Investigating Student Perceptions of Transformational Intent and Classroom Culture in Organic Chemistry Courses

Supplementary Information

Nonmutually Exclusive Coding

As mentioned in the article, the authors used a mutually exclusive coding scheme. However, in order to assess the viability of using mutually exclusive codes. One of the authors (RSB) did a quick analysis of the data using non-mutually exclusive codes in order to determine if the analysis changed. It is important to note that most of the responses could only be categorized in one way; however, some responses mentioned multiple ideas that technically qualified for two different categories (even though in many cases these responses primarily focused on one category). **Figure S1**, **S2**, and **S3** correspond to question 1 (expectations of thinking), question 2 (most difficult thing), and question 3 (assessment), respectively.

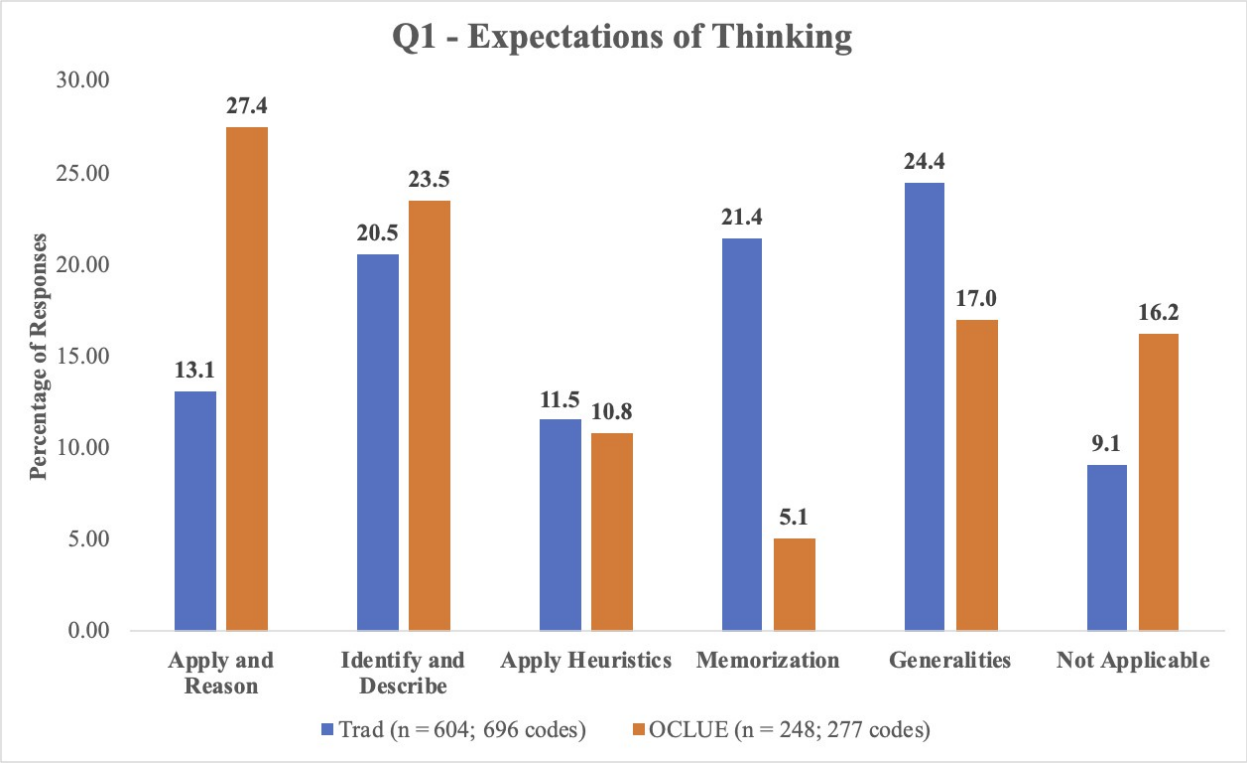

Figure S1. Percentage of student responses for each category/code for Question 1 using nonmutually exclusive codes.

