ETHICAL REASONING IN ACTION
2018-2019 ASSESSMENT REPORT

Caroline Prendergast, M.Ed.
Jeanne Horst, Ph.D.
Chi Hang Au, M.A.
Jessica Jacovidis, Ph.D.

Center for Assessment and Research Studies

James Madison University

Table of Contents

Executive Summary

The 2018 – 2019 school year marked six years of Ethical Reasoning in Action interventions. Similar to previous years, all first-year students experienced *The One Book* and *It's Complicated*, prior to taking the ethical reasoning assessments during Fall Assessment Day. There were no other required interventions, though students may have experienced ethical reasoning training in coursework and/or co-curricular experiences.

A comprehensive assessment protocol was used to evaluate students' facility with the 8KQs, as well as inform future interventions. Assessment data, aligned to each student learning objective, were collected on JMU's Assessment Day. In the Fall, before taking any coursework at JMU, all first-year students were assessed. Students were tested a second time as second-year students in their Spring semester. The collection of data at two time points from the same students allows for longitudinal comparisons to evaluate the extent to which students' ethical reasoning skills and abilities, as measured by the Ethical Reasoning in Action assessments, change over time. Results are highlighted below by instrument.

Survey of Ethical Reasoning (SER)

The Survey of Ethical Reasoning (SER) is a noncognitive measure designed to assess students' attitudes toward ethical reasoning.

- Both first-year and second-year students ranked ethical reasoning as an important skill, with 48% of first-year and 35% of second-year students selecting ethical reasoning in their top three most important desirable skills from a list of ten skills.
- First-year and second-year students reported that ethical reasoning skills are important (4.52 and 4.43 out of 5 points, respectively). Additionally, first-year and second-year students were confident in their ability to apply the ethical reasoning process (4.13 and 4.17 out of 5 points, respectively).
- Students' ratings of the importance of ethical reasoning decreased significantly over time (Fall 2017 to Spring 2019); however, students' confidence scores did not change significantly over time.

Ethical Reasoning Recall Test (ERRT)

The Ethical Reasoning Recall Test (ERRT) is a short, constructed-response measure that asks students to state the 8KQs and to provide a brief explanation of each key question.

- First-year students accurately <u>recalled</u> about 5 and <u>explained</u> about 2 of the 8KQs; second-year students accurately <u>recalled</u> and <u>explained</u> about 1 of the 8KQs.
- For both first-year and second-year students, Fairness and Responsibilities were the easiest 8KQs to <u>recall</u>. Fairness and Outcomes were easiest for first-year students to explain, and Fairness and Character were easiest for second-year students to explain.
- Character, Liberty, and Rights were the most difficult key questions for first-year students to <u>recall</u> and Responsibilities, Character, and Liberty were the most difficult for first-year students to <u>explain</u>. Outcomes, Character, Liberty, and Rights were the most difficult key questions for second-year students to <u>recall</u> and Liberty and Authority were the most difficult for second-year students to <u>explain</u>.
- Students <u>recalled</u> and <u>explained</u> significantly more of the 8KQs as first-year students than as second-year students, <u>recalling</u> at least four fewer key questions and <u>explaining</u> at least three fewer key questions as second-year students than as first-year students.

Ethical Reasoning Identification Test (ERIT)

The ERIT is a 50-item multiple-choice test that asks students to consider a scenario and choose the key question most applicable to the decision or rationale presented in the scenario.

- On average, first-year students scored about 68% correct on the ERIT, while second-year students scored about 66% correct.
- First-year and second-year students answered the most items correctly for Fairness, Outcomes, Character, and Empathy, suggesting these key questions may be easier for students to grasp.
- First-year and second-year students answered the least items correctly for Liberty and Rights, suggesting these key questions may be harder for students to grasp.
- Students' ERIT total scores changed statistically significantly from Fall 2017 to Spring 2019; however, these changes may not be practically significant.

Ethical Reasoning - Writing (ER-WR) Essay

The ER-WR essay is a performance assessment instrument that asks students to apply the 8KQ framework to an ethical dilemma. The original ER-WR measures students' ability to describe an ethical dilemma from their own lives and apply the 8KQ framework. The ER-WR2 and ER-WR3 essay prompts provide students

with a hypothetical scenario and ask them to apply the 8KQs. The modified version of the ER-WR is similar to the original ER-WR, but respondents are explicitly cued to use the 8KQ framework when writing their responses. Essays are scored using the **Ethical Reasoning Rubric** (located in Appendix B).

- Generally, scores were highest on element A (when this element was used) and lowest on elements C, D, and E.
- Students scored statistically significantly higher as first-year students than as second-year students for both versions of the ER-WR for which this longitudinal comparison could be made (original ER-WR and ER-WR2).
- Given that four versions of the ER-WR are currently in use (and their usage patterns differ across the past few years), not all analyses are possible for all versions of the measure.
- The average first-year scores for the ER-WR2 ($M = 1.57$), ER-WR3 ($M = 1.56$), and modified ER-WR ($M = 1.54$) are similar to each other and somewhat higher than average first-year scores for the original ER-WR ($M = 1.22$).
- The average second-year scores are very similar for the original ER-WR ($M = 1.00$), ER-WR2 ($M = 1.00$), and modified ER-WR ($M = 1.11$). No second-year scores were collected for the ER-WR3.

Overall Summary

Generally, students' **attitudes** toward ethical reasoning appear to remain stable or decrease slightly over time. Students' **ethical reasoning skills** appear to be either stable, or decreasing, over time. The Ethical Reasoning in Action team could consider why student knowledge appears to be decreasing over time. For example, given that students received the *It's Complicated* programming as part of their freshman orientation **and** students may not have had an additional exposure to the 8KQs, it seems reasonable that students would have forgotten some of what they learned. Additional interventions could be developed to ensure that students retain the information they learned through the *It's Complicated* programming.

Background, Objectives, & Interventions

The Ethical Reasoning in Action Annual Technical Report houses information regarding Ethical Reasoning assessment results. The Ethical Reasoning in Action team has set the foundational goals of

1) elevating the campus-wide understanding and discourse on ethical reasoning as a teachable, evaluative process;
2) providing a unifying framework that aligns campus efforts to teach and assess ethical reasoning; and
3) encouraging multiple avenues of intentional connection among personal, professional, and civic application of ethical reasoning skills in the classroom, co-curricular activities, and student life.

These goals are facilitated through the use of the Eight Key Questions (8KQs) ethical reasoning framework; see Appendix A for a list of the 8KQs.

The 2018-2019 academic year was the sixth year of the Ethical Reasoning in Action intervention plan. A comprehensive assessment protocol was used to evaluate students' facility with the 8KQs, as well as inform future interventions. The Ethical Reasoning in Action team has outlined seven student learning objectives (SLOs) that students should meet as a result of their participation in the ethical reasoning interventions:

Cognitive Learning Outcomes

1. Students will be able to state, from memory, all eight Key Questions (8KQs).
   Alternate assessment: From a list of ways of conceptualizing issues, students will correctly identify the eight Key Questions (8KQs).
2. When given a specific decision and rationale on an ethical issue or dilemma, students will correctly identify the Key Question most consistent with the decision and rationale.
3. Given a specific scenario, students will identify appropriate considerations for each of the eight Key Questions (8KQs).
   Alternate approach: Students will be able to provide the specific considerations raised or rationale implied when applying every Key Question to an ethical situation or dilemma.
4. For a specific ethical situation or dilemma, students will evaluate courses of action by applying (weighing and, if necessary, balancing) the considerations raised by Key Questions.
5. Students will apply SLO 4 to their own personal, professional, and civic ethical cases.
   NOTE: Implied within this SLO is the students' ability to identify an ethical situation, based on the belief that the process of ethical reasoning increases discriminatory capacities. This will be addressed via the assessment rubric.

Attitudinal Outcomes

6. Students will report that they view ethical reasoning skills as important.
7. Students will report increased confidence in their ability to use the ethical reasoning process.

Although students experience varying amounts of ethical reasoning interventions and exposure to the 8KQs while at James Madison University (JMU), there are two interventions that all students experience: *The One Book* and *It's Complicated*. Upon paying their deposit to attend JMU, all first-year students receive *The One Book*. This publication contains essential steps new students are required to complete to matriculate into the university. A two-page spread was dedicated to introducing incoming students to Ethical Reasoning in Action as well as explaining the purpose of Ethical Reasoning in Action, why it is important, and how it will affect their JMU learning experience. *The One Book* content links indirectly to SLO 6, the importance of ethical reasoning, as it emphasizes the overall program and its relevance to student learning.

During 1787 Orientation, all first-year students experience *It's Complicated*, the second Ethical Reasoning in Action intervention to which all students are exposed. During *It's Complicated*, faculty, staff, and administrators facilitate an ethical reasoning case scenario with small groups of first-year students over 75-minutes. Students are introduced to the 8KQs, watch a video depicting a case scenario, and grapple with a tough decision to be made as part of the case scenario. *It's Complicated* emphasizes the importance of ethical reasoning (SLO6), exposes students to the 8KQs in a way that they can understand the meaning behind each question (SLO1), and allows students the opportunity to identify the most relevant key questions for the case scenario (SLOs 2, 3, and 4).

Additionally, numerous General Education, Honors, and major-specific faculty are modifying their courses and assignments to include the 8KQs. Specifically, several faculty have participated in the Core Introduction workshop, curriculum development, course redesign and are infusing their courses with the 8KQs ethical reasoning framework. Moreover, Ethical Reasoning in Action has been the focus of several doctoral dissertation research projects. From these projects, we know that ethical reasoning is a skill that may be taught and learned, and with targeted ethical reasoning curricula, students improve in their ethical reasoning skills and abilities (Good, 2015; Holzman, 2018; Smith, 2017). Thus, the next step is to determine how to scale up the ethical reasoning interventions across the JMU campus community.

