Portland State University

PDXScholar

Chemistry Faculty Publications and Presentations

Chemistry

2-4-2020

Assessing Differences between Three Virtual General Chemistry Experiments and Similar Hands-On Experiments

Cory Hensen
Portland State University, chensen@pdx.edu

Gosia Glinowiecka-Cox
Portland State University

Jack Barbera
Portland State University, jbarbera@pdx.edu

Follow this and additional works at: https://pdxscholar.library.pdx.edu/chem_fac

Part of the Chemistry Commons, and the Science and Mathematics Education Commons

Let us know how access to this document benefits you.

Citation Details

Hensen, C., Glinowiecka-Cox, G., & Barbera, J. (2020). Assessing Differences between Three Virtual General Chemistry Experiments and Similar Hands-On Experiments. Journal of Chemical Education, 97(3), 616-625.

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Chemistry Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Assessing Differences Between Three Virtual General Chemistry Experiments and Similar Hands-On Experiments

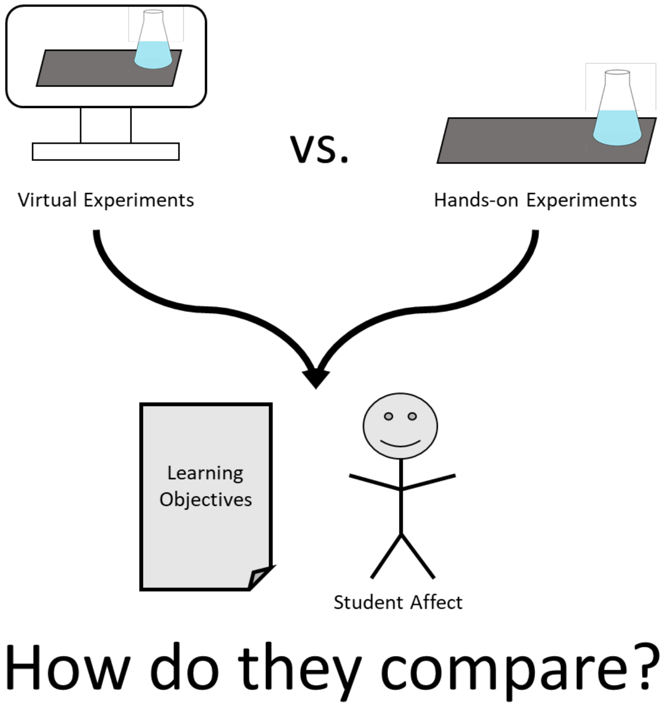
Cory Hensen, Gosia Glinowiecka-Cox and Jack Barbera

Department of Chemistry, Portland State University, Portland, Oregon, 97207-0751, United States

Abstract

To date the efficacy of virtual experiments is not well understood. To better understand what differences may exist between a hands-on learning environment and a virtual learning environment, three experiments were chosen for investigation. For each experiment, approximately half of the students completed a hands-on version of the experiment and the other half completed a virtual version. After completing the given experiment, students were compared on: their ability to meet the learning objectives for that experiment, their responses to six affective scales, and their grade on a laboratory report. Differences were found on four learning objectives. Two of these learning objectives were on the Beer's Law experiment and the other two were on the titration experiment whereas the calorimetry experiment had no differences between groups on learning objectives. However, all four differences are likely due to differences in procedures between environments and not due to the environment itself. Additionally, differences were found on two of the affective scales (usefulness of lab and equipment usability) across all three experiments indicating that the students who completed a virtual experiment found the experiment to be less useful and the virtual environment harder to use. Students that completed the virtual version of the titration experiment also reported that the experiment took less time as indicated by the difference on the open-endedness of lab scale. These differences are not representative of a students' individual experience, however. To capture individual experiences, latent profile analysis was conducted to determine what affective profiles existed within the population. There were three common profiles identified across the three experiments: low affective outcomes, medium affective outcomes, and high affective outcomes. These indicate that while the majority of the students have medium or high affective outcomes and do well on laboratory reports, there is anywhere from four to seventeen percent of the students completing a given experiment, that have low affective outcomes but still do equally well on the laboratory report as the other students. Future work should be conducted to assess why students report low affective outcomes and if a different type of laboratory learning environment or curriculum type would better serve them.

Graphical Abstract



Keywords
First-Year Undergraduate/General, Chemical Education Research, Computer-Based Learning, Distance Learning

Reference Information:
Cory Hensen, Gosia Glinowiecka-Cox, and Jack Barbera, “Assessing Differences Between Three Virtual General Chemistry Experiments and Similar Hands-On Experiments,” *Journal of Chemical Education*, 2020, 97, 3, 616-625, DOI:10.1021/acs.jchemed.9b00748.

Introduction

The general chemistry laboratory has historically been a place where students or apprentices learn valuable trade skills for their future career. While scientific thinking and fundamental laboratory skills are still essential for many careers, there has been a drastic increase in the career options students have. This, along with the fact that typically a wide variety of majors enroll in the chemistry sequence, creates a new challenge for designing a laboratory experience that adequately prepares all students for their future career.

Some institutions have accommodated the differing career goals by creating laboratory sections that cater to different populations of students. For example, students at the University of California San Diego that are pursuing a career in chemistry may opt to enroll in a laboratory course designed specifically for chemistry majors whereas students pursuing nursing at California State University, Sacramento may opt to enroll in a laboratory course with a pre-health focus. However, the ability to create multiple sections catering to different populations of students varies by institution and there is a lack of agreement as to whether non-science majors, or specifically non-chemistry majors, need to take a laboratory that teaches them chemistry-specific skills.[1-3] In fact, some have gone as far as suggesting that non-majors do not need the laboratory and question why institutions are spending money to teach them laboratory skills.[4] One challenge in offering multiple types of laboratory experiences is that the number of laboratory sections is often limited by space and staff availability. Some have met this challenge by creating a hybrid curriculum where students complete half of their experiments in a virtual environment and the other half in a traditional hands-on environment.[5] While this frees up physical laboratory space, questions remain on the efficacy of virtual experiments and as such the current ACS Guidelines for Bachelor Degree Programs recommends that the General Chemistry course remain primarily hands-on, supervised laboratory experiences.

Previous research on virtual experiments across STEM disciplines have generally found that students perform equally well on cognitive assessments regardless of the type of learning environment they completed the experiment in.[6-11] This trend holds true within chemistry where some studies found no difference in cognitive outcomes[5, 10] and others found that students completing the virtual experiment outperformed those that completed the hands-on experiment.[9, 11-12] Therefore, there exists an established body of evidence that laboratory coordinators can use to make well informed decisions about using virtual experiments. However, cognitive assessments do not measure if students learn the same skills for their career or if they had a positive experience in the laboratory. There has been significantly less research conducted on the psychomotor and affective domains of learning, leaving laboratory coordinators unsure if virtual experiments can truly provide an equal experience for students. Two studies that include a laboratory practical as part of their comparison between virtual and hands-on environments have found that students that learned a skill in the virtual environment are able to successfully perform the skill in person as well.[10, 12] Despite this, it is possible that students can learn the same content and perform the skill without having a favorable laboratory experience. In fact, one study focused on hybrid laboratories did include affective domain items and found that students who completed the virtual experiment had lower affective outcomes than the control group.[5]

This prior work highlights the need to further assess the affective domain when students are completing a virtual experiment. While this study used an instrument (the MLLI[13]) that measures the affective domain with eight general items, the affective domain is a broad domain that contains many constructs. As virtual experiments grow in popularity, it is imperative that information about how outcomes on specific constructs compare. This then allows laboratory coordinators to make more informed decisions. One affective construct that has been previously studied in the laboratory and can impact students' experience is anxiety. The Chemistry Laboratory Anxiety Instrument (CLAI)[14] was developed to specifically measure this construct in the chemistry laboratory environment. It is possible that students who complete the experiment virtually have differing levels of anxiety, as they do not need to worry about personal protective equipment (PPE) or chemical safety. In addition to anxiety, there may be other differences based on the specific environment. The Virtual and Physical Environment Questionnaire (VPEQ)[9] was designed to address specific differences between environments and addresses constructs of equipment usability, usefulness of lab, and open-endedness of lab. These three constructs measure students' *feelings* towards specific components of the laboratory. In addition to environment-specific differences, there may also be broader affective differences. One of the important goals of any science laboratory and especially chemistry is to improve students' attitude toward chemistry.[15] It is possible that the ability to improve students' attitude differs based on the learning environment.

