

Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory

Kelli R. Galloway and Stacey Lowery Bretz*

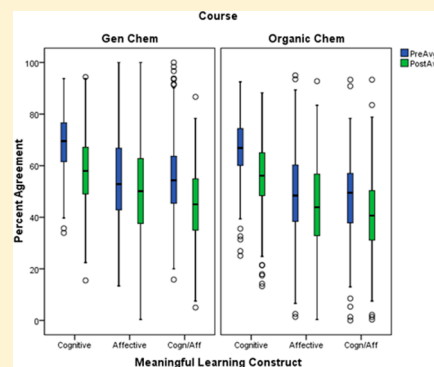
Department of Chemistry & Biochemistry, Miami University, Oxford, Ohio 45056, United States

S Supporting Information

ABSTRACT: Research on learning in the undergraduate chemistry laboratory necessitates an understanding of students' perspectives of learning. Novak's Theory of Meaningful Learning states that the cognitive (thinking), affective (feeling), and psychomotor (doing) domains must be integrated for meaningful learning to occur. The psychomotor domain is the essence of the chemistry laboratory, but the extent to which the cognitive and affective domains are integrated into the laboratory is unknown. For meaningful learning to occur in the undergraduate chemistry laboratory, students must actively integrate both the cognitive domain and the affective domains into the "doing" of their laboratory work. The Meaningful Learning in the Laboratory Instrument (MLLI) was designed to measure students' expectations before and after laboratory courses and experiences, in both the cognitive and affective domains, within the context of conducting experiments in the undergraduate chemistry laboratory. The MLLI was pilot-tested and modified based on an analysis of the pilot study data. The revised, 31-item MLLI was administered online both at the beginning and end of a semester to both general and organic chemistry laboratory students. Evidence for both the validity and reliability of the data, as well as comparisons between general and organic chemistry students' responses, are discussed.

KEYWORDS: Chemistry Education Research, Assessment, Laboratory Instruction, First-Year Undergraduate, Second-Year Undergraduate, Organic Chemistry, Hands-On Learning, Learning Theories

FEATURE: Chemical Education Research



INTRODUCTION

Bench scientists would rarely make a claim without supporting evidence. Yet, teaching scientists (in this case, teaching chemists) continually claim that the teaching laboratory is an essential aspect of the undergraduate chemistry curriculum because it is necessary to learn chemistry, all while lacking substantial evidence to support such a claim.^{1,2} Reviews on the research about the chemistry teaching laboratory frequently begin by pointing to the integral nature of the laboratory to chemistry, how chemistry is a laboratory science, and how chemists cannot imagine teaching without the laboratory.^{1,3–8} Yet, this claim is widely accepted despite the scant evidence for how and to what extent learning is taking place.

Hofstein and Lunetta's 1982 seminal review stated that (ref 3, pp 212)

Researchers have not comprehensively examined the effects of laboratory instruction on student learning and growth in contrast to other modes of instruction, and there is insufficient data to confirm or reject convincingly many of the statements that have been made about the importance and effects of laboratory teaching.

In their follow-up review 20 years later, Hofstein and Lunetta continue to point to the "sparse data" used to support the assumption that laboratory experiences help students learn.⁴ The few who do see through the unsupported claim question the time, resources, and money poured into laboratory courses

at the undergraduate level.^{6,9} There is ample reason to question why the laboratory is argued to be integral to learning chemistry.⁸ Chemists face the challenge of demonstrating that learning in the laboratory is complementary to, yet different from, learning outside the laboratory.

If the teaching laboratory does indeed provide the unique learning experiences presumed by so many, then research evidence is needed to demonstrate student learning.¹⁰ Although research has characterized faculty goals for laboratory learning, students need to be asked about their experiences.^{5,11–13} In the same way that instructors ought to measure students' prior content knowledge, students' ideas about learning in the laboratory need to be measured to design instructional materials to bridge the gap between instructor goals and student expectations. Such evidence should be generated from assessments of what and how students are learning in the laboratory.

A variety of assessment tools have previously been used to measure student perceptions of learning in the laboratory in some capacity. The Metacognitive Activities Inventory (MCAI) was designed to assess "students' metacognitive skillfulness while problem solving."¹⁴ Although problem solving is an oft-stated goal for laboratory learning, the design of the MCAI was intended for more general problem solving in chemistry, not to

focus primarily upon the laboratory. The Chemistry Expectations Survey (CHEMX) was designed to measure students' cognitive expectations for learning chemistry with seven factors, just one of which considered expectations regarding learning in the laboratory.¹⁵ To measure the affective domain, the Chemistry Self-Concept Inventory (CSCI) was designed to measure students' beliefs about their abilities to do chemistry, whereas the Attitude toward the Subject of Chemistry Inventory (ASCI) and its shortened counterpart (ASCIV2) were designed to measure student attitudes regarding the subject of chemistry.^{16–19} Similar to the ASCI, the Colorado Learning Attitudes about Science Survey (CLASS-Chem) explores student attitudes toward the discipline of chemistry and student beliefs about learning chemistry.²⁰ The aforementioned assessments, though not an exhaustive list, were intended to measure student perceptions of cognitive and affective experiences in chemistry (other than those that measure specific content knowledge) but did not specifically focus upon learning in the laboratory. Few assessments have been developed specifically for student perceptions of learning in the laboratory. Bowen developed the Chemistry Laboratory Anxiety Instrument (CLAI) as an affective measure for students in the lab, identifying when students experience anxiety regarding the lab (before, during, or after) and how that anxiety changes for different kinds of laboratory activities.²¹ Others have developed laboratory specific assessment tools to evaluate a particular, innovative laboratory curriculum.²² These types of questionnaires, however, tend to ask questions related to the new curriculum and not about learning chemistry or making meaning from laboratory work and are not broadly useful.

Although the assessments described above can provide some valuable information about students, their usefulness in the chemistry laboratory is limited because they were not developed to operationalize a learning theory in the specific context of learning in the chemistry laboratory. Research on how people learn within the domain of chemistry must be integrated into the development of assessment tools for evaluating learning.²³ Many learning theories could be used to develop a useful assessment tool for the chemistry laboratory, including Perry's Scheme of Intellectual Development,^{24,25} Kolb's Theory of Experiential Learning,²⁶ Distributed Cognition,⁵ Novak's Theory of Meaningful Learning,^{27–29} Mezirow's Theory of Meaning Making,^{30,31} or Piaget's Theory of Cognitive Development.^{32,33}

To evaluate student learning in the undergraduate chemistry laboratory, we chose to use Joseph Novak's Theory of Meaningful Learning and Human Constructivism to design the Meaningful Learning in the Laboratory Instrument (MLLI). Novak's theory states that (ref 27, pp18)

Meaningful learning underlies the constructive integration of thinking, feeling, and acting leading to human empowerment for commitment and responsibility.