As mentioned, a comprehensive assessment protocol is used to evaluate students' facilities with the 8KQs, as well as inform future interventions. Assessment data, aligned to each student learning objective, are collected on JMU's Assessment Day. In the fall, before taking any coursework at JMU, all first-year students are assessed. Students are tested a second time as second-year students in their spring semester. The collection of data at two time points from the same students allows for longitudinal comparisons to

evaluate the extent to which students' ethical reasoning skills and abilities, as measured by the Ethical Reasoning in Action assessments, change over time.

During the 2018-2019 academic year, four Ethical Reasoning in Action assessment instruments were administered. All objectives were covered by at least one assessment. Further description of each instrument, as well as assessment results for each instrument, are provided below.

About this Report

In the 2016-2017 reporting cycle, a new reporting format was implemented. Similar to previous years, this report is organized by each assessment instrument. However, for each instrument, results are now organized by questions relevant to Ethical Reasoning in Action stakeholders. Each section includes a brief description of the instrument, first-year student results, second-year student results, and longitudinal comparisons. Historical results for psychometric properties of the instruments, such as reliability and validity evidence, have been removed from the report. If stakeholders wish to evaluate the psychometric properties of scores across previous years, they may refer to the previous years' reports.

Analyses fall into three major categories: longitudinal (cohort) analyses, academic year analyses, and single semester analyses. Single semester analyses analyze either first-year students (using data from Fall Assessment Day) or second-year students (using data from Spring Assessment Day). Academic year analyses use data from both Assessment Days in a single year; therefore, they capture both freshmen and sophomores in a given academic year. See the diagram to the right for a visual representation of these analyses.

Longitudinal analysis, Fall 2017 cohort

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

2018-2019 Academic Year analysis

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

First-year students

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

Second-year students

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

Survey of Ethical Reasoning (SER)

The Survey of Ethical Reasoning (SER) is a noncognitive measure designed to assess students' attitudes toward ethical reasoning. Specifically, the assessment was created to explicitly measure SLOs 6 and 7. Measuring students' attitudinal SLOs is important because it may be difficult to enhance students' ethical reasoning skills if they perceive these skills as unimportant. Moreover, learning how to apply the ethical reasoning process to real life situations goes hand-in-hand with valuing ethical reasoning skills. Perhaps the more students value ethical reasoning, the more they will work toward improving these skills. Similarly, it is important to ensure students feel confident applying the ethical reasoning process to real life situations.

The SER is comprised of four sections that include rank-order items and Likert-scale items. The first section of the SER asks students to rank order 10 different skills, such as artistic skills, critical thinking skills, ethical reasoning skills, interpersonal skills, writing skills. Students are instructed to rank these skills from 1 (*Most Important*) to 10 (*Least Important*). The second section of the SER includes five statements about perceived importance of ethical reasoning and five statements about confidence in applying the ethical reasoning process. This section also includes six statements that correspond to the Ethical Reasoning in Action SLOs and the 8KQs (i.e., "*When faced with an ethical situation, I can correctly identify the most relevant key questions*"). Students are asked to indicate how much they agree with each statement using a five-point Likert scale (1 = *Strongly Disagree*, 2 = *Somewhat Disagree*, 3 = *Neither Agree nor Disagree*, 4 = *Somewhat Agree*, and 5 = *Strongly Agree*).[1] The third section of the SER describes five different behaviors related to applying, discussing, and engaging in ethical reasoning. Students are asked to indicate how frequently they engage in each of the five behaviors using a five-point Likert scale (1 = *Never*, 2 = *Every Few Months*, 3 = *Monthly*, 4 = *Weekly*, and 5 = *Daily*). The final section of the SER lists each of the 8KQs separately. Students are asked to indicate how important each key question is in their ethical reasoning process using a five-point Likert scale (1 = *Not At All Important*, 2 = *Slightly Important*, 3 = *Somewhat Important*, 4 = *Important*, and 5 = *Very Important*).

Similar to previous years, a confirmatory factor analysis indicated that a two-factor solution fit the data. The two factors are thought to represent "Importance" and "Confidence." As such, results are reported as two subscales, one for importance, and another for confidence. Reliability was adequate for the

---

[1] Between the importance and confidence items is one distractor item (i.e., "*I am not a JMU student.*") that is used to identify students who are responding carelessly.

importance subscale scores (α = .86 for Fall 2018; α = .88 for Spring 2019) as well as the confidence subscale scores (α = .84 for Fall 2018; α = .81 for Spring 2019).

How important are ethical reasoning skills to first-year students?

Fall 2017 ⟩ Spring 2018 ⟩ **Fall 2018** ⟩ Spring 2019

In Fall 2018, 346 first-year students completed the first ten items on the SER, on which they were asked to prioritize a set of desirable skills (e.g., artistic, critical thinking, etc.). A total of 163 first-year students (48%) indicated that ethical reasoning was in their top three most important desirable skills, with 52 first-year students (15%) indicating that ethical reasoning was their most desirable skill. Ethical reasoning was most often ranked the third most important skill. Figure 1 displays the percentage of first-year students who placed ethical reasoning at each importance ranking.

Percent of Students Ranking Ethical Reasoning at Each Level
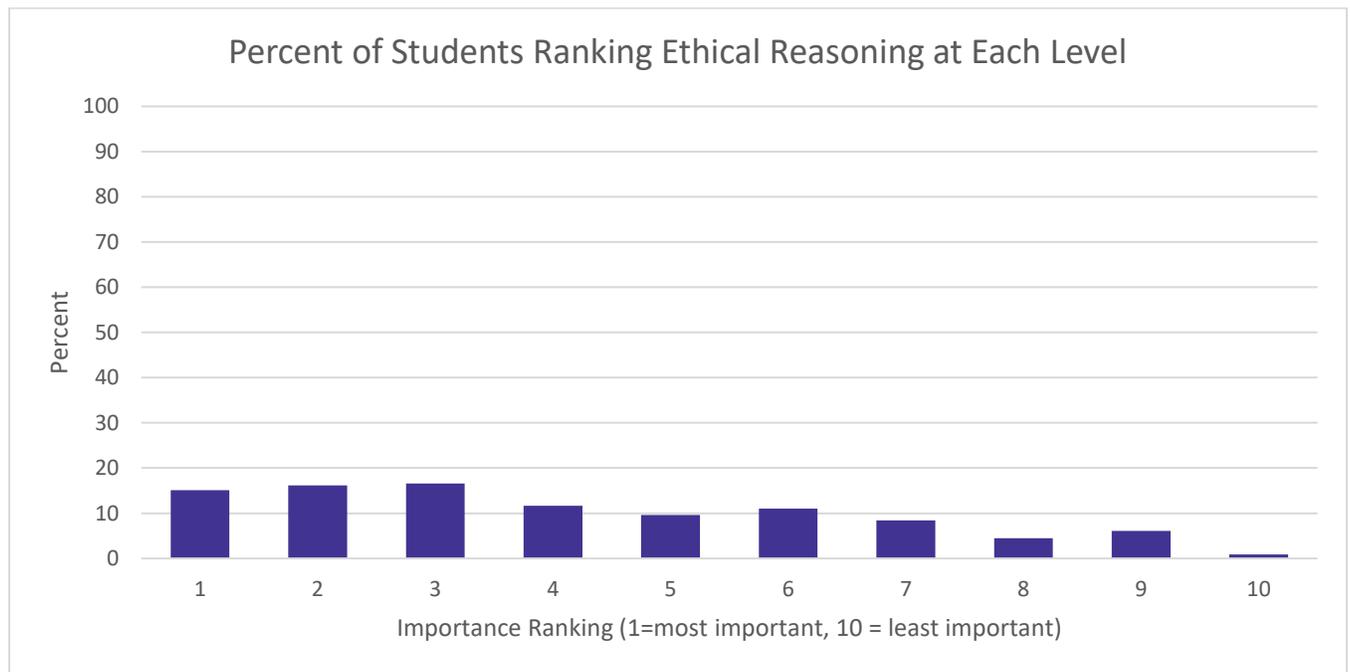
Figure 1. Percent of first-year students ranking ethical reasoning skills at importance levels

29.2% of first-year students identified critical thinking skills as the most important skill, followed by 15.0% of students placing ethical reasoning skills in this position. Oral communication skills were identified as the most important skill by 14.2% of responding first-year students.

In addition to rank-ordering desirable skills, first-year students were explicitly asked about their perceived importance of ethical reasoning skills through five Likert-type items. The distribution of average scores is

provided in Figure 2. On average, first-year students scored 4.52 out of 5 points on the importance subscale, suggesting that first-year students agree that ethical reasoning skills are important.
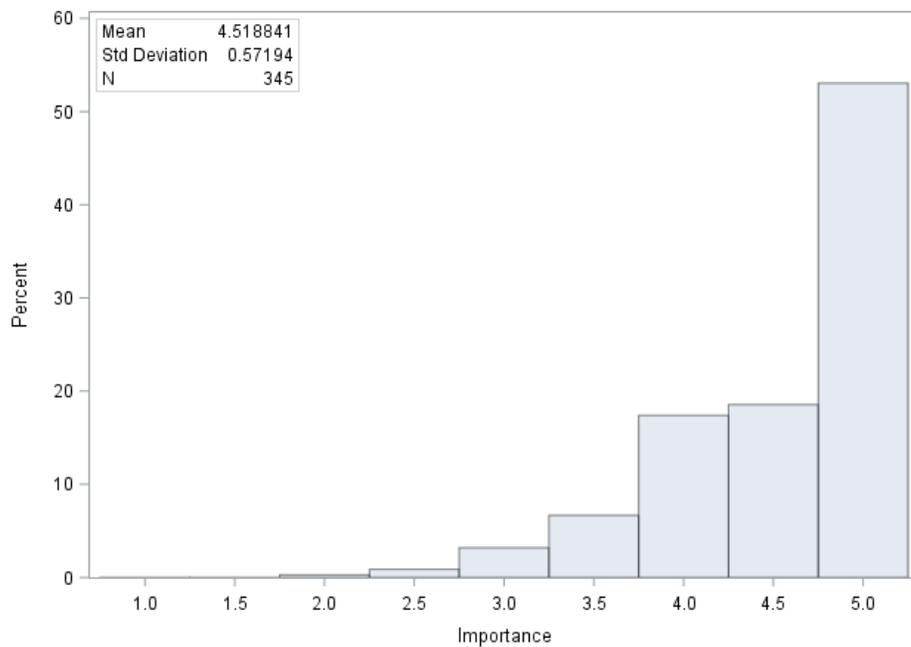
Figure 2. Distribution of average importance scores

How confident are first-year students in their ethical reasoning abilities?