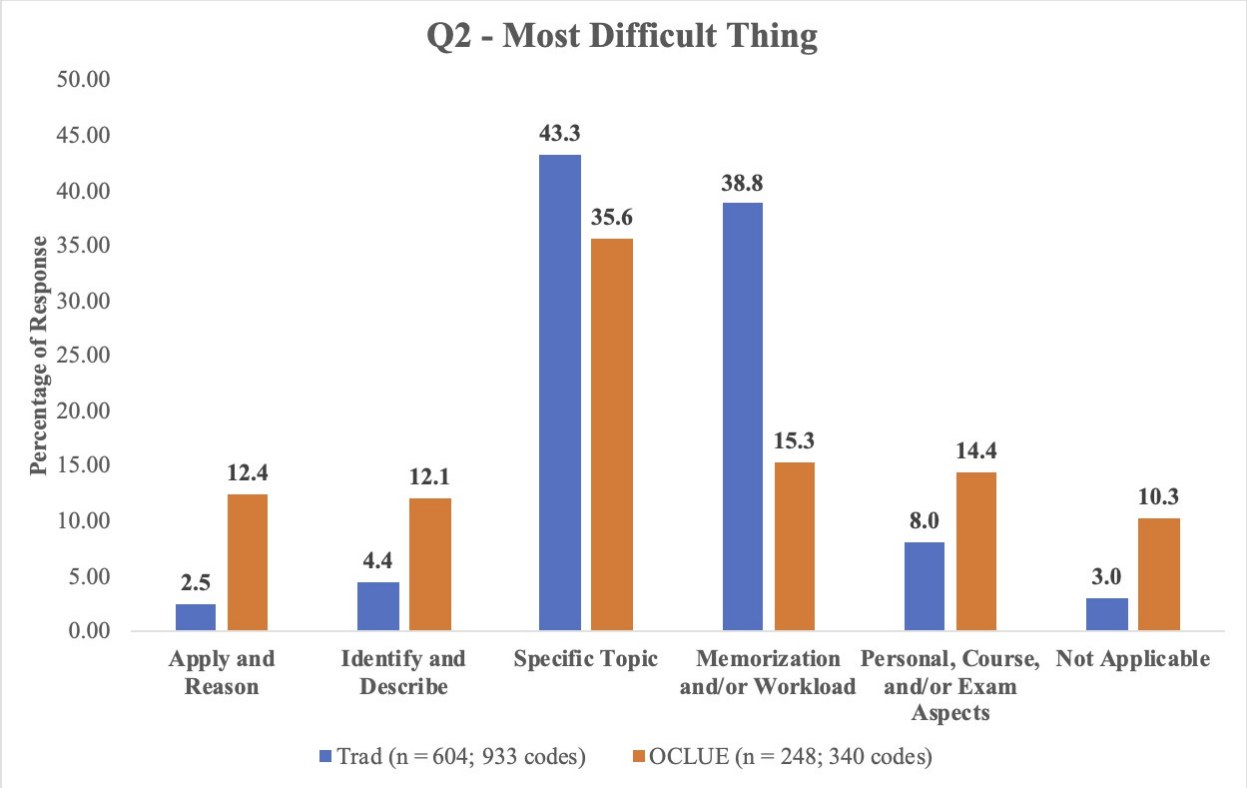

Figure S2. Percentage of student responses for each category/code for Question 2 using nonmutually exclusive codes.

