JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc

Chemical Education Research

Beyond Convenience: A Case and Method for Purposive Sampling in Chemistry Teacher Professional Development Research

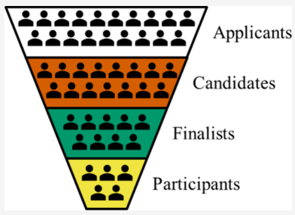
KatieMarie Q. Magnone and Ellen J. Yezierski*

Cite This: *J. Chem. Educ.* 2024, 101, 718–726

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: When designing a study, the sampling method for selecting research participants is an important decision with a host of considerations. When designing a professional development (PD) program with a limited number of spaces, the method of choosing participants from the applicants is also important. When research and professional development are entwined, sets of sampling criteria could conflict. Additionally, in a mixed methods study, such as the one undertaken by the researchers, additional trade-offs exist when considering the many potential methods of sampling and participant selection. In this report, we present a novel solution to the problem of multiple-criterion-focused selection of research and PD participants when the number of applicants to participate outnumbers the availability of resources. By using a weighted ranking system, we were able to incorporate multiple purposive sampling criteria in a simultaneous, rather than sequential, fashion. This allowed us to focus our evaluation of chemistry-specific free-response questions on a narrower pool of finalists and more consistently rate their responses. This multipurposive sampling method and a novel tool developed herein have broad implications for both sampling for chemistry education research and participant selection in professional development.



KEYWORDS: *High School/Introductory Chemistry, Graduate Education/Research, Continuing Education, Chemistry Education Research, Multimedia-Based Learning, Professional Development*

■ INTRODUCTION

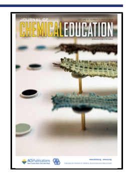
There are many broad types of sampling commonly used in research.[1−4] These range from intentionality-driven probabilistic and purposive sampling methods to availability-driven convenience sampling, with mixed method studies often employing a combination of these techniques. Probabilistic sampling methods include, but are not limited to, simple random, multistage random, systematic, stratified, and cluster sampling techniques.[2] Examples of such techniques are shown in Figure 1a. The purpose of probabilistic sampling methods is to remove as much researcher intent and bias as possible. This is contrasted with nonprobabilistic sampling methods, which are often driven by research intent. The researcher plays an active role in the selection and recruitment of participants. A nonexhaustive list of nonprobabilistic sampling methods used in qualitative and mixed methods research includes volunteer, convenience, snowball and a myriad of purposeful sampling techniques, such as maximum variation (heterogeneous case), homogeneous case, critical case, and typical case.[2,4] Examples of some of these techniques are shown in Figure 1b.

Descriptions of the techniques shown in Figures 1a and 1b are readily available in the literature. To focus on the techniques used in this report, three sampling techniques are defined here: volunteer, purposive, and quota. Volunteer sampling involves potential participants responding to a call for participants, and the researcher accepting all qualified volunteers into the sample. Purposive sampling involves intentionally selecting participants based on some predefined criteria. The two most common types of purposive sampling are heterogeneous case (also known as maximum variation) wherein the researcher attempts to select sample participants from a population to maximize their differences in some trait, and homogeneous case, wherein the research attempts to select for minimal variation in some trait among the population. There are many other kinds of purposive sampling; however, one type will be discussed in this paper (multi purposive sampling). Finally, quota sampling is used to ensure a minimum (or maximum, in this case) number of sample participants have a specific trait.

There are no universal guidelines to follow when it comes to determining a sampling technique to use for any given study; the technique(s) employed must be driven by the study itself, with the study employing one or more of the aforementioned techniques.[5] For example, a study whose intended population is first-year general chemistry students may use convenience sampling to determine the universities where the study will be undertaken and then use volunteer sampling to recruit students from the relevant courses to participate.

As Lawrie points out in a 2021 editorial published in *Chemistry Education Research and Practice* (*CERP*), "exemplars

Received: March 17, 2023
Revised: September 26, 2023
Accepted: January 29, 2024
Published: February 16, 2024



