging students to reflect on their learning process. For example, one instructor uses the "Muddiest Point" technique in class, allowing students to reflect on areas that they found difficult that day (King, 2011). Only three instructors (16%) reported using AaL-aligned assessment tools with each of them using

one type. It's worth noting that some activities or tasks categorized under AoL and AfL may contain reflective components for students, which could also be considered as AaL. For example, instructors might ask questions to students during class discussions to prompt them to reflect on their thought processes.

The academic culture and departmental norms tend to influence instructors' assessment practices

During the interviews, instructors were prompted to answer questions specifically addressing the external pressures or factors that they perceive to influence their assessment practices. These questions were designed to target the external pressures or factors at various levels, including the national level (i.e., American Chemistry Society), institutional level, department level, and classroom level. Additionally, some instructors voluntarily discussed their perceptions of external pressure when asked about the purposes of assessment. Responses to both prompted and unprompted questions were captured and analyzed.

Classroom influences. Five instructors (26%) perceived pressure or external factors at the classroom level. The first category at this level is instructors feeling pressured by students' comments or requests. Two instructors (11%) cited students' dissatisfaction with their assessment workload as a source of pressure, but they did not abide by that pressure. For example, instructor #10's students complained on the student evaluation that *"there's too much homework or there's too many quizzes;"* however, since other students mentioned the benefit of having many quizzes, instructor #10 *"tend to not follow through on their advice"*. The second category at this level is class size. Two instructors (16%) indicated that the large class size of general chemistry prevented them from using the type of assessment they would like to use or influenced the way they use the assessment, as stated by instructor #1:

"I guess the only factor that really affects assessment is volume, I would say volume. Multiple choice it needs to be Mastery content ... Like with 3,000 to 6,000 people taking the test at the same time, Volume, and we can't grade by ourselves, it should be computer gradable."

Departmental influences. Eight instructors (42%) reported experiencing pressure at the departmental level. The dominant code at this level is "pressure of aligning with departmental culture", which was cited by six instructors (32%). In particular, these instructors expressed the pressure to maintain consistency on the assessment across sections of the course, as instructor #18 explains:

1
2
3 *"So, there is definitely pressure that we do similar things like it doesn't work if I give four exams*
4 *and someone else gives seven; it doesn't work if I have 25% for formative assessments and they*
5 *have 10%. So there is pressure to conform in that way so that the classes feel the same ... we force*
6 *everybody to give the ACS exam. I mean, you have to, you have to do that. So there are norms that*
7 *group of faculty who teach general chemistry and in gen chem one, there will be between six and*
8 *nine of us in any given year."*
9

10
11
12
13
14
15 Another instructor expressed an internal pressure to fit into the departmental norms. Unlike the former
16 group, this instructor was not formally obligated to use strict assessment guidelines, but they felt pressure
17 to conform to their perceptions of the departmental model:
18
19

20
21
22 *"I want to fit in with the rest of the department. I want to be doing what they're doing. I want*
23 *to make sure their students are getting a fair shake. Like my students are getting. So, I will do*
24 *assessment because that's just part of the culture"* - Instructor #15.
25
26
27

28
29 Two instructors (11%) expressed the pressure of preparing students for their future courses due to the
30 foundational nature of the general chemistry course. For example, instructor #17 mentioned that
31

32
33 *"There's pressure from the department in the sense that like the instructors of the future courses*
34 *want to make sure that the students coming to them are prepared. It's very difficult for them. So ...*
35 *the real focus is, are we giving assessments that accurately predict whether they're prepared for*
36 *the next class?"*
37
38
39
40
41

42 **Institutional influences.** Five instructors (26%) reported having experienced pressure from the university.
43 Four of them (21%) mentioned that they were required by the university to have the same questions
44 covering certain topics or to have a standardized exam for the purpose of comparing students'
45 performance across sections. This is exemplified by responses from instructor #5 *"in the past we've used*
46 *the ACS general exam. Now basically it's a box to check for the university assessment people to show that*
47 *we've got some standardized measure that's across all the sections and all the professors."* Additionally,
48 one instructor (5%) mentioned that the university's provost has a requirement that every student has a
49 final exam, and every course has to meet a certain number of contact hours, including a three-hour final
50 exam period. However, there is freedom in terms of the format and the content covered in the exam.
51
52
53
54
55
56
57
58
59
60

Professional organization. Only two instructors (11%) cited elements of influence associated with their professional organization, i.e. the American Chemical Society. One instructor (#10) mentioned that they feel compelled to make students comfortable with seeing the multiple-choice question format as they use the ACS exam as their final exam. The other instructor (#6) indicated that they are *“pressured to do some assessment that allows us to compare one semester to another at some point, but that's really the only standard we have to conform to like to keep our ACS Accreditation.”*

Academic culture. Around half of the instructors (n=10; 55%) reported feeling pressure from the academic culture without identifying specific stakeholders. Six of them (32%) expressed using assessment because they feel obligated to assign grades to students. This perceived pressure is evident in the responses of instructor #19 when asked why they use assessment in general chemistry: *“I think because ... it has been communicated overtly and through non-verbal ways that I have to assess my students in one way or another and provide a letter grade.”* Instructor #12 echoed this sentiment, stating, *“why I use assessment in general chemistry? So, part of it is expectations, right? Like I've got to have assessments in there because I've got to give grades.”* Five of these ten instructors (26%) stated that they use assessments because this is typically what is done in general chemistry or education as a whole. This is exemplified by the responses of instructor #14, *“One is the, the not nice answer, because that's what everyone does.”* Similarly, instructor #4 stated *“Part of it is because that's the way that the courses have classically been, of course, that's definitely part of it.”* Although most instructors whose responses fall under this category stated they use assessment partially due to pressure from the academic culture, it was unclear from the interview how this pressure specifically influenced their assessment practices.

Overall, the dominant types of pressures felt by the general chemistry instructors were departmental ones and the academic culture in which they are embedded.

