JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

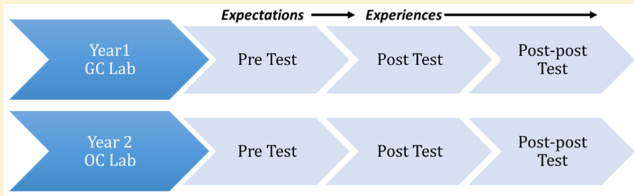
Measuring Meaningful Learning in the Undergraduate General Chemistry and Organic Chemistry Laboratories: A Longitudinal Study

Kelli R. Galloway and Stacey Lowery Bretz*

Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056, United States

Ⓢ *Supporting Information*

ABSTRACT: Understanding how students learn in the undergraduate chemistry teaching laboratory is an essential component to developing evidence-based laboratory curricula. The Meaningful Learning in the Laboratory Instrument (MLLI) was developed to measure students' cognitive and affective expectations and experiences for learning in the chemistry laboratory. Previous cross-sectional studies in both general and organic chemistry laboratory courses have shown trends where cognitive expectations go unfulfilled, and affective expectations and experiences are diverse among all students. On the basis of these previous findings, a longitudinal study was carried out to explore how students' ideas about laboratory learning change over two years of chemistry laboratory instruction. The data were analyzed using multiple visualizations, inferential statistics, and cluster analysis. Findings from this study supported previous findings from the cross-sectional studies. In addition, it was found that students "reset" their expectations for organic chemistry laboratory, meaning they indicated high expectations for learning despite unfulfilled expectations in their general chemistry laboratory. Further findings and their implications are discussed within the context of Novak's theory of meaningful learning.



KEYWORDS: *Chemistry Education Research, Testing/Assessment, Laboratory Instruction, First-Year Undergraduate/General, Second-Year Undergraduate, Organic Chemistry, Hands-On Learning/Manipulatives, Inquiry-Based/Discovery Learning, Learning Theories*

FEATURE: Chemical Education Research

■ INTRODUCTION

The Meaningful Learning in the Laboratory Instrument (MLLI) is a novel assessment tool designed to measure students' cognitive and affective expectations and experiences across a semester of an undergraduate chemistry laboratory courses.[1] Previously, the MLLI has been used to collect cross-sectional student data from general (GC) and organic chemistry (OC) laboratory courses both at a single university and across multiple universities.[1,2] Similar results were found both at the single institution and in the multiuniversity study, and the data from these studies showed interesting trends across a single semester of GC and OC laboratory courses. Namely, students have high cognitive expectations that go unmet from their laboratory experiences; students have disparate affective expectations and experiences where students change differently in their affective perceptions of learning across the semester.[1,2] Cluster analysis from both studies found student groupings in need of attention, especially two particular clusters: the students with low expectations and subsequent low experiences ("Low" cluster) and students with high expectations but unexpected low experiences ("Change" cluster).[2,3] The expectations for GC and OC students were similar for the cross-sectional study despite reporting experiences that did not meet those initial expectations. As both studies were cross-sectional (data were simultaneously collected from students in GC and OC), a question remained as to whether similar findings would exist in a truly longitudinal study, that is, by following the same students through GC and OC. In particular, the extent to which students' experiences in the second semester of GC lab aligned with their OC expectations was of interest, as was how their cognitive and affective perceptions of learning changed as they learned more chemistry and gained more experience in the undergraduate chemistry laboratory.

Longitudinal Studies

White and Arzi define a longitudinal study as "one in which two or more measures or observations of a comparable form are made of the same individuals or entities over a period of at least one year".[4] The focus of this definition is the length of time and measurements of the same nature. The suggested length of time for a longitudinal study is around one year due to the organization of the school schedule and to allow for enough time to see change and maturation.[4,5] Second, the observations should be comparable if not identical in nature. This is not to say that studies carried out over the course of a single semester with multiple measurements are not beneficial; those studies do have implications for teaching and research. The most important choice in designing a research study involves articulating a specific research question and choosing research methods that adequately align with the question. Longitudinal

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

studies do have unique challenges that can pose threats to the validity of the findings, including time, resources, and attrition.[4,5] When handled well, however, the benefits of longitudinal studies can outweigh the costs. As White and Arzi point out, "only longitudinal studies can show whether early change in learning or any other educational process is permanent or volatile, and whether it leads on to further development".[4]

By using this definition, few longitudinal studies have been documented in the science and chemistry education literature. The longitudinal studies in the literature can be categorized as measuring solely cognitive[6−13] or solely affective[14−17] variables, sampling adolescents[6−8,12,14] or undergraduate students,[9−11,13,15−17] and comparing an intervention[6,8−13] or exploring a phenomena.[7,14−17] Of this last category, White and Arzi call the former "experiments" as they seek to evaluate an intervention and the latter "descriptions" as they seek to gather information on "unfolding events or the development of knowledge or behavior".[4] With regard to the longitudinal studies that measure cognitive variables, some are experimental and others are descriptive. The descriptive studies tend to focus on conceptual change and how adolescents develop scientifically accurate ideas over time. For example, Øyehaug and Holt interviewed four middle school students over two years to learn how the students' ideas about matter and chemical reactions evolved over time.[12] The experimental studies measuring cognitive variables included the comparison of two teaching methods for middle school chemistry and evaluating the effectiveness of Peer-Led Team Learning across a year of GC.[6,11] These studies used final exams to assess the effectiveness of the reforms.

Several of the longitudinal studies measuring affective variables were descriptive in nature. In the social psychology literature, longitudinal studies tend to investigate motivation, self-esteem, or other affective variables from childhood through adolescence.[14] In science education, many studies focused on the development of underrepresented students in chemistry, including females and minorities, as these types of studies are useful in furthering research on retention in chemistry. Robnett et al. measured students' self-efficacy and identity as a scientist in relation to their research experience as an undergraduate across two years.[17] Findings from Robnett et al. showed that students' identities as scientists were predicted by their research experience and that the relationship between identity and research experience was mediated by science self-efficacy.[17] Marra et al. investigated women's experiences as engineering students.[15] The multiyear study measured self-efficacy twice over two years and found that some areas of self-efficacy increased while others decreased.[15] In chemistry education research, Villafañe et al. conducted a repeated measures study across one semester of chemistry.[16] While the study does not fit the White and Azri definition of longitudinal (because the length of the study was only one semester), Villafañe et al. did collect five measurements of identical nature throughout a single semester to measure how students' self-efficacy changes.[15] Villafañe et al. found a negative self-efficacy trend for Black and Hispanic males and a positive trend for Hispanic males with the implication to measure self-efficacy at different time points to be aware of different trends among different student groups.[16] The longitudinal design of these studies allowed for the observation patterns and trends that might not otherwise be observed during a one-time or cross-sectional study.

The literature on laboratory learning has continually called for rigorous research to understand how students are learning in the laboratory[18−23] and give evidence for the merit of the undergraduate chemistry laboratory.[24,25] While the research on laboratory learning has grown in the past few years, few longitudinal studies have been carried out in the undergraduate chemistry laboratory. Poock et al. followed 78 students across two semesters of general chemistry lab to analyze the implementation of the Science Writing Heurstic (SWH).[9] By using the total points earned in the course as the measure of academic performance, results showed that the SWH approach had a positive impact on student learning over the academic year and that a greater gain might be evident in the first semester than the second.[9] Szteinberg and Weaver conducted a longitudinal study where they followed students who had participated in a research-oriented general chemistry laboratory course to understand the effects two and three years later.[13] In comparison with the students in the more traditional laboratory setting, students who participated in the research-oriented course demonstrated a greater ability to remember and explain the work they did in their laboratory course.[13] Szteinberg and Weaver note the importance of the longitudinal study in uncovering the long-terms effects of pedagogical innovations.[13] While most call for longitudinal studies focus on topics such as conceptual change, transfer, persistence of misconceptions, and retention,[23] a study into how students' perceptions of learning change as they learn more chemistry and can offer unique evidence to support future design of laboratory curricula.