https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726

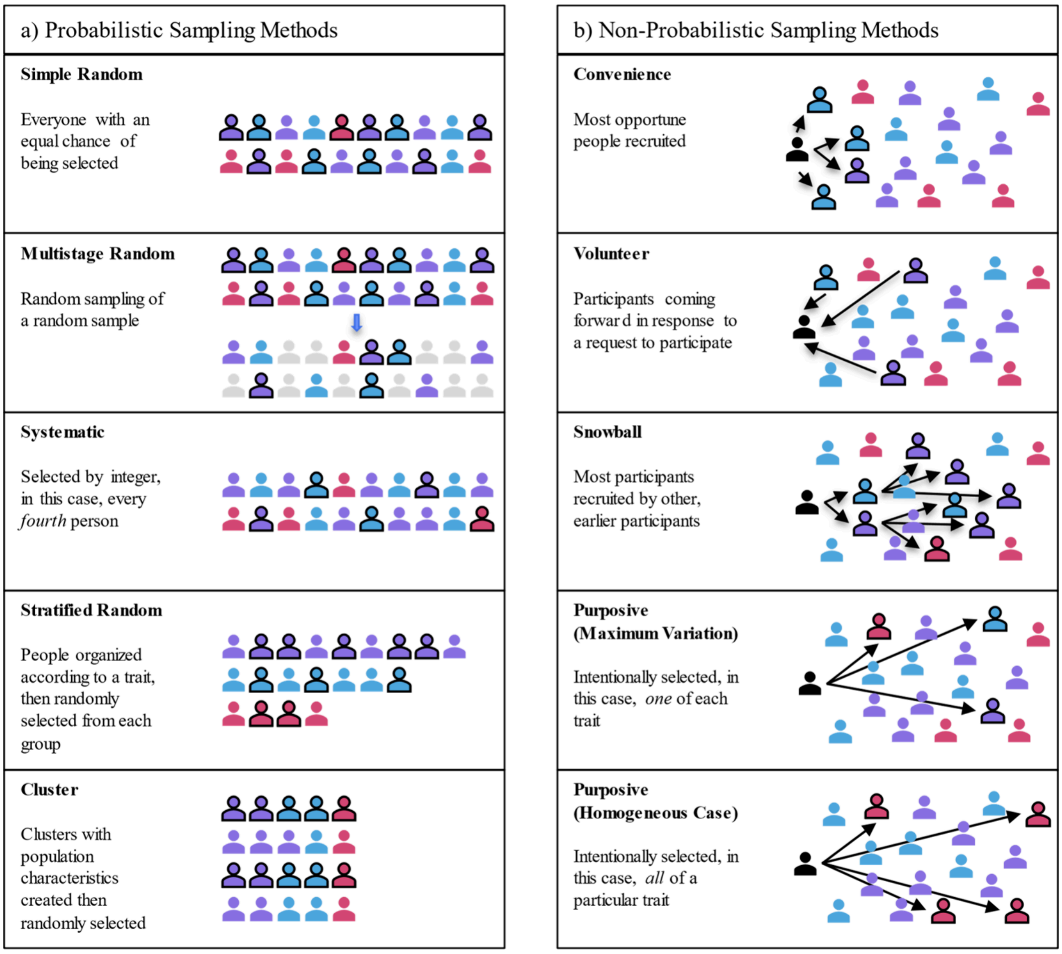


Figure 1. Examples of multiple types of sampling methods. Participants are shown with a black outline. a) Probabilistic sampling methods. b) Nonprobabilistic sampling methods, with the researcher shown in black. Arrows are used to show whether a subject was recruited (arrow emanating from the research/recruiter) or volunteered (arrow terminating at researcher).

from [a researcher's] own field" may provide the best insight into the sampling methods that should be employed in a given study, based upon the study context or research question.[6] As chemistry education researchers, we then look to the two preeminent journals that publish chemistry education research (CER) for English speakers, namely, this journal (*JCE*) and *CERP*. In the preparation of this report, we desired to examine a sample of recent literature (that is, all 95 CER papers published in these two journals in 2021) to see what kinds of sampling methods were employed. The goal of this review was to get a coarse idea of how researchers in the field describe their sampling methods rather than providing a comprehensive overview (requiring sampling from many more years of papers). In most of the sampled papers, the method of sampling was not explicitly named (63% of papers). In the remaining cases, the sampling method was volunteer sampling (17%), unspecified purposive sampling (7%), convenience sampling (5%), or whole group sampling (7%). Only one paper in the sample from either journal named and used a probabilistic sampling method. We present this lack of specificity in naming sampling methods as a discrepancy to be addressed by researchers in our writing: the method of sampling has potential implications for the validity of claims that are made from the results of these studies, and therefore, it is the responsibility of authors to contextualize findings by explicitly stating our sampling methods. Jeffrey Raker and colleagues carefully attend to defining the population along with describing the sampling methods used in their work:[7,8] however, this is rare in CER. Of the studies that we have identified on the topic, very few go into detail about their sampling protocols with even fewer papers methodologically focused on sampling in CER, as there are in other disciplines.[6,9−11]

The sampling methods employed in our study and professional development program are the basis for this report due to the unique nature of performing research during and after professional development for secondary chemistry teachers. The considerations that played a role in our purposeful sampling method are outlined ahead.

First, the broader societal impact goals of the professional development required that we attend to the economic needs of the student populations served by our potential participant teachers as well as the degree of historic marginalization of their student populations. Research shows that sustained PD in science education focused on concept learning, unsurprisingly, improves student achievement.[12] It is important to the research team, then, that the resources we made available through the PD be targeted at schools that have been historically under-resourced. This priority also aligns well with the broader impact merit criterion of the project's funder, the National Science Foundation.

Second, from both a researcher and PD developer perspective, we are interested in targeting teachers who are ideally positioned

https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726

for change in their teaching practices. Research has shown that many teachers leave the profession in the first three years of teaching, while others in this time period are "more concerned with dealing with, and surviving through, their first years" than with reflective teaching practices.[13] At the other end of the spectrum of experience, we likewise did not want to select teachers too close to retirement minima to maximize the impact of the PD during the subsequent years of teacher practice. Finally, we wanted to maintain some degree of representativeness of our population of applicants in our sample of participants. Additionally, as a goal of both research and PD, we wanted to maximize the number of students affected by implementation.

The multiple criteria that we desired to use to sample from our population of applicants put us in a difficult position: one that we have not seen reported in CER or other educational research literature. How does a team purposively sample, simultaneously and not sequentially, during participant selection? That is, how does a team consider and evaluate multiple criteria at once rather than sorting by a criterion and selecting the top few, then sorting those few by the second criterion, and so forth? This report answers that question by showcasing a novel sampling approach and evaluating its efficacy.

The VisChem Project

The VisChem project encompasses both a PD program for high school chemistry teachers, referred to as the VisChem Institute (VCI), and two "laboratories" for chemistry education study where data are collected, namely at the VCI and in teachers' classrooms following the VCI. The goals of the project are broadly to understand how undertaking the "VisChem Approach," as taught at the VCI, affects teacher and student understanding of molecular-level chemistry phenomena. Participating in the VCI was contingent on teachers applying to the program, which was advertised nationally through *ChemEdX*, the Target Inquiry listserv, and on social media.

In anticipation of more interested teachers than available spaces, we intended to pursue purposive sampling rather than convenience sampling, which is more frequently used in education research. Our priorities in participant selection were centered around the broader impact goals of the project that are focused on equity. These priorities included providing opportunities to teachers at minority serving and economically disadvantaged schools across the country as well as teachers who do not have as many opportunities for specialized PD. We also wanted a range of experience levels in teaching and the use of visualization in the classroom. These criteria influenced the application questions as well as the planned selection processes.

As of December 2019 when the first VCI opened for applications, the plan was to host 16 teachers at Miami University for a four-day, residential, intensive PD where they would learn about the VisChem Approach and how to implement it in their classrooms, as well as the general challenges associated with teaching about the molecular level in high school chemistry classrooms. We expected 50 or so applicants who could be screened for eligibility down to a more manageable number. As of the closing of applications in February 2020, 166 applications had been received. This large number of applications, as well as the intended short turnaround time to notify selected participants (planned for one month later, March 2020) necessitated a more efficient way to screen and sort candidates according to selection criteria. Additionally, when the decision was made to shift to a remotely delivered VCI due to COVID-19, applicants had to be rescreened for their willingness to participate in a remote, rather than in-person, PD.[14] This report details the selection process that was developed for the first VCI (2020, "Cohort 1") and subsequently modified and applied to the second VCI (2021, "Cohort 2"). This work has implications for sampling considerations in CER and provides a structure that can be applied in a variety of studies in chemistry education.