AoL dominated instructors' conceptions of assessment purposes

During the interview, instructors were prompted to answer three questions regarding their conceptions of assessment purposes. The questions posed were: *“What do you believe assessment is? What is the function of assessment in general chemistry? Why do you use assessment in general chemistry?”* (see Appendix 1 for interview details). The in vivo codes that emerged from the analysis were classified into four categories: AoL, AfL, AaL, and others, based on their alignment with Earl and Kraft's framework (2006) for assessment purposes.

The dominant purpose of assessment among instructors is AoL, with eighteen instructors (95%) describing a goal that falls under this category, reflecting their view of assessment as a tool for evaluating students' success. Sixteen instructors (84%) articulated that assessment allows them to assess students' mastery of knowledge and skills and their learning progress, as demonstrated by instructor #9:

"Assessment is a tool to see how well a student or a person ... understand a concept, or how well they can perform a skill. So, it's really a way to gauge where somebody is on a current, on some sort of goal that you have for them."

Furthermore, numerous instructors (n=12, 63%) highlighted the role of assessment in determining students' readiness for future chemistry courses. For example, instructor #10 emphasized the responsibility of instructors to *"prepare the students to take more chemistry courses...for biology major, environmental studies major"* and stated that *"they have to ask questions that at least attempt to prove their knowledge and competency of certain fundamental concepts they're going to encounter in the next course."* Similarly, instructor #17 underscored that *"the purpose of general chemistry is to get you ready for future chemistry (courses)"* and *"the purpose of assessment is to make sure that ... students have understanding and the ability that they need to progress to further chemistries (chemistry courses) or to progress in their major."* Two instructors (10%) mentioned using assessment to probe students' problem-solving ability in new scenarios, as noted by instructor #12: *"I think assessment for me ... it helps me figure out whether or not they can take the concepts to going to the class and then apply them in new situation(s)."*

The second most popular purpose of assessment was AaL with seven instructors (37%) describing goals that aligned with this purpose. AaL views assessment as a means for students to reflect on their learning process. Some instructors (n=5; 26%) highlighted the role of assessment in providing feedback to students, as evidenced by instructor #6: *"for intrinsically motivated student, the goal of the assessment is to provide them with feedback because students aren't always the best judges of how much they've learned and what their progresses (are)."* Four instructors (21%) underscored the role of assessment in helping students be aware of their learning process, promoting self-reflection and metacognition. For example, instructor #11 held a growth-oriented perspective with a belief in the value of mistakes and struggles within the assessment process as opportunities for learning and development: *"I want to get them to a point where they can make a mistake, or say gosh, there's something here that does not make sense; there's something here that doesn't fit with what I know so far. And maybe it makes sense when he talks about it, but I can't*

1
2
3 *get the problem to work in the right way. And then, and it's kind of like ... when you get to that point and*
4 *say okay, I know I'm struggling right here, that's when you can learn...I need to figure out something new*
5 *that's where I'm going to grow."*
6
7
8
9

10 Approximately a third of the instructors (n=6; 32%) viewed assessment as a means of gathering
11 information on students' learning, with the intention of informing and adapting instructional practices
12 (AfL). When demonstrating this idea, these instructors expressed viewing teaching as a dynamic process,
13 remaining open to adapting their teaching methods to enhance students' understanding based on
14 assessment insights. The following quote from instructor #14 illustrates this perspective: *"... if they're all*
15 *doing terrible on something that's maybe indicative of, I'm not doing something well, or maybe they're not*
16 *getting something. So, it's ideally a way to check their understanding so I can modify my methods so that*
17 *they can get that understanding."*
18
19
20
21
22
23
24

25 Finally, a minority of instructors (n=4; 21%) held perspectives not directly aligned with Earl and Kraft's
26 framework (2006). In this category, assessment was seen as a means of ensuring students' accountability
27 rather than focusing on students' learning outcomes or processes. This idea highlights the role of
28 assessment as a tool to drive or motivate students' learning, thereby promoting their engagement with
29 course materials and enhancing their learning efforts. For example, instructor #15 implied that students
30 may lack intrinsic motivation to learn so they view assessment as a tool to instill a sense of urgency and
31 commitment to studying the material: *"something to scare the students, so they study and learn the*
32 *material."* Similarly, instructor #6 believes that assessments with grades serve as a motivating factor for
33 some students and *"it's the assessments themselves that motivate them to try and, to learn and to*
34 *participate."*
35
36
37
38
39
40
41
42

43 Several instructors described more than one purpose for their assessment. In Fig. 3 we map the different
44 purposes that each instructor had. Eight instructors (42%) only described AoL purposes. Seven instructors
45 (37%) stated AoL alongside either AaL or AfL purposes. Only three instructors (16%) described all three
46 assessment purposes. The remaining instructor, whose responses did not align with either of these three
47 purposes, only emphasized the purpose of "ensuring students are studying and learning" during their
48 interview.
49
50
51
52
53
54
55
56
57
58
59
60

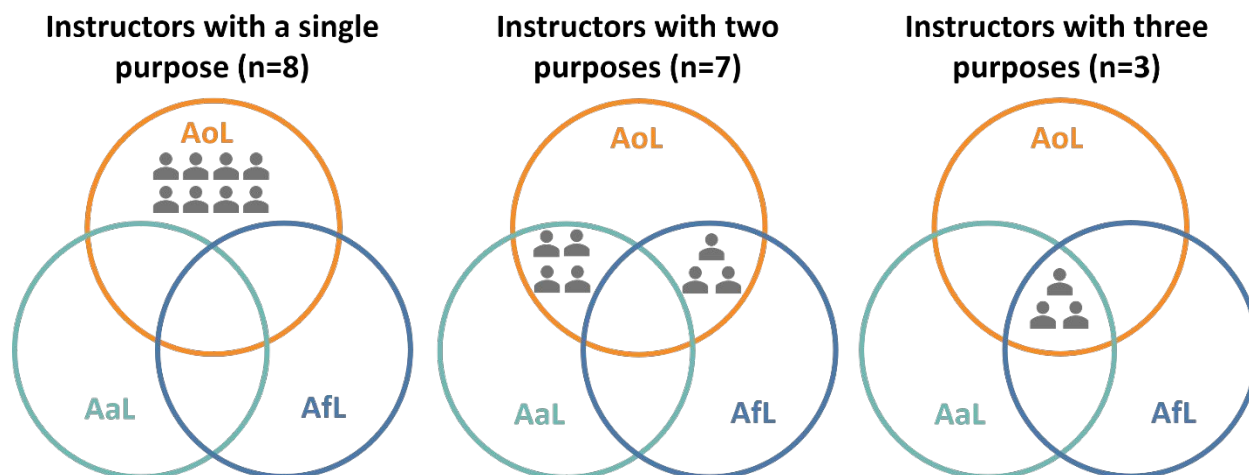


Fig. 3 Classification of instructors by their conceptions of assessment purposes

Instructors' assessment practices are related to their conceptions of assessment purposes