In addition to the lack of specific affective constructs studied with regard to virtual laboratories, there is also a need to further study the cognitive outcomes. Despite the number of studies finding no differences between environments, these studies have relied heavily on the use of multiple-choice test or quiz items to determine performance. Relying on these types of assessments inadequately captures whether students have the scientific thinking needed for many careers. In fact, there has been a recent push to incorporate curricula that focus on scientific writing instead of short post-lab items.[16-18] Thus, it is important that rather than compare students on multiple-choice assessments they are compared on their ability to meet the desired cognitive learning objectives of the experiment. However, to date, there is a lack of agreed upon experiment-specific learning objectives that can be used to assess the environments.[19] With specific learning objectives for each experiment, it would then possible to compare environments and determine if they meet them equally on an experiment by experiment basis.

If evidence is provided that students are meeting the same cognitive objectives and affective outcomes in a new experiment environment (i.e., virtual) as compared with the traditional environment, then laboratory coordinators can select the environment that best matches both the faculty members' goals at that institution and the intended student population for the curriculum. With a wide arsenal of experiments, both virtual and hands-on, that have established and measurable outcomes it would be possible to design multiple laboratory courses that align with the ranges of student expectations and career motivations without the limitation of physical laboratory space.

Research Questions

There is a need to better understand if students completing an experiment in an alternative environment, such as the virtual environment, are able to meet the same learning objectives and acquire similar affective outcomes as students in the traditional hands-on environment. The following research questions guided this study:

1. To what degree can experiments in a virtual environment meet the same learning objectives as similar traditional hands-on experiments?
2. How do students' affective outcomes compare when completing an experiment in different learning environments?
3. What effect do individual student differences have on any observed differences in either the affective or learning objective outcomes?

Methodology

Human Subject Research

This research was approved by the institutional review board at Portland State University. Participants were asked to provide informed consent and only data from those who consented were analyzed.

Selection of Experiments

There is wide variety in the experimental topics covered across different institutions with each institution selecting the topics that they value most. However, there are a number of common topics that are taught at most institutions. Previous work conducted by Reeves and Exton as part of the development of the ACS General Chemistry Laboratory Exam helped gain a better understanding of which experiments are commonly done.[20] They first compiled a list of laboratory manuals used at a range of institutions, which generated thirty-six unique sources, and reviewed each for the experimental topics included. After reviewing the manuals, the six most commonly covered topics were:

- Volumetric analysis (titrations)
- Stoichiometry
- Kinetics (determination of rate law)
- Spectrophotometry/Beer's Law
- Properties of Acids and Bases
- Calorimetry

These six experiments range in the level of laboratory skills required to complete the procedure and take place across the entire year of the general chemistry curriculum. To cover a range of skills and chemistry content, one experiment from each term of the general chemistry curriculum was selected for this study. Beer's Law was chosen as the experiment to investigate in the first term, calorimetry was chosen for the second term, and volumetric analysis (titrations) was chosen for the third term of a general chemistry laboratory course based on a quarter system.

Establishing Learning Objectives

Five faculty members from three institutions in the Pacific Northwest were interviewed in a semi-structured format to capture the specific objectives each had for the chosen experiments. Three faculty were from two different community colleges and two were from a doctoral granting university with high research activity.[21] One of the community colleges used inquiry-based experiments while the other community college and the university used expository experiments. Including different institutions and types of curricula in the targeted population allowed for different perspectives on the learning objectives to be captured. The interviews took place the week prior to the experiment being done at the respective institution. Participants were asked to explain the procedure for each experiment and what they hoped students would gain by completing the experiment. As participants had different levels of understanding of what "experimental objectives" meant, the question "If students missed today's experiment, what would they miss out on?" was also asked. This question allowed participants to better articulate what important objectives they had for their students. For the full interview protocol, see Supporting Information.

For a given experiment, each faculty member's learning objectives were listed and then compared with the others' objectives. With variety in the types of experiments done at institutions over the same topic, it was expected that not all learning objectives would be shared across participants. Therefore, to capture the most salient objectives of each experiment (i.e., those that faculty agreed upon) only the common learning objectives across all participants were used to assess differences in laboratory environments. Once the common objectives were established for each experiment, they were used to generate rubrics to score student's laboratory reports based on how well they met those objectives.

Student Population

Students enrolled in the general chemistry laboratory sequence during the 2018-2019 academic year at Portland State University were the targeted population. This convenience population was chosen as it provided several important features including: 1) the ability to provide significant input to the structure of the laboratory sections, 2) multiple sections that could be easily split by environment type, and 3) the ability to directly work with the university office of information technology to set-up and monitor the functioning of the virtual experiments on the laboratory computers. The sections of the laboratory courses were split approximately in half for each of the three experiments, with some of the sections completing the traditional hands-on experiment and the other sections completing the experiment in a virtual environment. All enrolled students in a given section conducted the same experiment and generated the same cognitive and affective data as part of their normal laboratory requirements for that day. As students enrolled in whichever section best fit their schedule and did not know ahead of time which sections would conduct the experiment in a virtual environment, there was approximately random grouping. Further information about self-report demographics by grouping can be found in Table SI1 in the Supporting Information. No institutional data was provided by the university. The virtual environment used for all three experiments was the LearnSmart Labs by McGraw-Hill Education. Students completed the virtual environment procedure in their normal laboratory room working with a laboratory partner and with their teaching assistant (TA) present.

Rubrics and Scoring of Laboratory Reports

Students completed a formal laboratory report after each of the experiments included in this study. Identifying information was removed from the reports and each was assigned a random identification number prior to analysis. Coders were not aware of which environment a student completed the experiment in when scoring their report. As the codes were pre-determined based on the faculty members' learning objectives, this was a deductive analysis. A primary and secondary coder individually scored seven student reports at a time for each experiment and then met to discuss their scoring and calculated a percent agreement. This process repeated until the coders reached 100% agreement. Consensus scoring is one method used to establish inter-rater reliability and with a high consensus score indicates the rubrics were interpreted and utilized in a similar way for the student reports.[22] Initially, the rubric was scored using categories of 'Does not meet' and 'Meets' to mark if a student met the learning objective, however, after preliminary testing of the rubric a third category of 'Partially Meets' was added for cases where students demonstrated only limited evidence of meeting a learning objective. After reaching consensus on a set of reports, the coders individually scored the remaining reports and met regularly to clarify any questions that arose. The rubric scores were then used to compare if students in both environments met the learning objectives to the same degree.

For each experiment, chi-square tests were conducted for individual learning objectives to determine if there were significant differences between rubric scores by learning environment. A 2x3 chi-square test was used to compare scores across two groups (i.e., hands-on and virtual) on a variable with three category options (i.e., meets, partially meets, and does not meet).[23] A non-significant chi-square test indicated that no statistical differences between learning environments were detected for a given learning objective. Chi-square tests were conducted using version 26 of SPSS.

Measuring Differences in the Affective Domain

Immediately upon the completion of an experiment, six affective scales were administered to students through a Qualtrics survey. The scales measured the constructs of anxiety, intellectual accessibility, emotional satisfaction, equipment usability, usefulness of lab, and open-endedness of lab. Evidence for the reliability and validity of the data generated by these scales, in these learning environments and with this specific population, has been previously reported.[24] The reported validity data included response process validity interviews, conducted to ensure students are interpreting the items in a similar manor as is intended, as well as measurement invariance, establishing that each scale functioned similarly for students in both learning environments. A multivariate analysis of variance (MANOVA) was conducted on the affective scale data from each experiment to detect differences between the learning environments. A MANOVA is an appropriate test to compare two groups of students for multiple outcomes.[25] Significant findings in the MANOVA would indicate differences in the affective outcomes between learning environments for the given experiment. MANOVA results from the Beer's Law experiment have been previously reported after checking all assumptions for running a MANOVA.[24] The assumptions were tested again for the calorimetry and titration data sets as they contain a number of different students than the Beer's Law data set. The MANOVAs were conducted using version 26 of SPSS.

Latent Profile Analysis

Students have many different expectations about the laboratory experience, which have been previously shown to relate to students' affective outcomes.[26] To explore what underlying groups, or profiles, of students were present in this study, a cluster analysis was performed on the data generated for each experiment. The model-based cluster analysis for latent variables is called latent profile analysis or latent class analysis depending on the type of data used.[27] A model-based approach has the advantage of generating fit indices that are then used to directly compare different models and groupings of the data. Typical fit indices that are reported in latent profile analysis include the Bayesian information criterion (BIC), the Akaike information criterion (AIC), and the log-likelihood.