AIMS

The aims of this portion of our work were as follows:

1. Develop a ranked system that considers several criteria for participant selection.
2. Use this system to select participants and to allow researchers to rapidly exchange for alternates as participants dropped out or failed to complete the preinstitute work required of them.

METHODS

The sampling scheme developed for this project encompasses multiple sampling processes, including volunteer sampling, purposive sampling, and, to a degree, quota sampling. We outline how each was used in various stages of the selection process below.

First, volunteer sampling was used in the form of teachers opting to apply to the VCI. A summary table of the application questions from the 2020 and 2021 application years is shown in the Supporting Information in Table S1. The 2020 application included questions and a supplemental questionnaire that requested the number of students taught in each of several areas and teacher interest in a remotely delivered VCI. Certain questions did not elicit the intended responses in the 2020 application year and were modified or removed from the 2021 application. Additionally, a question designed to probe the student-centeredness of teachers' practices was added in 2021. After compiling the complete applications each year, a three-stage selection process was undertaken to screen applicants for participation. This purposive sampling process is outlined in Figure 2.

Applicants to Candidates

To advance from the applicant stage to the candidate stage, applicants had to meet three eligibility criteria: 1) teach at least some chemistry students at a public high school in the United States (determined through the NCES database); 2) have between 3 and 20 years of teaching experience (identified by the research team as the range for which pedagogical change is likely to occur); and 3) be willing to participate in a remotely delivered VCI. Screening for these criteria narrowed the pool to 70 and 94 candidates in 2020 and 2021, respectively. Most applicants who failed to advance did not meet the first (not teaching high school chemistry at a public high school in the United States) or third (unwilling to participate in a remotely delivered PD) criteria. A description of the shift to a remotely delivered PD program is detailed in Wu et al., 2021.[14]

Candidates to Finalists

To narrow the pool of candidates into finalists, a weighted ranking system was developed to sort candidates according to certain desirable traits, as determined by the broader research and societal impact goals of the project. Using data from the National Center for Education Statistics (NCES), candidates' school demographic information was evaluated. Racial/ethnic

https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726

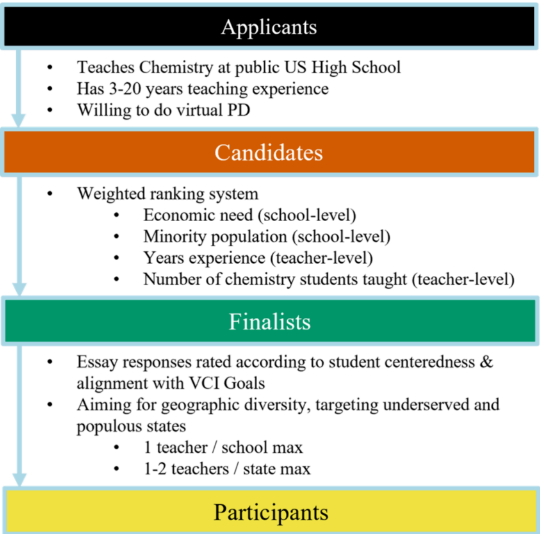


Figure 2. Process for selection of participants from the applicant pool. This three-stage process includes initial screening of applicants for eligibility criteria to become candidates, a weighted ranking system to advance eligible candidates to finalists, and analysis of free-response questions for eligible finalists to be selected as participants.

breakdowns and free and reduced lunch eligibility were used to assign points to each candidate. In addition to the NCES data, years of teaching experience and the number of chemistry students taught were also used to assign points. These are henceforth termed the weighted variables.

The specific manner of assigning points was a multistage approach, seen in Figure 3. First, the ranges of continuous values for each weighted variable were binned into brackets (e.g., similar to age or tax brackets elsewhere in the literature). These were preliminary brackets, as they might have changed with the weighting process. Second, a score was assigned to each backet based on its desirability for the PD. For example, while all teachers *must* have between 3 and 20 years of teaching experience, we had a slight preference for teachers with 6−15 years of teaching experience as more ideal ranges for pedagogical change. As such, scores assigned to those brackets are slightly higher than to the 3−6 and >15 brackets. These were also preliminary scores, as they could change during the weighting process. Third, the scores were converted to normalized points such that the sum of all the points for a variable is one. This was to allow us to highly vary the points within a weighted variable without it dramatically affecting the between-criteria weighting. For example, the scores for minority serving percentage in 2020 ranged from 0 to 5, while the scores for teacher experience ranged from 0 to 2. By converting to a normalized point system, we could mitigate the effects of having such a high score assigned for 100% minority serving compared to, for example, 6 years of teaching experience (a score of 5 compared to a score of 2 but both values in the ideal ranges). Because these points were based on a calculation of score (eq 1) for each weighted variable, they were not seen as weighting variables but only the result of the score variable. Finally, the score points were multiplied by the between-group weights (highlighted in yellow in Figure 3) to produce the weighted score (eq 2) that could be assigned to a candidate based on their specific values. These between-category weights were considered variables in the weighting process, as well. In summary, three factors were available to be modified in the weighting process: 1) the ranges of the brackets,

Economic Need Weight: 3

Brackets (F&R Lunch %)	Score	Score Points	Weighted Score
0 to 19.9%	0	0.000	0.000
20 to 39.9%	1	0.067	0.200
40 to 59.9%	2	0.133	0.400
60 to 79.9%	3	0.200	0.600
80 to 99.9%	4	0.267	0.800
100%	5	0.333	1.000

Minority Serving Weight: 4

Brackets (Minority %)	Score	Score Points	Weighted Score
0 to 19.9%	0	0.000	0.000
20 to 39.9%	1	0.067	0.267
40 to 59.9%	2	0.133	0.533
60 to 79.9%	3	0.200	0.800
80 to 99.9%	4	0.267	1.067
100%	5	0.333	1.333

Teacher Experience Weight: 2

Brackets (Years Experience)	Score	Score Points	Weighted Score
3 to 5.9	1	0.125	0.250
6 to 8.9	2	0.250	0.500
9 to 11.9	2	0.250	0.500
12 to 14.9	2	0.250	0.500
15 to 17.9	1	0.125	0.250
18 to 20	0	0.000	0.000

Chemistry Students Weight 3

Brackets (Number of Students)	Score	Score Points	Weighted Score
0 to 24	0	0.000	0.000
25 to 49	0	0.000	0.000
50 to 74	1	0.100	0.300
75 to 99	2	0.200	0.600
100 to 149	3	0.300	0.900
150+	4	0.400	1.200

Figure 3. 2020 weighted scoring brackets for all four weighted variables. The within-group weighting is determined by the value of the score (higher represents a more desirable bracket value) and converted to normalized score points (that sum to 1 for each variable) before being multiplied by the between-group weight (highlighted in yellow) to produce a weighted score for each bracket.