We explored the relationship between instructors' conceptions of assessment purposes and their assessment practices from two perspectives. First, we analyzed the variation in the average number of assessment types across the three groups of instructors created based on the combinations of the assessment purposes they described (see groups in Fig. 3). As Fig. 4 demonstrates, overall, instructors who describe a greater variety of purposes tend to utilize a greater variety of assessment types. Specifically, instructors describing all three purposes of assessment reported using an average of around six different assessment types, while instructors who described two purposes reported using around five different assessment types, and instructors who described a single purpose reported using around four different assessment types (Fig. 4). Second, we investigated variations across the three groups of instructors by assessment types (those that align with AoL, AfL, and AaL purposes). As Fig. 4 indicates, the average number of assessment tools aligned with AoL purpose remains consistent across the three groups of instructors while the number of assessment tools aligned with AfL and, to some extent, AaL increases as the number of purposes described by the instructors increases. In other words, instructors who reported more assessment purposes tend to utilize more assessments that align with AfL purpose such as pre- and in-class assessments as well as quizzes (Fig. 5).

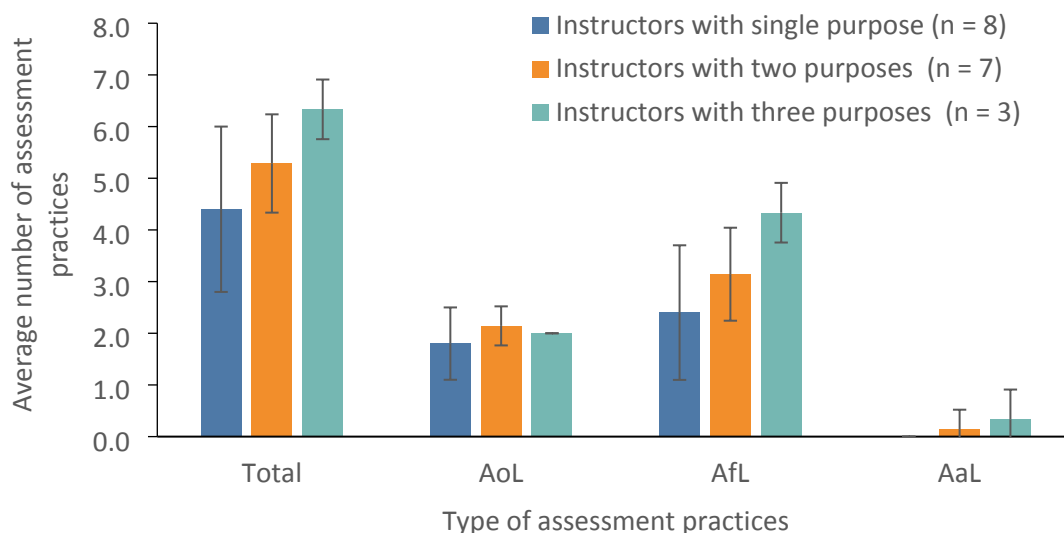


Fig. 4 Average number for each type of assessment practices by instructor's group

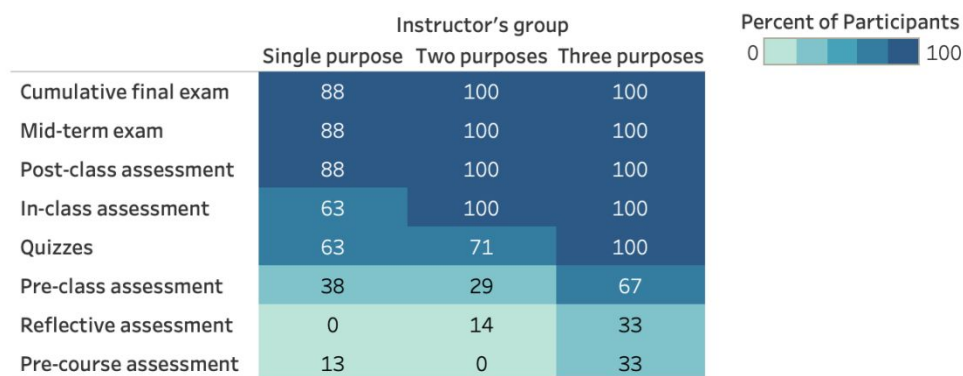


Fig. 5 Percent of instructors (%) within each group for each assessment

DISCUSSION AND IMPLICATIONS

The main goal of this study is to provide a baseline description of chemistry instructors' assessment practices, factors influencing them, and their thinking about assessment with a specific focus on their conceptions of assessment purposes. This baseline description is intended to inform the design of assessment reform efforts in chemical education.