Meaningful Learning and Human Constructivism

The theoretical framework guiding the research design and analysis for this study was Novak's theory of meaningful learning and human constructivism.[26−28] Built upon Ausubel's theory, meaningful learning requires a learner to have relevant prior knowledge, for the new material to be presented in a meaningful way, and for the learner to choose to nonarbitrarily connect the new knowledge to the existing knowledge.[26,28,29] Ausubel contrasted meaningful learning with rote learning where the learner chooses to memorize new knowledge rather than integrate it into prior knowledge.[29] The human systems at work in the brain to make sense of prior and new knowledge include cognitive learning, affective learning, and psychomotor learning.[26] Each of these systems is distinct and unique, but they are also connected and interactive in their roles for making sense of human experiences.[26] Current research in cognitive psychology indicates that human emotion is not limited to just one area of the brain.[30,31] Instead, different areas of the brain encode different parts of emotions where the combination of those systems give rise to the emotion.[31] When a student is engaged in a learning activity and attempting to make sense of a new experience, the brain is inherently recalling previous feelings as well as previous thoughts and actions. Therefore, a student's prior experiences have a great influence on how they choose to act in their chemistry laboratory course.

Because the "doing" of chemistry laboratory work is obvious and visible to students and instructors, the MLLI was designed to investigate the less obvious and visible domains of thinking and feeling while performing chemistry laboratory experiments.[1] The MLLI was developed using a pre/post format where students are asked about their expectations for learning prior to conducting laboratory work for the semester and then asked about their learning experiences toward the end of the semester. MLLI items were constructed to ask about a cognitive

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

experience (e.g., Q3 to make decisions about what data to collect), an affective experience (e.g., Q9 to be nervous about making mistakes), or an experience that is both cognitive and affective in nature (e.g., Q20 to worry (affective part) about the quality (cognitive part) of my data). Students are asked to indicate their percent agreement with each MLLI item from 0% (completely disagree) to 100% (completely agree). Composite scores are calculated for the cognitive, affective, and cognitive/affective scales by reserve coding the negatively worded items and averaging the item responses for the items on each scale. Together, MLLI items measure students' cognitive and affective perceptions about their learning experiences in the undergraduate chemistry laboratory. (The evidence gathered for the validity and reliability of the data generated by the MLLI has already been reported.[1]) As previous studies characterized students' cognitive and affective expectations and experiences in learning chemistry in the laboratory during one semester both a single institution[1,3] and across multiple institutions,[2] this study sought to characterize how students' expectations and experiences change over two years of undergraduate chemistry instruction.

■ RESEARCH QUESTIONS

The main research question for this part of the larger research study was: how do students' cognitive and affective perceptions of learning, as measured by MLLI, change as they learn more chemistry? The research design and analysis described in this manuscript sought to answer four subquestions to the overall research question:

R1: How do students' cognitive and affective perceptions of learning in the undergraduate chemistry laboratory change over GC and OC laboratory courses?

R2: How do students' responses to individual MLLI items change over GC and OC laboratory courses?

R3: What is the relationship between students' initial expectations for laboratory learning with how their perceptions change over time?

R4: What happened to the students in the "Change" cluster from the GC pre and post cluster analysis?

■ METHODS

Data Collection

The goal of this study was to investigate how students' cognitive and affective experiences in the undergraduate chemistry laboratory changed across GC and OC laboratory courses. To do so, the MLLI was administered six times over two years using Qualtrics Survey software. Starting in general chemistry, the MLLI was administered during the first week of the fall semester prior the start of the laboratory experiments to measure students' expectations for learning (pre). The MLLI was then administered at the end of the fall semester (post) and end of the spring semester (postpost) to measure students' experiences regarding their laboratory work in the first semester and second semester GC laboratory courses, respectively. The same administration format was followed for organic chemistry. IRB approval was obtained prior to the start of data collection.

Students were asked to give their university email addresses to match their responses over the two years. (After students' responses were matched, the data file was stripped of identifiable information prior to analysis to ensure the confidentiality of the participants.) The matching of students responses allowed for direct comparison of students' expectations and experiences over the two courses and six points of data collection. During the 2013−2014 and 2014−2015 academic years, 268 students took the MLLI at least once as a GC student and at least once as an OC student. This manuscript will focus on the responses of the students who took the MLLI all six times ($N = 61$). While analysis presented in this manuscript could also examine the students with missing data, the goal of this study was to examine those students who did take the MLLI all six times across the two chemistry laboratory courses. A power analysis was conducted to know whether sizable, significant changes over time could be detected with a sample of $N = 61$. By using the free G*Power software, the necessary sample size to detect a significant medium effect with a power of 0.8 over six different measurements was calculated to be 20.[32] Thus, the decision was made to move forward with the sample size of 61 to examine how these students' cognitive and affective perceptions of laboratory learning changed over time. This sample of 61 students was 75% female (greater than the course enrollment of approximately equal ratio male to female) and ≥80% white (similar to the university profile).

Sample Description

This research took place at a midsize liberal arts university in the midwestern United States. The GC laboratory course is a two-semester sequence with both courses offered on and off sequence (during the fall and spring semesters). The OC laboratory course is also a two-semester sequence, but the first semester is only offered in the fall, and the second semester is only offered in the spring of the academic year. Both GC and OC offer separate courses for the chemistry majors and for the nonchemistry majors. Concurrent enrollment in the laboratory and lecture course is not required for GC or OC, but it is encouraged. The target courses for this study were on-sequence GC and OC laboratory courses, including chemistry and nonchemistry majors. The semesters of each laboratory course will be distinguished by GC1, GC2, OC1, and OC2.

During the GC laboratory sequence, students performed 10 experiments in each of the 15-week semesters. Students conducted experiments that were a mixture of confirmatory and structured inquiry with content focusing on stoichiometry, acid/base, oxidation−reduction, thermochemistry, quantitative analysis, and properties of gases.[33,34] The students worked both individually and in small groups throughout the course, and they were expected to complete the experiments in the allotted 3-h lab time. Students completed individual lab reports due the week following the experiment. The format of the lab reports was a summary sheet with one formal lab report during each semester. Each lab room held a maximum of 42 students with two teaching assistants per lab room.

During the OC laboratory sequence, the students completed nine experiments in each of the 15-week semesters. The majority of the experiments for OC were structured inquiry, with some guided inquiry at the end of the semester.[33,34] The topics of the experiments focused on teaching the techniques of extraction, separation, purification, recrystallization, TLC, IR, distillation, and reflux with many experiments having explicit real-world connections. Students performed experiments in pairs with frequent collaborations in larger groups of 3−4 pairs. Lab work was expected to be completed within the 3-h time block. Lab reports consisted of written responses to laboratory questions due within a week with the exception of two formal

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 1. Descriptive Statistics for MLLI Responses over Four Semesters

Scale	Response Means[a] (SD) by Measurement Time (N = 61)					
	GC Pre	GC Post	GC Postpost	OC Pre	OC Post	OC Postpost
Cognitive	71.6 (8.7)	64.4 (10.4)	61.4 (12.3)	68.8 (10.2)	58.6 (13.6)	56.0 (14.2)
Affective	53.2 (14.4)	56.4 (18.3)	51.0 (18.0)	50.0 (16.7)	46.6 (21.2)	45.5 (17.9)
Cognitive/Affective	54.2 (12.5)	48.6 (13.3)	41.0 (13.5)	48.8 (13.4)	43.1 (16.0)	40.1 (13.9)

[a]Scale responses range from 0−100%.

reports, which were due within 2 weeks. Each lab room held 30 students with one teaching assistant per lab room.

Data Analysis

The first step in the analysis was calculating the descriptive statistics for each MLLI scale (cognitive, affective, and cognitive/affective) for each time point. A variety of plots were constructed to visualize the data including boxplots and scatterplots. Visualizations were constructed for composite variables as well as on an item level. For each item, additional plots were constructed, including boxplots and scatterplots that compared consecutive time points (GC pre vs GC post, GC post vs GC postpost, GC postpost vs OC pre, etc.). These plots allowed for exploration of how students' perceptions of specific learning experiences changed over time.