Humans construct meaning from their experiences based on how they interact with the experience and the context of the experience.²⁸ The meaning a person makes from an experience depends upon the combination of thinking, feeling, and acting.^{27–29} A person chooses to act a certain way based on how they think and feel about the experience.²⁸ Likewise, in the undergraduate teaching laboratory, how a student chooses to act (psychomotor) in the lab depends on how they think about (cognitive) and feel toward (affective) their laboratory experiences. Rarely do instructors or teaching assistants allow

students to stand idle in the lab. Therefore, if *doing* is inherent in the laboratory, the question must be asked, to what extent do students integrate their thinking and feeling with the *doing*. To measure students' thinking and feeling, the MLLI was designed to be a tool that instructors and researchers can use to measure evidence of meaningful learning within the context of the undergraduate chemistry laboratory.

RESEARCH QUESTIONS

Three research questions guided the development of the MLLI:

1. Are the cognitive expectations of students fulfilled by their experiences in an undergraduate chemistry laboratory course?
2. Are the affective expectations of students fulfilled by their experiences in an undergraduate chemistry laboratory course?
3. In what ways do these expectations and experiences change as students learn more chemistry, that is, move from general chemistry to organic chemistry?

The purpose of this article is to describe (a) the development of the MLLI, (b) the validity and reliability of the data produced, and (c) the findings regarding the research questions.

METHODS

Instrument Development

The goal in developing the Meaningful Learning in the Laboratory Instrument (MLLI) was to design an assessment tool to measure students' expectations and experiences related to the cognitive and affective dimensions of their learning in an undergraduate chemistry laboratory course. To do this, a literature review was conducted to identify existing assessments that were designed to measure either students' affective domain in learning or student's ideas about laboratory work including the MCAI,¹⁴ CHEMX,¹⁵ CSCI,¹⁶ ASCI,^{17–19} CLASS-Chem,²⁰ CLAI,²¹ and the CASPiE Undergraduate Questionnaire.²²

Initially, the idea was that each MLLI item would be categorized a priori as either cognitive, affective, or psychomotor, with an equal number of items in each domain. Before individual MLLI items were written, the format of the instrument was chosen to use one common stem for each item. To narrow the focus of the instrument, the stem was written to set the context as the "doing" of experiments (as opposed to preparing for laboratory, or conducting analyses or writing lab reports after the laboratory). For the alpha version of MLLI, the response format was a four-point Likert scale from Strongly Disagree (1) to Strongly Agree (4) in order to avoid a neutral middle option and to minimize the number of answer choices. Thus, the focus of the instrument was for students to indicate to what extent they expected to have these cognitive, affective, or psychomotor experiences during their laboratory course.

Existing inventories were considered to inspire items that could be modified to fit the context of "doing" chemistry (see examples in Supporting Information). Each item on the alpha version of MLLI was reviewed and revised to facilitate categorization as a cognitive, affective, or psychomotor item, with attention paid to the number of positively and negatively worded items. The alpha version of MLLI had 79 items, including one indicator item intended as a check to ensure students were reading each item. Although the intent was to have an equal distribution for each theoretical category, as well

as balance the number of positive and negative items, the coding of the items developed showed otherwise. There was an excess of cognitive items and relatively few affective or psychomotor. Extra cognitive items were deleted and the wording of other items revised. Items that appeared to be opposites (e.g., to be excited vs to be bored; to feel comfortable when using equipment vs to feel anxious when using equipment) were kept to examine the reliability of students' responses.

Prior to data collection, this research was approved by the Institutional Review Board. The first page of the MLLI contained the consent form.

To measure student *expectations* for learning in the chemistry laboratory, the MLLI was administered during the first week of the fall semester, prior to the first experiment. Then, at the end of the semester, the MLLI was given at the time of the last experiment to measure student *experiences*. For the post-semester test, the items on MLLI were modified to change the verb tenses from future (*expectations*) to past tense (*experiences*). The content and context of the MLLI remained constant from presemester (pretest) to postsemester (post-test) administration.

Pilot Study

The alpha version of the MLLI was administered to general chemistry (GC) and organic chemistry (OC) laboratory students at both a midsize liberal arts university and a research university in the Midwest. The MLLI was administered online using Qualtrics at the research university for both the pre- ($N = 739$) and the post-test ($N = 185$). At the liberal arts institution, the MLLI was administered online using Checkbox Prezza for the pretest ($N = 184$) and then on paper for the post-test ($N = 321$). The decision to change the administration format at the liberal arts university was to increase student participation. For the pretest, an email was sent to all GC and OC laboratory students with a link to the survey. For the post-test, the MLLI was given on paper, and students were asked to complete it during the first 10 min of their lab period.

To analyze the data, scores on individual items were summed together for a composite score that could range from 78 (answering Strongly Disagree to each item) to 312 (answering Strongly Agree to each item). Inversely worded items were recoded so that a higher score indicated a positive contribution to meaningful learning. Scores were compared between the pre- and post-test. For the pilot study, a repeated measures analysis could not be calculated, however, because student responses could not be matched from pretest to post-test.

Using the data from the pilot test, the MLLI was revised using (1) the Meaningful Learning theoretical framework and (2) the statistical technique of exploratory factor analysis (EFA). The pretest data from the RI institution were used for the EFA because they had the larger sample size. Given that there were only four answer choices for each item, a multivariate normal distribution could not be assumed.³⁴ Thus, Principal Axis Factoring was used as the factor extraction method. A Promax rotation was used to allow for the factors to be correlated given that meaningful learning would necessitate an integration of thinking, feeling, and doing. The conventional use of EFA is to identify which items on an assessment group together to measure a latent construct and then delete the remaining items.³⁴ For this analysis, however, EFA was used to identify redundant items that might be eliminated, as they would factor together. Items that did not appear to load onto a

factor were also of interest to be retained because perhaps they represented expectations/experiences that were not captured by the other items.³⁴

Other considerations when revising items included using more precise language and modifying wording to constrain students to consider only their time in the chemistry laboratory course and not time spent working on the report or other outside of lab activities. For example, the item "to worry about getting the right answer" was changed to "to worry about the quality of my data" because using the words "an answer" could indicate a response on a lab report prepared outside of the lab room, whereas a student collects data in real time in the lab room.

Initially, the items were coded independently by three chemistry education researchers using the meaningful learning framework using any combination of the three domains. The initial interrater agreement for this process was 63%. After coding the items, the raters discussed their disagreements and found challenges arose with the ambiguity of integrating the psychomotor domain into each item. During the analysis of the pilot test data, it became apparent that the psychomotor domain was inherent in the MLLI stem (when *performing* experiments...). Every item was already situated in this context. Thus, there would be no purely psychomotor items but rather items that connect the affective with the psychomotor or the cognitive with the psychomotor. Taking the inherent psychomotor aspect of each item into consideration, consensus was reached on the coding of the items for the pilot version of the MLLI.