General chemistry instructors do not often use AfL-aligned assessment practices that monitor students' prior knowledge and rarely use AaL-aligned assessment practices

All but one of the general chemistry instructors in this study used AoL-aligned assessment practices, such as final exams and mid-terms. This prevalence resonates with findings from a recent national survey study exploring postsecondary chemistry instructors' assessment practices (Gibbons *et al.*, 2022), in which they found that over 90% of chemistry instructors use mid-terms and final exams. The majority of their participants also used AfL-aligned assessments, such as post-class assessments and quizzes. However, in our study, while AfL-aligned assessments that monitor student learning during the learning process (e.g., in-class and post-class assessments) were used by a majority of the participants, AfL-aligned assessments that target new content, including pre-class and pre-course assessments, were used by less than half of our participants. This latter type of assessment is intended to provide feedback to instructors regarding where students stand with their understanding of topics that will be explored in the upcoming sessions. With this knowledge in hand, instructors can tailor the sessions to focus on and address students' difficulties. These assessments thus allow the instructors to judiciously use their time with their students and maximize students' opportunities for learning. Several potential reasons exist for the limited use of this type of assessment by our instructor population. First, there may be a limited availability of resources. Indeed, resources that instructors often rely on for AfL-aligned assessments, such as post-class assessments, are provided by textbook publishers. However, publishers typically market their assessment package as a homework system. While instructors could leverage the package to create pre-class assessments, it does not seem that instructors engage in this approach. Second, instructors may not feel that they have enough time to process the results from pre-class assessments and refine their plan for the learning sessions based on these results. Time is one of the most common barriers to instructional innovation mentioned across studies exploring influences on adopting innovative instructional methods (Sturtevant and Wheeler, 2019). Third, instructors may not be aware of the principles of constructivism and the role that prior knowledge plays in the learning process. The relationship between assessment practices and pedagogical knowledge (e.g., knowledge about learning theories) is an underexplored area of research in chemical education education and one that ought to be explored further.

The least commonly used type of assessments, by general chemistry instructors, were those aligned with the purpose of AaL (e.g., self-assessment and exam wrappers). The scarcity of AaL-aligned assessments supports the findings from an interview study of pharmacy instructors' assessment practices (Postareff *et al.*, 2012). The lack of use of AaL-aligned assessments could be explained by the limited availability of

resources. Another potential reason would be that instructors may not be aware of their existence. Indeed, research on instructors' adoption of innovative practices has demonstrated that choices to integrate new instructional practices are often grounded in instructors' past experiences as students (e.g., Kraft *et al.*, 2024). Without experiencing these practices as students, instructors would have to be exposed to them via pedagogical training. Yet a recent national survey of general chemistry instructors shows a substantial lack of pedagogical training (Wang *et al.*, 2024). For instructors who are in a position to include AaL in their courses, we encourage them to explore extant resources in education research that help students develop metacognition and self-regulated skills by actively engaging in assessments and acting on feedback from instructors (e.g., Tanner, 2012; Muteti *et al.*, 2021; Stanton *et al.*, 2021; Swanson *et al.*, 2024).

Changing assessment practices necessitates changes in instructors' conceptions of assessment purposes

Our results indicate that nearly all instructors (95%) recognized the purpose of AoL, whereas only around a third were aware of the purposes of AfL and AaL. Only 16% mentioned all three purposes. Generally, instructors tended to view assessment as a means to evaluate students rather than to support teaching and learning. These findings contrast with a previous study on higher education instructors' conceptions of assessment practices in New Zealand (Fletcher *et al.*, 2012). The authors collected data using the Conceptions of Assessment instrument with 877 instructors and found that instructors viewed assessment as serving two purposes, namely improving student learning and informing teaching practices. The contrasting results may be attributed to the difference in context. Indeed, prior research highlights the context-dependency of assessment literacy (Willis *et al.*, 2013). These findings suggest that the chemical education community needs to raise general chemistry instructors' awareness of the critical role that assessment plays in teaching and learning. Importantly, supporting instructors in developing broader conceptions of assessment could lead to enhanced assessment practices. For example, in this study, general chemistry instructors' conceptions of assessment purposes were related to their assessment practices. Specifically, instructors who perceived the purposes of AfL and AaL alongside AoL tended to employ more AfL-aligned assessment practices. This alignment between conceptions of assessment and assessment practices is in line with extant literature in both secondary and postsecondary education (Brown *et al.*, 2009; Postareff *et al.*, 2012). Therefore, the findings of this study related to both conceptions of assessment purposes and assessment practices suggest opportunities for change agents, including institutional administrators, professional development facilitators, education researchers and instructors, to drive assessment reform in chemistry. Hutchings (2010) pointed out that one of the

greatest barriers to involving faculty in assessment has been a lack of training. Given the limited professional development experiences reported among general chemistry instructors nationally (Wang *et al.*, 2024), urgent action is needed to provide targeted professional development opportunities, such as workshops and pedagogical courses, focused on assessment conceptions and practices. Additionally, such change initiatives can be adopted through more collaborative practices such as a learning community. As Feola *et al.* (2023) highlighted, instructors' assessment literacy and their assessor identity can be influenced through interactions with peers in a community of practices.

Future endeavors on assessment reform need to address academic culture and departmental norms

The findings of this study uncover potential barriers to change in assessment practices. Firstly, general chemistry instructors commonly perceived a culture within academia that obligates them to implement assessment practices in order to provide grades. When they were asked in the interview why they assess, over half of the instructors provided a version of the following answer: "because we have to." Second, instructors perceived departmental norms regarding assessment. For example, they felt pressured to maintain consistency in assessment practices across course sections or to follow the same assessment standards as their departmental peers. The findings from this study are in dissonance with a recent national survey that captured general chemistry instructors' level of autonomy with different aspects of their teaching, including instructional methods and exams (Wang *et al.*, 2024). In this study, about 80% of the respondents indicated that they were the sole decision-makers on these two aspects of their teaching. The differences in the descriptions of influences between the two studies demonstrate the complexity of the factors influencing instructional practices and the need to develop a more comprehensive baseline of chemistry instructors' instructional decision-making processes and influencers.

With respect to practical implications, institutional and departmental leaders should provide more flexible mechanisms and freedom around assessment to enable assessment growth among their instructors. Indeed, an assessment culture can be cultivated through the empowerment of stakeholders, including institutional leaders, assessment administrators, and instructors (Simper *et al.*, 2022).