To analyze how the students' responses changed over time, separate repeated-measures (RM) ANOVA models were performed for each MLLI scale. Prior to the analyses, the assumptions for the RM ANOVA were examined including independent random sampling, multivariate normal distributions, homogeneity of variance, and homogeneity of covariance (sphericity).[35] Normality was assessed with the Shapiro−Wilk test, and the homogeneity of variance was assessed with Levene's test. Both showed some deviation, but the RM ANOVA is not as sensitive to departures from normality or to some heterogeneity of variance.[35] Homogeneity of covariance was assessed using Mauchly's test of sphericity. The results showed violations in sphericity for each scale, and the Greenhouse−Geisser corrected degrees of freedom were used to determine the critical F value.[35] Prior to conducting the RM ANOVA, the decision was made to conduct six posthoc dependent t tests of the 15 total possible comparisons. Because of the sphericity violation, a conservative posthoc correction was applied using the Bonferroni adjustment. The Bonferroni corrected alpha for the posthoc comparisons was 0.0083. Adjusting alpha for pairwise comparisons helped to control for Type-1 error.[35] If the alpha were not adjusted with multiple statistical tests conducted on the same data, then the Type 1 error rate would be arbitrarily inflated, and significant results could be found when there actually were not any. The Bonferroni adjustment is one of the most conservative techniques[35] but was chosen in this case due to the violation of sphericity and the choice to conduct only six of the 15 possible pairwise comparisons. Appropriate effect sizes were calculated for the omnibus test and the pairwise comparisons to indicate the degree of difference in the measurements. Students voluntarily participated in the study, so conclusions should be drawn cautiously.

Previous work has shown that students' cognitive and affective perceptions of their learning change differently over a semester of GC and OC.[2,3] As these students were followed across both courses in this study, the question arose as to whether similar patterns remained. The previous work clustered students based only on their pre and post averages for one semester as those were the only data points available for those students. One major finding from those studies was that students' expectations governed what the students experienced in their chemistry laboratory courses, especially for the affective scale.[2,3] This finding has also been reported in science education and social psychology research.[36,37] Thus, because the current study analyzed how students perceive their learning, the decision was made to cluster students based only on their GC expectations. In this way, analysis sought to investigate how students responded to the MLLI throughout GC and OC based on their initial expectations. By using students' pre cognitive and affective averages, students were clustered using a hierarchical agglomerative procedure with Ward's method as the linkage technique and squared Euclidean distance for the dissimilarity measure.[38,39] Hierarchical agglomerative clustering allows for the creation of clusters when theory does not suggest a specific a number of clusters a priori.[38] Because each step of the hierarchical agglomerative clustering is irrevocable, Ward's method was used to minimize the within groups differences and maximize the between groups differences.[38,39] Squared Euclidean distance was selected at the dissimilarity measure because as the clustering variables are continuous in nature, the squared Euclidean distances can be interpreted as physical distances between points in Euclidean space.[38] The solution was analyzed for distinctiveness and interpretability of the clusters.[40−46] Then, a mixed-methods ANOVA was conducted using cluster membership as the between groups variable to find out if these clusters with distinct expectations for learning change their perceptions over their two years of laboratory course. The assumptions for these tests are similar to those of the RM ANOVA and were assessed prior to analysis.[35] An additional power analysis was conducted prior to this inferential test as well. By using G*Power, to detect significant medium effects with a power of 0.8 over six different measurements with multiple groups, the total sample size was calculated to be 20−32 (depending on the number of clusters as suggested by the cluster analysis).[32]

The results from the previous cluster analysis of MLLI student responses, both at a single institution and across multiple institutions, revealed four clusters of GC students based upon their pre- and post-test responses.[2,3] Three clusters were sequential in nature as one had low responses overall ("Low"), one had high responses overall ("High"), and one had responses between Low and High ("Mid"). The fourth cluster had high expectations for their laboratory experiences that went largely unfulfilled by their experiences ("Change"). The results from these studies raised additional questions, namely about the Change cluster. Accordingly, the data collected for this study examined whether those students who continued in taking chemistry laboratory courses and participated in the MLLI data collection reported similar expectations and experiences in year 2 or recalibrated their expectations for future chemistry laboratory courses. Those students belonging to the Change cluster who participated in MLLI data collection

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

all six times were identified, and their responses patterns were analyzed.

RESULTS AND DISCUSSION

RQ1: How Do Students' Cognitive and Affective Perceptions of Learning in the Undergraduate Chemistry Laboratory Change over GC and OC Laboratory Courses?

Descriptive Statistics. Table 1 displays the descriptive statistics for the cognitive, affective, and cognitive/affective scales across for each of the six time points. The distributions of the responses are shown by boxplots in Figure 1. Additional scatterplots are available in the Supporting Information. From the summary statistics and the boxplots, the cognitive scores have the smallest spread of the three scales, but the cognitive variation does appear to increase over time. The affective scores have the greatest variations across all time points, while the spread for the cognitive/affective scores remain similar over time. OC post test has the greatest variation for affective and cognitive/affective, and OC postpost test shows the lowest averages for all three variables.

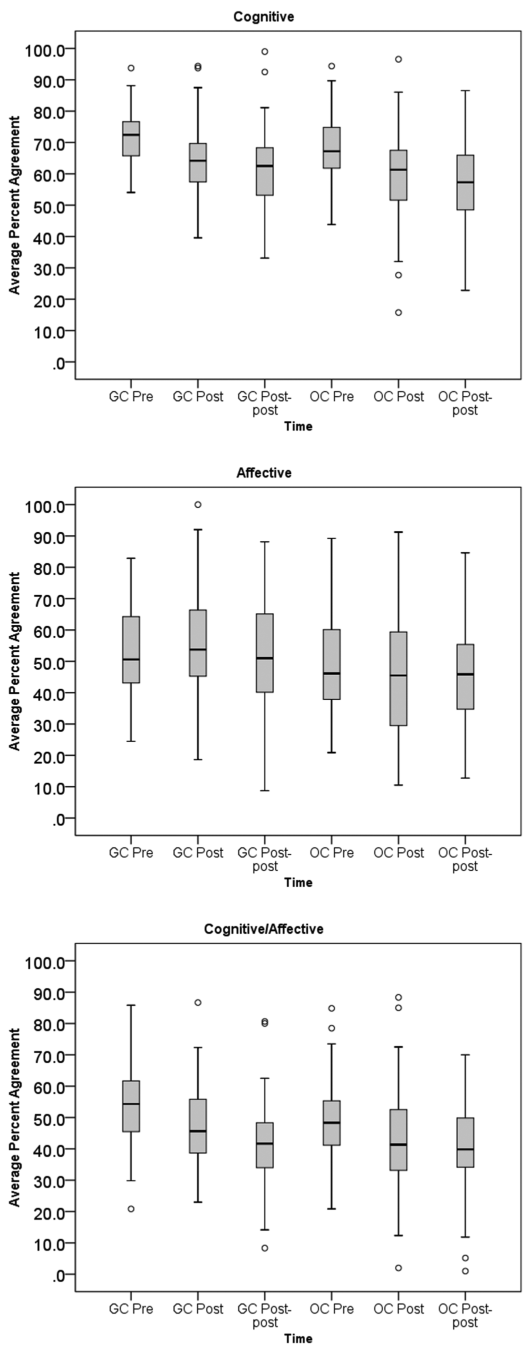


Figure 1. Boxplots of cognitive, affective, and cognitive/affective averages across GC and OC.

Analysis of Change over Time. A RM ANOVA was conducted for each scale to analyze the change over time. The omnibus test showed significant decreases for each variable with effect sizes ranging from small for affective to large for cognitive and cognitive/affective (Table 2). To examine the changes between each consecutive pair of time points, six posthoc dependent t tests were conducted with a Bonferroni corrected alpha of 0.0083. The selected posthoc analyses compared consecutive points (pre vs post, post vs postpost, etc.) as well as expectations for GC vs OC. The rationale for comparing GC postpost and OC pre was to see how students' expectations for OC were influenced by their prior knowledge (i.e., their experiences in GC laboratory).