Upon shortening and revising the MLLI, a second interrater procedure was carried out with two raters who recoded the items as cognitive, affective, or cognitive/affective. This procedure had an initial 83% agreement. Here, the disagreements came from discerning the overlap between the cognitive and the affective domains. In order to reach agreement, the cognitive and affective parts of each item were explicitly identified. Items were deemed purely cognitive if they only dealt with thinking about the concepts and not having to do with feelings or attitudes about the laboratory work (e.g., Q5 "to make decisions about what data to collect"). The classification of cognitive/affective was used for items that explicitly included both domains (e.g., Q4 "to *feel* (affective part) unsure about the *purpose of the procedures* (cognitive part)"). After discussion, 100% agreement was reached.

Full Study

The revised (and final) version of the MLLI consists of 31 items, including 1 indicator item. There are 16 cognitive, 8 affective, and 6 cognitive/affective items, of which 16 are positively worded items and 14 are negatively worded items. The answer format was modified following the pilot study. The 4-point Likert scale was replaced with a slider bar (Figure 1) ranging from 0% (Completely Disagree) to 100% (Completely Agree). As with the pilot version, students were asked to indicate their percent agreement with each statement.

For the full study, the MLLI was administered to GC and OC laboratory students using Qualtrics survey software both at the beginning and end of the fall 2013 semester. At the liberal arts university, over 800 students took the instrument during the pre- and the post-test, and 614 students had matched responses from both the pre and the post. Of the 614 students, there were 436 GC students and 178 OC students. (In order to increase the response rate and participation, some lab

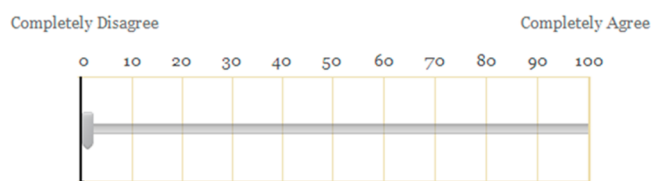


Figure 1. Screenshot of the slider bar students used to indicate their percent agreement with each statement during the full study. Students clicked and dragged the bar to their desired percent. If students wished to mark 0%, they had to drag the bar out and then back to ensure they were interacting with the instrument.

instructors offered nominal extra credit to students who answered the MLLI both at the beginning and end of the semester.)

Item scores were averaged together to calculate a composite score that ranged from 0 to 100. Composite scores were calculated not only for the total score but also for the cognitive items, the affective items, and the cognitive/affective items. To calculate the composite scores, negatively worded items (experiences that inhibited meaningful learning) were reverse coded so that higher scores represented expectations/experiences that contributed to meaningful learning and lower scores those that hindered meaningful learning. The data from the GC and OC courses were analyzed separately as the courses have inherently different goals and curricula.

Exploratory factor analysis was conducted for the final version of the MLLI to examine the internal structure of the instrument. Principal axis factoring was chosen again as the factor extraction method due to deviations from normality for each item.³⁴ Promax rotation was selected to allow for the factors to be correlated given that meaningful learning would necessitate an integration of thinking and feeling with the doing of lab work.³⁴ The EFA was carried out twice (once for the pretest $N = 524$ and once for the post-test $N = 501$) with the GC data in order to ensure a sufficiently large number of responses needed for the EFA. The results of both EFAs were analyzed and are discussed below.

To examine how students interpreted the item wording, validation interviews were conducted with 13 GC and OC laboratory students as part of a larger qualitative research protocol. The MLLI item validation was the second phase of a three-phase interview protocol in this separate study. For this phase, each MLLI item was printed on a 3×5 index card and displayed on a table in front of the student. The interviewer asked each student to explain what each item meant in their own words, and then asked the student to give an example of when the circumstance occurred for them during their chemistry laboratory course. Because there were a possible 30 items, no student was asked to discuss all of the items. The

number of items discussed during each interview ranged from 5 to 17, with an average of 11 items per interview.

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 summarizes the descriptive statistics for the MLLI full study (i.e., pre vs post, GC vs OC, cognitive vs affective vs cognitive/affective scales). Figure 2 shows the distributions of the responses on each scale. Scatterplots were constructed to examine pre- vs post-test scores and the spread of scores. There does not appear to be a ceiling or floor effect, that is, an excess of scores at the high or low ends of the scales.

Figure 2 also reveals different distributions for each of the three scales. The GC cognitive scale shows a narrow distribution of responses that appear to change together from pre- to post-test, whereas the affective distribution is diverse. This apparent disparity between the cognitive and affective scales can be explained by considering how students are selected for admission to the university. High achieving individuals are admitted based on their cognitive ability, but not based upon their affective ideas about learning. This phenomenon is not as apparent for OC students, but these students have completed one year of chemistry laboratory courses upon which to form expectations about organic chemistry lab. The cognitive/affective plots fall between those of the cognitive and affective scales, with more variation than the cognitive scales, but not as much as the affective scales. Each scatterplot includes a $y = x$ line to visualize how students change from pre- to post-test. If students' scores remained unchanged from pre- to post-test, they fall on the $y = x$ line. Students whose experiences surpassed their expectations fall above the line (an increase in percent agreement), whereas those whose expectations went unfulfilled by their experiences fall below the line (a decrease in percent agreement). For both GC and OC, there appears to be a decrease in scores for cognitive and cognitive/affective scores. The affective plots show approximately equal numbers of students above and below the $y = x$ line.

Scatterplots were constructed to examine the relationship between the cognitive and affective scales. Figure 3 shows the GC and OC scatterplots for affective vs cognitive for both pre- and post-test administrations. These plots also include a $y = x$ line. Here, this line represents the point at which cognitive and affective scores are equal (not a change in score as in Figure 2). Points above the line indicate relatively higher affective scores, whereas points below the line indicate relatively higher cognitive scores. Each plot in Figure 3 shows a larger number of students with higher cognitive responses than affective. Few students are engaging their thoughts *and* feelings while doing their laboratory experiments.