LIMITATIONS

We would like to acknowledge several limitations of this work. First, while we aimed to recruit instructors with diverse demographic backgrounds, the generalizability of our findings was limited by the small sample size. Additionally, the nature of the general chemistry courses they teach, commonly considered gateway courses for STEM majors, may shape disciplinary culture differently from that of upper-level

chemistry courses, where instructors may perceive different purposes of assessment. Specifically, some of the general chemistry instructors mentioned evaluating students' preparedness for upper-level courses, a purpose possibly not shared by instructors of upper-level chemistry courses. We therefore advocate for further research to explore this topic in a broader context with a larger sample size (e.g., a national survey of instructors). Furthermore, although data collection occurred during the COVID-19 pandemic, the study's goal was to explore instructors' perspectives on assessment practices under normal circumstances (i.e., in-person instruction). Thus, we specifically prompted instructors to discuss their assessment practices prior to the pandemic. However, it is possible that instructors' assessment practices and their perspectives were influenced by the pandemic, during which they were compelled to conduct assessments online. Therefore, additional research is needed to examine this topic in the post-pandemic era. Lastly, we gathered information on instructors' assessment practices through self-reported data, which may not fully reflect their actual classroom practices. The reported assessment practices were categorized based on their alignment with one of the three assessment purposes. However, instructors may implement the same assessment practice in diverse ways. Future research could explore instructors' perceptions of the nature of assessment practices and how they implement them in real classroom settings.

CONCLUSIONS

Assessment is an essential component of instructional design and plays a critical role in providing equal opportunities for students to learn. Recent studies have indicated the inadequacies of assessment practices in general chemistry courses and the need for reform. Effective reforms require baseline knowledge of the practices and ways of thinking of the targeted population. Unfortunately, few studies have explored chemistry instructors' assessment practices and ways of thinking. This work addresses this gap by capturing the assessment practices and conceptions of assessment purposes of 19 general chemistry instructors. The study shows that few general chemistry instructors identified all three purposes of assessment described by Earl and Kraft (2006). Almost all saw assessment as a means to evaluate students (AoL); a third felt assessment was intended to provide feedback to both students and instructors (AfL), while another third thought assessment was intended to promote students' self-regulated learning skills (AaL). Interestingly, those who identified purposes beyond AoL tended to employ a broader range of assessment practices in their courses. Moreover, the study highlights that academic culture and departmental norms may play an important role in influencing instructors' assessment practices. These

findings offer valuable insights for change agents and educational stakeholders seeking to support instructors' engagement in assessment reform.

CONFLICT OF INTEREST

There are no conflicts to declare.

AUTHOR CONTRIBUTIONS

L.S. and M.S. conceptualized the study. L.S. conducted the semi-structured interviews. All authors participated in data analysis. Y.W. and M.S. wrote the manuscript with input from all authors. All authors read, edited, and approved the final manuscript.

ACKNOWLEDGEMENTS

We would like to acknowledge the support of the National Science Foundation under Career Award 1552448, 2021491. We would also like to thank all the general chemistry instructors who participated in this study and our research team, including Dr. Brandon Yik, Emily Kable, Haleigh Machost, for their contribution to the data interpretation and constructive feedback on previous versions of the manuscript.

REFERENCES

- Anney V. N., (2014), Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria, *J. Emerg. Trends Educ. Res. Policy Stud.*, **5**, 272-281.
- Apkarian N., Henderson C., Stains M., Raker J., Johnson E. and Dancy M., (2021), What really impacts the use of active learning in undergraduate STEM education? Results from a national survey of chemistry, mathematics, and physics instructors, *PLoS One*, **16**, e0247544.
- Barbera J., Harshman J. and Komperda R., (2023), The Chemistry Instrument Review and Assessment Library (CHIRAL): A New Resource for the Chemistry Education Community, *J. Chem. Educ.*, **100**, 1455-1459.
- Barnes N., Fives H. and Dacey C. M., (2017), U.S. teachers' conceptions of the purposes of assessment, *Teach. Teach. Educ.*, **65**, 107-116.
- Black P., Harrison C. and Lee C., (2003), *Assessment for learning: Putting it into practice*, McGraw-Hill Education (UK).
- Black P., Harrison C., Lee C., Marshall B. and Wiliam D., (2004), Working inside the black box: Assessment for learning in the classroom, *Phi Delta Kappan*, **86**, 8-21.
- Black P. and Wiliam D., (1998), *Inside the black box: Raising standards through classroom assessment*, Granada Learning.
- Brown G. A., Bull J. and Pendlebury M., (2013), *Assessing student learning in higher education*, Routledge.