One of the most important decisions when conducting a cluster analysis is which variables to include. If too many variables are included the resulting profiles have no meaningful interpretation whereas if not enough variables are included then there is not enough variance in the data to detect meaningful profiles. Scores from the emotional satisfaction, intellectual accessibility, usefulness of lab, equipment usability, and open-endedness of lab scales were used as the clustering variables to generate student profiles based on the overall affective outcomes. Anxiety was not included as there were few differences on this scale between environments in all three experiments and as such did not add information toward meaningful profiles. As part of the interviews conducted in a previous study, students in both environments frequently reported that working with chemicals was much less a source of anxiety as compared with the social anxiety of working with other people.[24] Thus, it was unsurprising that there were few differences seen between environments on anxiety despite different equipment used. For this study, the latent profile analysis was conducted using the expectation-maximization algorithm and maximum likelihood estimates. The latent profile analyses were conducted using version 5.4.3 of the mclust package in version 3.5.3 of R.[28]

Results and Discussion

Generating experiment-specific rubrics based on learning objectives

The list of experiment-specific learning objectives, generated through faculty interviews, was analyzed to determine which objectives were shared by the majority of the faculty members interviewed. As seen in Table 1, there were three common learning objectives for the Beer's Law experiment, four for the calorimetry experiment, and four for the titration experiment. For more information about individual faculty member's objectives, see Table SI2 in the Supporting Information

Table 1: Common learning objectives across faculty interviewed

Experiment	Learning Objective	Abbreviation
Beer's Law	Understand the relation between absorbance and concentration	BL1
	Prepare solutions	BL2
	Determine an unknown concentration using the relation between absorbance and concentration	BL3
Calorimetry	Predict the sign of the change in enthalpy for a given reaction	C1

	Determine the enthalpy change for a given reaction	C2
	Understand how to calculate a change in enthalpy from a temperature change	C3
	Understand the difference between endothermic and exothermic and how it relates to the sign of the enthalpy change	C4
Titration	Visually identify a change in pH during a titration using a mixture of indicators	T1
	Identify key points on a titration curve	T2
	Determine the pKa of an unknown analyte using a titration curve	T3
	Determine the molar mass (or mass) of an unknown analyte using a titration curve	T4

These learning objectives were then used to assess the students' ability to demonstrate evidence of meeting them in their laboratory report. To do this, a rubric was generated for each experiment. As an example, the Beer's Law rubric is shown in Table 2.

Table 2: Rubric used to score student laboratory reports for the Beer's Law experiment in each learning environment

Learning objective		***Does Not Meet***	***Partially Meets***	***Meets***
Understand the relation between absorbance and concentration	BL1			
Prepare solutions	BL2			
Determine an unknown concentration using the relation between absorbance and concentration	BL3			

Student Population

For the Beer's Law experiment, 174 students completed the hands-on experiment and 216 students completed the virtual experiment. The following term for the calorimetry experiment, 129 students completed the hands-on experiment and 152 students completed the virtual experiment. Finally, in the last term for the titration experiment, 72 students completed the hands-on experiment and 117 students completed the virtual experiment. For more information on the student population and demographics see Table SI1 in the Supporting Information.

Assessing Differences in Learning Objectives

The laboratory reports of study participants were carefully read and the coders looked for any evidence of the students meeting the stated learning objectives noted on each rubric. For the first Beer's Law objective (BL1), an example of a student report that received a score of 'Meets' is "*A substances concentration and it's absorbance are directly proportional. A high-concentration solution absorbs more light and a low-concentration solution absorbs less light*". This student demonstrated that they fully understood the relation. For comparison, a student report that received a score of 'Partially Meets' is "*Beer's law, which states A=ebc, lets one use the relationship between absorbance to create a calibration curve*". This student seems to have some understanding of how to use the relation but does not provide further evidence that they understand it and do not simply just understand the experimental steps. The score 'Does Not Meet' was given for any report that provided no evidence of understanding the relation. The three scoring categories were used in a similar fashion for all other learning objectives. Table 3 contains the results of scoring the reports and the significance of the chi-square results when comparing an objective between environments. The N/A category was used when students did not include a relevant section in the report as there was no way of judging a missing section.

Table 3: Comparative Results of Students Meeting Learning Objectives and Chi-Square Values for All Learning Objectives by Environment Type

	Objectives' Status in Hands-On Environment, %					*Objectives' Status in Virtual Environment, %*				
LO[a]	N	Does Not Meet	Partially Meets	Meets	N/A	N	Does Not Meet	Partially Meets	Meets	N/A
BL1	137	19.7	6.6	69.3	4.4	176	21.0	6.8	72.2	0.0
BL2[b]		0.0	0.0	5.1	94.9		0.0	0.0	0.0	100.0
BL3[c]		26.3	0.0	73.7	0.0		17.0	0.0	83.0	0.0
C1	110	76.4	0.0	23.6	0.0	140	84.3	0.7	15.0	0.0
C2		0.9	0.0	99.1	0.0		2.1	0.0	97.9	0.0
C3		12.7	2.7	84.5	0.0		15.0	10.0	75.0	0.0
C4		13.6	10.9	75.5	0.0		12.1	6.4	81.4	0.0
T1[b]	64	0.0	0.0	37.5	62.5	90	0.0	0.0	0.0	100.0
T2		9.4	25.0	65.6	0.0		12.2	25.6	62.2	0.0
T3		34.4	6.3	59.4	0.0		28.9	3.3	67.8	0.0
T4[b]		7.8	1.6	90.6	0.0		58.9	3.3	37.8	0.0

[a] *LO: Learning objective; for definitions of these Beer's Law, Calorimetry, and Titration objectives, see Table 1.*
[b] *Significant at p = 0.01.* [c] *Significant at p = 0.05.*

Most of the learning objectives showed no statistical difference between environments. However, as noted in Table 3 with asterisks, there were significant differences on four learning objectives. Two of these, BL2 and T1, were skill-based objectives that explicitly addressed a procedural step. Thus, it is not surprising that only a few students in the traditional hands-on environment included evidence of meeting these objectives and none of the students in the virtual environment included evidence of meeting them, as students typically do not include details about specific procedural steps in their report. Interestingly, the majority of hands-on students that did meet BL2 did not meet BL1. It is likely that these students were only able to summarize the procedural steps they conducted rather than understand and document why they conducted them. Thus, the score of 'Meets' on these skill-based objectives may not be an indication of whether students learned the skill but rather their ability to write a complete laboratory report.

Students also differed on their ability to meet learning goal T4. This difference could be due to a function of the design of the LearnSmart Labs. In the traditional hands-on titration experiment students started with an unknown solid and were asked to identify the unknown by calculating the molar mass. However, in the virtual environment students started with an unknown solution and were asked to identify the unknown by calculating the pKa and then asked to calculate how much mass was initially dissolved to make the solution. While students in both environments were asked to use the equation: $Molar\ Mass = \frac{grams}{Molarity * volume}$, the virtual students frequently did not provide evidence of calculating the initial mass dissolved. Instead, the students stopped once they were able to get the identity of the acid with the pKa, as only that finding had to be reported to the TA before they could leave for the day.

The fourth objective that students differed on was their ability to use Beer's Law to calculate the unknown concentration (BL3), with a higher proportion of students that completed the experiment in the virtual environment meeting this objective. It was observed by the first author and the TAs that students in the virtual environment had more time to do the calculations, as the experiment itself did not take as long as the hands-on counterpart did. Therefore, extra time students in the virtual environment had to work on the calculations with their lab partner and/or TA could explain this higher percentage. It is possible that if each student had an equivalent amount of time to work on the calculations with assistance from a partner and/or TA that this difference would be minimized. Additionally, this finding was not significant using the stricter p-value of 0.01 to correct for multiple comparisons.

Assessing Differences in Affective Outcomes

In addition to the learning objectives, affective outcomes were also compared across environments. After checking the assumptions for running a MANOVA, there were normality and homoscedasticity violations. However, MANOVAs are robust to violations in these assumptions.[29] For the skewness and kurtosis values see Table SI3 in the Supporting Information. MANOVAs were conducted to compare the scale scores for the anxiety, emotional satisfaction, intellectual accessibility, usefulness of lab, equipment usability, and open-endedness of lab scales. Table 4 consists of the results of these MANOVAs and the respective effect sizes as measured by partial eta squared. A bolded p-value indicates a significant result. A partial eta of 0.01 represents a small effect, a value of 0.06 represents a medium effect, and a value of 0.14 represents a large effect.[30] See Table SI4 in the Supporting Information for the averages of all six scales by experiment and environment type.