Table 2. One-Way Repeated Measures ANOVA Models for Each MLLI Scale

Scale	RM ANOVA
Cognitive	Wilks's Λ = 0.40, $F(5, 300) = 30.3$, $p \leq 0.0001$, $\varepsilon = 0.69$, $\eta^2_p = 0.34$
Affective	Wilks's Λ = 0.67, $F(5, 300) = 6.55$, $p \leq 0.0001$, $\varepsilon = 0.85$, $\eta^2_p = 0.10$
Cognitive/Affective	Wilks's Λ = 0.45, $F(5, 300) = 17.8$, $p \leq 0.0001$, $\varepsilon = 0.79$, $\eta^2_p = 0.23$

The results for the posthoc comparisons are reported in Table 3. For the cognitive scale, there were large, significant differences between GC pre and GC post and between OC pre and OC post. Students indicated high expectations for cognitive learning in their laboratory courses that went unfulfilled by their experiences. Interestingly, there was a large, significant increase from GC postpost to OC pre. Despite students' responses that their cognitive experiences were less than they expected in GC, they reported high expectations for OC. While there is a medium effect size for the difference between GC expectations and OC expectations, this difference was not significant with the Bonferroni corrected alpha of 0.0083.[47,48] Students "reset" their cognitive expectations for OC despite their unmet expectations during both semesters of GC, but not to the level of expectation they had at the start of GC. Thus, at the start of a new course in their second year, students expected to carry out behaviors consistent with those necessary for meaningful learning to occur, that is, they expected to engage cognitively while performing their laboratory experiments as indicated by their high cognitive expectations. The changes from post to postpost for both courses were not significant, indicating that students reported similar experiences in both semesters of the laboratory course.

The posthoc comparisons for the affective scale showed no significant changes within the boundaries of the Bonferroni corrected alpha. There was a medium-sized decrease from GC post to GC postpost, but this change was not significant by the Bonferroni corrected alpha to protect from Type-1 error.[47,48] As shown in Figure 1, the spread for the affective averages remained large with a median consistently between 45 and 55% agreement across all time points. One explanation for this result

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 3. Post-Hoc Comparisons between Courses

Pairwise Comparison	Cognitive			Affective			Cognitive/Affective		
	t	p	η^2	t	p	η^2	t	p	η^2
GC pre vs GC post	5.624	0.0001[a]	0.345	−1.429	0.158	0.033	3.113	0.003[a]	0.139
GC post vs GC postpost	2.406	0.019	0.088	2.619	0.011	0.103	4.753	0.0001[a]	0.274
GC postpost vs OC pre	−5.236	0.0001[a]	0.314	0.493	0.624	0.004	−5.252	0.0001[a]	0.315
OC pre vs OC post	5.989	0.0001[a]	0.374	1.561	0.124	0.039	3.131	0.003[a]	0.140
OC post vs OC postpost	1.636	0.107	0.043	0.515	0.608	0.004	1.860	0.068	0.055
GC pre vs OC pre	2.715	0.009	0.109	1.589	0.117	0.040	3.095	0.003[a]	0.138

[a]Significant p value at the Bonferroni corrected alpha of 0.0083.

Table 4. Descriptive Statistics for Items with Different Patterns across GC and OC

Items	Response Means[a] (SD) by Measurement Time (N = 61)					
	GC Pre	GC Post	GC Postpost	OC Pre	OC Post	OC Postpost
#7—Learn critical thinking skills	84.3 (18.5)	53.8 (23.9)	65.2 (25.0)	78.9 (18.1)	66.2 (23.4)	60.3 (25.7)
#8—Excited to do chemistry	75.1 (19.9)	58.9 (27.4)	47.8 (32.1)	61.3 (26.9)	49.9 (32.5)	44.8 (32.4)
#15—Procedures to be simple to do	37.3 (23.5)	59.7 (22.9)	50.2 (27.7)	33.5 (22.5)	50.3 (28.2)	49.5 (23.4)

[a]Scale responses range from 0−100%.

Table 5. Descriptive Statistics for Items with Similar Patterns across GC and OC

Items	Response Means[a] (SD) by Measurement Time (N = 61)					
	GC Pre	GC Post	GC Postpost	OC Pre	OC Post	OC Postpost
#4—Feel unsure about the purpose of the procedures	37.7 (24.1)	38.6 (27.2)	56.6 (28.5)	53.8 (26.6)	54.0 (28.5)	56.0 (25.5)
#24—Focus on procedures, not concepts	40.9 (24.0)	47.5 (22.6)	56.7 (26.0)	44.5 (25.4)	60.9 (23.4)	64.3 (24.4)

[a]Scale responses range from 0−100%.

is that while the omnibus ANOVA detected individual changes over time, within each pairwise comparison, the number of students who increased in their affective average was nearly equal to the number of students who decreased. (Scatterplots comparing consecutive MLLI administrations can be found in the Supporting Information.)

For the cognitive/affective pairwise comparisons, significant changes were detected for all conducted tests except one. Cognitive/affective responses decreased throughout GC. For the GC to OC transition, students indicated higher expectations for learning than they reported experiencing in the previous course, but not as high as their GC expectations. During the first semester of OC, these high expectations went unfulfilled. The only nonsignficant pairwise comparison for the cognitive/affective scale showed that students' experiences remain unchanged across both semesters OC lab.

RQ2: How Do Students' Responses to Individual Items Change over GC and OC Laboratory Courses?

Analysis of each MLLI item using various visualizations revealed some items with similar response patterns across GC and OC, while other items showed that students' responses changed over time. Example items are discussed further.

Items with similar patterns across both GC and OC indicated that students had similar expectations and similar experiences for both courses. Items with this response pattern included items 7 (learn critical thinking skills), 8 (excited to do chemistry), and 15 (procedures to be simple to do) (descriptive statistics are provided in Table 4). For item 7, students had high expectations about learning critical-thinking skills. While students indicated an overall agreement for this experience during both semesters of GC and OC, the average percent agreement for the experience was 15−20% below the expected. Students expected to learn critical-thinking skills during their chemistry laboratory courses, but they did not report meeting those expectations during GC or OC. Item 8 revealed even more pronounced differences between expectations and experiences. Again, students started both semesters with high expectations to be excited to do chemistry. However, when students reported their actual experiences in the chemistry laboratory, the majority of students reported experiences that did not meet their expectations. Interestingly, the high expectations to learn critical-thinking skills and to be excited to chemistry in OC, despite not experiencing them in GC, is promising as it suggests that students may remain open to new experiences. Both of these items showed an increased spread of responses for the post and the postpost of each year indicating diverse perceptions of their experiences despite similar expectations and being students in the same laboratory courses. The similar patterns of these two items could indicate an opportunity lost by both students *and* instructors. The students may lose an opportunity for meaningful learning, while the instructor may lose the opportunity to capitalize upon students' expectations to learn meaningfully. If students have expectations to be excited to do chemistry and to learning critical thinking skills, then the instructor has a responsibility to follow through by designing the curriculum in such a way that the student can choose to engage in those experiences.

MLLI item 15 also revealed similar responses across GC and OC, but with a different pattern than items 7 and 8. Item 15 was an unexpected experience for many students. The majority of students did not expect the procedures to be simple to do for either course indicated by an average percent agreement for these expectations of less than 40%. However, students reported on the post-tests for both courses that the procedures were in fact simple to do, or at least more simple than they had expected them to be. For the GC postpost, students'

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

perceptions of the simplicity of the procedures decreased in comparison to the end of GC1, and they were still similar at the end of OC2. The increased perception of the simplicity of the procedures, in conjunction with decreased reports of critical thinking and excitement about doing chemistry, could be indicative of a disproportionate emphasis on the doing of laboratory work rather than thinking about the hows and whys of the experiments.