2) the score associated with each bracket, and 3) the overall category weight. Score points (eq 1) and weighted score (eq 2) are simply the result of the computations of the values associated within and between weighted variables. An example of an alternative study using different variables is given in the Supporting Information, where these values can all be manipulated to see the resulting effects on finalist distributions.

$$\text{Score Points} = \frac{\text{Score}}{\Sigma \text{Within Variable Scores}} \quad (1)$$

$$\text{Weighted Score} = \text{Score Points} \times \text{Variable Weight} \quad (2)$$

To determine the values to assign to the brackets, scores, and variable weights, multiple variables were monitored as we varied these values. In addition to the four weighted variables, the other monitored variables included school state, school locale, and school size. Additionally, responses to two of the most salient free response questions (3 and 6) were rated on a 0−2 scale by one (2020) and two (2021) members of the research team based on student-centeredness and applicability of response. Distributions of the free-response scores were also monitored with the others.

Varying the values for the brackets, scores, and weights was also a multistep process. First, we set up the brackets for each criterion and assigned preliminary scores to them. At this stage,

https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726

all criteria weights remained 1. The first objective was to examine the skew of each variable, with left skewed data (with the mean value to the left of the mode value, meaning there are more individuals on the right-hand side of the graph) being heavily favored for economic need % and minority serving %. We examined the distributions and determined that the distributions of economic need % and minority serving % were still heavily right skewed (with more potential finalists toward the lower end of the values and fewer trailing toward the top). Consequently, we increased the weight of those weighted variables and reexamined the distributions of the monitored variables. We repeated this process iteratively until the research team was satisfied with the distributions across the monitored variables. This process involved making preference decisions. For example, we had to determine which was more preferable: more high economic needs schools or more variability in geography and locale; more teachers in the ideal experience range, or more student-centered responses to free response questions. There were trade-offs in the process that forced us to consider competing interests of closing equity gaps for the PD while assuring maximum variation for research. Simultaneously optimizing all selection criteria such that all monitored variables were "ideal" was impossible; ultimately, we had to prioritize. For the most part, the intravariable values (brackets and scores) did not significantly change from their starting values, except for years of experience, which saw score adjustments from 0, 1, 2, 1, 0, 0 (refer to the brackets of teacher experience from Figure 3) to the final scores of 1, 2, 2, 2, 1, 0 (Figure 2). Most of the changes occurred in the intercriteria weights.

Once favorable distributions (idiosyncratic to our values as a research and professional development team) were established in each monitored variable, the top ranked candidates were evaluated as finalists. To select a pool of about 30 finalists to examine, the point totals were rank-ordered from the highest to lowest and examined for a break in points. In 2020, 34 finalists were selected; in 2021, this increased to 40 due to the natural point break values to ensure that arbitrary breaks did not govern finalist choices. An example of the small point break is shown in Figure 4. In the first year of the project (2020), the initial monitoring process was done for a static "Top 20," assuming that these 20 would likely be our 20 participants. However, more

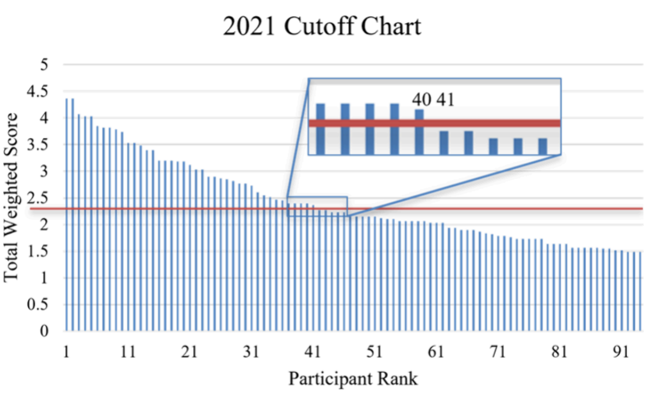


Figure 4. 2021 finalist cutoff chart used to determine that the finalist cutoff would be 40. After the 40th ranked candidate, there is a slight drop in score. The potential other location for this cutoff would have been after the 31st ranked candidate, but we chose to increase our alternate pool based on difficulties selecting qualified alternates in the prior year. A callout shows this drop in scores between the 40th and 41st ranked candidate.

finalists at the "lower" end of the list turned out to have excellent free responses, and some toward the top of the list dropped out or stopped responding to emails. In 2021, we drew from the 2020 experience, and we kept a closer eye on the point breaks at each variation. Because the data were nearly continuous in 2021, the point break is small and was calculated as the difference in the n^{th} and n^{th} + 1 candidate score. The largest two such point breaks occurred at 31 and 40 candidates. Due to the difficulties experienced in 2020 (teachers' schedules and professional lives were far less predictable for them), we decided that 40 would be a more optimal number of finalists to visualize at the last variation in weighted scores.

Finalists to Participants

To be selected as one of the 20 participants (up from 16 as originally planned, as remote PD costs were lower and allowed for a slightly larger cohort to be accepted), finalists were evaluated according to their specific responses to the free response questions. At this stage, all research team members were involved in the review of responses and came to agreement upon selections. In response to question 3, which asked about approaches in the classroom, student-centered approaches were favored, as these were aligned with a quality implementation of the VisChem Approach. In response to question 4, which probed the experience levels of applicants with various molecular level visualizations, a mix of experiences was desirable, as teachers more experienced with molecular-level visualizations would be a benefit to others who were less familiar with molecular-level visualizations. In response to question 6, which inquired about goals for attending the VCI, stated goals that aligned with the learning goals of the VCI were favored.