As seen in Table 4, for the Beer's law experiment, many of the affective scale outcomes were significantly different between environments and both the emotion satisfaction and equipment usability scales were approaching a medium effect size. The differences highlighted in orange indicate that the hands-on students had the significantly higher average whereas the difference highlighted in purple indicates the virtual students had the significantly higher average. However, this Beer's Law data was previously analyzed[24] and an instructor-effect was detected.

Table 4: Comparative MANOVA Results of Affective Differences across Laboratory Environments

Affective Scale	***Beer's Law***		***Calorimetry***		***Titration***	
	p-value	Effect Size[a]	p-value	Effect Size[a]	p-value	Effect Size[a]
Anxiety	0.237	0.004	0.512	0.002	0.477	0.003
Emotional Satisfaction	<0.001[b,c]	0.049[c]	0.478	0.001	0.110	0.003
Intellectual Accessibility	0.001[b,c]	0.027[c]	0.681	0.002	0.489	0.014
Usefulness of Lab	0.001[b,c]	0.028[c]	0.013[b,c]	0.022[c]	0.017[b,c]	0.030[c]
Equipment Usability	<0.001[b,c]	0.056[c]	<0.001[b,c]	0.043[c]	<0.001[b,c]	0.067[c]
Open-endedness of Lab	0.971	0.000	0.194	0.006	0.034[b,d]	0.024[d]

[a]A partial eta result of 0.01 represents a small effect; 0.06 represents a medium effect; 0.14 represents a large effect: see ref 30. [b]Result is significant. [c]Higher average for students in the hands-on environment. [d]Higher average for students in the virtual environment.

In a previously reported analysis of the Beer's Law data, Hensen and Barbera[24] noted that four TAs that taught the virtual experiment had sections with much lower averages on the emotion satisfaction scale than the other four TAs that taught the virtual experiment. As part of their analysis, a MANOVA was run with three groupings (Hands-on, Virtual A - higher emotional satisfaction, and Virtual B - lower emotional satisfaction) instead of just by learning environment. With these TA groupings, none of the affective scale results were significantly different between students in the hands-on sections and the Virtual A group. However, the emotional satisfaction, intellectual accessibility, usefulness of lab, and equipment usability scales were significantly different between students in the hands-on sections and those in the Virtual B group. No evidence of an instructor effect was found for the calorimetry or titration experiment. As both the calorimetry and titration experiments take place in later terms and the Virtual B group consisted of mostly first-year TAs with limited teaching experience, the instructor effect could have been minimized as the TAs gained experience. However, no generalizations about the effect of teaching experience can be made from this study as TAs rotate in and out of teaching general chemistry laboratories throughout the academic year and each quarter consisted of a different combination of TAs.

For both the calorimetry and titration experiments, data from the affective scales of usefulness of lab and equipment usability showed differences between environments with the traditional hands-on students reporting higher averages for both scales (noted in orange in Table 4). For all experiments, the effect size of the usefulness of lab was small but the effect size of equipment usability was medium indicating that the students had minor differences on how *useful* they thought the experiment was but larger differences on their *perceived ability* to use the equipment. However, when accounting for multiple comparisons, the usefulness of lab differences are not significant at a corrected p-value of 0.01 and thus there is not enough power in this sample to make definitive conclusions about that scale. It is possible that if students utilized the virtual environment more often that they may begin to feel more comfortable using it as it does take time to get oriented with the program.

Additionally, the open-endedness of lab scale was significantly different with a small effect size for the titration experiment and the virtual students having a higher average (noted in purple in Table 4). However, similar to the usefulness of lab differences, when accounting for the multiple comparisons made, this finding was not significant at the stricter p-value of 0.01.

Latent Profile Analysis

The affective comparisons noted above do not evaluate differences between specific students but rather differences between environments. Therefore, latent profile analyses were conducted to investigate what groupings of students existed based on their affective characteristics. These analyses indicated that the Beer's Law and calorimetry data had four profiles (groupings of students) and the titration data had three, as shown in Table 5. Each analysis was run ten times, with a random order of the data, to ensure that the solutions were stable.[31] The profiles were named based on the defining characteristics of the affective scale scores. More detailed information on the process of selecting the best fitting profiles using mclust is contained in Table SI5 in the Supporting Information.

Table 5: Distribution of Students by Profile

Experiment	***Latent profile***	**Students, N**

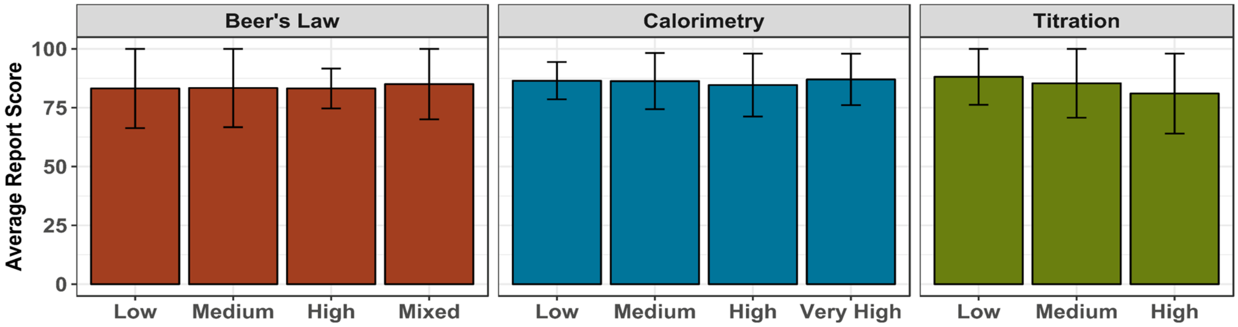
Beer's Law	Low	83
	Medium	209
	High	78
	Mixed	20
Calorimetry	Low	22
	Medium	67
	High	111
	Very High	81
Titration	Low	33
	Medium	100
	High	56

For each experiment, there were three similar groupings: low, medium, and high affective outcomes. For more information on the scale averages by grouping see Table SI6 in the Supporting Information. These groupings are similar to previous cluster analysis results found by Galloway and Bretz.[26] There were also groupings that were unique to an experiment. For the Beer's Law experiment, there was a grouping of students that had low averages on the emotional satisfaction and intellectual accessibility scales but high averages on the usefulness of lab, equipment usability, and open-endedness of lab scales. This indicated mixed outcomes where the students thought the experiment worked well and was useful but still found it to not be accessible or emotionally satisfying. Also, as noted earlier, the calorimetry experiment had a 'very high' profile. It was unsurprising that many students reported high affective outcomes for the calorimetry experiment because both the hands-on and virtual versions of this experiment involved relatively few experimental steps and were shorter than other experiments conducted that term.

While there was a range of affective outcomes across each experiment, interestingly, as seen in Figure 1, the average report score across profiles was consistent indicating that it may not be possible to identify which students had poor affective outcomes solely based on their academic performance in the laboratory. In other words, a student that did very well on the laboratory report may still have had low affective outcomes and vice versa. To investigate this further, the profiles were examined by individual learning objective rather than an overall grade for better resolution.

Figure 1: Average report score by profile and experiment

The percent of students in each rubric category for each learning objective are shown in Figure 2. Learning objectives BL2 and T1 were not included in their respective analyses as no students in the virtual environment, and few students in the traditional hands-on environment met them. Despite providing a more detailed view of the cognitive outcomes, the lack of differences in learning objectives was similar to the lack of differences seen in the report scores, adding more evidence that it was not possible to identify which students were in each grouping based on their laboratory reports. For example, on objective C3 (understand how to calculate a change in enthalpy from a temperature change) there were approximately equal percentages of students that either met or partially met the objective despite differences in affective outcomes. Similarly, there were no major differences between the students' ability to meet the learning objectives for the majority of the objectives (BL1, C2, C3, C4, T2, and T3).