Table 1. Descriptive Statistics for MLLI pre- and post-test administrations

		Pretest			Post-test		
		Cognitive	Affective	Cogn/Aff	Cognitive	Affective	Cogn/Aff
General Chem $N = 436$	Mean (SD)	69.9 (10.8)	54.7 (17.0)	55.1 (14.7)	57.9 (12.9)	50.3 (18.1)	44.8 (14.6)
	δ	0.97	0.98	0.98	0.97	0.98	0.98
	α	0.77	0.79	0.63	0.76	0.80	0.60
Organic Chem $N = 178$	Mean (SD)	65.9 (12.1)	48.6 (18.6)	47.3 (15.4)	55.8 (14.2)	44.6 (17.1)	40.9 (15.5)
	δ	0.96	0.98	0.97	0.97	0.98	0.98
	α	0.79	0.82	0.63	0.81	0.78	0.62

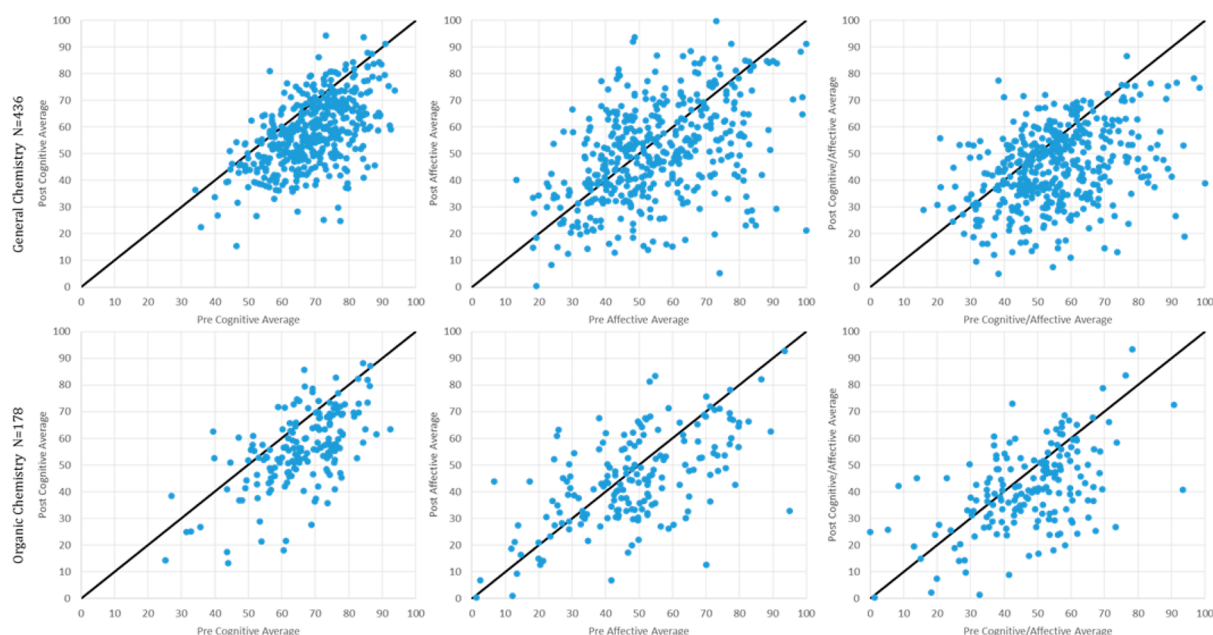


Figure 2. Scatterplots comparing pre- and post-test responses on the cognitive, affective, and cognitive/affective scales. The top row shows the general chemistry students ($N = 436$) and the bottom row shows organic chemistry students ($N = 178$).

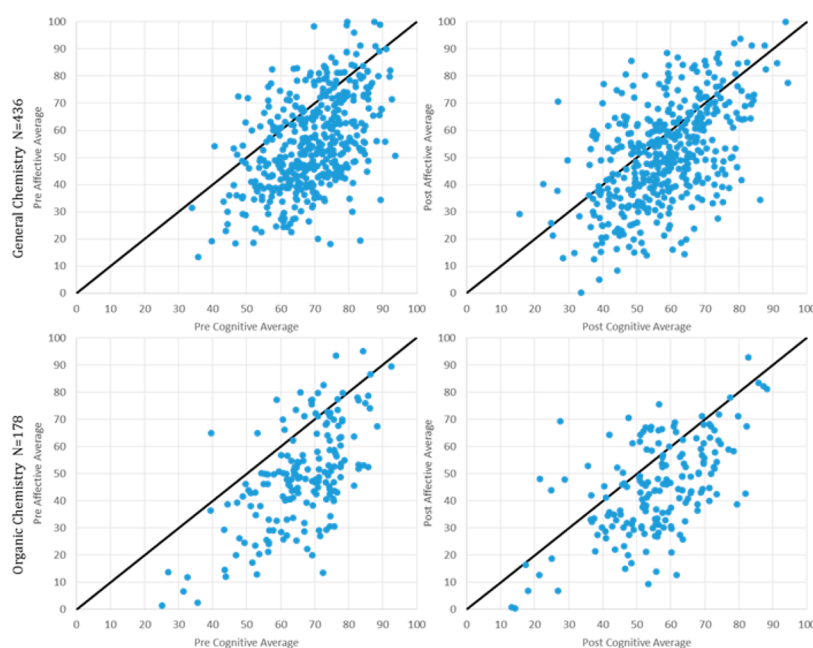


Figure 3. Scatterplots of affective vs cognitive responses for pre- and post-tests. The top row shows the general chemistry students ($N = 436$) and the bottom row shows organic chemistry students ($N = 178$).

Reliability

Table 1 shows Cronbach α and Ferguson's δ for test reliability. These statistics were calculated for the MLLI's three subscales. The traditionally targeted threshold for Cronbach α is 0.7. For both GC and OC, the pre- and post-test cognitive and affective scales have an α greater than 0.7. The GC and OC cognitive/affective scales for both pre- and post-test are less than 0.7. The Cronbach α is a measure of the average of correlation of all the split-halves of a test or scale.³⁵ In nontechnical terms, Cronbach α is a measure of how correlated students' responses are on a set of items. Responses on the cognitive and affective items appear to be internally consistent, but the cognitive/affective

items do not yield the same evidence. One explanation for the lower α is that the cognitive and affective domains are not integrated in the mind of the student. Instead, the cognitive and affective ideas are more segregated in the students' minds as seen by the higher α for those individual scales. In the same way that low α values on a diagnostic assessment can indicate fragmented knowledge in the mind of the learner, a lower α on the cognitive/affective scale can indicate disconnected ideas about learning.^{27,36} The α values do appear to be consistent from the pre- to post-test, indicating consistency (reliability).

Ferguson's δ was calculated as a measure of test discrimination. This statistic is a measure of the distribution of scores compared to the possible range of scores.^{37,38}

Generally accepted values of Ferguson's δ are greater than 0.9. Values for the MLLI exceed 0.96, indicating that the MLLI can discriminate among students' differing expectations and experiences.