The chemistry specificity of this sampling method heavily relies on analysis of the free response questions in the final stage. In particular, the specific pedagogies in chemistry evidenced in questions 3 and 4 helped us to determine who to accept to the VCI. Some finalist teachers already had a high degree of both knowledge and usage of the molecular level in their classrooms. Other teachers expressed interest in pursuing professional development on teaching approaches centered on the molecular level because they lacked specific pedagogies to incorporate it. The former group likely would not benefit as much from the VCI as the latter. As teaching approaches focused on the molecular level are key to understanding many chemical phenomena, we aimed to develop teachers who would derive the most benefit from molecular-level focused PD. While we do not have IRB approval to share examples of responses to the application questions, representative composites of how we would have analyzed responses is in the Supporting Information in Table S2.

While participants were selected, the distributions of prior variables were continuously monitored. An effort was made to select teachers from a wide variety of geographic locations, with no more than two teachers from any given state and no teachers from duplicate schools. This was the quota sampling portion of the process. Medians for each of the four weighted variables at each stage were also calculated. Once participants were selected and notified, the remaining finalists were accepted as alternates and notified so they could also complete the VCI prework. If replacements needed to be made for any reason, the same process for selecting a participant was followed, but participants were selected from the pool of alternates who completed the prework, while monitoring for the weighted variable distributions and medians.

https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726

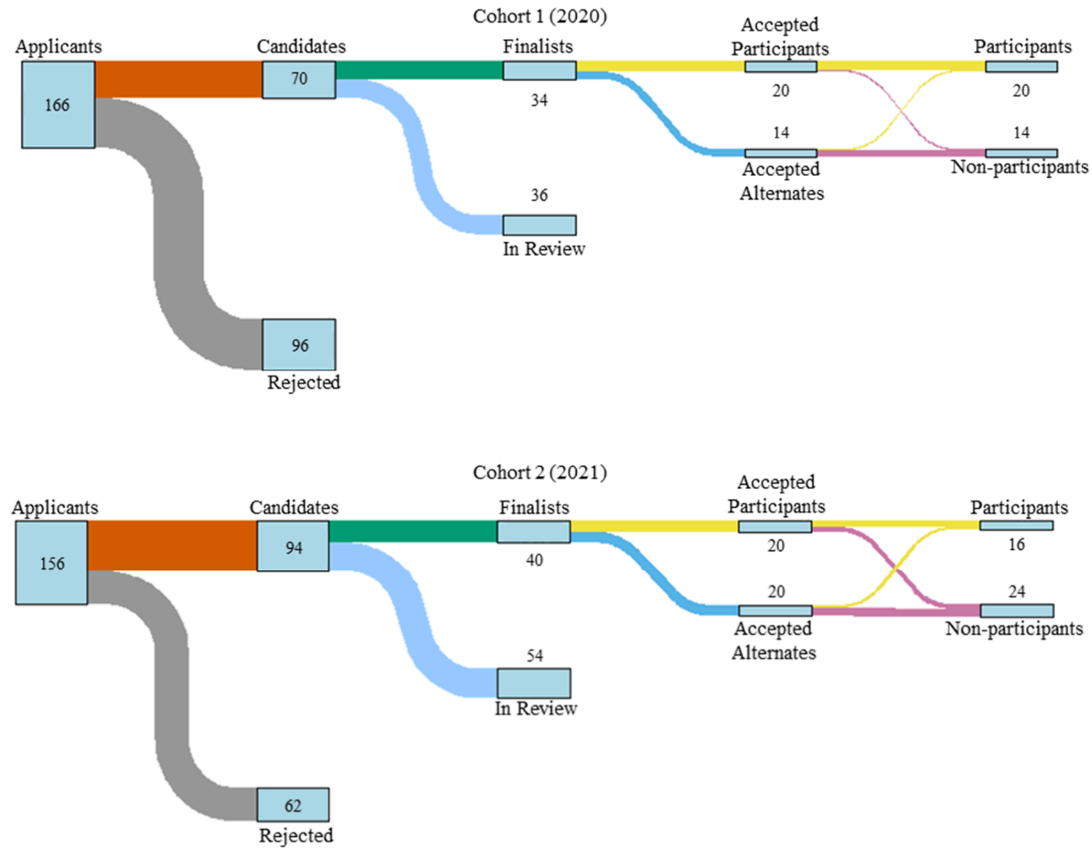


Figure 5. Sankey diagrams showing teacher flow through the selection process for each of the two cohorts.