While the majority of the learning objectives had no differences based on learning profile, three differences were observed. As can been seen in Figure 2 for objective BL3 (determining an unknown concentration), the 'low' affective group had the highest percentage of students meeting the objective. However, this difference may be an artifact of the low grouping itself having a higher percentage of virtual students, seen in Figure SI1. As noted earlier and seen in Table 3, the virtual experiment was observed to be shorter, which led to the virtual students meeting this goal more often. The fact that the virtual students may have had more time to work through the calculations with their lab partner and/or TA, combined with the fact that more virtual students are in the low profile, provide a possible reason for the higher percentage of students in the low-profile meeting BL3. Similarly, the virtual students did not meet learning objective T4 (determining molar mass) as often as the hands-on students. As seen in Table 3, the majority of virtual students failed to meet this learning objective whereas the majority of hands-on students did meet this objective due to procedural differences. Thus, the higher percentage of virtual students in the low affective group (as seen in Figure SI1) explains a possible reason why the low affect group did not meet this learning objective as often as the other groups. The third difference was seen with learning objective C1 (predict the sign of the change in enthalpy for a given reaction). This difference is likely a function of the small sample size for the low affective profile. There were only 22 students in the low affective profile for this experiment, which means that each student represents 4.5% of the data plotted in Figure 2. Given the lack of differences on the majority of objectives and laboratory report scores, the difference seen in learning goal C1 is most likely contributed to sample size limitations. It is possible that the difference may not be observed in studies with a larger sample size.

Overall, the majority of learning objectives, in addition the laboratory report scores, showed no difference between affective groups. This highlighted that solely relying on differences in cognitive outcomes to determine if an intervention is successful fails to differentiate between students with low and high affective outcomes. While there is a body of literature that has found students in the virtual experiment are able to meet the cognitive outcomes similarly[5, 10] or outperform the hands-on students[9, 11-12], future research should ensure that affective outcomes are measured and focus on how to identify the students in the low affective profiles in order to target laboratory interventions and ensure all students are having a positive laboratory experience.

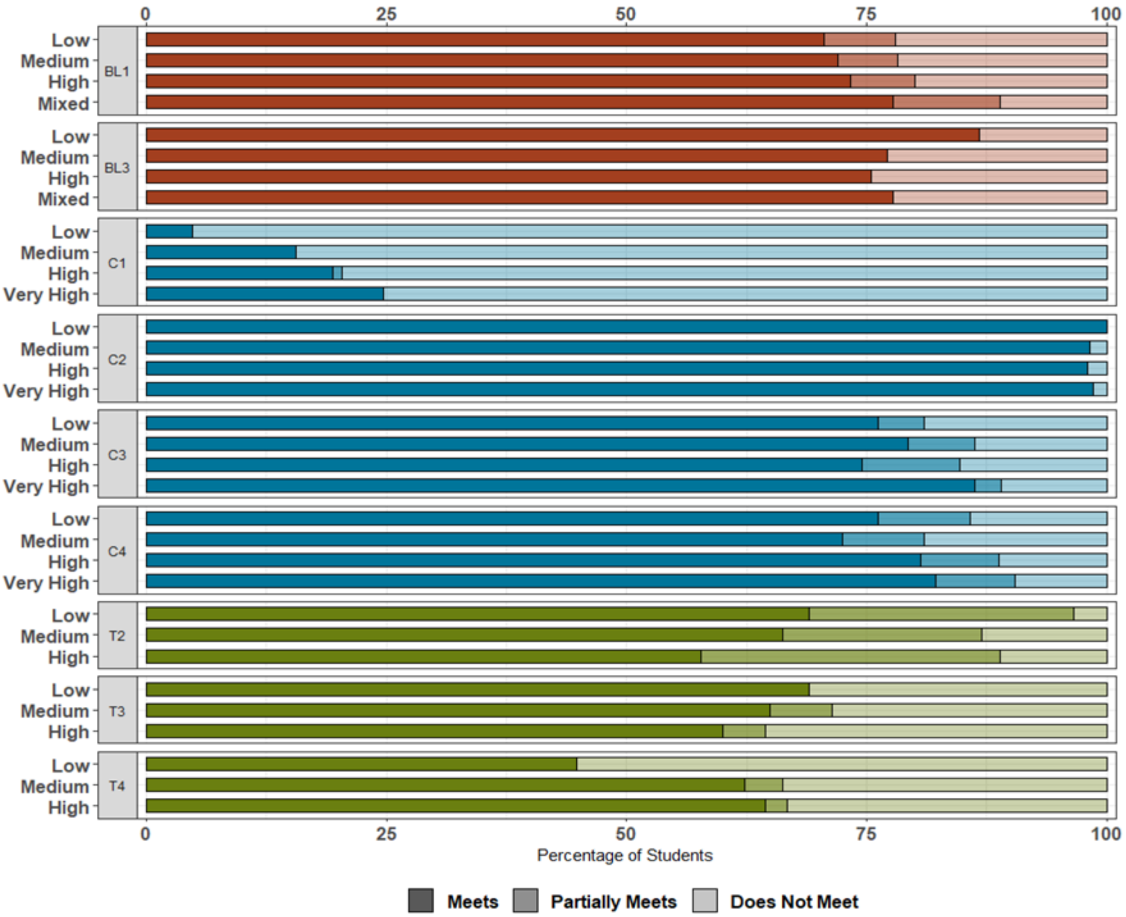


Figure 2: Percent of students in each rubric category by learning objective and affective profile. Colors match experiments as shown in Figure 1, the darkest shade represents the 'Meets' category, the medium shade represents the 'Partially Meets' category, and the lightest shade represents the 'Does Not Meet' category.

Conclusion

After comparing student outcomes across three experiments conducted in both a virtual and a hands-on environment, differences were detected on four of the eleven common learning objectives. Two of the differences were on skill-based objectives and the other two were on objectives related to the outcome of specific calculations within an experiment. While statistically different outcomes were detected, the results are likely due to alignment issues with experimental procedures and report requirements. For the Beer's Law experiment, differences were seen on learning objectives BL2 and BL3. The differences seen on objective BL2 could be contributed to the report requirements as this was a skill-based objective and the report requirements did not include having students explicitly write about the procedural steps they completed. Additionally, in the Beer's Law experiment students in the virtual environment were observed to take less time to complete the experiment which freed up more time to work through the analysis of data with their lab partner and/or TA and thus they met learning objective BL3 more often. Similar to the Beer's Law experiment, students in the titration experiment also did not provide evidence of meeting a skill-based objective, T1, and an objective that had procedural differences between the two learning environments, T4. Overall, the students in the virtual environment consistently struggled to provide evidence of meeting skill-based learning objectives and outcomes designed around specific procedural steps due to differences in the procedures between environments and the report requirements. This result is similar to previous findings that specific differences between procedures in the learning environments account for the differences observed.[10] Therefore, careful design of the experiments and assessments should take place to ensure that students have the opportunity to equally meet the desired experimental learning objectives. If students cannot meet the learning objective in a given learning environment, then that environment should not be used for that experiment. Overall, if students had equal time to work on processing the data and identical procedures in both environments, the differences found in these learning objectives would be greatly minimized.

In addition to investigating differences in the learning objectives between environments, six different affective outcomes (anxiety, intellectual accessibility, emotional satisfaction, equipment usability, usefulness of lab, and open-endedness of lab) were monitored. For most, no detectable differences were found. However, across all three experiments, students in the virtual environment reported lower averages on the equipment usability and usefulness of lab scales. This finding provides unique insight into what differences may exist between learning environments. Historically, the affective domain is understudied, and thus, past studies have focused on cognitive outcomes. While very minor cognitive differences were found, the affective differences highlight larger discrepancies between the learning environments. While an instructor effect was found for the Beer's Law experiment, no evidence of an instructor effect was found for either the calorimetry or titration experiments. This effect was either minimized as TAs gained experience or that the instructor effect was specific to individual TAs who did not teach laboratory sections in subsequent terms.

As the result of a latent profile analysis, there were also differences on affective outcomes based on individual students regardless of which environment they completed the experiment in. For each experiment, the majority of students had medium and high affective outcomes. However, there were still a fair number of students that had low affective outcomes. This may be a function of the wide variety of students that enroll in the chemistry laboratory with many different backgrounds and career paths. The one-size-fits-all approach may work for a large majority of the students but it is possible that select students may benefit from different types of laboratories.