Fall 2017 ⟩ Spring 2018 ⟩ Fall 2018 ⟩ Spring 2019

In Fall 2018, 344 first-year students completed the confidence subscale on the SER. The confidence subscale items are intended to measure students' perceived confidence in applying the ethical reasoning process. The distribution of average subscale scores is provided in Figure 3. On average, first-year students scored 4.13 out of 5 points on the confidence subscale, suggesting that first-year students agree that they have confidence to apply the ethical reasoning process.

Mean: 4.134302
Std Deviation: 0.638501
N: 344
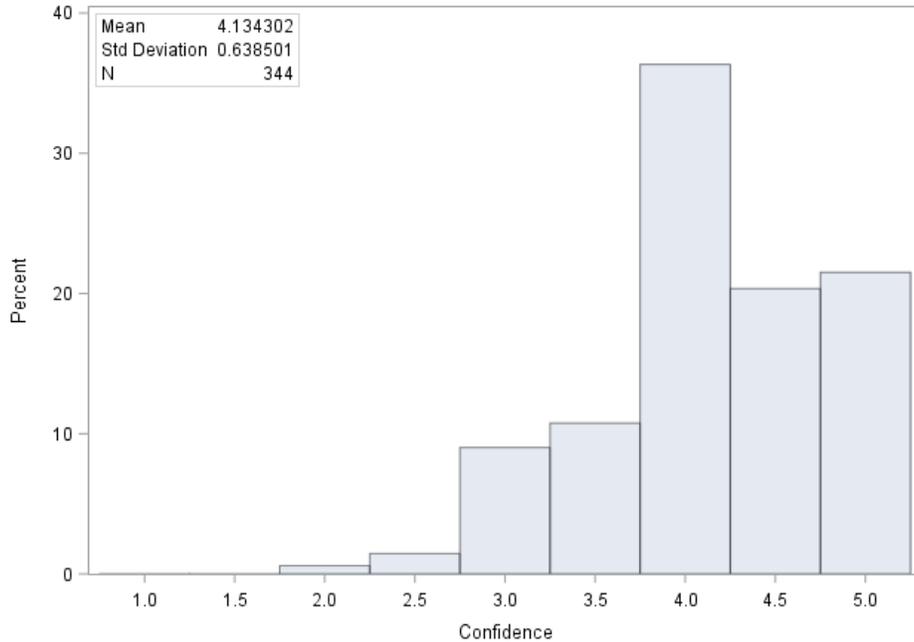
Figure 3. Distribution of average confidence scores

How important are ethical reasoning skills to second-year students?

Fall 2017 > Spring 2018 > Fall 2018 > **Spring 2019**

In Spring 2019, 422 second-year students completed the first ten items on the SER, on which they were asked to prioritize a set of desirable skills, (e.g., artistic, critical thinking, etc.). A total of 146 second-year students (35%) indicated that ethical reasoning was in their top three most important desirable skills, with 45 second-year students (11%) indicating that ethical reasoning was their most desirable skill. Ethical reasoning was most often ranked the 4th or 5th most important skill. Figure 4 displays the frequency of second-year students who placed ethical reasoning at each importance ranking.

Percent of Students Ranking Ethical Reasoning at Each Level

Importance Ranking (1=most important, 10 = least important)
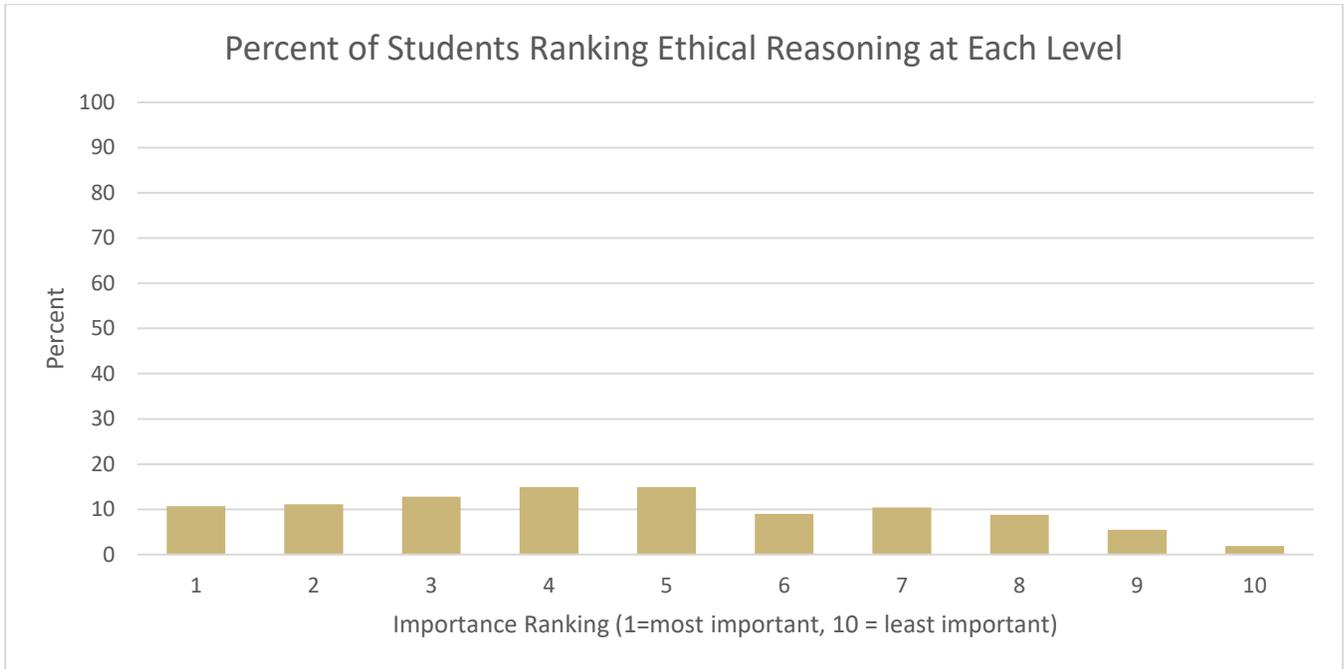
Figure 4. Percent of second-year students ranking ethical reasoning skills at importance levels

20.9% of second-year students identified critical thinking skills as the most important skill, followed by 19.7% of students placing oral communication skills in this position. Interpersonal skills were identified as the most important skill by 16.1% of responding first-year students.

In addition to rank-ordering desirable skills, second-year students were explicitly asked about their perceived importance of ethical reasoning skills through five Likert-type items. The distribution of average scores is provided in Figure 5. On average, second-year students tend to score 4.43 out of 5 points on the importance subscale, suggesting that second-year students agree that ethical reasoning skills are important.

Mean   4.431905
Std Deviation   0.660101
N   420
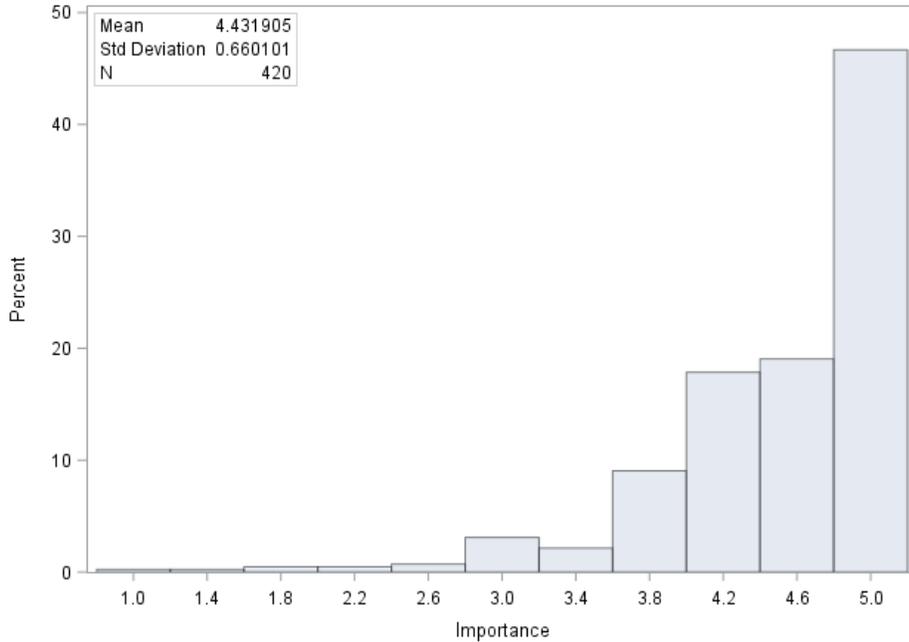
Figure 5. Distribution of average importance scores

How confident are second-year students in their ethical reasoning?