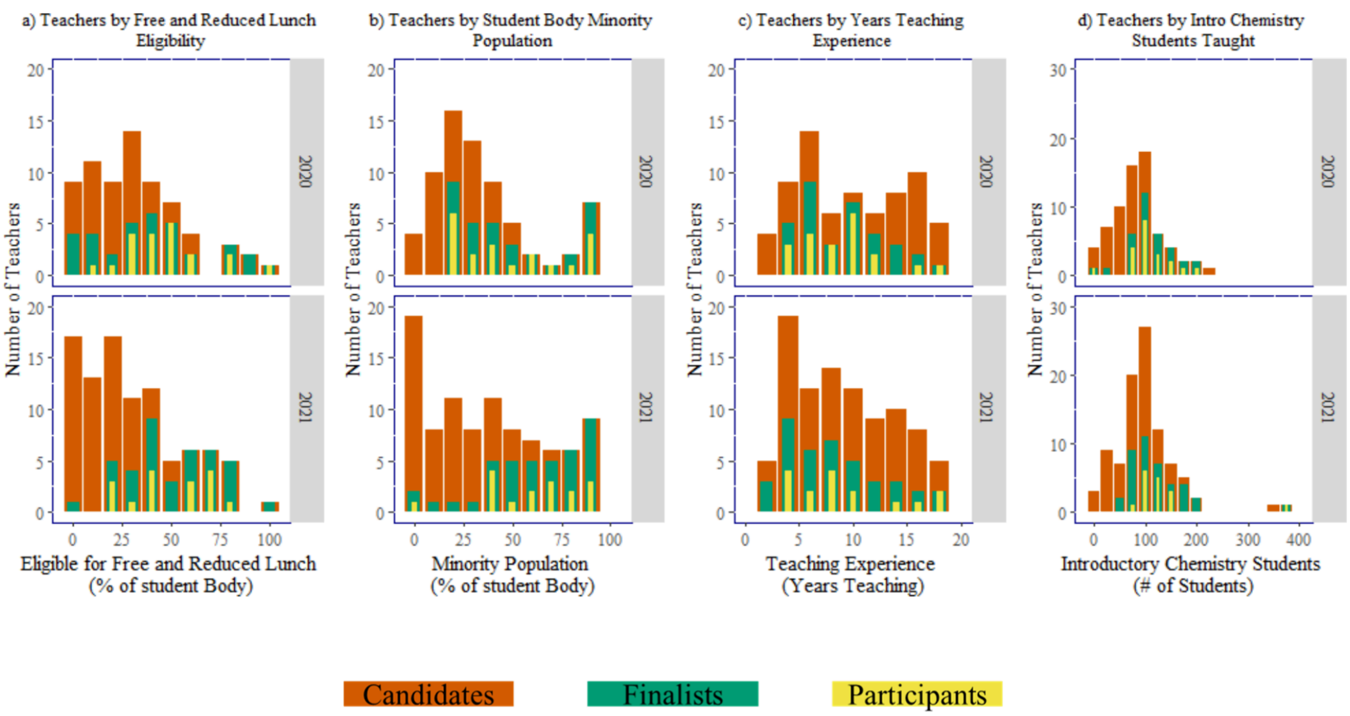


Figure 6. Monitored variables through each stage of the selection process from Candidate (orange) to Finalist (green) to Participant (yellow). Monitored variables include a) economic need; b) minority population; c) years of teaching experience; and d) number of introductory chemistry students taught.

All processing took place using Microsoft Excel as a robust tool for performing calculations, reading free response answers, and visualizing distributions using one software package.

https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726

Table 1. Medians of Monitored Variable at Each Stage of the Selection Process

	Cohort 1			Cohort 2		
Variable	Candidates	Finalists	Participants	Candidates	Finalists	Participants
Economic Need (%)	32.9	42.6	48.4	28.6	52.3	56.2
Minority (%)	33.0	49.7	47.5[a]	40.3	69.1	67.8[a]
Years of Experience	10	9.5	9.5	9	8	9
Chemistry Students	90	120	120	100	120	128

[a]Indicates a decrease from finalist to participant stage.

RESULTS AND DISCUSSION

Ultimately, 20 teachers were selected for each cohort, with 14 and 20 alternates in 2020 and 2021, respectively. The large number of potential alternates allowed us a great deal of flexibility in the case of a teacher being unable or unwilling to complete the VCI or the necessary and required preparation work. Applicant flow diagrams shown in Figure 5 summarize the number of teachers in each phase of the selection process. In 2020, a total of 3 replacements were necessary. In 2021, 5 replacements were made, though a total of 4 additional teachers dropped out right before the PD and were unable to be replaced. In both years, either shifting plans due to the pandemic or inability to complete the prework in the classroom before the end of the school year was the primary driving force behind replacements. Through the replacement process, 20 teachers were selected and satisfied the requirements for cohort 1 (2020), and 16 were selected and completed the necessary prework for cohort 2 (2021). For the remainder of this report, these actual participants, rather than those originally selected but who dropped out or failed to complete the prework, are known as the participants.

Weighted Variables

Economic Need, as determined by the percentage of the student body population eligible for the federal free and reduced lunch program, was monitored during variable weighting. Monitoring focused on the goal of shifting a right-skewed distribution (favoring lower values) to a more normal or left-skewed distribution (favoring higher values). Eligibility for free and reduced lunch was chosen as a weighted variable due to its freely accessible nature through the NCES database and due to the metric's correlation with community economic status, which is likely to impact the availability of school funding. A summary of the distributions at each of the selection stages is shown in Figure 6a.

Minority Serving Status, as determined by the percentage of the student body population that does not identify as non-Hispanic white, was monitored during variable weighting. Monitoring focused on the goal of shifting a right-skewed distribution to a more normal or left-skewed distribution. Minority, using this definition, includes American Indian, Alaskan Native, Asian, Black, Hispanic, Native Hawaiian, Pacific Islander, and mixed-race students. A summary of the distributions at each of the selection stages is shown in Figure 6b.

Years of Teaching Experience, as self-reported by the applicants, was used as a weighting variable for two reasons: (1) to ensure an adequate distribution of experience levels approximate to candidates' demographics and (2) to target teachers in our ideal teacher experience range (targeting midcareer teachers with 6−14 years teaching experience). A summary of the distributions at each of the selection stages is shown in Figure 6c.

Number of Introductory Chemistry Students Taught, as self-reported by the applicants, was the final weighted variable. The goal was to prioritize teachers with higher numbers of students to potentially impact our PD. A summary of the distributions at each of the selection stages is shown in Figure 6d.

Weighting and Outcomes

While the intravariable brackets and scores remained the same from cohort 1 to cohort 2, the intervariable weights changed. For cohort 1, the weights assigned to each variable were 3 (economic need), 4 (minority serving), 2 (years teaching experience), and 3 (number of chemistry students), while in cohort 2 the weights were 5, 3, 2, and 3, respectively. These changes were made not only to reflect the changing demographics of the pool of candidates from cohort to cohort but also to prioritize participation from teachers from low-income schools in 2021 compared to 2020. Using these weightings, we were still able to acquire finalists and participants from low income and high minority serving schools in both years. The weightings from years of teaching experience and number of chemistry students remained the same and allowed us to acquire teachers in the middle career range with primarily 100 or more chemistry students taught in a year. After the monitored variable distributions were deemed adequate for the pool of finalists, analysis of free response questions allowed us to select participants with ideal student-centered characteristics. Because of the finalists' distributions being monitored, we were able to make replacements as necessary without altering the demographic breakdowns of the group of selected teachers too substantially.