Recently, there has been a call to conduct more research on the laboratory environment and what the role of the laboratory is.[32] A unique challenge of the laboratory is that it often consists of multiple sections taught by multiple instructors and the student population is made of diverse majors. Thus, to study the laboratory effectively requires researchers to carefully consider how to control for a wide range of confounding variables that exist in the natural setting of a laboratory course rather than conduct controlled studies that rely on volunteers that do not necessarily represent the average student population. Once more research that carefully controls for the confounding variables present in the laboratory setting is conducted, there may be a better sense of which students benefit from the current model of the laboratory and which do not.

Limitations

To minimize changes to the curriculum at Portland State University, learning objectives were assessed using the assessments already in place. This meant that the tactile learning objectives were evaluated using the laboratory report instead of a laboratory practical. It is possible that the differences seen in the skill-based learning objectives would be different if a laboratory practical was utilized. Additionally, McGraw-Hill generously allowed us to use the

LearnSmart Labs as the virtual platform. However, this meant that there was no control over the elements of the procedural design in the virtual environment. It is possible that a different virtual environment made to specifically target desired learning outcomes could produce different findings. The ability (or lack of it) to control procedural design could also impact learning objectives that are specific to an institution. For example, one institution uses nanomaterials for their Beer's Law experiment and has learning objectives directly related to using nanomaterials. Thus, if the institution uses a virtual environment that is not customizable it may not be possible for students to meet institution-specific learning objectives.

Beyond experimental limitations, this research took place at Portland State University, an urban Pacific Northwest university, and as such the findings should not be generalized to other settings without future work being conducted. Future studies would benefit from the inclusion of other settings, such as community college populations or those where students have more exposure to the virtual learning environment. While two of the five faculty interviewed were professors at the institution the data was collected at, the learning objectives reported by the five faculty members are not comprehensive, therefore, it is possible that faculty members at other institutions place different value on the objectives presented. Additionally, previous work found an instructor effect existed for the affective outcomes in the Beer's Law experiment. While this effect may be minimized as TAs gain experience, it is also possible that it was specific to individual TAs. Therefore, instructor effect should be examined or controlled for in future research to ensure that outcomes are not a result of who is teaching the section.

With the current sample it was not possible to further investigate the characteristics of the profiles based on demographics. With a more adequate sample size, it would possible to evaluate for measurement invariance by demographic group to ensure that members across groups of interest are interpreting the items in a similar fashion. Once measurement invariance is established, the profiles could be further compared on the demographic variable of interest.

Implications for Future Research

This study expanded on previous research to investigate the learning objectives and affective outcomes for a range of experiments. Based on the findings, there is a need to conduct future studies using laboratory practical exams to investigate the tactical learning objectives. Previous research[10, 12] has found no differences on the students' ability to complete skill-based learning objectives but more work in this area is warranted. As a possible instructor effect was found with one experiment but not the other experiments, future research should choose research designs that allow for a true treatment-control study to be conducted where the same instructor is teaching in both environments. Additionally, there is a need for qualitative studies to further investigate the affective grouping of students. These studies could help identify the nature of the defining characteristics within the groupings. With this information, curriculum reform could then take place to target these groupings to ensure that more students have a laboratory experience that produces positive affective, cognitive, and psychomotor outcomes.

Associated Content

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.XXXXXXX. Interview protocol; List of learning objectives; Demographics, skew, and kurtosis values; Affective averages by environment/experiment; BIC indices; Affective averages by profile/experiment (DOCX)

Author Information

Corresponding Author

*E-mail: jbarbera@pdx.edu

Acknowledgments

We would like to thank Medina Glenn and Amanda Schmidt for their help with this study. We would also like to thank the students, graduate teaching assistants, the laboratory coordinator, and stockroom staff involved with this work. Without these people, this study would not have happened. Also, many thanks to McGraw-Hill Education who generously allowed the use of the LearnSmart Labs platform for this study.

References

1. Wartell, M., A new general chemistry laboratory scheme: Observation, deduction, reportage. *J. Chem. Educ.* **1973,** *50* (5), 361-362.
2. Chittleborough, G. D.; Treagust, D. F.; Mocerino, M., Achieving greater feedback and flexibility using online pre-laboratory exercises with non-major chemistry students. *J. Chem. Educ.* **2007,** *84* (5), 884-888.
3. Tro, N. J., Chemistry as general education. *J. Chem. Educ.* **2004,** *81* (1), 54-57.
4. Hawkes, S. J., Chemistry Is Not a Laboratory Science. *J. Chem. Educ.* **2004,** *81* (9), 1257.
5. Enneking, K. M.; Breitenstein, G. R.; Coleman, A. F.; Reeves, J. H.; Wang, Y.; Grove, N. P., The Evaluation of a Hybrid, General Chemistry Laboratory Curriculum: Impact on Students' Cognitive, Affective, and Psychomotor Learning. *J. Chem. Educ.* **2019,** *96* (6), 1058-1067.
6. De Jong, T.; Linn, M. C.; Zacharia, Z. C., Physical and virtual laboratories in science and engineering education. *Science* **2013,** *340* (6130), 305-308.
7. Ma, J.; Nickerson, J. V., Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Computing Surveys (CSUR)* **2006,** *38* (3), 1-24.
8. Tatli, Z.; Ayas, A., Effect of a virtual chemistry laboratory on students' achievement. *Journal of Educational Technology & Society* **2013,** *16* (1), 159-170.
9. Pyatt, K.; Sims, R., Virtual and physical experimentation in inquiry-based science labs: Attitudes, performance and access. *J. Sci. Educ. Technol.* **2012,** *21* (1), 133-147.
10. Hawkins, I.; Phelps, A. J., Virtual laboratory vs. traditional laboratory: Which is more effective for teaching electrochemistry? *Chem. Educ. Res. Pract.* **2013,** *14* (4), 516-523.
11. Tatli, Z.; Ayas, A., Virtual laboratory applications in chemistry education. *Procedia-Social and behavioral sciences* **2010,** *9*, 938-942.
12. Winkelmann, K.; Keeney-Kennicutt, W.; Fowler, D.; Macik, M., Development, Implementation, and Assessment of General Chemistry Lab Experiments Performed in the Virtual World of Second Life. *J. Chem. Educ.* **2017,** *94* (7), 849-858.
13. 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* (7), 1149-1158.
14. Bowen, C. W., Development and score validation of a chemistry laboratory anxiety instrument (CLAI) for college chemistry students. *Educ. Psychol. Meas.* **1999,** *59* (1), 171-185.
15. Hofstein, A., The role of Laboratory in Science Teaching and Learning. In *Science Education: An International Course Companion*, Taber, K.S., Akpan, B., Eds.; Sense Publishers: Rotterdam, Netherlands, 2017; pp 357-368.
16. Stephenson, N.; Sadler-McKnight, N., Developing critical thinking skills using the science writing heuristic in the chemistry laboratory. *Chem. Educ. Res. Pract.* **2016,** *17* (1), 72-79.
17. Burke, K.; Greenbowe, T. J.; Hand, B. M., Implementing the science writing heuristic in the chemistry laboratory. *J. Chem. Educ.* **2006,** *83* (7), 1032-1038.
18. Greenbowe, T. J.; Poock, J. R.; Burke, K.; Hand, B. M., Using the science writing heuristic in the general chemistry laboratory to improve students' academic performance. *J. Chem. Educ.* **2007,** *84* (8), 1371-1379.
19. Hofstein, A.; Lunetta, V. N., The laboratory in science education: Foundations for the twenty-first century. *Sci. Educ.* **2004,** *88* (1), 28-54.
20. Reeves, J. H.; Exton, D., Developing the first online general chemistry laboratory exam. *Innovative Uses of Assessments for Teaching and Research* **2014,** *1182*, 181-191.
21. Carnegie Commission *The Carnegie Classification of Institutions of Higher Education*; Bloomington, IN, **2015**.