Examples of items with different response patterns across GC and OC were items 4 (feel unsure about the purpose of the procedures) and 24 (focus on procedures, not concepts) (descriptive statistics are reported in Table 5). In general, these items revealed diverse expectations, diverse experiences, and little convergence on responses. The findings reported here discuss the overall trend for this longitudinal sample. On item 4, students generally reported not expecting to feel unsure about the purpose of the procedures. That expectation was met in GC1 (meaning they reported that they did not feel unsure), but the experience was largely reported for GC2. OC expectations for this item were too diverse to identify an overall expectation for the sample. Comparing students' experiences in both semesters of OC suggests that students reported similar experiences for this item. The spread and shifts in the responses on this item demonstrate that different curricula and different types of experiments could have an effect on how students feel and think while they perform experiments.

Similarly, students did not expect to focus on procedures rather than concepts (item 24), but they reported increasing experiences for both GC1 and GC2. At the start of OC1, students' responses show a "reset" of their expectations by once again reporting that they did not expect to focus on procedures over concepts despite their previous experiences. Unfortunately, students reported similar experiences for both OC1 and OC2 as they unexpectedly focused on procedures over concepts. The combination of the results from items 4 and 24 shows that even with increased experience as a student in the chemistry laboratory, students increasingly focus on execution of the experiments despite being unsure of the purpose of the procedures.

RQ3: What Is the Relationship between Students' Initial Expectations for Laboratory Learning with How Their Perceptions Change over Time?

Cluster analysis of the students' responses on their initial cognitive and affective expectations (GC pre) suggested four distinct clusters. A scatterplot was constructed to display the students' affective versus cognitive expectations (the clustering variables) to show the four distinct clusters (Figure 2). Two clusters had higher cognitive expectations than affective (Cluster 1 and Cluster 2), and two clusters had similar expectations for both cognitive and affective (Cluster 3 and Cluster 4). One cluster had the highest cognitive and affective expectations (Cluster 4), while another had the lowest expectations for both scales (Cluster 1). There were two clusters in the middle, one with higher cognitive expectations (Cluster 2) and one with higher affective expectations (Cluster 3). The clusters were given descriptive names based on their relative expectations. Table 6 contains the descriptive statistics for each of the four clusters across all six time points.

A series of two-way RM ANOVAs were conducted to analyze how students' perceptions of learning changed over time in relation to their initial expectations. The between group variable was the cluster membership, and the within group variable was time. An interaction between cluster and time would indicate that the responses of students in different clusters changed differently over the GC and OC laboratory courses. Table 7 shows the results from these analyses. The between group differences were significant for all three scales with large effect sizes for each demonstrating the differences between the four clusters based on their initial expectations. The within groups differences were also significant for all three scales, which is not surprising given the results from the RM ANOVA for the overall sample (Table 2). The interaction between cluster membership and time was not significant for the cognitive scale, indicating that the clusters do not change differently over time. The only significant interaction was found for the affective scale with a medium effect size.[47,48] While a power analysis was conducted and it was determined that the sample was large enough to detect differences, the cognitive/affective scale did not detect a significant interaction but did have the same medium effect size as the affective scale. The line plots showing the estimated marginal means from the ANOVA models display the cluster change over time for each scale (Figure 3). While interactions were detected for affective and cognitive/affective, similar overall trends can be observed for all clusters. Thus, despite the students initial expectations for laboratory learning, they can be influenced by the curricular and pedagogical design of the laboratory course.

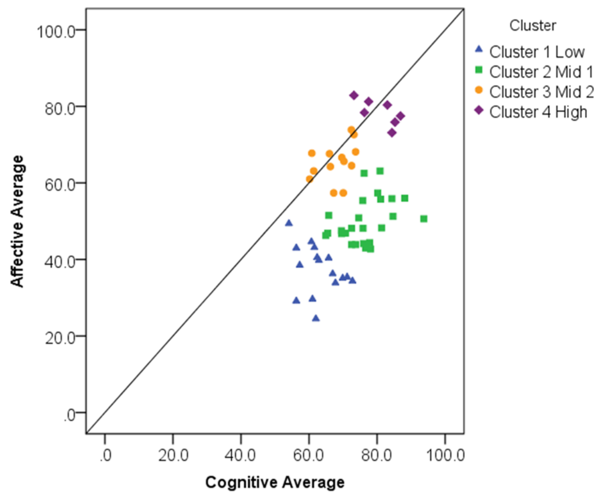


Figure 2. Scatterplot of GC1 pre affective average versus GC1 pre cognitive average, color-coded by cluster to demonstrate the distinctness of the clusters based on students' initial expectations for GC1 chemistry lab.

RQ4: What Happened to the Students in the "Change" Cluster from the GC Pre and Post Cluster Analysis?

The initial cluster analysis with the GC students who participated both in the pre and post-test administration ($N = 436$) suggested four clusters of students.[3] While three of the clusters had sequential responses where their expectations and experiences fell within the same general area on the scale, a fourth cluster reported experiences that were misaligned with their expectations. From this "Change" cluster, 42 of the 99 students participated in MLLI data collection at least once more, and six students participated all six times. While this sample was too small for inferential statistics, the patterns of responses for these six students were qualitatively analyzed for how they changed after their first semester of GC lab.

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 6. Descriptive Statistics for MLLI Responses over Four Semesters for the Four Clusters Generated by Clustering on GC1 Expectations

Cluster	Scale	Response Means[a] (SD) by Measurement Time					
		GC Pre	GC Post	GC Postpost	OC Pre	OC Post	OC Postpost
Cluster 1 Low $N = 16$	Cognitive	66.0 (5.6)	59.9 (9.2)	57.6 (10.4)	64.1 (8.2)	55.0 (7.3)	50.0 (14.5)
	Affective	37.3 (6.4)	48.6 (12.7)	42.3 (16.2)	39.5 (8.7)	37.7 (15.4)	37.2 (16.6)
	Cognitive/Affective	42.9 (8.3)	42.7 (9.8)	34.9 (10.8)	38.4 (9.5)	36.4 (12.4)	33.5 (14.7)
Cluster 2 Mid 1 $N = 25$	Cognitive	76.4 (7.0)	64.8 (7.7)	62.7 (10.5)	70.4 (10.3)	60.5 (11.4)	58.4 (11.5)
	Affective	50.0 (5.8)	54.3 (19.1)	50.1 (18.1)	50.1 (16.9)	48.9 (20.0	46.4 (15.0)
	Cognitive/Affective	56.1 (9.5)	47.8 (12.2)	34.9 (10.8)	38.4 (9.5)	36.4 (12.4)	40.7 (10.7)
Cluster 3 Mid 2 $N = 13$	Cognitive	67.9 (4.8)	64.3 (12.3)	59.5 (11.5)	65.9 (9.4)	51.6 (17.6)	55.6 (17.2)
	Affective	65.4 (5.0)	62.4 (16.2)	53.2 (12.8)	51.9 (11.4)	41.4 (19.7)	46.5 (18.9)
	Cognitive/Affective	59.2 (16.2)	53.6 (18.7)	41.3 (12.3)	49.7 (11.6)	38.3 (18.0)	42.8 (15.5)
Cluster 4 High $N = 7$	Cognitive	80.9 (5.2)	73.4 (13.8)	69.1 (20.7)	69.1 (20.7)	73.8 (12.9)	56.0 (15.0)
	Affective	78.2 (3.3)	71.4 (21.4)	70.8 (16.4)	70.3 (20.5)	68.3 (24.0)	59.8 (22.3)
	Cognitive/Affective	64.2 (16.1)	56.3 (18.7)	57.9 (16.7)	65.7 (13.1)	63.3 (18.2)	48.5 (16.4)

[a]Scale responses range from 0−100%.