- 1
2
3 Brown G. T., (2006), Teachers' conceptions of assessment: validation of an abridged version,
4 *Psychol. Rep.*, **99**, 166-170.
- 5 Brown G. T., Hui S. K., Flora W. and Kennedy K. J., (2011), Teachers' conceptions of
6 assessment in Chinese contexts: A tripartite model of accountability, improvement, and
7 irrelevance, *Int. J. Educ. Res.*, **50**, 307-320.
- 8 Brown G. T. and Remesal A., (2012), Prospective teachers' conceptions of assessment: a cross-
9 cultural comparison, *Span. J. Psychol.*, **15**, 75-89.
- 10 Brown G. T. L., Gebril A. and Michaelides M. P., (2019), Teachers' Conceptions of Assessment:
11 A Global Phenomenon or a Global Localism, *Front. Educ.*, **4**.
- 12 Brown G. T. L., Kennedy K. J., Fok P. K., Chan J. K. S. and Yu W. M., (2009), Assessment for
13 student improvement: understanding Hong Kong teachers' conceptions and practices of
14 assessment, *Assess. Educ. Princ. Policy Pract.*, **16**, 347-363.
- 15 Darling-Hammond L. and Snyder J., (2000), Authentic assessment of teaching in context, *Teach.*
16 *Teach. Educ.*, **16**, 523-545.
- 17 Deneen C. C., Fulmer G. W., Brown G. T. L., Tan K., Leong W. S. and Tay H. Y., (2019),
18 Value, practice and proficiency: Teachers' complex relationship with assessment for
19 learning, *Teach. Teach. Educ.*, **80**, 39-47.
- 20 Dwyer C. A., (2006), Assessment and Classroom Learning: theory and practice, *Assess. Educ.*
21 *Princ. Policy Pract.*, **5**, 131-137.
- 22 Earl L. M. and Katz M. S., (2006), *Rethinking classroom assessment with purpose in mind:*
23 *Assessment for learning, assessment as learning, assessment of learning*, Manitoba
24 Education, Citizenship & Youth.
- 25 Eichler J. F. and Peebles J., (2013), Online Homework Put to the Test: A Report on the Impact of
26 Two Online Learning Systems on Student Performance in General Chemistry, *J. Chem.*
27 *Educ.*, **90**, 1137-1143.
- 28 Emenike M. E., Schroeder J., Murphy K. and Holme T., (2013), Results from a National Needs
29 Assessment Survey: A View of Assessment Efforts within Chemistry Departments, *J.*
30 *Chem. Educ.*, **90**, 561-567.
- 31 Emenike M. E., Schroeder J. D., Murphy K. L. and Holme T., (2011), A Snapshot of Chemistry
32 Faculty Members' Awareness of Departmental Assessment Efforts, *Update*, **23**.
- 33 Fang Y., Ren Z. H., Hu X. E. and Graesser A. C., (2019), A meta-analysis of the effectiveness of
34 ALEKS on learning, *Educ. Psychol.*, **39**, 1278-1292.
- 35 Feola S., Lemons P. P., Loertscher J. A., Minderhout V. and Lewis J. E., (2023), Assessor in
36 action: assessment literacy development in a biochemistry context, *Chem. Educ. Res.*
37 *Pract.*, **24**, 914-937.
- 38 Fletcher R. B., Meyer L. H., Anderson H., Johnston P. and Rees M., (2012), Faculty and
39 Students Conceptions of Assessment in Higher Education, *High Educ.*, **64**, 119-133.
- 40 Fuller M., Henderson S. and Bustamante R., (2015), Assessment leaders' perspectives of
41 institutional cultures of assessment: a Delphi study, *Assess. Eval. High. Educ.*, **40**, 331-
42 351.
- 43 Fuller M. B., Skidmore S. T., Bustamante R. M. and Holzweiss P. C., (2016), Empirically
44 Exploring Cultures of Assessment in Higher Education, *Rev High Educ*, **39**, 395-429.
- 45 Fulmer G. W., Lee I. C. H. and Tan K. H. K., (2015), Multi-level model of contextual factors
46 and teachers' assessment practices: an integrative review of research, *Assess. Educ. Princ.*
47 *Policy Pract.*, **22**, 475-494.
- 48
49
50
51
52
53
54
55
56
57
58
59
60

- Gess-Newsome J., Southerland S. A., Johnston A. and Woodbury S., (2003), Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching, *Am. Educ. Res. J.*, **40**, 731-767.
- Gibbons R. E., Reed J. J., Srinivasan S., Murphy K. L. and Raker J. R., (2022), Assessment Tools in Context: Results from a National Survey of Postsecondary Chemistry Faculty, *J. Chem. Educ.*, **99**, 2843-2852.
- Gibbons R. E., Reed J. J., Srinivasan S., Villafañe S. M., Laga E., Vega J., Murphy K. L. and Raker J. R., (2018), Assessment in Postsecondary Chemistry Education: A Comparison of Course Types, *Assess. Update*, **30**, 1-16.
- Glaser R., Chudowsky N. and Pellegrino J. W., (2001), *Knowing what students know: The science and design of educational assessment*, National Academies Press.
- Heitink M. C., Van der Kleij F. M., Veldkamp B. F., Schildkamp K. and Kippers W. B., (2016), A systematic review of prerequisites for implementing assessment for learning in classroom practice, *Educ. Res. Rev.*, **17**, 50-62.
- Holme T., Bretz S. L., Cooper M., Lewis J., Paek P., Pienta N., Stacy A., Stevens R. and Towns M., (2010), Enhancing the role of assessment in curriculum reform in chemistry, *Chem. Educ. Res. Pract.*, **11**, 92-97.
- Hutchings P., (2010), Opening doors to faculty involvement in assessment, *NILOA Occasional Paper*, **4**.
- King D. B., (2011), Using Clickers To Identify the Muddiest Points in Large Chemistry Classes, *J. Chem. Educ.*, **88**, 1485-1488.
- Kraft A. R., Atieh E. L., Shi L. and Stains M., (2024), Prior experiences as students and instructors play a critical role in instructors' decision to adopt evidence-based instructional practices, *Int. J. STEM Educ.*, **11**.
- 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. and Cooper M. M., (2016), Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol, *PLoS One*, **11**, e0162333.
- Lincoln Y. S. and Guba E. G., (1985), *Naturalistic inquiry*, sage.
- Medland E., (2014), Assessment in higher education: drivers, barriers and directions for change in the UK, *Assess. Eval. High. Educ.*, **41**, 81-96.
- Mertler C. A. and Campbell C., *Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory"*, 2005.
- Moss C. M., Brookhart S. M. and Long B. A., (2013), Administrators' Roles in Helping Teachers Use Formative Assessment Information, *Appl. Meas. Educ.*, **26**, 205-218.
- Muteti C. Z., Zarraga C., Jacob B. I., Mwarumba T. M., Nkhata D. B., Mwavita M., Mohanty S. and Mutambuki J. M., (2021), I realized what I was doing was not working: the influence of explicit teaching of metacognition on students' study strategies in a general chemistry I course, *Chem. Educ. Res. Pract.*, **22**, 122-135.
- Offerdahl E. G. and Tomanek D., (2011), Changes in instructors' assessment thinking related to experimentation with new strategies, *Assess. Eval. High. Educ.*, **36**, 781-795.
- Park E. S., Wilton M., Lo S. M., Buswell N., Suarez N. A. and Sato B. K., (2024), STEM Faculty Instructional Beliefs Regarding Assessment, Grading, and Diversity are Linked to Racial Equity Grade Gaps, *Res. High. Educ.*
- Postareff L., Virtanen V., Katajavuori N. and Lindblom-Ylänne S., (2012), Academics' conceptions of assessment and their assessment practices, *Stud. Educ. Eval.*, **38**, 84-92.