Median values for each of the four monitored variables are shown in Table 1 for each stage of the selection process. In both years, the median percent minority population decreased from the finalist stage to the participant stage but both still retained a marked improvement over the median at the candidate stage. It is important to note that for each of these cohorts the weightings and corresponding distributions are relative to the starting distributions of the candidates. The shape of each variable's distribution that we aimed for in cohort 2's selection process was therefore vastly different from cohort 1. As an example, minority serving percent has a substantially different shape at the candidate stage between cohorts and therefore very different distributions at the finalist stage. The accomplished goal was not to replicate the distributions from one cohort to the next but rather to give preference to particular demographics relative to the starting population.

After we completed the second sampling iteration, we modeled how the distributions would have looked had we chosen a different sampling method. We chose to model a random sample and purposive sampling solely on the basis of economic need or minority serving status. Single criterion quota sampling was rejected, as our goals required the consideration of multiple criteria. In the random model, finalists were selected by

https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726

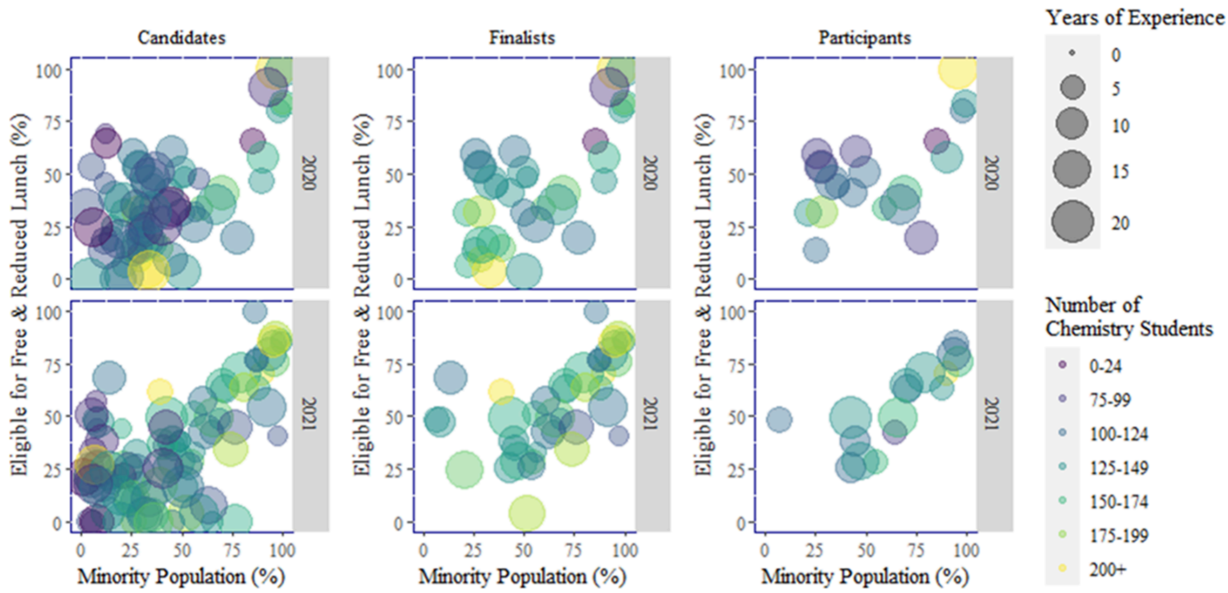


Figure 7. Bubble plot showing economic need by minority serving status, along with years of experience indicated by bubble size and number of chemistry students taught indicated by bubble color for each stage of the selection process.

rank ordering the candidates according to a random number assigned to each based on Excel's RAND function that randomly assigns a value between 0 and 1. The number of finalists to select became a conundrum, as the nature of the RAND function meant that finding a natural point break was difficult. We therefore selected 40, as was the case in 2021. The distributions of the monitored variables closely aligned with the candidate distributions and thus did not satisfy our sampling priorities. We therefore concluded that a random selection of finalists or participants would not have been appropriate for our PD goals. For the two purposive models, each based on one of our weighted variables, we rank ordered the candidates based on the variable of interest. For each of the two variables in each of two years, we observed a more dramatic shift to a left skewed distribution for the selected variable, but the other distributions did not see similarly favorable shifts. The comparative failures of these sampling models are encouraging for the success of our multipurposive sampling method.

Using a multipurposive sampling method to select finalists may not yield the "highest possible scoring" in any one variable, but it yields high scoring across the multiple variables. For our purposes, this outcome was deemed adequate. In other research settings or contexts in which multiple variables associated with participants are required to select research participants, a similar weighted ranking system (substituting the appropriate variables, brackets, scores, and weights) could be used to rank candidates for selection. See Implications for a discussion of applications in other research settings.

Limitations

The primary limitation of this work lies in the fact that while all variables were monitored simultaneously, they were monitored separately (i.e., distributions were monitored, but variables were not plotted versus one another). It would therefore have been possible to obtain school characteristics at opposite extremes of our two broader impact needs criteria (e.g., a school with a very high economic need but low minority serving). As a posthoc analysis, we plotted these two variables against each other to ensure that we had acquired teachers at schools that met both needs criteria. This is shown, alongside the other two monitored variables (Students Taught and Years of Experience) in Figure 7. If this method of sampling is to be used in the future, we recommend not only plotting the current distributions by variable but also plotting the variables versus one another. This will allow the researcher to more actively select for participants who meet all of their criteria rather than just one criterion.

■ CONCLUSIONS

Using a tiered, multipurposive sampling method, we were able to develop a ranked system that uses several criteria simultaneously in participant selection. This method began with volunteer sampling by having teachers apply to participate in the VCI. Upon receiving a larger than expected number of teacher applicants, a ranked system was used to purposively order candidates and select finalists. From the finalists, participants were selected according to maximum quota sampling and analysis of free-response questions. Using this sampling method allowed us to not only select the original participants but also make rapid replacements from the remaining alternates when participants dropped out or failed to complete the pre-VCI work required of them. We have shown that using this approach it is possible to select participants by applying multiple criteria simultaneously. Where a purely research driven approach might have resulted in the use of simple random or even stratified sampling, we have shown the value of a tiered, multipurposive method for meeting both our research goals and our broader societal impact goals.