Table 7. Results from Two-Way Repeated Measures ANOVA for How the Clusters Based on Students' Initial Expectations for GC1 Changed over Time for Each Scale

Factor	Two-Way RM ANOVA
Cognitive	
Cluster	Wilks's Λ = 0.731, $F(3, 57) = 6.99$, $p \leq 0.0001$, $\eta^2_p = 0.27$
Time	Wilks's Λ = 0.453, $F(5, 285) = 24.54$, $p \leq 0.0001$, $\eta^2_p = 0.30$, $\varepsilon = 0.66$
Time*Cluster	Wilks's Λ = 0.676, $F(15, 285) = 1.23$, $p = 0.25$, $\eta^2_p = 0.06$, $\varepsilon = 0.66$
Affective	
Cluster	Wilks's Λ = 0.621, $F(3, 57) = 11.61$, $p \leq 0.0001$, $\eta^2_p = 0.38$
Time	Wilks's Λ = 0.636, $F(, 285) = 7.10$, $p \leq 0.0001$, $\eta^2_p = 0.11$
Time*Cluster	Wilks's Λ = 0.667, $F(15, 285) = 1.78$, $p = 0.04$, $\eta^2_p = 0.08$
Cognitive/Affective	
Cluster	Wilks's Λ = 0.647, $F(3, 57) = 10.37$, $p \leq 0.0001$, $\eta^2_p = 0.35$
Time	Wilks's Λ = 0.508, $F(, 285) = 14.05$, $p \leq 0.0001$, $\eta^2_p = 0.19$, $\varepsilon = 0.77$
Time*Cluster	Wilks's Λ = 0.626, $F(15, 285) = 1.67$, $p = 0.05,7$ $\eta^2_p = 0.08$, $\varepsilon = 0.77$

Figure 4 shows scatterplots comparing these six students' pre and post responses from GC1 through OC1. Students' affective averages were plotted against their cognitive averages with pre and post for one course on the same plot to visualize how the students' perceptions changed during one course. "Vectors" were drawn between individual students' pre and post responses to indicate how the students changed over the course.[3] The plot on the left shows how the students changed for GC, and the plot on the right shows how the students changed during OC. Individual students are marked by the same color in both plots. In GC, these students had similar pre- and post-test responses as can be seen from the heads and tails of the vectors all being near one another. For OC, the heads and tails of the vectors that represent the changes in these students' responses are not localized on the scatterplot. These students who indicated similar expectations and experiences for the first semester of GC lab reported disparate expectations and experiences for OC1. Some students were better able to align their expectations with their experiences as shown by the smaller vectors. Other students lowered their expectations for OC1 based on their GC experiences and then reported even lower experiences for OC1. For those students who belonged to the Change cluster, some modified their expectations for OC1 based up on their experiences in GC, while it appears that others could have modified their actions in the laboratory based upon their differently shaped vectors indicated different experiences in OC1 than GC1.

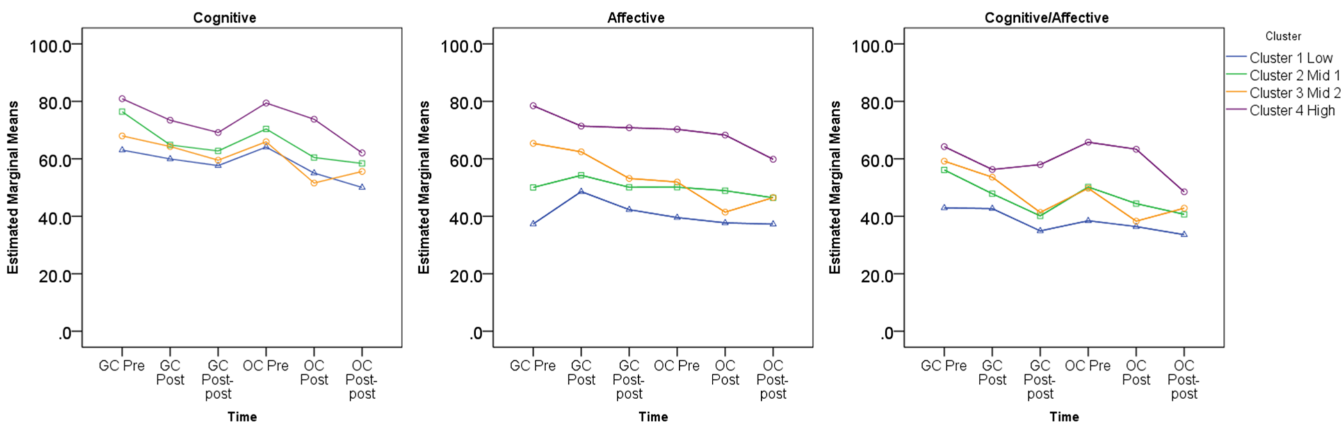


Figure 3. Line plots of estimated marginal means for cognitive, affective, and cognitive/affective averages for the cluster based on students' initial expectations showing their change in perceptions of learning from GC1 through OC2.

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

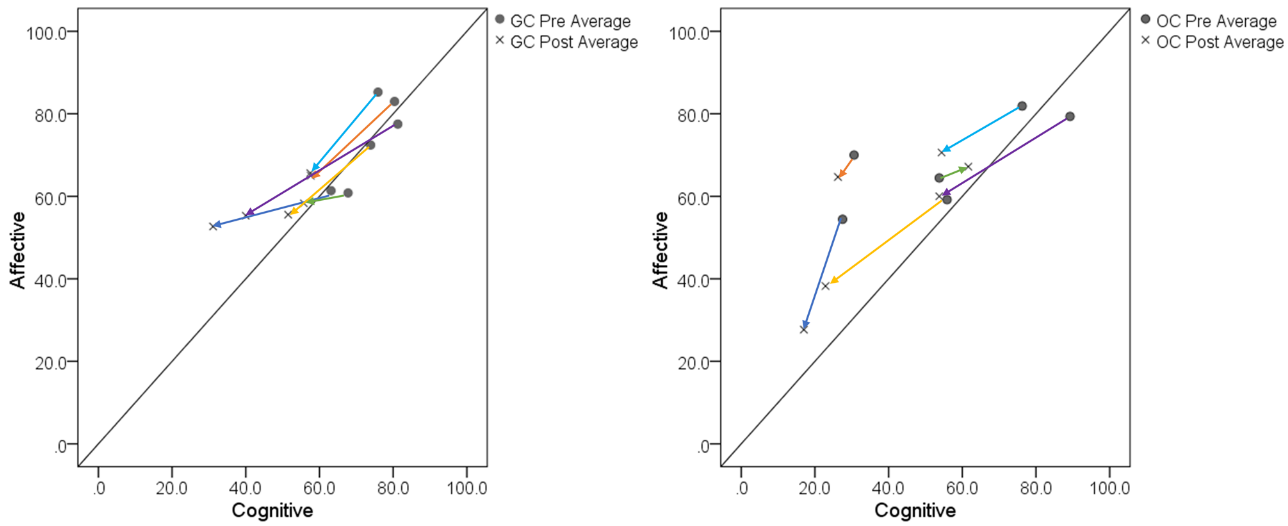


Figure 4. "Vector" plots of cognitive and affective averages comparing expectations and experiences for GC1 and OC1 for the six students from the "Change" cluster who responded all six times in the longitudinal data collection.

■ CONCLUSIONS

This study followed 61 students from GC1 through OC2 laboratory courses to measure their cognitive and affective expectations and experiences related to their laboratory learning. Analysis of changes over time was conducted using inferential statistics. The cognitive and cognitive/affective scales showed similar patterns across the two years as students started GC1 with high expectations that went unmet by their experiences during both semesters of GC. For OC1, students appeared to "reset" and reported high expectations for cognitive and cognitive/affective experiences despite the experiences they had in GC1 and GC2. These experiences then went unmet in OC1 and OC2 as well. The students' affective averages remained relatively constant over time. This result does not mean that students' affective perceptions of learning remained constant over time. Instead, as has been found in previous research, reasonably even numbers of students increased in affective averages as decreased.[1,2] When analyzing changes over time at the item level, distinct categories of items emerged. One category of items had similar trends across GC and OC where students responded similarly to those items across both courses, which indicated that students had similar expectations and similar experiences for both courses. The other category of items exhibited a change in response patterns for students from GC to OC.