Fall 2017 〉 Spring 2018 〉 Fall 2018 〉 Spring 2019

In Spring 2019, 421 second-year students completed the confidence subscale on the SER. The confidence subscale items are intended to measure students' perceived confidence in applying the ethical reasoning process. The distribution of average subscale scores is provided in Figure 6. On average, second-year students scored 4.16 out of 5 points on the confidence subscale, suggesting that second-year students agree that they have confidence to apply the ethical reasoning process.

14
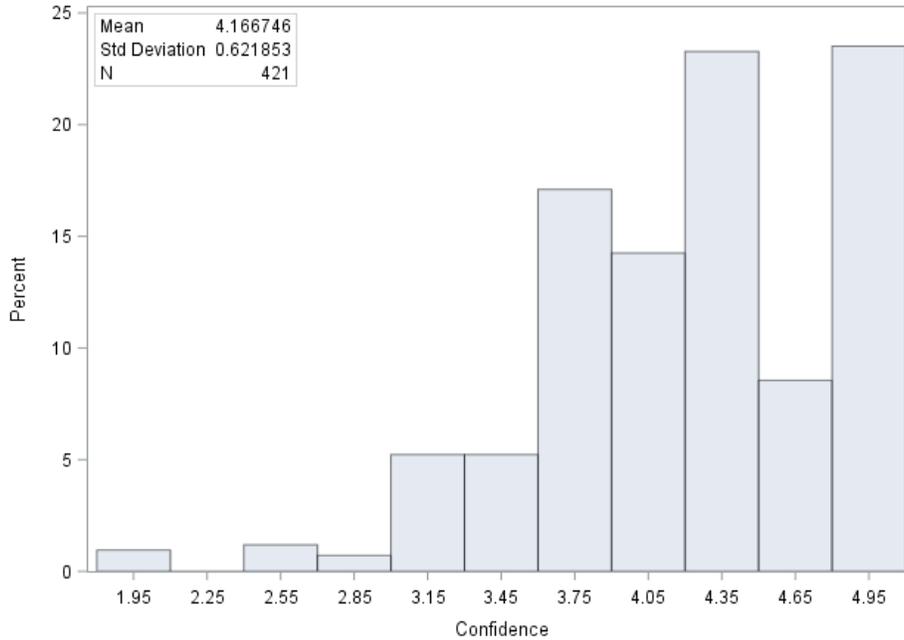
Figure 6. Distribution of average confidence scores

How does students' perceived importance and confidence change over time?

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

Assessment Day data collection methodologies allow for longitudinal comparisons (e.g., comparing a student's pretest score to that same student's posttest score). Thus, students' SER scores as entering first-year students (Fall 2017) scores were compared to their SER scores as second-year students (Spring 2019). Only students who completed the SER in Fall 2017 **and** Spring 2019 were included in the analyses. Descriptive information is presented in Table 1 below. On average, students' ratings of the importance of ethical reasoning was statistically significantly lower in Spring 2019 than in Fall 2017, [$t(414) = -2.57$, $p = .0107$, $d = .18$]. However, although statistically significant, the difference was not meaningfully different, as evidenced by the low Cohen's d value. Ethical Reasoning in Action stakeholders may consider whether this is a meaningful decrease in importance scores. Students' confidence scores did not change significantly from Fall 2017 to Spring 2019, [$t(414) = -1.48$, $p = .139$, $d = .10$].

15

Table 1. *Subscale-Level Descriptive Statistics for the Survey of Ethical Reasoning (SER) Average Scores: Matched Longitudinal Sample*

Cohort	Subscale	Min	Max	Mean	*SD*
Fall 2017	Importance	1.00	5.00	4.54	0.58
	Confidence	1.40	5.00	4.11	0.65
Spring 2019	Importance	1.00	5.00	4.43	0.66
	Confidence	1.80	5.00	4.17	0.61

Note. Subscales ranged from 1 to 5.

Ethical Reasoning Recall Test (ERRT)

The Ethical Reasoning Recall Test (ERRT) is a short, constructed-response measure designed to explicitly address SLO 1. The test asks students to state the 8KQs and to provide a brief explanation of each key question. Thus, the ERRT consists of two subscales: KQ recall and KQ explanations. Student responses are scored by two raters. The KQ recall subscale items are scored as correct or incorrect. The KQ explanations are scored on a three-point scale (0 = *incorrect*, .5 = *partially correct*, and 1 = *correct*). Both subscales range from 0 to 8.

To evaluate the extent to which scores reflect students' abilities, rather than rater characteristics or other random error, we employed a generalizability analysis. There are two types of reliability estimates generated from a generalizability analysis: relative and absolute. Relative estimates are represented by the G-coefficient and are most useful when comparing students to one another. The G-coefficient is more appropriate for the desired ERRT comparisons, and thus is reported here. The KQ recall scores for first-year students had similar reliability (G = 0.99) to second-year students (G = 0.99). For KQ explanation scores, first-year students had slightly lower reliability (G = 0.77) than the second-year students (G = 0.80). Though reliability was lower for the KQ explanation scores than the KQ recall scores, reliability was still acceptable, supporting inferences that scores are predominately representative of students' abilities.

Five transfer students completed the ERRT in the Spring 2019 semester. Although their scores were retained for reliability analyses, these students' scores were removed from the dataset for the remaining ERRT analyses. In future years, if more transfer students complete the measure during the spring semester, transfer student scores should be compared to scores of students who entered JMU as first-year students.

How many 8KQs do first-year students recall?

The distribution of first-year students' scores is provided in Figure 7. On average, first-year students accurately recalled about 5 out of the 8 KQs; these results are slightly lower than the results from previous years.

Number of KQs correctly recalled by first-year students

Number of KQs correctly recalled	Percent
0	1.2
1	4.79
2	5.99
3	9.58
4	10.18
5	19.16
6	19.76
7	17.37
8	11.98
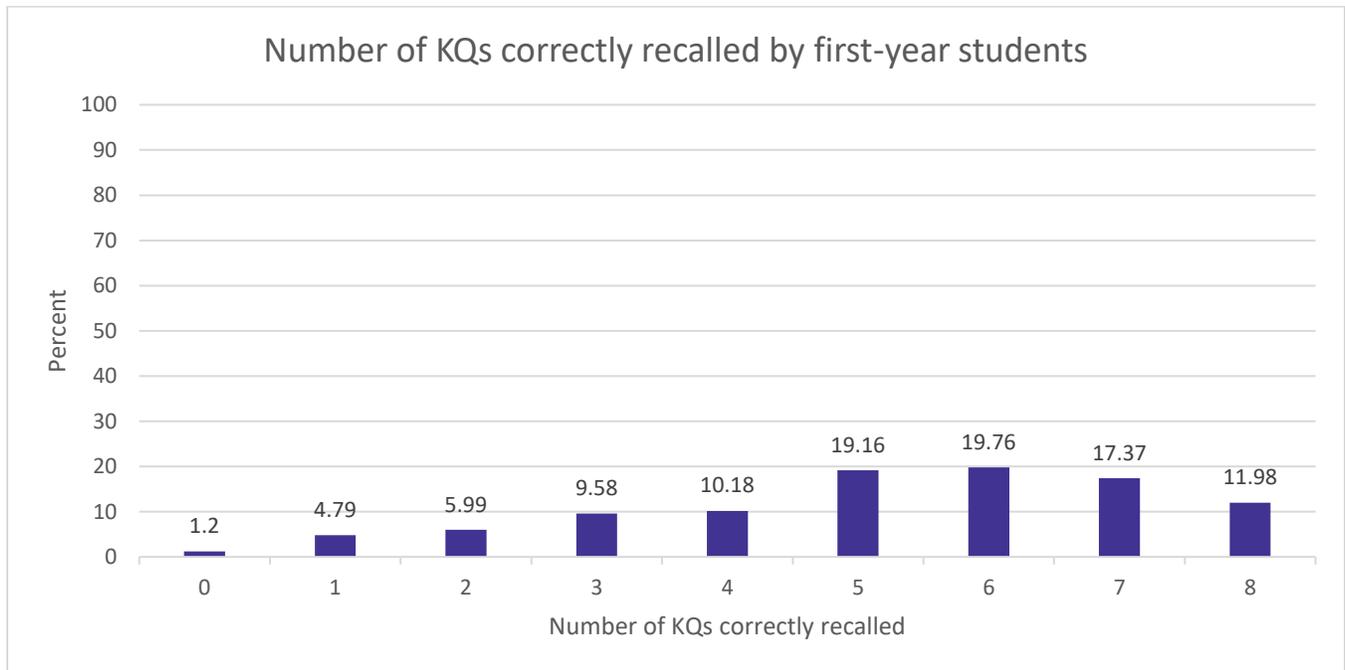
Figure 7. Percentage of first-year students correctly recalling KQs on the ERRT.

Table 2 provides the percentage of first-year students recalling each of the 8KQs. It appears that Fairness and Responsibilities were the easiest 8KQs to recall, as at least 70% of first-year students correctly recalled these KQs. It appears that Character, Liberty, and Rights were the most difficult key questions to recall, with less than 60% of first-year students able to recall each of these 8KQs.

Table 2. *Difficulty of Key Question Recall (Fall 2018)*

Key Question	Percent of Students Recalling KQ
Fairness	85.03%
Outcomes	62.28%
Responsibilities	71.26%
Character	45.51%
Liberty	55.69%
Empathy	68.86%
Authority	68.26%
Rights	58.68%

Note. N = 167

How many 8KQs do first-year students explain?