Validity

Validity can be examined in different ways, but establishing validity of the data is necessary in order to draw conclusions based upon it.^{39,40} The validity of the MLLI data was assessed by examining both student response processes and the internal structure of the instrument. Again, the goal of the MLLI is to understand students' perceptions of learning in their undergraduate chemistry laboratory courses. Novak defines meaningful learning as an integration of thinking (cognitive), feeling (affecting), and doing (psychomotor).²⁷ Although students are innately *doing* in the chemistry laboratory, the MLLI assesses the extent to which students think and feel about their laboratory work and the connections between thinking and feeling. Therefore, establishing the validity of MLLI data requires an examination of how the data can be interpreted within this learning theory.

Student response processes were analyzed through student interviews. The purpose of the item validation phase of the interview was to elicit how students interpreted the wording of each item. When the MLLI was revised from the pilot to the full version, the wording of the items was modified to minimize the use of pronouns and ambiguity. During the validation interviews, students demonstrated understanding and interpretation of the items as they were intended. No items were considered ambiguous, nor revised, as a result of the interviews.

To illustrate this process, an example regarding interview responses about item 4 "to feel unsure about the purpose of procedures" is provided. Five of the 13 students discussed this question during their interview. Each student discussed how being unsure about the purpose of the procedures means not knowing *why* the procedures should be done. (All students were assigned a pseudonym in the transcription and reporting of their ideas.) Jan (OC) cited that she did not need to understand why something had to be done a certain way (a specific set up or order of steps) in order for an experiment to work properly. Angela (GC) talked about how she can understand the goal of an experiment but sometimes does not understand the individual steps that add up to achieving that goal. Another GC student, Pam, offered a nonchalant response saying that her strong high school chemistry background helped her to understand the procedures pretty well. The sum of these students paraphrasing the item and describing personal experiences demonstrates adequate student interpretation of this MLLI item.

The validation interviews also provided evidence for how an item was classified via the meaningful learning theory. For example, before analyzing the validation interviews, the researchers debated whether item 27, "to be intrigued about the instruments," should be cognitive, affective, or cognitive/affective. Namely, did the word "intrigued" have both a cognitive and an affective component, or just one, but not both. Analysis of students' interpretations of this item showed that students used both cognitive words (understand, learn, etc.) and affective words (cool, fancy, appreciate, etc.) when discussing this item. These interpretations give evidence that item 27 was best classified as a cognitive/affective item because of the overlap of cognitive and affective descriptions.

The internal structure of the MLLI was examined by conducting an exploratory factor analysis (EFA). Using the eigenvalues and pattern matrix to identify which items grouped together, the most logical structure, for both pre- and post-test, was a two-factor structure where one factor contained the items contributing to meaningful learning (positively worded items) and the second factor contained the items inhibiting meaningful learning (negatively worded items). In addition, there was one item (Q15) that did not load on either factor and one item (Q17) that loaded on both factors.

Previous work has demonstrated the difficulty of using EFA to measure the internal validity of assessments as well as the apparent presence of a two-factor solution with positively and negatively worded items in separate factors.^{14,41} Incorporating the use of positively and negatively worded items is helpful for preventing acquiescence bias, but they can pose challenges with EFA results.⁴¹ These EFA solutions did not invalidate the proposed theoretical structure of the MLLI. Rather, these solutions revealed important information about how students interpreted the MLLI items and how factor analysis alone is limited in discerning internal structure of assessments. The results of any EFA can only uncover response patterns on the assessment; therefore, if the theoretical structure employs domains that are not made explicit to the respondent, then it is not unsurprising to obtain results indicating an absence of such a structure. These solutions suggest that students were either not connecting their thinking and feeling with their doing of chemistry lab or they were unaware of their thinking and feeling while conducting their chemistry laboratory experiments. These results have implications for how laboratory work is currently presented and how it could be presented to students. Namely, it could be beneficial to make explicit to students the necessity to engage both their thinking and feeling with their laboratory work for meaningful learning to occur.

Cronbach α was calculated for each scale as the items were sorted a priori based on learning theory. An analysis of α "if item deleted" was also conducted to examine how well the items fit together within the proposed structure. In such an analysis, if α were to increase when an item was removed, then that item could be interpreted as not contributing as much to the construct of the factor as the other items in the scale. Two MLLI items were identified for possible deletion in the α "if item deleted" analysis. For the cognitive scale, Q15 (the procedures to be simple to do) and Q17 (to "get stuck" but keep trying) both increased α if deleted for both the pre- and post-test data. However, examination of the student validation interviews revealed that these items are influenced by students' prior experiences. Students could view these experiences differently, contributing to their *apparent* disconnect with the other cognitive items. Some students could view the procedures as too simple (e.g., they had already mastered these skills in high school) and wanted to be challenged, whereas other students without such prior experiences may appreciate the simple, straightforward nature of some procedures. The same could be true for Q17: some students could appreciate the challenge of "getting stuck" and working through difficulties, whereas other students could fear "getting stuck." Interpretation of the α "if deleted" alone is not an adequate reason to remove these items nor to assign them to a different scale. Rather, these examples demonstrate the complexity of item function within the context of measuring student perceptions of learning in the laboratory.

Change Over Time

Inferential statistics were used to analyze both changes in students' responses from pre- to post-test and differences between GC and OC students. First, a two-way repeated measure analysis of variance (ANOVA) was conducted to investigate whether GC and OC students' ideas about learning in their chemistry laboratory courses changed differently over the course of a semester. The dependent variable was the composite score on a subscale of the MLLI, so three two-way repeated measures ANOVAs were conducted to examine each scale individually. The between groups factor was course with two levels (GC and OC), and the within factor was time (pre- and post-test). GC students were expected to demonstrate greater change in their ideas about learning in the chemistry lab due to the fact that GC students have limited prior laboratory experiences and the diversity of students who take general chemistry is greater than those who take organic chemistry. OC students could be considered to be a less diverse population because fewer majors require OC than GC, and there is often a high attrition rate from GC to OC. Also, OC students have already taken a year of chemistry lab, so it could be expected that their ideas about learning in lab would change less than GC students due to their common prior experiences. The two-way repeated measures ANOVA allows for the testing of these differences.

The assumptions for this test include independent random samples, multivariate normal distributions, homogeneity of variance, and homogeneity of covariance across groups. Students self-selected to take the MLLI, so conclusions must be drawn carefully about chemistry students at the university where the study occurred. Normality was assessed with the Shapiro–Wilk test. While some deviation from normality was detected, the repeated measures design is robust to departures from normality.⁴² Levene's Test for homogeneity of variance and Box's M for homogeneity of covariance were all nonsignificant, indicating no differences in variance and covariance.