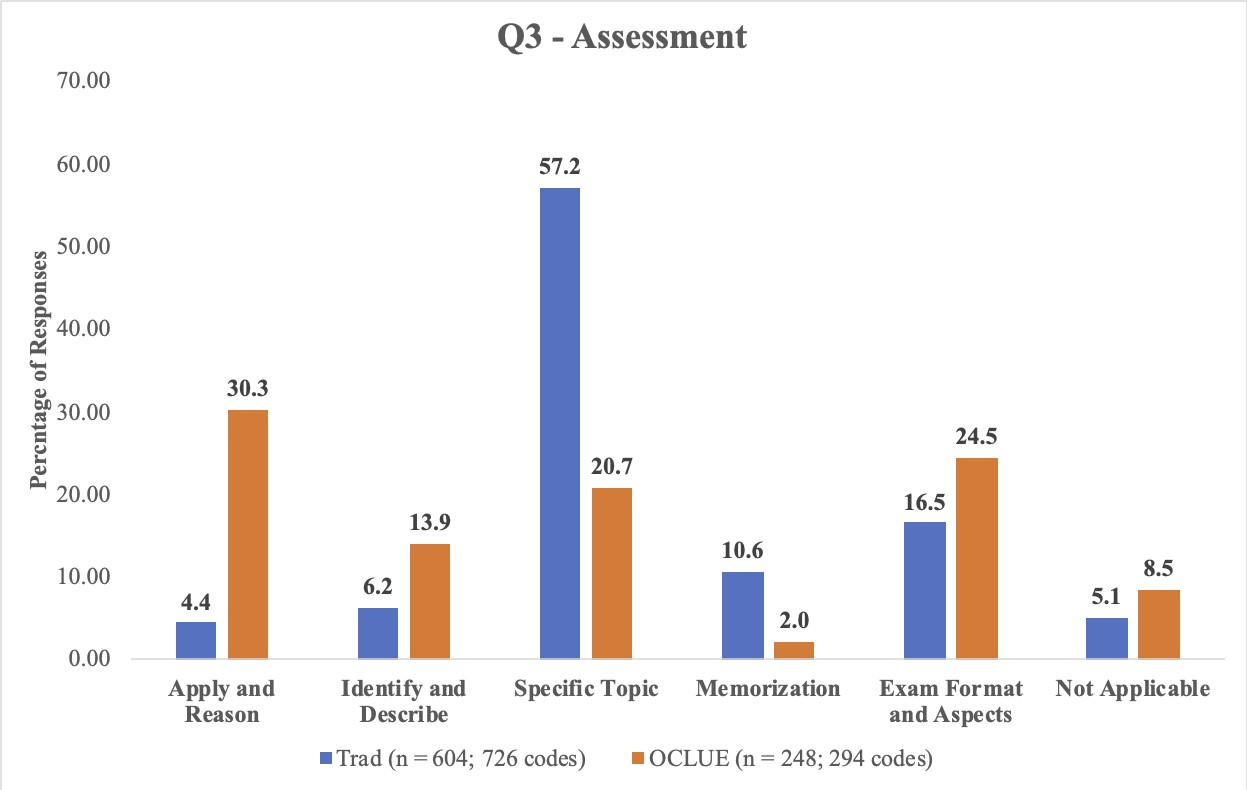

Figure S3. Percentage of student responses for each category/code for Question 3 using nonmutually exclusive codes.

In the case of these photos, the percentages are calculated out of the total number of codes assigned to each course, either traditional or OCLUE. In the graph legends, we have listed the "n" values for the total number of students for traditional (Trad) and OCLUE. Beside that we have also included the total number of nonmutually exclusive codes assigned to the data set. As can be noted, the same patterns are recognized when using mutually exclusive coding.

Pearson's Chi-Square Tests and Post-Hoc Analyses

It's important to note that the calculation of these statistics did include the data in the "Other" theme which we largely do not discuss throughout the paper. Since each question yielded different patterns and codebooks, they will be discussed individually. For question 1

(expectations of thinking), it was noted in the qualitative results that over 21% more OCLUE students perceived they were expected to use their knowledge while over 20% more traditional students perceived they were expected to rely more on rote knowledge. The Pearson's chi-square test further supported the differences noted in the qualitative data at the alpha level of 0.05 (χ^2 = 47.247, df = 2, p < 0.001; see Table 4). This indicates that there is a statistically significant difference in student perceptions of how they are expected to think between the two cohorts, albeit with a relatively small effect size. For question 2 (most difficult thing), almost 25% more students in OCLUE perceived the most difficult aspect was the use of knowledge while almost 42% more traditional students perceived the most difficult aspect was related to the rote knowledge theme (i.e., memorization or a discrete specific topic). Once again, the Pearson chi-square test supported these differences at an alpha level of 0.05 (χ^2 = 144.220, df = 2, p < 0.001; see Table 4), indicating that there is a difference in student perceptions of the most difficult aspects between the OCLUE and traditional organic chemistry students with a medium-to-large effect size.

Finally, for question 3 (assessment), almost 37% more students in OCLUE perceived they were assessed on their use of knowledge, while over 49% more students in traditional perceived they were assessed on their rote knowledge. The chi-square test supported the differences we noted in our coding (χ^2 = 203.605, df = 2, p < 0.001; see Table 4) which indicates that the perceptions of what is assessed between the two cohorts are different with a medium-to-large effect size.