Fall 2017 〉 Spring 2018 〉 **Fall 2018** 〉 Spring 2019

The distribution of first-year students' scores is provided in Figure 8. On average, first-year students accurately explained about 2 out of the 8KQs; these results are slightly lower than the results from previous years.


Number of KQs correctly explained by first-year students
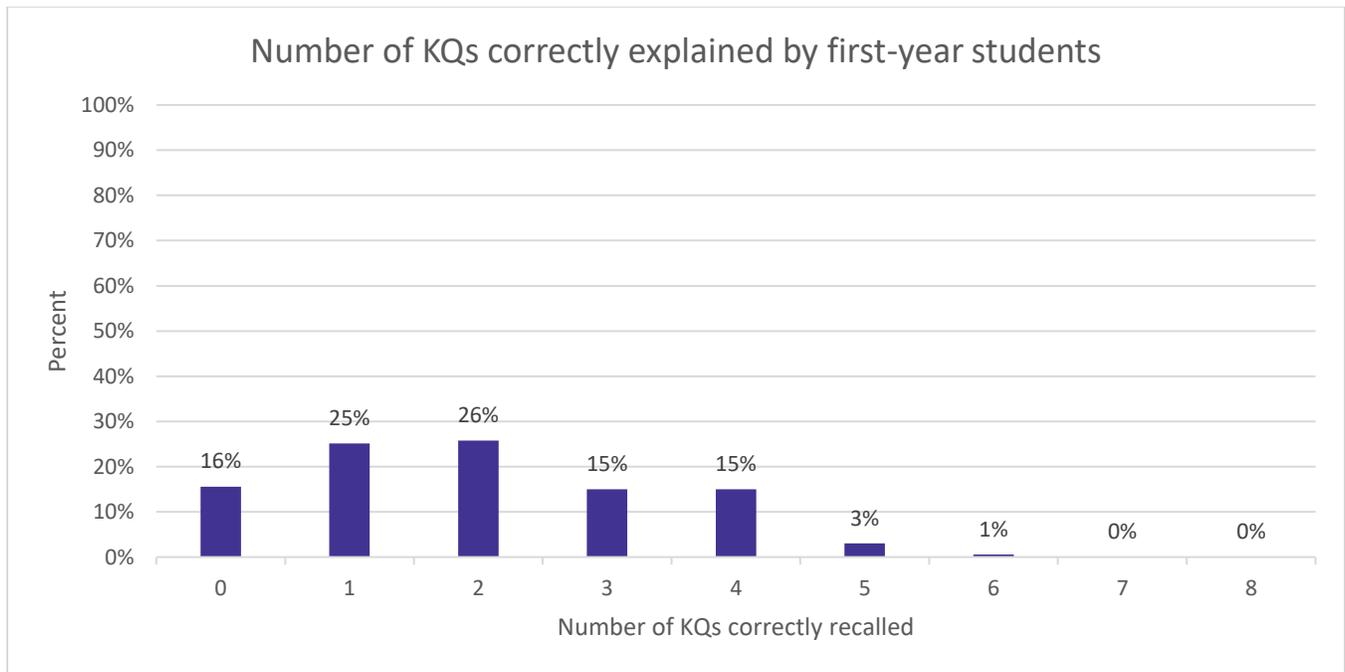
Figure 8. Percentage of first-year students correctly explaining KQs on the ERRT.

Table 3 provides the percentage of first-year students explaining each of the 8KQs. Recall that students may receive partial credit on the explanation subscale. Thus, information is provided regarding the percentage of students who received **partial** credit, percentage of students who received **full** credit, and the percentage of students who received **any** credit. It appears that Fairness and Outcomes were the easiest key questions to explain, with at least 60% of first-year students receiving partial or full credit for their explanations. It appears that Responsibilities, Character, and Liberty were the most difficult key questions to explain, as less than 50% of first-year students received partial or full credit for their explanations of these KQs.

Table 3. *Difficulty of Key Question Explanation (Fall 2018)*

Key Question	Percentage of Students Receiving <u>Partial</u> Credit	Percentage of Students Receiving <u>Full</u> Credit	Percentage of Students Receiving <u>ANY</u> Credit
Fairness	69.46%	8.98%	78.44%
Outcomes	50.30%	14.97%	65.27%
Responsibilities	35.33%	8.98%	44.31%
Character	38.92%	3.59%	42.51%
Liberty	36.53%	7.78%	44.31%
Empathy	53.89%	5.99%	59.88%
Authority	46.11%	7.19%	53.29%
Rights	46.71%	6.59%	53.29%

Note. N=167

How many 8KQs do second-year students recall?

Fall 2017 〉 Spring 2018 〉 Fall 2018 〉 Spring 2019

The distribution of second-year students' scores is provided in Figure 9. On average, second-year students accurately recalled about 1 out of the 8KQs. Over 45% of second-year students did not recall any of the 8KQs; about 20% of second-year students recalled only one of the 8KQs.
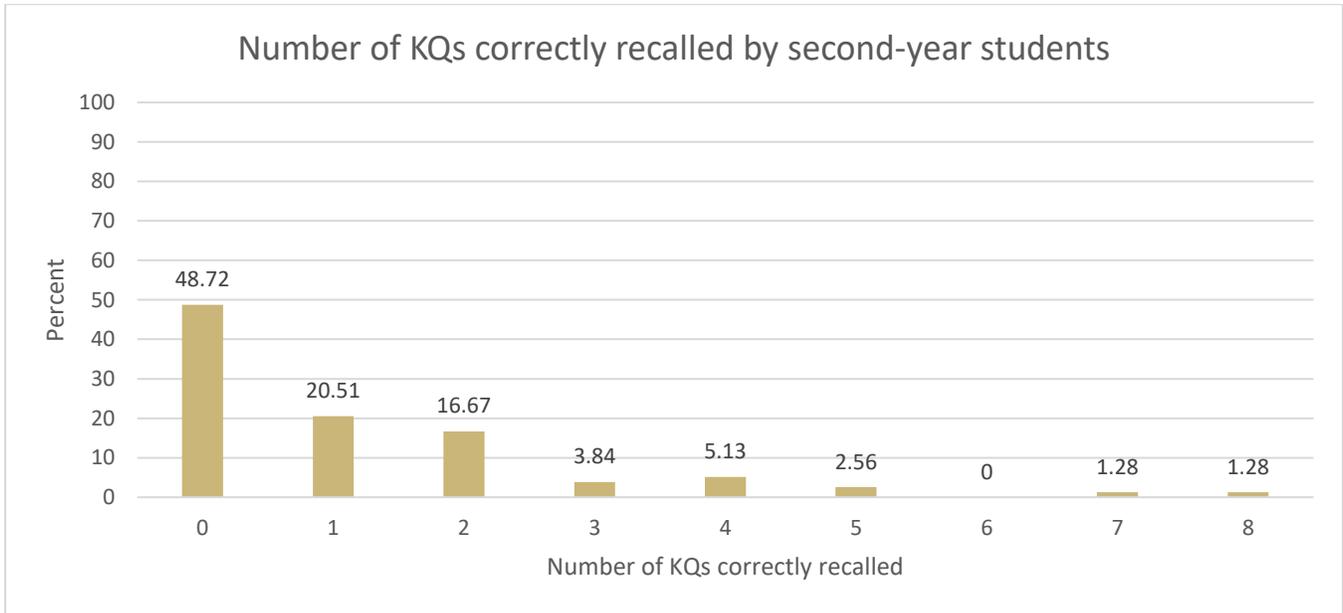
Figure 9. Percentage of second-year students recalling KQs correctly on the ERRT.

Table 4 provides the percentage of second-year students recalling each of the 8KQs. It appears that Fairness and Responsibilities were the easiest 8KQs to recall, with more than 20% of second-year students correctly recalling these 8KQs. It appears that Outcomes, Character, Liberty, and Rights were the most difficult key questions to recall, as less than 10% of second-year students were able to recall each of these 8KQs.

Table 4. *Difficulty of Key Question Recall (Spring 2019)*

Key Question	Percent of Students Recalling KQ
Fairness	20.51%
Outcomes	6.41%
Responsibilities	28.21%
Character	8.97%
Liberty	8.97%
Empathy	15.38%
Authority	16.67%
Rights	7.69%

Note. N = 78

How many 8KQs do second-year students explain?