Table 2 displays the results from the three two-way repeated measures ANOVAs. The main effects for course and time are

Table 2. Results from Mixed Model ANOVA

Factor	Two-Way RM ANOVA
Cognitive	
Course	Wilks's $\Lambda = 0.985$, $F(1, 612) = 7.04$, $p = 0.008$, $\eta^2_p = 0.015$
Time	Wilks's $\Lambda = 0.596$, $F(1, 612) = 414.67$, $p \leq 0.0001$, $\eta^2_p = 0.40$
Time*Course	Wilks's $\Lambda = 0.999$, $F(1, 612) = 0.873$, $p = 0.873$, $\eta^2_p = 0.00$
Affective	
Course	Wilks's $\Lambda = 0.969$, $F(1, 612) = 19.2$, $p \leq 0.0001$, $\eta^2_p = 0.03$
Time	Wilks's $\Lambda = 0.957$, $F(1, 612) = 27.6$, $p \leq 0.0001$, $\eta^2_p = 0.04$
Time*Course	Wilks's $\Lambda = 1.00$, $F(1, 612) = 0.037$, $p = 0.848$, $\eta^2_p = 0.00$
Cognitive/Affective	
Course	Wilks's $\Lambda = 0.947$, $F(1, 612) = 27.3$, $p \leq 0.0001$, $\eta^2_p = 0.05$
Time	Wilks's $\Lambda = 0.821$, $F(1, 612) = 133.2$, $p \leq 0.0001$, $\eta^2_p = 0.18$
Time*Course	Wilks's $\Lambda = 0.989$, $F(1, 612) = 7.0$, $p = 0.008$, $\eta^2_p = 0.01$

significant for each scale. For the between factor Course, this significant result means that GC students scored higher than OC students on cognitive ($\eta^2_p = 0.015$), affective ($\eta^2_p = 0.03$), and cognitive/affective scales ($\eta^2_p = 0.05$).^{43,44} For the within factor Time, this significant result means that all students decreased in scores from pre- to post-test with $\eta^2_p = 0.44$ for

cognitive, $\eta^2_p = 0.04$ for affective, and $\eta^2_p = 0.18$ for cognitive/affective, indicating small to large effects.^{43,44} Stated another way, students had higher expectations for their chemistry laboratory courses that went unmet by the end of the semester.

A significant interaction term would indicate that GC students changed differently than OC students from pre- to post-test. The only significant interaction was for the cognitive/affective scale but with a trivial effect size ($\eta^2_p = 0.01$). This significant result indicates that GC students had a greater decrease in their cognitive/affective scores than OC students. The effect size, however, indicates that any difference in how general chemistry students changed compared to the organic chemistry students is too small to be meaningful. The nonsignificant interactions for the cognitive and affective scales are interesting to note because these results indicate that a student is likely to decrease in their ideas about learning in the lab no matter if they are in general or organic chemistry lab.

A one-way repeated measure ANOVA was used to analyze the change from pre- to post-test for each course individually (see Table 3). The decrease for cognitive and cognitive/affective

Table 3. Results from Repeated Measures ANOVA for GC and OC separately

Course	RM ANOVA Results Comparing Pre- to Post-Test on Three Scales	Pairwise Comparisons p , η^2		
		Cognitive	Affective	Cogn/Aff
GC $N = 436$	Wilks's $\Lambda = 0.465$, $F(4, 433) = 166.2$, $p \leq 0.0001$, $\eta^2 = 0.53$	<0.0001, 0.48	<0.0001, 0.05	<0.0001, 0.28
OC $N = 178$	Wilks's $\Lambda = 0.565$, $F(3, 175) = 44.8$, $p \leq 0.0001$, $\eta^2 = 0.43$	<0.0001, 0.42	<0.001, 0.06	<0.0001, 0.15

affective was a large effect, with a medium effect for the decrease in the affective scale for both GC and OC. Comparing the ANOVA results to the scatterplots (Figure 2) reveals that the small effect for the decrease in affective scores is actually the result of an almost equal number of students increasing in affective scores as the number of students who decreased. This same phenomenon is not seen, however, with the cognitive scores. Rather, the sample appears to shift to lower scores as a whole. When laboratory curricula are developed, the focus is on what students will do and think in lab but typically does not consider how students will feel in lab. As a result, an entire course may have similar cognitive and psychomotor experiences but have a diversity of feelings toward what they're thinking and doing. In order to move toward meaningful learning experiences for students in the undergraduate chemistry laboratory, the affective domain must also be attended to when designing laboratory curricula. These findings echo a call from the NCR DBER report that "researchers and instructors should not consider cognitive and affective development apart from each other."²

Individual items were also analyzed for differences between GC and OC using boxplots of the students' responses. Some items showed a similar pattern for both GC and OC, whereas others suggested a difference between the two courses. For example, the boxplot for item 11 (Figure 4), "to think about what the molecules are doing," showed similar responses at the pre- and post-test for GC and OC. Students expected to think at a molecular level, but this expectation was clearly not met. Item 28, "to feel intimidated," is an example of how GC and

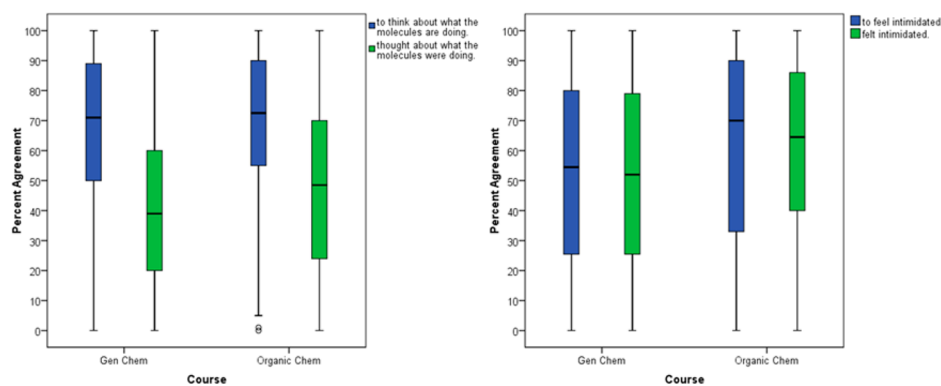


Figure 4. Boxplots of MLLI items 11 (left) and 28 (right).

OC responded differently (Figure 4). Both courses show a diversity of responses to this item, but OC clearly indicates a higher expectation to feel intimidated. The OC response does decrease a small amount at the post-test, but the median response remains higher than the GC median response. The median for this item remains almost constant for GC.

CONCLUSIONS

The MLLI was designed using meaningful learning theory to measure students' *cognitive* and *affective* expectations and experiences within the context of *doing* in their undergraduate chemistry laboratory courses. This article describes the development, pilot testing, revisions, and analyses to support the claim that MLLI can generate valid and reliable data. The MLLI is able to discriminate among a variety of students' expectations and experiences, as well as between GC and OC students. The scatterplots and the Ferguson's δ values demonstrate that students use the entire range of the 0–100% slider bar scale. The Cronbach α values provide evidence for the theoretical scales, rather than the two factor structure proposed by the exploratory factor analysis. Results from the inferential statistics show that GC and OC students change in a similar way from pre- to post-test for cognitive, affective, and cognitive/affective experiences. The OC students consistently scored lower on the MLLI than GC students, but students in both courses responded that their experiences that failed to meet their expectations.