- Raker J. R., Emenike M. E. and Holme T. A., (2013), Using Structural Equation Modeling To Understand Chemistry Faculty Familiarity of Assessment Terminology: Results from a National Survey, *J. Chem. Educ.*, **90**, 981-987.
- Raker J. R. and Holme T. A., (2014), Investigating Faculty Familiarity with Assessment Terminology by Applying Cluster Analysis To Interpret Survey Data, *J. Chem. Educ.*, **91**, 1145-1151.
- Reimann N. and Sadler I., (2016), Personal understanding of assessment and the link to assessment practice: the perspectives of higher education staff, *Assess. Eval. High. Educ.*, **42**, 724-736.
- Robinson O. C., (2014), Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide, *Qualitative Research in Psychology*, **11**, 25-41.
- Sadler I. and Reimann N., (2017), Variation in the development of teachers' understandings of assessment and their assessment practices in higher education, *High. Educ. Res. Dev.*, **37**, 131-144.
- Saldaña J., (2021), The coding manual for qualitative researchers, *The coding manual for qualitative researchers*, 1-440.
- Schultz M., Lawrie G. A., Bailey C. H., Bedford S. B., Dargaville T. R., O'Brien G., Tasker R., Thompson C. D., Williams M. and Wright A. H., (2017), Evaluation of diagnostic tools that tertiary teachers can apply to profile their students' conceptions, *Int. J. Sci. Educ.*, **39**, 565-586.
- Shepard L. A., (2000), *The Role of Classroom Assessment in Teaching and Learning*, Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, and Center for Research on Education, Diversity and Excellence, University of California, Santa Cruz.
- Sim J. and Wright C. C., (2005), The kappa statistic in reliability studies: use, interpretation, and sample size requirements, *Phys. Ther.*, **85**, 257-268.
- Simper N., Mårtensson K., Berry A. and Maynard N., (2022), Assessment cultures in higher education: reducing barriers and enabling change, *Assess. Eval. High. Educ.*, **47**, 1016-1029.
- Skidmore S. T., Hsu H. Y. and Fuller M., (2018), A person-centred approach to understanding cultures of assessment, *Assess. Eval. High. Educ.*, **43**, 1241-1257.
- Stanton J. D., Sebesta A. J. and Dunlosky J., (2021), Fostering Metacognition to Support Student Learning and Performance, *CBE Life Sci. Educ.*, **20**.
- Stiggins R., (2005), From formative assessment to assessment FOR learning: A path to success in standards-based schools, *Phi Delta Kappan*, **87**, 324-328.
- Stiggins R. J., (1994), *Student-centered classroom assessment*, Merrill New York.
- Stowe R. L. and Cooper M. M., (2017), Practicing what we preach: assessing "critical thinking" in organic chemistry, *J. Chem. Educ.*, **94**, 1852-1859.
- Stowe R. L., Scharlott L. J., Ralph V. R., Becker N. M. and Cooper M. M., (2021), You Are What You Assess: The Case for Emphasizing Chemistry on Chemistry Assessments, *J. Chem. Educ.*, **98**, 2490-2495.
- Sturtevant H. and Wheeler L., (2019), The STEM Faculty Instructional Barriers and Identity Survey (FIBIS): development and exploratory results, *Int. J. STEM Educ.*, **6**, 1-22.
- Swanson H. J., Ojutiku A. and Dewsbury B., (2024), The Impacts of an Academic Intervention Based in Metacognition on Academic Performance, *Teaching and Learning Inquiry*, **12**.
- Tanner K. D., (2012), Promoting Student Metacognition, *CBE Life Sci. Educ.*, **11**, 113-120.

- 1
2
3 Wang Y., Apkarian N., Dancy M. H., Henderson C., Johnson E., Raker J. R. and Stains M.,
4 (2024), A National Snapshot of Introductory Chemistry Instructors and Their
5 Instructional Practices, *J Chem Educ*, **101**, 1457-1468.
6
7 Willis J., Adie L. and Klenowski V., (2013), Conceptualising teachers' assessment literacies in
8 an era of curriculum and assessment reform, *Aust. Educ. Res.*, **40**, 241-256.
9
10 Woodbury S. and Gess-Newsome J., (2002), Overcoming the paradox of change without
11 difference: A model of change in the arena of fundamental school reform, *Educ. Policy*,
12 **16**, 763-782.
13
14 Yan Z., Li Z. Q., Panadero E., Yang M., Yang L. and Lao H. L., (2021), A systematic review on
15 factors influencing teachers' intentions and implementations regarding formative
16 assessment, *Assess. Educ. Princ. Policy Pract.*, **28**, 228-260.
17
18 Yik B. J., Raker J. R., Apkarian N., Stains M., Henderson C., Dancy M. H. and Johnson E.,
19 (2022), Association of malleable factors with adoption of research-based instructional
20 strategies in introductory chemistry, mathematics, and physics, *Front. Educ.*, **7**.
21
22 Zeng W. J., Huang F. Q., Yu L. and Chen S. Y., (2018), Towards a learning-oriented assessment
23 to improve students' learning-a critical review of literature, *Educ. Assess. Eval. Account.*,
24 **30**, 211-250.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix 1: Interview Protocol

Course information and teaching experience:

1. How long have you been teaching Gen Chem?
2. What is the structure of the course? (lecture, lecture + lab)
3. How many students are enrolled in this course?
4. What is the classroom layout? (amphitheater, round table, moveable desks)
5. Do you have access to technologies and tools that could be used in the class (e.g. Clickers, whiteboard)
6. Did you co-teach this course?
7. If YES: Do the instructors design the assessment together? Which one (homework exams, clicker questions)? How?
8. Can you walk me through a typical day in your class in general chemistry?