The distribution of second-year students' scores is provided in Figure 10. On average, second-year students accurately explained less than one of the 8KQs. About 76% of second-year students did not explain any of the 8KQs correctly.
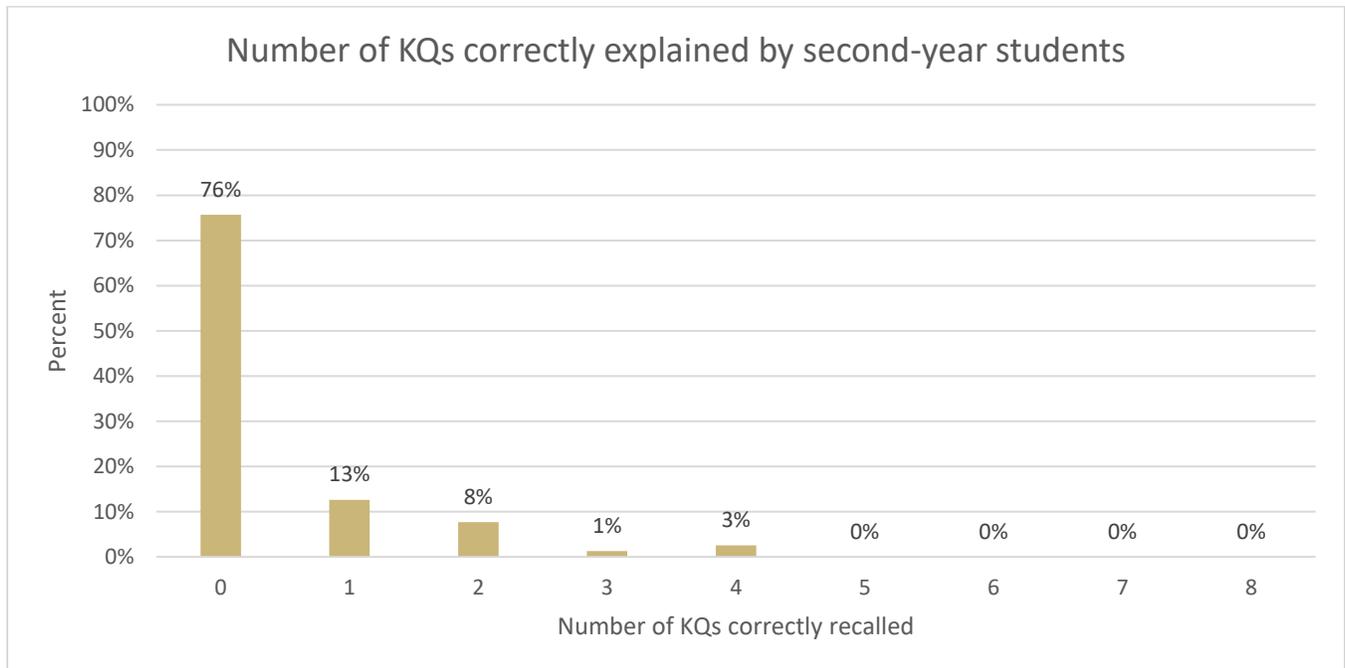

Figure 10. Percentage of second-year students correctly recalling KQs on the ERRT.

Table 5 provides the percentage of second-year students explaining each of the 8KQs. Recall that students may receive partial credit on the explanation subscale. Thus, information is provided regarding the percentage of students who received **partial** credit, percentage of students who received **full** credit, and the percentage of students who received **any** credit. It appears that Fairness and Character were the easiest key questions to explain, with more than 20% of second-year students receiving partial or full credit for their explanations. It appears that Liberty and Authority were the most difficult key questions to explain, as less than 15% of second-year students received partial or full credit for their explanations of these KQs.

Table 5. *Difficulty of Key Question Explanation (Spring 2019)*

Key Question	Percentage of Students Receiving <u>Partial</u> Credit	Percentage of Students Receiving <u>Full</u> Credit	Percentage of Students Receiving <u>ANY</u> Credit
Fairness	24.36%	3.85%	28.21%
Outcomes	14.10%	1.28%	15.38%
Responsibilities	15.38%	1.28%	16.67%
Character	28.21%	1.28%	29.49%
Liberty	5.13%	1.28%	6.41%
Empathy	15.38%	0.00%	15.38%
Authority	10.26%	1.28%	11.54%
Rights	14.10%	2.56%	16.67%

Note. N=78

Do students' abilities to recall and explain the 8KQs change over time?

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

Assessment Day data collection methodologies allow for longitudinal comparisons (e.g., comparing a student's pretest score to that same student's posttest score). Thus, students' ERRT scores as entering first-year students (Fall 2017) scores were compared to their ERRT scores as second-year students (Spring 2019). Only the 40 students who completed the ERRT in Fall 2017 **and** Spring 2019 were included in the analyses.

On average, students' ERRT recall scores were statistically significantly higher as first-year students than as second-year students, $[t(39) = 10.97, p < .0001; d = 2.37]$. Specifically, second-year students recalled at least four fewer key questions ($M = 1.33$) than they did as first-year students ($M = 5.98$). Figure 11 presents the percentage of students correctly recalling each of the 8KQs as first-year students and second-year students.

Percent of Students Recalling KQs
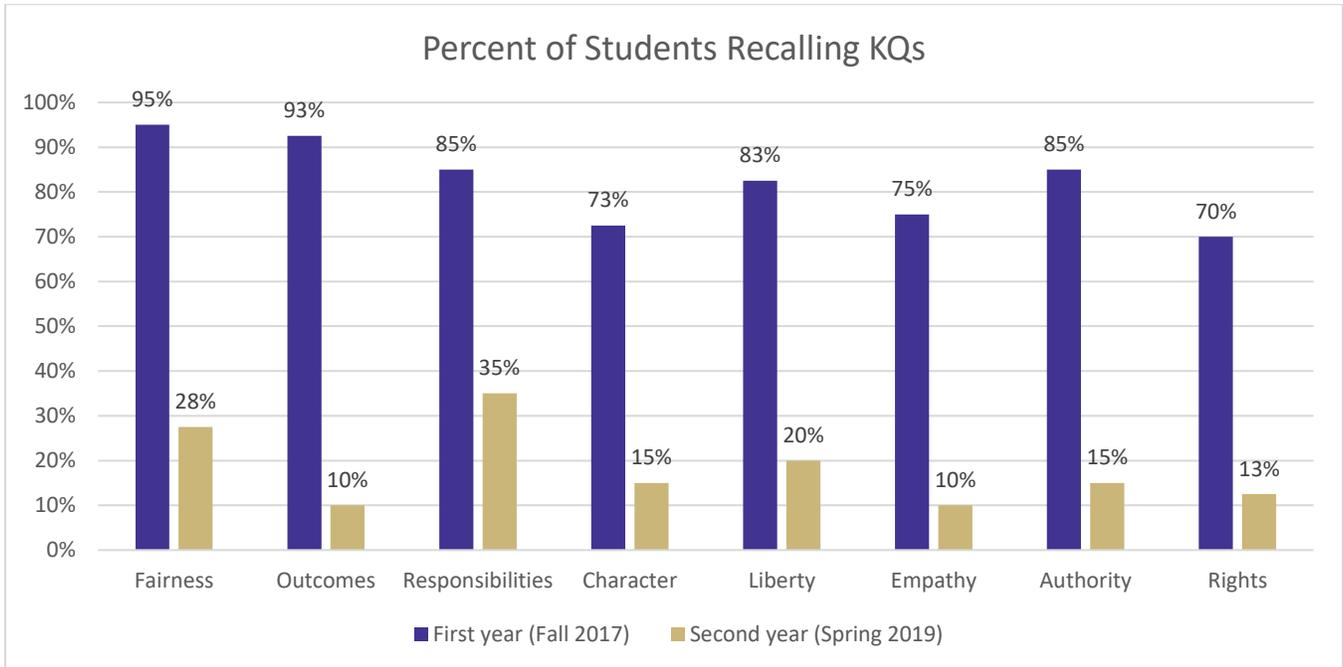
Figure 11. Percent of students recalling each KQ.

Additionally, students' ERRT explanation scores were statistically significantly higher as first-year students than as second-year students, [$t(39)$ = 10.20, p < .0001; d = 1.61]. Students are explaining at least three fewer key questions as second-year students (M = 0.79) than as first-year students (M = 3.83). Figure 12 presents the percentage of students correctly explaining each of the 8KQs as first-year students and second-year students.

Percent of Students Explaining KQs

| KQ	First year (Fall 2017)	Second year (Spring 2019)
Fairness	93%	33%
Outcomes	90%	18%
Responsibilities	80%	23%
Character	78%	40%
Liberty	78%	13%
Empathy	70%	5%
Authority	58%	5%
Rights	68%	20%
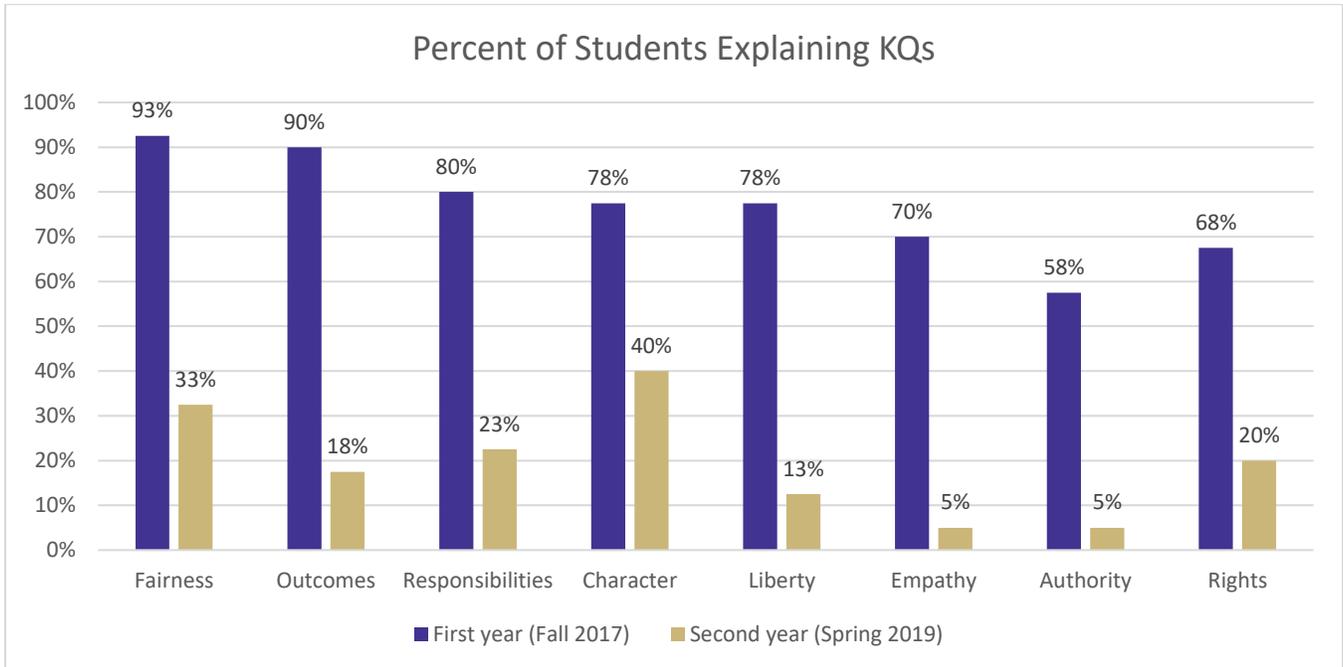
Figure 12. Percent of students receiving credit for explaining each KQ.