After analyzing the students' responses as a whole across GC and OC, the students' responses were analyzed multiple times with cluster analysis. First, students were clustered based on their initial expectations for GC, and four clusters were found to characterize students' expectations. A mixed-methods ANOVA was conducted to analyze whether students' MLLI responses changed differently based on cluster membership. Initial expectations did not appear to influence how cognitive perceptions of learning changed, but initial expectations did have a medium effect on how affective and cognitive/affective perceptions changed. Second, students who were classified into the "Change" cluster from the GC1 pre and post cluster analysis[3] were analyzed to see how they modified their perceptions of learning based on their first semester that resulted in largely unfulfilled expectations. Of the original 99 students in the "Change" cluster, the six students who participated in MLLI data collection all six times showed different patterns of responses beyond the first semester. It appears that some students kept their high expectations but adjusted their behavior, while others lowered their expectations to align with the negative experiences they had the first semester. These findings suggest that students' initial expectations and experiences do not create boundaries for how they perceive their learning throughout all of GC and OC laboratory courses. Put another way, students' perceptions of learning in GC do not appear to determine their perceptions of learning in OC.

In comparison with our cross-sectional studies,[1−3] there are some similar pieces of evidence and some new information generated uniquely by the longitudinal study. The "reset" is apparent in both the cross-sectional and the longitudinal where students have high expectations in OC despite reporting low experiences in GC. In the previous studies, there were no differences between courses, only significant differences over time. Thus, students were reporting high expectations for GC that went unmet and then setting the same high expectations for OC only to go unmet again. The unique information offered from the longitudinal study shows the continued decrease in the second semester of lab for both cognitive and cognitive/affective. For cognitive, there was a larger change from pre to post for both GC and OC than from post to postpost. For cognitive/affective, the larger change was from post to postpost for both GC and OC. In both cases, students started the course with high expectations for both cognitive and cognitive/affective scales that went largely unmet during both semesters of lab. Additional research is needed to understand why students "reset" their expectations for OC despite their previously unfulfilled expectations for GC.

Implications for Research

This longitudinal study followed 61 students across two years (four semesters) of their undergraduate chemistry laboratory courses to track changes in their cognitive and affective perceptions of learning. The results reported in this manuscript speak to the experiences of students at one university, which leads to the question of how do students perceptions of learning change at universities different than the one described here. (In our previous studies, similar results were found

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

between this single institution and the national, multiple institutional study.) Further, the MLLI could be used to measure the long-term effects of evidence-based laboratory curricula as this study has shown that the MLLI can measure unique changes in students' ideas about learning over time.

There are many factors that could influence the evolution of students' perceptions of learning in the laboratory that were not examined in this study. First, sex differences were not explored in this study as two-thirds of the sample was female. Future work could seek to sample an equal proportion males and females to analyze any differences in how the two groups change over time. Another area for educational research could explore whether females are more compliant than males to consistently respond to repeated requests for research participation over time. A second area that was not explored in this study was student retention. As we analyzed the "Change" cluster students who participated all six times, the question was raised as to who dropped out of chemistry and when and who was only required to take one semester of GC or OC for their major. For our study, we were unable to discern between those who dropped out versus those who just did not respond to the request to participate in the research study. These questions would require a new IRB to obtain enrollment records and course rosters from the university registrar. In addition, the body of research on expectancy shows how a person's expectations can be based upon perception of ability, effort expenditure, degree of difficulty, chance, and motivation.[49−51] The exploration of these psychological factors could give insight into how students create their expectations for their laboratory courses and in turn offer insight into how to best address students' expectations upon entrance to the course as well as when students' expectations go unmet or when they encounter unexpected experiences. The design of the curricula and the choice of pedagogy matters as it does have an influence on the students. The students' perceptions of learning do change, just not strictly dependent on their initial expectations. Thus, care should be taken in the development and choice of laboratory curricula and pedagogy to consider how students learn and the factors influencing their learning.

Implications for Teaching

While our previous studies that clustered students on their expectations and experiences found that students' expectations tended to govern their experiences, this study found few interactions between cluster membership and how the students changed their perceptions of learning over time. This finding indicates that perhaps students are malleable in our courses and able to be influenced by the course. Students' change in responses showed similar patterns over time demonstrating how the course was affecting the students in similar ways over two years. This analysis also showed that students have a wide range of expectations when they enter the GC laboratory course. We cannot assume that our students come in with similar backgrounds or perceptions of learning. Rather, this evidence shows a full range of expectations: some with low cognitive and affective expectations, some high cognitive and affective, and some with a combination of high and low cognitive and affective expectations. Since positive integration of thinking and feeling with the doing of laboratory work is necessary for meaningful learning in the laboratory, then it would be helpful for the course instructor to be aware of the range of cognitive and affective expectations that students bring to their chemistry laboratory course. The design of laboratory curriculum ought to address students' incongruent or low cognitive and affective expectations to offer them opportunities to positively integrate their thinking and feeling with their doing of their laboratory work. The results from exploring the students who continued from the Change cluster also show that students with initial unfulfilled expectations do not necessarily sustain negative experiences through the rest of their chemistry career.

Limitations

Interpretations and conclusions should be made by carefully considering the context within which the study took place. The obvious limitation to this longitudinal study was the attrition of the students. Only a fraction of those who participated in the study in GC1 continued through the end of OC2. Many factors contributed to the attrition of the students including lack of support from laboratory instructors to encourage their students to participate. While this sample is less representative of the population of chemistry students at this university, there were some similar trends in responses to the cross-sectional study. The information gleaned from this sample does give new information about how students' perceptions of learning change over time as they progress through chemistry.

Additionally, the methodological choices made for the data analyses impacted the results as well. Different clustering variables, objects, algorithms, distance measures, and linkage techniques might have produced different cluster solutions. Our decision to go with the four rather than three cluster solution yielded clusters with greater diversity, which in turn led to unique, pedagogically useful conclusions about how the clusters change over time.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.5b00754.

Additional scatterplots (PDF, DOCX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bretzsl@miamioh.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Volwiler Family Endowment to the Miami University Department of Chemistry and Biochemistry and by Grant No. 0733642 from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank Ellen Yezierski for the invaluable feedback on the research and statistics advice.

■ REFERENCES

(1) Galloway, K. R.; Bretz, S. L. Development of an Assessment Tool to Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2015**, *92*, 1149−1158.

(2) Galloway, K. R.; Bretz, S. L. Measuring Meaningful Learning in the Undergraduate Chemistry Laboratory: A National, Cross-Sectional Study. *J. Chem. Educ.* DOI: 10.1021/acs.jchemed.5b00538.