Implications for Teaching

The development of the MLLI offers laboratory instructors a unique opportunity to assess meaningful learning in the context of their specific undergraduate laboratory instruction. Lab instructors can use the MLLI to assess the current experiences of students and use that data as evidence to improve the curriculum and/or pedagogy to bring about more meaningful learning, that is, increased cognitive experiences, increased affective experiences, and increased integration of these two domains with one another. Analysis of local results could also include a holistic examination of which expectations went unmet by students' experiences, whether students enter with unrealistic expectations, and whether instructors ought to explicitly dialogue with students about their expectations.¹⁵ The question of whether students' expectations of learning actually affect their experiences and learning merits investigation. Certainly laboratory curricula are planned in advance and would be challenging to modify based on students' pretest expectations. It is appropriate to analyze results by comparing pre- to post-test and looking at changes in responses for

measuring cognitive and affective expectations.^{14,15,19,20,44–46} Such analyses allow for the instructor to make informed curricular and pedagogical decisions for subsequent years. Changes to any curriculum should not be made without first reflecting on the current learning goals for the course both explicitly stated in the syllabus and implicitly included in each experiment. With such goals in mind, instructors should consider what responses they might expect from their students on the MLLI. Then the results from the MLLI can help instructors see how their intended curriculum is operationalized by the students and instructors. These results can then point instructors in the direction of where to make changes.

Implications for Research

Although the MLLI is not the first assessment to target some dimension of learning in the chemistry laboratory, it is the first to focus solely on learning in the laboratory and to expressly operationalize a theory of learning. Effects from modifying curriculum or pedagogy can only be meaningfully interpreted when those modifications are measured by an instrument grounded in a theory that explains how people learn. Using a learning theory framework to develop assessment tools allows for measurement of specific experiences that promote or inhibit learning. Researchers can use MLLI to design experiments to assess innovative laboratory curricula and pedagogies beyond measuring how much the students "liked" the experiments. By creating an instrument whose results can be interpreted through the lens of how people learn, the data analyses are now directly applicable to how laboratory pedagogy and curriculum could be improved. The development of the MLLI can serve as a model for assessment development using learning theory and lead to the development of future assessments regarding learning in the undergraduate chemistry laboratory using other learning theories. For example, an instrument grounded in Kolb's active experimentation vs reflective observation²⁶ would be very useful in identifying the strengths, weaknesses, and needed improvements in the undergraduate chemistry laboratory.

The results of this research study also point toward the value of multiple representations for data visualization beyond tables and histograms. Visualizing the MLLI data through scatterplots facilitated multiple interpretations and conclusions to be drawn from just a few plots. Histograms were constructed, but the scatterplots offered more meaningful comparisons of more than one variable (two time points or two scales). Scatterplots and boxplots were constructed to analyze how responses on individual items changed, but the boxplots more clearly displayed a shift from pre- to post-test. More subtle

interpretations were possible from the data visualizations because of interest in understanding the *range* of responses rather than just focusing upon single points by comparing means and standard deviation.

Limitations

The claims made as a result of MLLI data collection should be situated in the context by which the data were collected. For both the pilot and the full studies with the MLLI, data were collected at a midsize liberal arts university. The pedagogy and curriculum were not explicitly characterized as a focus of this research; therefore, the results from this study cannot be extrapolated to all pedagogies and all curricula. Students were emailed the link to the assessment, and although some instructors did offer nominal extra credit points for participating, students were not required to participate. Students had the opportunity to complete the survey at their convenience, which could have been midafternoon or midnight. The choice to administer the MLLI online with the slider bar allowed for students to actively interact with the survey and to give their honest perspective. Attrition does constrain interpretation of the data because students chose whether to participate and complete the MLLI. Perhaps only the students with extreme experiences (positive or negative) chose to take the survey. Nonetheless, in this study, visual examination of the data shows that students who completed the MLLI had a range of experiences. Although there was attrition in the sample size from pre- to post-test, the results indicate that the MLLI is able to discriminate between different student experiences.

In this study, criterion validity was not analyzed. The absence of measures of learning in the undergraduate chemistry laboratory was a primary motivator for creating the MLLI. Examining criterion validity would require stipulating another outcome measure that is known to generate reliable and valid data. No such measure was available. However, now that the MLLI exists, it could be used as an outcome measure in the development of other assessments to examine criterion validity.

Analyses of MLLI data are intended to compare student responses over time, but for this study, students were only measured at two time points within the context of one chemistry course. The general to organic chemistry comparison is made using cross-sectional data, not longitudinal data. Students would need to be tracked from general through organic chemistry and beyond to make more robust conclusions about decreased responses in organic chemistry.

Lastly, the findings from this study should not be extrapolated from the undergraduate teaching laboratory to undergraduate research experiences. Such experiences merit their own investigation to understand how students learn through them. Because the MLLI was designed for the undergraduate teaching laboratory, it would clearly need revision to be used to measure students' perceptions of learning through undergraduate research experiences.

■ ASSOCIATED CONTENT

§ Supporting Information

Examples of how items were constructed and modified, along with a full list of the MLLI items (worded for both pretest expectations and post-test experiences), including their categorization as cognitive, affective, or cognitive/affective, is provided. This material is available via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bretzsl@miamioh.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported 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 would like to thank Ellen Yezierski for her invaluable feedback on the research and statistics advice.