22. Stemler, S. E., A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Pract. Assess., Res. Eval.* **2004,** *9* (4), 1-19.
23. Coolican, H., *Research Methods and Statistics in Psychology*. Psychology Press: New York, 2017.
24. Hensen, C.; Barbera, J., Assessing Affective Differences Between A Virtual General Chemistry Experiment and a Similar Hands-On Experiment. *J. Chem. Educ.* **2019,** *96* (10), 2097.
25. O'Brien, R. G.; Kaiser, M. K., MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychol. Bull.* **1985,** *97* (2), 316-333.
26. 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* (4), 879-892.
27. Vermunt, J. K.; Magidson, J., Latent class cluster analysis. *Applied latent class analysis* **2002,** *11*, 89-106.
28. Scrucca, L.; Fop, M.; Murphy, T. B.; Raftery, A. E., mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal* **2016,** *8* (1), 289-317.
29. Olson, C. L., Comparative robustness of six tests in multivariate analysis of variance. *J. Am. Stat. Assoc.* **1974,** *69* (348), 894-908.
30. Cohen, J., A power primer. *Psychol. Bull.* **1992,** *112* (1), 155-159.
31. Scrucca, L.; Raftery, A. E., Improved initialisation of model-based clustering using Gaussian hierarchical partitions. *Advances in data analysis and classification* **2015,** *9* (4), 447-460.
32. Bretz, S. L., Evidence for the Importance of Laboratory Courses. *J. Chem. Educ.* **2019,** *96* (2), 193-195.

Supporting Information for

Assessing Differences Between Three Virtual General Chemistry Experiments and Similar Hands-On Experiments

Cory Hensen, Gosia Glinowiecka-Cox and Jack Barbera

Department of Chemistry, Portland State University, Portland, Oregon, 97207-0751, United States

Interview Protocol:

1) I first wanted to just remind you of who I am and what my dissertation project is on. I am Cory Hensen and I am currently working on my Ph.D. under Jack Barbera. My dissertation project is looking at the efficacy of virtual laboratories. We are currently starting year 1 of the preliminary data collection before we move on to starting with students, we want to first understand where faculty are coming from through these interviews. Before I can begin looking at virtual laboratories, I first want to understand the learning objectives for the specific experiments I am interested in. Currently you are teaching (coordinating) Chem [course number] which covers the [experiment name] experiment in which I am interested.
    a. If you are okay with being interviewed, I would like to go over the informed consent [informed consent details].
    b. Thank you for signing that form. I am now going to turn on the audio recorder if you are okay with that.
2) I first want to start with asking how long you have been a faculty member at this institution?
3) How many of those years have you been involved in the general chemistry laboratory?
    a. In what capacity are you involved in the general chemistry laboratory?
4) Now I wanted to get into asking about a specific laboratory experiment. This term the students are doing an experiment over [topic]. Here is a copy of the procedure in case you need it. I wanted to ask you what learning objectives, or things you want your students to get out of this lab, you have?
    a. How many of these are assessed?
    b. If students missed today's experiment, what would they miss out on?

Table SI1: Demographics

		Beer's Law	*Calorimetry*	*Titration*
Total Enrollment (N)		630	484	355
Consented (N)	Hands-on	174	129	72
	Virtual	216	152	117
**Female (%)*	Hands-on	61.5	55.0	56.0
	Virtual	55.0	57.2	65.8
**White (%)*	Hands-on	57.5	49.6	44.0
	Virtual	49.1	57.2	49.6
**Biology Major (%)*	Hands-on	36.2	40.3	41.3
	Virtual	25.9	35.5	42.7

*These categories represent the plurality for all experiments and sections for both the consented and overall course populations

Table SI2: List of overarching learning goals and experiment-specific learning objectives by faculty member

Faculty	Overarching Goals	Beer's Law Objectives	Calorimetry Objectives	Titration Objectives
	After completing this course, students will be able to do:	After doing this experiment, students will be able to:		
A	○ Graphical analysis	○ Understand and use the relationship between absorbance and concentration ○ Prepare solutions from both a stock solution and a solid ○ Calculate the molarity of a given solution	○ Experimentally determine and feel enthalpy changes ○ Use Hess's Law to predict the enthalpy change for a given reaction ○ Understand the relationship between energy and enthalpy at a constant pressure ○ Understand the relationship between energy and temperature	○ Successfully preform a titration ○ Identify key points on a titration curve ○ Use a pH titration curve to determine the concentration of a solution containing an acid ○ Identify the Brønsted-Lowry acids and bases present in solution and which of these substance(s) control the pH
B	○ Error analysis ○ Measurement	○ Visualize concentration strength in a serial dilution ○ Derive graphically the relationship between absorbance and concentration ○ Use the relationship between absorbance and concentration to solve for an unknown concentration	○ Experimentally determine the thermal energy (q) for a given reaction ○ Use thermal energy to calculate the enthalpy change of a given reaction ○ Describe the relationship between a measured temperature change and an enthalpy change	○ Visually identify a change in pH during a titration ○ Use a titration curve to identify the molar mass and pKa of an unknown analyte
C	○ Comparison with literature values ○ Unit analysis	○ Graphically determine the relationship between absorbance and concentration ○ Determine an unknown concentration using the relationship between absorbance and concentration ○ Successfully prepare a calibration curve ○ Prepare standard solutions from a stock solution	○ Experimentally determine the enthalpy of neutralization of phosphoric acid ○ Compare the experimental value with the literature value and determine percent error ○ Apply and understand the first law of thermodynamics	○ Identify key points on a titration curve ○ Determine the pKa and molar mass of an unknown analyte using a titration curve ○ Visualize pH changes using a mixture of indicators

D	○ Graphing ○ Collaboration	○ Determine graphically the relationship between absorbance and concentration ○ Use the relationship to solve for an unknown concentration ○ Understand how light interacts with matter to produce the maximum wavelength ○ Understand real-world applications of spectroscopy	○ Experimentally determine the enthalpy of dissolution ○ Predict the sign of the change in enthalpy from a temperature change ○ Calculate heat energy by using a temperature change ○ Relate enthalpy changes to bond formation	○ Determine the pKa and identify of an unknown acid using a titration curve ○ Predict the pH at the equivalence point ○ Identify key points on a titration curve ○ Predict which acid-base species are present at various points throughout a titration
E	○ Graphing	○ Prepare calibration standard solutions ○ Understand the relationship between absorbance and percent transmittance ○ Understand the interaction of light and matter at the nano level ○ Use a calibration curve to determine an unknown concentration	○ Experimentally determine the change in enthalpy given a temperature change ○ Understand the relationship between mass and heat energy ○ Understand the difference between exothermic and endothermic reactions ○ Predict the sign of the change in enthalpy from a temperature change	○ Identify key points on a titration curve ○ Identify the unknown analyte using the calculated pKa value ○ Understand the reaction of a weak acid with a strong base ○ Understand real-world applications of titrations

The faculty members were not asked explicitly about any broad learning goals; however, some learning goals were still mentioned in the course of the interview. These were noted separately and were not included in any analysis as this study was focused on experiment-specific learning objectives.

Table SI3: Skew and Kurtosis values

		Hands-On		*Virtual*	
		Skewness	Kurtosis	Skewness	Kurtosis
Beer's Law	Anxiety	0.466	-0.648	0.329	-0.602
	Emotional Satisfaction	**-1.099**	**1.074**	-0.503	-0.688
	Intellectual Accessibility	-0.759	0.158	-0.251	-0.786
	Usefulness of Lab	-0.625	0.160	-0.536	-0.598
	Equipment Usability	**-1.277**	**2.077**	-0.764	-0.235
	Open-endedness of Lab	-0.488	0.163	-0.395	-0.127
Calorimetry	Anxiety	0.724	-0.657	0.903	-0.174
	Emotional Satisfaction	**-1.580**	**1.823**	**-1.399**	**1.217**
	Intellectual Accessibility	**-1.487**	**1.501**	**-1.571**	**1.796**
	Usefulness of Lab	-0.875	1.009	-0.622	-0.126
	Equipment Usability	**-1.009**	**0.329**	**-1.487**	**2.905**
	Open-endedness of Lab	-0.296	-0.744	-0.400	-0.553
Titration	Anxiety	0.798	-0.177	0.311	-0.765
	Emotional Satisfaction	-0.976	0.640	-0.548	-0.551
	Intellectual Accessibility	-0.802	0.436	-0.624	-0.411
	Usefulness of Lab	-0.335	-0.208	-0.459	-0.615
	Equipment Usability	**-1.404**	**2.539**	**-1.010**	0.232
	Open-endedness of Lab	0.153	0.019	-0.464	-0.169

Table SI4: Affective averages by environment and experiment

		Anx	**ES**	**IA**	**U**	**EU**	**OE**
Beer's Law	Hands-On	32.71	72.28	66.10	3.78	4.21	3.54
	Virtual	35.68	60.33	57.80	3.47	3.75	3.54
Calorimetry	Hands-On	23.56	78.12	77.32	3.88	4.66	4.07
	Virtual	21.72	75.83	78.56	3.62	4.41	3.95
Titration	Hands-On	32.08	69.10	69.58	3.73	4.29	3.23
	Virtual	33.12	63.50	68.25	3.37	3.76	3.50