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

(3) Galloway, K. R.; Bretz, S. L. Using Cluster Analysis to Characterize Meaningful Learning in a First-Year University Chemistry Laboratory Course. *Chem. Educ. Res. Pract.* **2015**, *16*, 879−892.
(4) White, R. T.; Arzi, H. J. Longitudinal Studies: Designs, Validity, Practicality, and Value. *Res. Sci. Educ.* **2005**, *35*, 137−149.
(5) Arzi, H. A. From Short- to Long-Term: Studying Science Education Longitudinally. *Stud. Sci. Educ.* **1988**, *15*, 17−53.
(6) Zarotiadou, E.; Tsaparlis, G. Teaching Lower-Secondary Chemistry with a Piagetian Constructivist and an Ausbelian Meaningful-Receptive Method: A Longitudinal Comparison. *Chem. Educ. Res. Pract.* **2000**, *1*, 37−50.
(7) Eskilsson, O.; Helldén, G. A Longitudinal Study on 10−12-Year-Olds Conceptions of the Transformations of Matter. *Chem. Educ. Res. Pract.* **2003**, *4*, 291−304.
(8) Sarantopoulos, P.; Tsaparlis, G. Analogies in Chemistry Teaching as a Means of Attainment of Cognitive and Affective Objectives: A Longitudinal Study in a Naturalistic Setting, Using Analogies with a Strong Social Content. *Chem. Educ. Res. Pract.* **2004**, *5*, 33−50.
(9) Greenbowe, T. J.; Poock, J. R.; Burke, K. A.; Hand, B. M. Using the Science Writing Heuristic in the General Chemistry Laboratory to Improve Students' Academic Performance. *J. Chem. Educ.* **2007**, *84*, 1371−1379.
(10) Trundle, K. C.; Atwood, R. K.; Christopher, J. E. A Longitudinal Study of Conceptual Change: Preservice Elementary Teachers' Conceptions of Moon Phases. *J. Res. Sci. Teach.* **2007**, *44*, 303−326.
(11) Mitchell, Y. D.; Ippolito, J.; Lewis, S. E. Evaluating Peer-Led Team Learning across the Two Semester General Chemistry Sequence. *Chem. Educ. Res. Pract.* **2012**, *13*, 378−383.
(12) Øyehaug, A. B.; Holt, A. Students' Understanding of the Nature of Matter and Chemical Reactions − a Longitudinal Study of Conceptual Restructuring. *Chem. Educ. Res. Pract.* **2013**, *14*, 450−467.
(13) Szteinberg, G. A.; Weaver, G. C. Participants' Reflection Two and Three Years After an Introductory Chemistry Course-embedded Research Experience. *Chem. Educ. Res. Pract.* **2013**, *14*, 23−35.
(14) Gottfried, A. E.; Fleming, J. S.; Gottfried, A. W. Continuity of Academic Intrinsic Motivation from Childhood through Late Adolescence: A Longitudinal Study. *J. Educ. Psych* **2001**, *93*, 3−13.
(15) Marra, R. M.; Rodgers, K. A.; Shen, D.; Bogue, B. Women Engineering Students and Self-Efficacy: A Multi-Year, Multi-Institution Study of Women Engineering Student Self-Efficacy. *J. Eng. Educ.* **2009**, *98*, 27−38.
(16) Villafañe, S. M.; Garcia, C. A.; Lewis, J. E. Exploring Diverse Students' Trends in Chemistry Self-Efficacy throughout a Semester of College-Level Preparatory Chemistry. *Chem. Educ. Res. Pract.* **2014**, *15*, 114−127.
(17) Robnett, R. D.; Chemers, M. M.; Zurbriggen, E. L. Longitudinal Associations among Undergraduates' Research Experience, Self-Efficacy, and Identity: Research Experience, Self-Efficacy, and Identity. *J. Res. Sci. Teach.* **2015**, *52*, 847−867.
(18) Hofstein, A.; Lunetta, V. N. The Role of the Laboratory in Science Teaching: Neglected Aspects of Research. *Rev. Educ. Res.* **1982**, *52* (2), 201−17.
(19) Nakhleh, M. B.; Polles, J.; Malina, E. Learning Chemistry in a Laboratory Environment. In *Chemical Education: Towards Research-Based Practice*; Gilbert, J. K., De Jong, R., Justi, D. F., Treagust, J. H., Van Driel, J. H., Eds.; Kluwer: Dordrecht, Netherlands, 2002; pp 69−94.
(20) Hofstein, A.; Lunetta, V. N. The Laboratory in Science Education: Foundations for the Twenty-first Century. *Sci. Educ.* **2004**, *88*, 28−54.
(21) Reid, N.; Shah, I. The Role of Laboratory Work in University Chemistry. *Chem. Educ. Res. Pract.* **2007**, *8*, 172−185.
(22) Elliott, M. J.; Stewart, K. K.; Lagowski, J. J. The Role of the Laboratory in Chemistry Instruction. *J. Chem. Educ.* **2008**, *85*, 145−149.
(23) National Research Council. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; National Academies Press: Washington, D.C., 2012.
(24) Hilosky, A.; Sutman, F.; Schmuckler, J. Is Laboratory Based Instruction in Beginning College-Level Chemistry Worth the Effort and Expense? *J. Chem. Educ.* **1998**, *75*, 100−104.
(25) Hawkes, S. J. 2004. Chemistry Is Not a Laboratory Science. *J. Chem. Educ.* **2004**, *81*, 1257.
(26) Novak, J. D. *Learning, Creating, and Using Knowledge*; Taylor & Francis Group: New York, 2010.
(27) Novak, J. D. Human Constructivism: A Unification of Psychological and Epistemological Phenomena in Meaning Making. *Inter. J. Pers. Const. Psych.* **1993**, *6*, 167−193.
(28) Bretz, S. L. Novak's Theory of Education: Human Constructivism and Meaningful Learning. *J. Chem. Educ.* **2001**, *78*, 1107.
(29) Ausubel, D. *Educational Psychology: A Cognitive View*; Holt, Rinehart, and Winston Inc.: New York, 1968.
(30) Niedenthal, P. M. Embodying Emotion. *Science* **2007**, *316*, 1002−1005.
(31) Touroutoglou, A.; Lindquist, K. A.; Dickerson, B. C.; Barrett, L. F. Intrinsic Connectivity in the Human Brain Does Not Reveal Networks for 'Basic' Emotions. *Soc. Cogn. Aff. Neurosci.* **2015**, *10* (9), 1257−1265.
(32) Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* **2007**, *39*, 175−191.
(33) Fay, M. E.; Grove, N. P.; Towns, M. H.; Bretz, S. L. A rubric to characterize inquiry in the undergraduate chemistry laboratory. *Chem. Educ. Res. Pract.* **2007**, *8*, 212−219.
(34) Bruck, L. B.; Bretz, S. L.; Towns, M. H. Characterizing the level of inquiry in the undergraduate chemistry laboratory. *J. Coll. Sci. Teach.* **2008**, *87*, 1416−1424.
(35) Cohen, B. *Explaining Psychological Statistics*, 3rd ed.; Wiley: New York, 2008.
(36) Kirschner, P.; Meester, M.; Middelbeek, E.; Hermans, H. Agreement Between Student Expectations, Experiences, and Actual Objectives of Practicals in the Natural Sciences at the Open University of The Netherlands. *Int. J. Sci. Educ.* **1993**, *15*, 175−197.
(37) Könings, K. D.; Brand-Gruwel, S.; van Merriënboer, J. J. G; Broers, N. J. Does a New Learning Environment Come Up to Students' Expectations? A Longitudinal Study. *J. Educ. Psych.* **2008**, *100*, 535−548.
(38) Everitt, B. S.; Landau, S.; Leese, M.; Stahl, D. *Cluster Analysis*, 5th ed.; Wiley: Chichester, UK, 2011.
(39) Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236−244.
(40) Norusis, M. J. Cluster Analysis, In *Statistical Procedures Companion*; Prentice Hall: Upper Saddle River, NJ., 2011; pp 375−404.
(41) Mooi, E.; Marko, S. Cluster analysis. In *A Concise Guide to Market Research*; Springer: Berlin, 2011; pp 237−284.
(42) Burns, R. P.; Burns, R. Cluster analysis. In *Business Research Methods and Statistics Using SPSS*; Sage Publications: London, 2009; pp 552−567.
(43) Caliński, T.; Harabasz, J. A Dendrite Method for Cluster Analysis. *Commun. Statist.* **1974**, *3*, 1−27.
(44) Milligan, G. W.; Cooper, M. C. An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika* **1985**, *50*, 159−179.
(45) Aldenderfer, M. S.; Blashfield, R. K. *Cluster Analysis*; Sage Publications: Beverly Hills, CA, 1984.
(46) Kaufman, L.; Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley: New York, 1990.
(47) Cohen, J. Eta-Squared and Partial Eta-Squared in Fixed Factor ANOVA Designs. *Educ. Psychol. Meas.* **1973**, *33*, 107−112.
(48) Levine, T. R.; Hullett, C. R. Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Comm. Res.* **2002**, *28*, 612−625.
(49) Phares, E. J. Expectancy Changes in Skill and Chance Situations. *J. Abnorm. Soc. Psychol.* **1957**, *54*, 339−342.

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX

(50) Kukla, A. Foundations of an Attritubion Theory of Performance. *Psych. Rev.* **1972**, *79*, 454−470.
(51) Roese, N. L.; Sherman, J. W. Expectancy. In *Social Psychology: A Handbook of Basic Principles*; Jruglanski, A. W., Higgins, E. T., Eds.; Guilford, Press: New York, 2007; Vol. *2*, pp 91−115.

DOI: 10.1021/acs.jchemed.5b00754
J. Chem. Educ. XXXX, XXX, XXX−XXX