Ethical Reasoning in Action stakeholders may consider whether these are meaningful decreases in KQ recall and KQ explanation scores and discuss why student knowledge appears to be decreasing over time. For example, given that students received the *It's Complicated* programming as part of their freshman orientation **and** students may not have had an additional exposure to the 8KQs, it seems reasonable that students would have forgotten some of what they learned. Additional interventions could be developed to ensure that students retain the information they learned through the *It's Complicated* programming.

Ethical Reasoning Identification Test (ERIT)

The ERIT is a 50-item multiple-choice test designed to address SLOs 2 and 3. Specifically, the ERIT contains 42 single items, each presenting a simple scenario that is independent from the scenarios in other items. Additionally, there are two complex scenarios provided, each with four items related to evaluating the scenario. For each item or group of four items, students are asked to consider a short scenario and from a list of the 8KQ are asked to choose the key question most applicable to the decision or rationale presented in the scenario.

Similar to previous years, reliability for the ERIT scores was adequate (α = .84 for Fall 2018; α = .87 for Spring 2019) for the 50-item test. This relatively high reliability suggests that students' responses are consistent across the items, and the majority of variability in scores is due to differences in students' abilities, rather than other random factors (i.e. error). Additionally, similar to previous years, a

24

unidimensional factor structure appeared to adequately fit the data, supporting the creation of a single total score rather than a subscale score for each key question.

How many questions do first-year students correctly answer on the ERIT?

A total of 365 first-year students responded to the ERIT. Figure 13 shows the distribution of Fall 2018 ERIT scores. On average, first-year students scored about 68% correct. The majority of students scored between 53% and 83% correct on the ERIT.
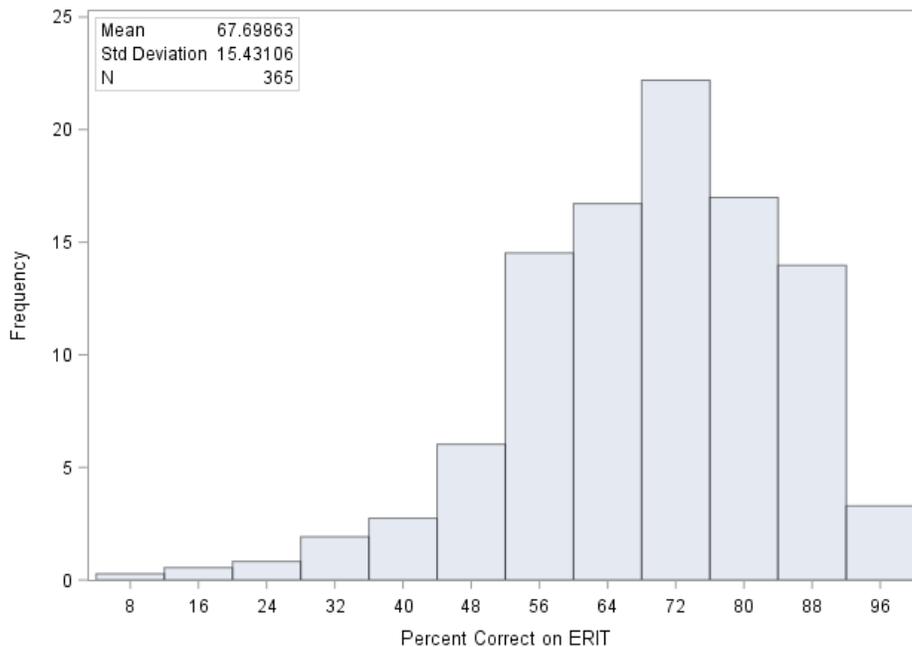
Figure 13. Distribution of ERIT Percent Correct Scores.

Table 6 provides the average percentage of items answered correctly for each key question. On average, first-year students answer the most items correctly for Outcomes, Fairness, Character, and Empathy, suggesting these key questions may be easier for students to grasp. Liberty and Rights appear to be the most difficult key questions for first-years students, as these key questions had the lowest average percent correct.

Table 6. *Percent Correct Scores by Key Question for the ERIT*

Key Question	Average % of items answered correctly	Standard Deviation *(SD)*
Fairness	71.78%	22.27
Outcomes	72.19%	19.51
Responsibilities	63.25%	22.60
Character	81.78%	18.67
Liberty	56.71%	29.08
Empathy	72.88%	17.93
Authority	70.82%	25.29
Rights	54.89%	25.61
TOTAL	**67.70%**	**15.43**

Note. N = 365

How many questions do second-year students correctly answer on the ERIT?

Fall 2017 〉 Spring 2018 〉 Fall 2018 〉 Spring 2019

A total of 449 second-year students completed the ERIT. Figure 14 shows the distribution of Spring 2019 ERIT scores. On average, second-year students scored about 66% correct. The majority of students scored between 49% and 83% correct on the ERIT.
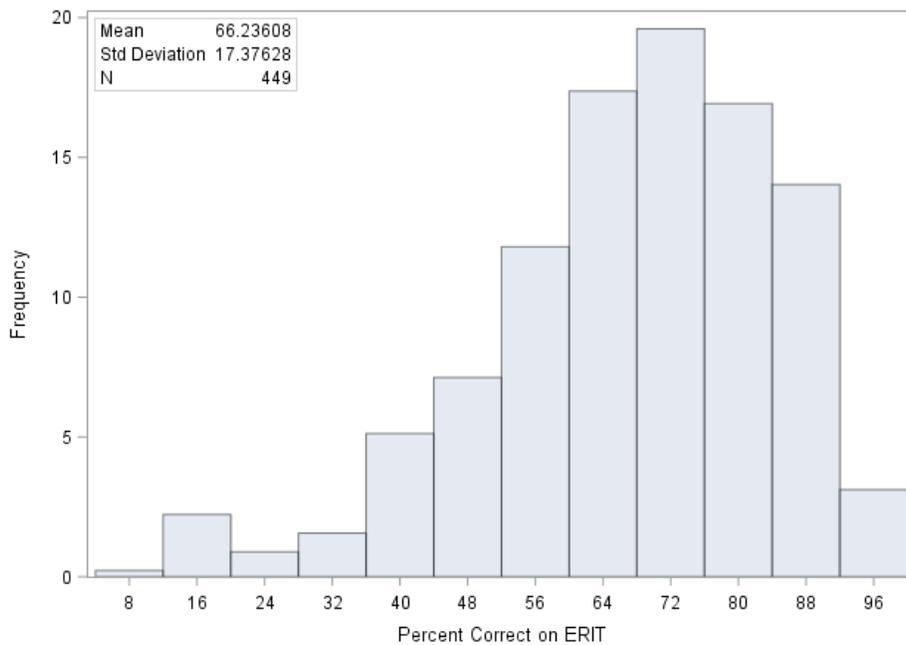

Figure 14. Distribution of ERIT percent correct scores.

26

Table 7 provides the average percent of items answered correctly for each key question. On average, second-year students answered the most items correctly for Fairness, Outcomes, Character, and Empathy, suggesting these key questions may be easier for students to grasp, very similar to the questions first-year students answered correctly. Liberty and Rights appear to be the most difficult key questions for second-year students, as these key questions had the lowest average percent correct.

Table 7. *Percent Correct Scores by Key Question for the ERIT*

Key Question	Average % of items answered correctly	Standard Deviation (SD)
Fairness	74.57%	22.90
Outcomes	69.23%	22.50
Responsibilities	62.31%	24.36
Character	79.92%	22.29
Liberty	55.93%	28.70
Empathy	71.08%	21.21
Authority	67.26%	26.51
Rights	52.08%	25.38
TOTAL	**66.24%**	**17.38**

Note. N = 449

How do students' scores on the ERIT change over time?

Fall 2017 ❯ Spring 2018 ❯ Fall 2018 ❯ Spring 2019

Assessment Day data collection methodologies allow for longitudinal comparisons (e.g., comparing a student's pretest score to that same student's posttest score). Thus, students' ERIT scores as entering first-year students (Fall 2017) scores were compared to their ERIT scores as second-year students (Spring 2019). Only students who completed the ERIT in Fall 2017 **and** Spring 2019 were included in the analyses. Students' ERIT total scores changed significantly from Fall 2017 to Spring 2019, [$t(271) = 1.98$, $p = .049$, $d = .10$]. However, averages were nearly identical, with students scoring 35.28 points on average as a first-year student and 34.48 points as a second-year student. Thus, these differences may not be *practically* significant. These results are similar to prior years' results. Table 8 provides descriptive information for the change in scores across time.

Table 8. *Descriptive information for student scores across time*

Maximum decrease	Maximum increase	Average change	Standard Deviation
32.00	15.00	0.81	6.71

Recall that *The One Book* and *It's Complicated* are the only required interventions for all students, and both occur just before first-year students take the ERIT. Though students may be exposed to the 8KQs in co-curricular or academic experiences, there is no other required Ethical Reasoning in Action intervention between students assessments. Thus, it may be expected that students' ERIT scores would not improve from Fall 2017 to Spring 2019.