Conceptions of assessment purposes

9. What do you believe the assessment is?
10. What is the function of assessment in general chemistry?
11. Why do you use assessment in general chemistry?

Contextual factors:

12. Do you experience external influences that affect the way you assess your students in general chemistry courses? How?

Follow-up with whatever is not addressed in the answer:

13. **National level:** ACS exams
14. **Institutional level:** does the institution force you to assess in a particular way, to do a certain thing or force a certain learning goal? Do other departments at your institution influence assessment because many students from their departments are enrolled in your course?
15. **Student level:** Do you feel pressure from the students to assess in a particular way or do specific types of assessment?

Assessment Practices:

16. How do you assess your students in general chemistry?
 - 1) What assessments did you use? Or “based on the course artifacts that you provided, here are the assessments XXX that you mentioned, is that true?”
 - 2) When do you use each of these assessments?
 - 3) How long have you used each of these assessments in your course?
 - 4) Why did you decide to use them? (should go through each of them)
 - a) How do you design the assessment? What resource did you use to design/choose each of these assessments?
 - b) What criteria did you use when designing or writing a question for these assessments?
 - c) When do you prepare assignments?
 - 5) What does feedback each of the assessments provide you? To what extent do you use this feedback? (do it for each assessment)
 - 6) What feedback does each of the assessments provide students?
 - 7) How does each assessment count toward students’ grades? Why are you using this grading scheme?

Appendix 2: Codebook for assessment practices

Assessment practices	Definition	Participants, n	Classification
Cumulative Final	Exams that are provided at the end of the semester to measure students' learning of ALL the topics that have been taught during the course	18	AoL
Mid-terms	Exams that are provided in the midst of a semester to measure students' understanding of the course material	18	AoL
Post-class Assessment	Questions associated with tasks that students are asked to complete after class, which targets the topics that have been covered in the past classes.	18	AfL
In-class Assessment	Tasks that are assigned to students to work on during regular class time or group work sessions, including poll questions, worksheets, problems, and questions that instructors orally asked during group discussions or to the whole class.	16	AfL
Quizzes	Tests that are assigned to students prior to class, at the beginning of class or after class on a regular basis (e.g., each unit, each chapter, on class days, daily, weekly, biweekly, and etc.)	14	AfL
Pre-class Assessment	Questions associated with tasks that students are asked to complete before class (e.g., reading tasks, tutorial videos), which are expected to prepare them for the class content.	7	AfL
Reflective Assessment	Tasks that probe students to reflect on their own learning process (e.g., Muddiest point at the end of a class, reflections on corrections for exams or homework, surveys that allow students to self-assess their learning gains)	3	AaL
Pre-course Assessment	Tasks (e.g., questions and worksheets) that students are required to complete before the course to help them prepare for the course	2	AfL
Group Exams	Take home exams that students are tasked to complete collaboratively in groups	1	AoL

Appendix 3: Codebook for perceived pressure and external influences

Codes	Definitions	Participants, n
National level--- influence by ACS		
No influence or pressure	Instructors don't use ACS exam.	17
ACS certification's requirement of having the same questions on exams	ACS certification requires instructors to have same questions on exams to monitor growth	1
ACS exam format	Instructors are required to have an ACS exam so they would like to get students familiar with ACS exam format	1
Institution level		
No influence or pressure	Instructors do not feel pressure or feel no external influence by university or other departments	14
University's requirement of having same questions/topics or standardized tests	The university requires instructors to have same questions or questions covering same topics in exams, or have standardized tests	4
University's requirement of having final exams	The university requires instructors to have final exams	1
Department level		
No influence or pressure	Instructors do not feel pressure or feel no external influence by their home department	11
Pressure of aligning with departmental norms	Instructors feel pressure of fitting in an established department, where they have to make decisions on the assessment collectively; or feel pressure to make decisions regarding what to cover as they consider their performance compared to other professors as a reflection of their teaching ability.	6
Pressure of preparing students for future courses	Instructors feel pressure because they want students to be prepared before their future courses	3
Classroom level		
No influence or pressure	Instructors do not feel pressure or feel no external influence by students	15
Class size	The class size prevents instructors from using the type of assessment they prefer	2
Pressure of having students' negative feedback on workload	Instructors feel pressure due to students' negative feedback regarding the number of assignments or quizzes in the course	2
Cultural pressure		
Assigning grade is mandatory	Instructors feel mandatory to assign students grades	6

Having done it this way in discipline or education	Instructors feel that they have to stick with the norms or tradition in chemistry courses or a broader field such as the education	5
--	--	---

Appendix 4: Codebook for conceptions of assessment purposes

Purposes of assessment	Definition	Participants, n	Classification
Opportunity for instructors to evaluate students' learning	Assessment provides opportunities for instructors to evaluate what knowledge and skills that the students have learned from the course, and which stages students are at in their learning	16	AoL
Probe students' preparation for next course or major	Assessment is a way for instructors to measure students' readiness for future courses, including upper-level chemistry courses (e.g., organic chemistry, physical chemistry) and courses that required by students' major (e.g., biology, engineering)	12	AoL
Opportunity for instructors to get feedback from students and modify their instruction	Assessment provides opportunities for instructors to reflect on their instructional practices, and identify ways to better help students to learn	6	AfL
Opportunity for students to get feedback on their understanding	Assessment provides opportunities for instructors to provide feedback on students' learning progress	5	AaL
To ensure students are studying and learning	Assessment is a way for instructors to ensure that students are learning the material along the semester and are keeping up with the progress of the course	4	Other
Opportunity for students to evaluate their learning	Assessment provides an opportunity for students to reflect on what they have learned from the course and what they still don't know about the course, and to improve metacognition	4	AaL
Probe students' ability to solve problems in new scenario	Assessment is a way for instructors to gauge if students can apply what they learned in a new scenario	2	AoL

Data availability

The authors confirm that the data supporting the findings of this study are available within the article. Raw data that support the findings of this study are not publicly available due to information that could compromise the privacy of research participants.