Implications for Research and Professional Development

The method reported here has broad implications for both research and professional development. First, as researchers, we have shown that it is possible to systematically target a sample based on multiple criteria and still maintain a degree of diversity in the sample. This has broader implications when considering diversifying the body of Diversity, Equity, Inclusion, and Respect (DEIR) work in the realm of chemistry education. We give an example of how this could be applied to alternative research contexts. In cases where purposeful sampling techniques are employed, we hope that researchers will feel empowered to incorporate multiple criteria if it serves the research design. For example, in instances where two or more diverse variables are of interest to selection, this multipurposive weighted ranking system could be of great use.

We have provided one such example in the Supporting Information. The idiosyncrasies of this method for our work lie only in the nature of the variables and values and not in the method itself. For example, in a study of samples from undergraduate chemistry students who complete a survey before selection, researchers could implement this three-stage process to select participants if the number of volunteers outnumbers the number of available interview slots. The first stage, in the hypothetical example, limits the study to students who have undergone a specific curricular change at a sample university. Then, the weighted variables could include demographic variables, the semester in school, and scores on various given instruments. We recommend first setting up the intravariable brackets and scores first, leaving all intervariable weights at 0. Researchers can then examine how each set of brackets and scores influences that variable's distribution and modify value until the changes are in the direction they desire. Having repeated this for each variable individually, set all intervariable weights to 1 and see what effect that has on all the distributions. The next stage will require modification of inter- or intravariable values to acquire satisfactory distributions and rankings, which are, again, idiosyncratic to the study at hand. In the example given in the Supporting Information, racial, ethnic, and gender minorities are given preference, as are those with scores on two instruments in a particular range. Finally, either the top number of finalists or some other criteria could be used to select participants at that point. The example given uses the natural break point method employed in our study and then purposive sampling from finalists. The strength of this approach lies primarily in the second stage that employs a multipurposive sampling method to rank candidates.

For professional development, we have shown that multiple key criteria can be targeted for societal impact if these variables are weighted and monitored during the selection process. In cases where the availability of participants is not the underlying limiting factor, we have demonstrated a technique that not only allowed us to select participants in an equitable fashion but also to make replacements in an efficient manner when originally selected participants dropped out of the VCI. In a multiple-year project, developing a system to screen applicants and select participants in the first year saved us time and effort in subsequent years.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.3c00217.

Application Questions (PDF, DOCX)

Free Response Composites (PDF, DOCX)

Example of Hypothetical Study Parameters and Weighting System (XLSX)

AUTHOR INFORMATION

Corresponding Author

Ellen J. Yezierski − *Department of Chemistry & Biochemistry, Miami University, Oxford, Ohio 45056, United States;* orcid.org/0000-0002-7067-7944; Email: yeziere@miamioh.edu

Author

KatieMarie Q. Magnone − *Department of Chemistry & Biochemistry, Miami University, Oxford, Ohio 45056, United States;* orcid.org/0000-0002-0015-9623

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.3c00217

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors deeply thank Gwendolyn Lawrie for raising the important issue of sampling within the chemistry education research community, particularly *via* her 2021 editorial[6] and its inspiration for this report's Figure 1. We also thank the members of the Yezierski and Bretz research groups at Miami University for their integral feedback during the methods development and reporting process. Particularly, we are grateful to Meng-Yang Matthew Wu for his involvement in gathering NCES data and analyzing the free response questions in 2021. This material is based upon work supported by the National Science Foundation under Grant DRL-1908121.

REFERENCES

(1) Teddlie, C.; Yu, F. Mixed Methods Sampling: A Typology With Examples. *J. Mix Methods Res.* **2007**, *1* (1), 77−100.

(2) Onwuegbuzie, A.; Collins, K. A Typology of Mixed Methods Sampling Designs in Social Science Research. *Qualitative Report* **2007**, *12* (2), 281−316.

(3) Taherdoost, H. Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *International Journal of Academic Research in Management* **2016**, *5* (2), 18−27.

(4) Creswell, J.; Poth, C. *Qualitative Inquiry and Research Design Choosing Among Five Approaches*, 4th ed.; SAGE Publications, Inc.: 2018.

(5) Tashakkori, A.; Creswell, J. W. The New Era of Mixed Methods. *J. Mix Methods Res.* **2007**, *1* (1), 3−7.

(6) Lawrie, G. Considerations of Sample Size in Chemistry Education Research: Numbers Do Count but Context Matters More! *Chemistry Education Research and Practice* **2021**, *22* (4), 809−812.

(7) Gibbons, R. E.; Reed, J. J.; Srinivasan, S.; Murphy, K. L.; Raker, J. R. Assessment Tools in Context : Results from a National Survey of Postsecondary Chemistry Faculty. *J. Chem. Educ.* **2022**, *99*, 2843−2852.

(8) Connor, M. C.; Pratt, J. M.; Raker, J. R. Goals for the Undergraduate Instructional Inorganic Chemistry Laboratory When Course-Based Undergraduate Research Experiences Are Implemented: A National Survey. *J. Chem. Educ.* **2022**, *99*, 4068.

(9) Cohen, R. D. How the Size of a Random Sample Affects How Accurately It Represents a Population. *J. Chem. Educ.* **1992**, *69* (3), 200−203.

(10) Etikan, I.; Musa, S. A.; Alkassim, R. S. Comparison of Convenience Sampling and Purposive Sampling Related Papers. *American Journal of Theoretical and Applied Statistics* **2016**, *5* (1), 1−4.

(11) Fernald, A. Getting beyond the "Convenience Sample" in Research on Early Cognitive Development. *Behavioral Brain Science* **2010**, *33* (2-3), 91−92.

(12) Johnson, C. C.; Kahle, J. B.; Fargo, J. D. A Study of the Effect of Sustained, Whole-School Professional Development on Student Achievement in Science. *J. Res. Sci. Teach* **2007**, *44* (6), 775−786.

(13) Southerland, S. A.; Sowell, S.; Enderle, P. Science Teachers' Pedagogical Discontentment: Its Sources and Potential for Change. *J. Sci. Teacher Educ* **2011**, *22* (5), 437−457.

(14) Wu, M. Y. M.; Magnone, K.; Tasker, R.; Yezierski, E. J. Remote Chemistry Teacher Professional Development Delivery: Enduring Lessons for Programmatic Redesign. *J. Chem. Educ.* **2021**, *98* (8), 2518−2526.

https://doi.org/10.1021/acs.jchemed.3c00217
J. Chem. Educ. 2024, 101, 718−726