Scales on a 0-100 semantic differential scale: Anx: anxiety, ES: emotional satisfaction, IA: intellectual accessibility
Scales on a 0-5 point Likert-type scale: U: usefulness of lab, EU: equipment usability, OE: open-endedness of lab

Latent Profile Analysis:

Once the clustering variables were selected as: emotional satisfaction, intellectual accessibility, usefulness of lab, open-endedness of lab, and equipment usability, the R package mclust was used to conduct a latent profile analysis. The anxiety scale was not selected as a clustering variable. A latent profile analysis has an advantage over traditional distance-based cluster analysis as it allows competing models to be compared with a fit index to determine the best clustering for the data. There are fourteen different types of models compared and each of these types had nine sub-models that were used to determine the number of profiles. There were four different categories that the models could be different on: the distribution of the data within each grouping, the volume of the grouping, the shape of the grouping, and the orientation of the grouping. The first letter of the model represents whether the volume was forced to be equal between the groupings (E) or if there was variation allowed in the volume (V). The second letter of the model indicates whether the shape of the model was forced to be equal between the groupings (E) or if there was variation allowed in the shape (V). The third letter of the model specifies whether the orientation of the model was on the coordinate axes (I), forced to be equal between groups (E), or allowed to vary (V). There are two models that do not follow this lettering. EII is for spherical groups with equal volume and equal shape and VII is for spherical groupings with variable volume and equal shape. For the Beer's Law data, the r function mclustBIC was used to compare all the models on the BIC fit index:

Table SI5: BIC indices for all possible models for Beer's Law data

	EII	**VII**	**EEI**	**VEI**	**EVI**	**VVI**	**EEE**	**EVE**	**VEE**	**VVE**	**EEV**	**VEV**	**EVV**	**VVV**
1	-16501.4	-16501.4	-10475.4	-10475.4	-10475.4	-10475.4	-9655.02	-9655.02	-9655.02	-9655.02	-9655.02	-9655.02	-9655.02	-9655.02
2	-15153	-15076.6	-9880.17	-9755.95	-9877.55	-9755.02	-9620.04	-9547.8	-9458.01	-9468.32	-9571.09	-9484.68	-9572.07	-9489.55
3	-14574.2	-14422.3	-9798.39	-9594.37	-9737.39	NA	-9484.54	-9573.11	-9461.59	-9430.9	-9582.6	-9485.44	-9611.83	-9520.89
4	-14237.5	-14144.4	-9605.86	-9501.16	-9643.88	NA	-9490.44	-9540.49	**-9408.39**	-9445.8	-9586.7	-9484.16	-9625.05	-9560.75
5	-13978.2	-13905.1	-9603.75	-9497.37	-9645.41	NA	-9505.78	-9562.34	-9408.89	-9433.96	-9641.8	-9551.22	-9689.47	-9551.73
6	-13909.1	-13792.4	-9613.43	-9461.04	-9639.73	NA	-9551.32	-9587.87	-9435.61	-9467.29	-9661.99	-9576.45	-9775.14	-9626.42
7	-13765.9	-13675.8	-9568.55	-9493.52	-9686.25	NA	-9577.47	-9649.66	-9453.16	-9511.17	-9720.75	-9631.47	-9855.15	-9721.86
8	-13674.6	-13450	-9576.26	-9501.44	NA	NA	-9587.03	-9663.83	-9484.48	-9547.29	-9807.52	-9711.39	-9853.02	-9788.98
9	-13616.4	-13322.9	-9612.14	-9492.6	NA	NA	-9622.77	-9672.88	-9492.94	-9567.88	-9840.68	-9729.3	-9987.25	NA

The best fitting model is the one that produces the highest BIC, since BIC is calculated to be maximized in mclust. Therefore, the best fitting model was VEE with 4 profiles, as shown in bold in Table SI5. The grouping with five profiles had a similar fit but ultimately four was chosen as it was slightly higher and presents the simpler case. The more profiles that are selected, the harder it is to make meaningful comparisons between the profiles. This means that the groups were ellipsoidal with varying volume but equal shape and orientation. This process repeated in a similar fashion for the other two experiments. For the calorimetry experiment, the solution of five profiles had the highest BIC but after looking at the profiles, two profiles had very similar characteristics and were collapsed into one profile, resulting in four profiles used for interpretation. For the titration experiment, the solution of three profiles had the highest BIC and was selected as the best fitting.

Table SI6: Affective averages by profile and experiment

		***Anx**	**ES**	**IA**	**U**	**EU**	**OE**
Beer's Law	Low	53.73	38.26	37.76	2.59	2.45	2.67
	Medium	31.45	71.98	66.93	3.70	4.16	3.54
	High	16.57	92.85	82.99	4.35	4.84	4.35
	Mixed	53.64	7.35	19.59	3.95	4.60	3.98
Calorimetry	Low	59.28	5.95	13.18	3.82	4.41	4.09
	Medium	31.54	57.48	66.10	3.15	4.03	3.36
	High	18.03	88.08	86.61	3.67	4.54	3.82
	Very High 1	8.57	99.38	98.38	4.57	5.00	5.00
	Very High 2	12.14	96.19	92.35	4.22	4.92	4.71
Titration	Low	54.98	26.21	46.75	2.69	2.23	2.73
	Medium	34.42	64.80	66.93	3.45	4.12	3.18
	High	17.90	90.36	84.98	4.10	4.70	4.18

Scales on a 0-100 semantic differential scale: Anx: anxiety, ES: emotional satisfaction, IA: intellectual accessibility
Scales on a 0-5 point Likert-type scale: U: usefulness of lab, EU: equipment usability, OE: open-endedness of lab
*Anxiety was not used as a clustering variable and is only presented here to inform the reader of the average scale score by profile. Similarly, the two "Very High 1" and "Very High 2" profiles were combined from the 5-profile solution to form the "Very High" profile seen in Table 5.

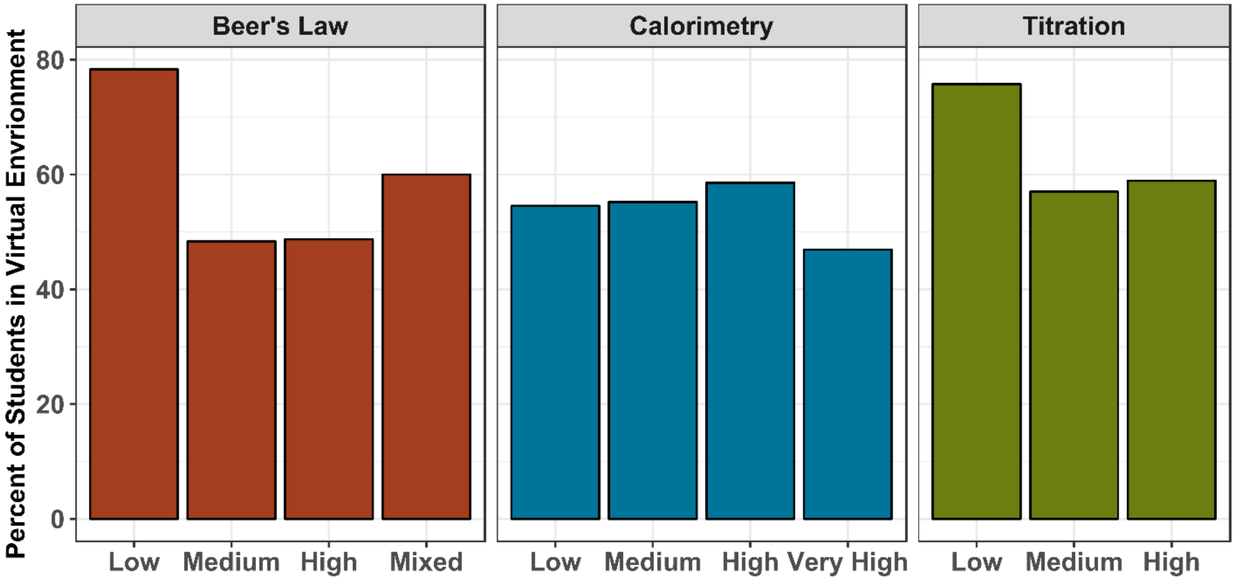


Figure SI1: Percent of students that completed the experiment in the virtual environment by profile