Table S1
Pearson's Chi-Square Tests

Question	Pearson Chi-Square Value	*df*	*p*-value (2-sided)	Cramer's V
Q1: Expectations of Thinking	47.247	2	$p < 0.001$	0.235
Q2: Most Difficult Thing	144.220	2	$p < 0.001$	0.411
Q3: Assessment	203.605	2	$p < 0.001$	0.489

For our chi-square analysis, we used an alpha of 0.05, this means that if the absolute value of the standardized residuals from the post-hoc analysis are greater than the critical value of plus-or-minus 1.96 then that particular theme ("Use of Knowledge", "Rote Knowledge", and/or "Other") is contributing to the statistically significant result from the Pearson's chi-square test. That is, that particular theme is a driving force for the differences between the two cohorts of Traditional and OCLUE. Furthermore, the positive and negative values associated with each standardized residual allows us to determine if the observed number of responses in a particular theme is greater than expected (a positive value) or lower than expected (a negative value). Finally, we can also consider the magnitude of the standardized residuals relative to one another to provide further insights.

For the first question (expectations of thinking), the standardized residual for the "Use of Knowledge" theme was -2.5 for Traditional and 3.9 for OCLUE (see Table 5). The absolute value of both standardized residuals is larger than 1.96 which indicates these two themes are influencing the Pearson's chi-square test. Furthermore, it can be noted that the Traditional standardized residual is negative while the OCLUE value is positive. This means that the number of responses categorized as "Use of Knowledge" for Traditional is lower than expected while it is higher than expected for OCLUE. For the "Rote Knowledge" theme, the standardized residual for Traditional was 2.7 and -4.2 for OCLUE. These values indicate that the "Rote Knowledge"

theme is another driver for significance in the Pearson's chi-square test for question 1 (expectations of thinking). Since the standardized residuals for the "Other" theme are below 1.96, this indicates that this theme is not a significant driver of the chi-square test for this question.

In terms of the second question (most difficult thing), we note similar patterns to question 1 (expectations of thinking), albeit more pronounced. For the "Use of Knowledge" theme, the standardized residuals were -4.6 for Traditional and 7.2 for OCLUE (see Table 5). On the other hand, the "Rote Knowledge" theme standardized residuals were 3.7 and -5.8 for Traditional and OCLUE, respectively. Similar to question 1 (expectations of thinking), these values indicate that both the "Use of Knowledge" and "Rote Knowledge" themes are influencing the outcome of the chi-square test. For this question, though, the standardized residuals for the "Other" theme appear to also be a driver of the chi-square test. However, as previously mentioned, the magnitude of the standardized residuals can provide further insights. For both cohorts (Traditional and OCLUE), the standardized residuals for "Use of Knowledge" and "Rote Knowledge" are larger in magnitude than the standardized residuals of the "Other" theme which sits at -2.7 and 4.2 for Traditional and OCLUE, respectively. In particular, the standardized residual of 7.2 for "Use of Knowledge" in OCLUE is fairly large. This indicates that while the "Other" theme is an influence in the initial chi-square test, the "Use of Knowledge" and "Rote Knowledge" themes are likely stronger influences overall.

Finally, for the third question (assessment), we once again note the same patterns in the standardized residuals as the other questions. For the "Use of Knowledge" theme, the standardized residuals were -5.6 for Traditional and 8.8 for OCLUE. Then, for the "Rote Knowledge" theme, the standardized residuals were 5.0 and -7.8 for Traditional and OCLUE,

respectively. Here, the standardized residuals for "Other" were -1.6 and 2.5 for Traditional and OCLUE. Therefore, similar to the previous results for the other questions, these results indicate that both the "Use of Knowledge" and "Rote Knowledge" categories are primary influences in Pearson's chi-square test, though the standardized residual for OCLUE was higher than 1.96 for the "Other" theme. This indicates that the "Other" theme was a primary driver for the chi-square test for OCLUE (alongside the other themes), but this was not the case for Traditional. As noted in question 2 (most difficult thing), although the "Other" theme in OCLUE was influencing the significant result of the chi-square test, it was far smaller in magnitude than "Use of Knowledge" or "Rote Knowledge".

Table S2
Standardized Residuals from Post-Hoc Analyses

Question	Course	Use of Knowledge Theme	Rote Knowledge Theme	Other Theme
Question 1: Expectations of Thinking	Traditional	-2.5	2.7	0.2
	OCLUE	3.9	-4.2	-0.2
Question 2: Most Difficult Thing	Traditional	-4.6	3.7	-2.7
	OCLUE	7.2	-5.8	4.2
Question 3: Assessment	Traditional	-5.6	5.0	-1.6
	OCLUE	8.8	-7.8	2.5