■ REFERENCES

- (1) Elliott, M. J.; Stewart, K. K.; Lagowski, J. J. The Role of the Laboratory in Chemistry Instruction. *J. Chem. Educ.* **2008**, *85*, 145–149.
- (2) National Research Council. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; National Academies Press: Washington, DC, 2012.
- (3) 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.
- (4) Hofstein, A.; Lunetta, V. N. The Laboratory in Science Education: Foundations for the Twenty-first Century. *Sci. Educ.* **2004**, *88*, 28–54.
- (5) 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.
- (6) Reid, N.; Shah, I. The Role of Laboratory Work in University Chemistry. *Chem. Educ. Res. Pract.* **2007**, *8*, 172–185.
- (7) Sevan, H.; Fulmer, G. W. Student Outcomes from Innovations in Undergraduate Chemistry Laboratory Learning: A Review of Projects Funded by the US National Science Foundation Between 2000–2008. *Educación Química* **2012**, *23*, 149–161.
- (8) NSTA Board of Directors. “NSTA Position Statement, The Integral Role of Laboratory Investigations in Science Instruction”. <http://www.nsta.org/about/positions/laboratory.aspx> (accessed Feb 2015).
- (9) Hawkes, S. J. Chemistry Is Not a Laboratory Science. *J. Chem. Educ.* **2004**, *81*, 1257.
- (10) 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.
- (11) Bretz, S. L.; Fay, M.; Bruck, L. B.; Towns, M. H. What Faculty Interviews Reveal About Meaningful Learning in the Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2013**, *90*, 281–88.
- (12) Bruck, L. B.; Towns, M. H.; Bretz, S. L. Faculty Perspectives of Undergraduate Chemistry Laboratory: Goals and Obstacles to Success. *J. Chem. Educ.* **2010**, *87*, 1416–24.
- (13) Bruck, A. D.; Towns, M. H. Development, Implementation, and Analysis of a National Survey of Faculty Goals for Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2013**, *90*, 685–693.
- (14) Cooper, M. M.; Sandi-Urena, S. Design and Validation of an Instrument to Assess Metacognitive Skillfulness in Chemistry Problem Solving. *J. Chem. Educ.* **2009**, *86*, 240–245.
- (15) Grove, N.; Bretz, S. L. CHEMX: An Instrument to Assess Students' Cognitive Expectations for Learning Chemistry. *J. Chem. Educ.* **2007**, *84*, 1524–1529.
- (16) Bauer, C. Beyond “Student Attitudes”: Chemistry Self-Concept Inventory for Assessment of the Affective Component of Student Learning. *J. Chem. Educ.* **2005**, *82*, 1864–1870.

- (17) Bauer, C. Attitude Towards Chemistry: A Semantic Differential Instrument for Assessing Curriculum Impact. *J. Chem. Educ.* **2008**, *85*, 1440–1445.
- (18) Xu, X.; Lewis, J. E. Refinement of a Chemistry Attitude Measure for College Students. *J. Chem. Educ.* **2011**, *88*, 561–568.
- (19) Brandriet, A. R.; Xu, X.; Bretz, S. L.; Lewis, J. E. Diagnosing Changes in Attitude in First Year College Chemistry Students with a Shortened Version of Bauer's Semantic Differential. *Chem. Educ. Res. Pract.* **2011**, *12*, 271–278.
- (20) Barbera, J.; Adam, W. K.; Wieman, C. E.; Perkins, K. K. Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *J. Chem. Educ.* **2008**, *85*, 1435–1439.
- (21) Bowen, C. W. Development and Score Validation of a Chemistry Laboratory Anxiety Instrument (CLAI) for College Chemistry Students. *Educ. Psych. Meas.* **1999**, *59*, 171–185.
- (22) Russell, C. B. *Development and Evaluation of a Research-Based Undergraduate Laboratory Curriculum*. Ph.D. Dissertation. Purdue University, 2006.
- (23) Bretz, S. L. Navigating the Landscape of Assessment. *J. Chem. Educ.* **2012**, *89*, 689–691.
- (24) Finster, D. C. Developmental Instruction: Part I. Perry's Model of Intellectual Development. *J. Chem. Educ.* **1989**, *66*, 659–661.
- (25) Finster, D. C. Development Instruction: Part II. Application of the Perry Model to General Chemistry. *J. Chem. Educ.* **1991**, *68*, 752–756.
- (26) Towns, M. H. Kolb for Chemists: David A. Kolb and Experimental Learning Theory. *J. Chem. Educ.* **2001**, *78*, 1107.
- (27) Novak, J. D. *Learning, Creating, and Using Knowledge*; Taylor & Francis Group: New York, NY, 2010.
- (28) Novak, J. D. Human Constructivism: A Unification of Psychological and Epistemological Phenomena in Meaning Making. *J. Constructivist Psych.* **1993**, *6*, 167–193.
- (29) Bretz, S. L. Novak's Theory of Education: Human Constructivism and Meaningful Learning. *J. Chem. Educ.* **2001**, *78*, 1107–1116.
- (30) Mezirow, J. *Transformative Dimensions of Adult Education*; Jossey-Bass: San Francisco, CA, 1991.
- (31) Wink, D. J. Reconstructing Student Meaning: A Theory of Perspective Transformation. *J. Chem. Educ.* **2001**, *78*, 1107.
- (32) Herron, J. D. Piaget for Chemists. *J. Chem. Educ.* **1975**, *52*, 146–150.
- (33) Bunce, D. M. Does Piaget Still Have Anything to Say to Chemists? *J. Chem. Educ.* **2001**, *78*, 1107.
- (34) Fabrigar, L. R.; Wegener, D. T. *Exploratory Factor Analysis*; Oxford University Press: New York, NY, 2012.
- (35) Cronbach, L. J. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika* **1951**, *16*, 297–334.
- (36) Bretz, S. L.; McClary, L. Students' Understanding of Acid Strength: How Meaningful Is Reliability When Measuring Alternative Conceptions? *J. Chem. Educ.* **2015**, *92* (2), 212–219.
- (37) Ferguson, G. On the Theory of Test Discrimination. *Psychometrika* **1949**, *1*, 61–68.
- (38) Hankins, M. Questionnaire Discrimination: (Re)-Introducing Coefficient δ . *BMC Medical Research Methodology* **2007**, *7*, 19.
- (39) American Educational Research Association; American Psychological Association; National Council on Measurement in Education; Joint Committee on Standards for Educational and Psychological Testing (US) *Standards for Educational and Psychological Testing*; American Educational Research Association: Washington, DC, 1999.
- (40) Arjoon, J. A.; Xu, X.; Lewis, J. E. Understanding the State of the Art for Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *J. Chem. Educ.* **2013**, *90*, 536–545.
- (41) Marsh, H. W.; Scalas, L. F.; Nagengast, B. Longitudinal Tests for Competing Factor Solutions for the Rosenberg Self-Esteem Scale: Traits, Ephemeral Artifacts, and Stable Response Scales. *Psych. Assess.* **2010**, *22*, 366–381.
- (42) Cohen, B. *Explaining Psychological Statistics*, 3rd ed.; Wiley: New York, 2008; pp 528–547.
- (43) Cohen, J. Eta-Squared and Partial Eta-Squared in Fixed Factor ANOVA Designs. *Educ. Psych. Meas.* **1973**, *33*, 107–112.
- (44) Levine, T. R.; Hullett, C. R. Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Commun. Res.* **2002**, *28*, 612–625.
- (45) Brandriet, A. R.; Ward, R. M.; Bretz, S. L. Modeling meaningful learning in chemistry using structural equation modeling. *Chem. Educ. Res. Pract.* **2013**, *14*, 421–430.
- (46) Chan, J. K. Y.; Bauer, C. F. Identifying at-risk students in general chemistry via cluster analysis of affective characteristics. *J. Chem. Educ.* **2014**, *91*, 1417–1425.