Ethical Reasoning – Writing (ER-WR) Essay

The ER-WR essay is a performance assessment instrument designed to address SLO 5. Students respond to an essay prompt in which they are asked to evaluate an ethical dilemma. Students are expected to provide the considerations from which they analyzed the dilemma and explain the decision they came to and why. Four versions of the ER-WR were administered during the 2018-2019 academic year: the original version, a modified version, and two scenario-based versions ("Blood Bank" and "Tip Jar"). In the original and modified versions, students were asked to describe a dilemma from their own lives. The modified version cues students to use the 8 Key Questions to evaluate the dilemma while the original version does not offer this specific cue. The two scenario-based versions of the measure provide students with an ethical dilemma to evaluate. Student responses are scored by trained raters using the ER-WR rubric (see Appendix B). Scores for the original and modified versions are assigned to five rubric elements on a five-point scale (0 = Insufficient, 1 = Marginal, 2 = Good, 3 = Excellent, and 4 = Extraordinary). Scores for the two scenario-based versions are scored on four rubric elements (using the same five-point scale) because Element A, the element indicating proper identification of an ethical dilemma, is irrelevant when the dilemma is provided.

A total of 925 essays (392 from Fall 2018 and 533 from Spring 2019) were rated. Both veteran and new ERiA raters attended the rating session. We provided a half-day rater training to raters to ensure that all raters were adequately calibrated to the ER-WR rubric. This training session included three practice essays that were rated as group and individually to ensure rubric calibration. On the second day, raters participated in a recalibration, in which they rated and discussed one essay at the start of the rating session. Seventeen raters participated in the rating session, three of whom were graduate students and the rest were faculty members or JMU staff. Raters were assigned to rater pairs to ensure each student response was scored independently by two different raters. Each rater pair was assigned a different subgroup of essays to rate such that all 17 raters did not rate all student ER-WR responses. Each rater evaluated and rated their assigned subgroup of student essays in reverse order to counteract fatigue effects.

The tables below also contain information about "makeups" (examinations taken in a make-up session by students who did not attend their assigned Assessment Day session). The makeups, administered only in the spring, were the modified version of the ER-WR. Table 8 below displays the number of people who completed each version of the ER-WR during fall 2018 and spring 2019. Note that the Tip Jar version was not administered in Spring 2019.

Table 8. *Sample Sizes by ER-WR Version and Semester*

Semester	Original	Blood Bank	Tip Jar	Modified	Makeups
Fall 2018	140	69	91	92	-
Spring 2019	107	93	-	31	302
Total	247	162	91	123	302

Table 9 below displays the average element scores and overall scores for each of the versions (including makeups, which were the Modified version in Spring 2019) for Fall 2018 and Spring 2019. These values represent the average of the subscores on the ER-WR rubric across all students who completed the version. For all versions with data for both Fall 2018 and Spring 2019, scores were higher in the fall than in the spring. This aligns with findings from previous years.

The table below displays the subscores (followed by standard deviations in parentheses) for each of the five rubric elements, the four versions (plus makeups), and both semesters. The average score and standard deviations of the average score are included in the bottom row.

Table 9. *Average Scores (Standard Deviations) by ER-WR Version and Semester*

| Element | Original | | Blood Bank | | Tip Jar | | Modified | | Makeups | |
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
A	2.13 (.76)	2.01 (0.83)	-	-	-	-	2.24 (0.86)	2.14 (0.73)	-	2.00 (0.88)
B	1.13 (.66)	0.90 (0.63)	1.75 (0.90)	1.06 (0.61)	1.73 (0.84)	-	1.55 (0.79)	0.90 (0.47)	-	0.81 (0.60)
C	0.96 (.61)	0.73 (0.58)	1.51 (0.81)	0.92 (0.59)	1.50 (0.76)	-	1.32 (0.73)	0.78 (0.42)	-	0.66 (0.51)
D	0.91(.70)	0.67 (0.61)	1.50 (0.85)	0.94 (0.64)	1.51 (0.85)	-	1.33 (0.84)	0.81 (0.52)	-	0.60 (0.53)
E	0.95 (.71)	0.70 (0.58)	1.53 (0.84)	0.94 (0.64)	1.50 (0.85)	-	1.26 (0.79)	0.94 (0.53)	-	0.68 (0.56)
Average	1.22 (.59)	1.00 (0.56)	1.57 (0.79)	0.97 (0.58)	1.56 (0.78)		1.54 (0.70)	1.12 (0.45)	-	0.95 (0.52)

Table 10 below displays the percentages of students meeting and not meeting the standard (average score of 2) for each semester and prompt version. For each version, note that the percentage of students meeting the standard in the fall far exceeds the percentage of students meeting the standard in the spring. This follows the pattern noted in previous years and aligns with the information in the average score tables above, which demonstrate lower scores in the spring than in the fall. Note, too, the very small percentage of students in the "makeup" group meeting the standard (3.3%).

Table 10. *Percent of Students Reaching Benchmark by ER-WR Version and Semester*

	Original		Blood Bank		Tip Jar		Modified		Makeups	
	Fall	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**
Met standard	12.1%	4.7%	31.9%	5.4%	28.6%	-	29.3%	9.7%	-	3.3%
Did not meet standard	87.9%	95.3%	68.1%	94.6%	71.4%	-	70.7%	90.3%	-	96.7%
N	140	107	69	93	91	-	92	31	-	302

How do transfer students score on the ER-WR compared to non-transfer students?

Fall 2017 ⟩ Spring 2018 ⟩ Fall 2018 ⟩ **Spring 2019**

This section describes the differences in scores and standard attainment for transfer students and non-transfer students. The following tables are not disaggregated by semester because transfer students are only included in the Spring 2019 sample. Table 11 below displays the percentages of students meeting and not meeting the standard (average score of 2) for each prompt version, disaggregated by transfer status. No transfer students met the standard for any of the prompt versions. Note that some samples are very small (e.g., only one transfer student completed the modified prompt). Results should therefore be interpreted cautiously.

Table 11. *Percent of Students Reaching Benchmark by ER-WR Version and Transfer Status*

	Original		Blood Bank		Tip Jar		Modified		Makeups	
	Non-transfer	**Transfer**	**Non-transfer**	**Transfer**	**Non-transfer**	**Transfer**	**Non-transfer**	**Transfer**	**Non-transfer**	**Transfer**
Met standard	9.2%	-	17.9%	-	28.6%	-	24.6%	-	4.0%	-
Did not meet standard	90.8%	100.0%	82.1%	100.0%	71.4%	-	75.4%	100.0%	96.0%	100.0%
N	238	8	151	10	91	-	122	1	250	52

Table 12 below contains the average scores and standard deviations for each of the prompt versions disaggregated by transfer status (note that transfer status was missing for three people in the dataset, so aggregate counts do not exactly match those in prior tables). Note that the standard deviation could not be computed for transfer students completing the modified version, as this category contained a single student. Transfer students' scores were notably lower than non-transfer students' scores for the original and blood bank prompts. Although the modified values look similar for transfer and non-transfer students, it is again worth noting that a single transfer student completed this prompt. Scores are similar

for transfer and non-transfer students for the makeup category, but both are fairly low; note that makeup scores are generally lower than non-makeup scores. Given that the average score for each group is less than one, it is possible that the similarity between transfer and non-transfer students in makeup scores reflects floor effects, rather than truly similar student ability in these groups. A t-test was conducted to compare the overall scores for transfer and non-transfer students across all prompts. This test was statistically significant (t[921] = 3.95, p < .0001, M_{diff} = .32), indicating that transfer students scored significantly lower than non-transfer students. This is somewhat unsurprising, given that transfer students do not experience the primary ethical reasoning intervention experienced by other students (which takes place during August Orientation). The finding provides some evidence that the *It's Complicated* intervention has some impact on student learning.

Table 12. *Means (Standard Deviations) by ER-WR Version and Transfer Status*

	Original	Blood Bank	Tip Jar	Modified	Makeups
Non-transfer	1.13 (0.59)	1.26 (0.74)	1.56 (0.78)	1.43 (0.67)	0.96 (0.53)
Transfer	0.91 (0.30)	0.67 (0.49)	-	1.40 (-)	0.90 (0.50)

Because transfer students performed significantly worse than non-transfer students, coupled with their lack of exposure to the primary ethical reasoning intervention, they will be excluded from most of the following analyses in this report. However, they will be retained for reliability analyses in the G-studies, as information about ratings for transfer students still provides important information about the reliability of the measures and should be unaffected by the differences in means. Table 13 therefore lists the means and standard deviations for ratings of students' responses to each of the prompts after transfer students' scores have been removed.

Table 13. *Subscores (Standard Deviations) with Transfer Students' Scores Removed*

Rubric Element	Original (1) Fall N = 140	Spring N = 98	Blood Bank (2) Fall N = 69	Spring N = 82	Tip Jar (3) Fall N = 91	Spring N = 0	Modified (4) Fall N = 92	Spring N = 30	Makeups (5) Fall N = 0	Spring N = 250
Element A	2.13 (0.76)	1.99 (0.85)	-	-	-	-	2.24 (0.86)	2.11 (0.73)	-	2.01 (0.87)
Element B	1.13 (0.66)	0.91 (0.64)	1.75 (0.90)	1.09 (0.61)	1.73 (0.84)	-	1.55 (0.79)	0.90 (0.48)	-	0.82 (0.61)
Element C	0.96 (0.61)	0.74 (0.59)	1.51 (0.81)	0.95 (0.60)	1.50 (0.76)	-	1.32 (0.73)	0.78 (0.43)	-	0.68 (0.52)
Element D	0.91 (0.70)	0.67 (0.62)	1.50 (0.85)	0.98 (0.65)	1.51 (0.85)	-	1.33 (0.84)	0.81 (0.53)	-	0.62 (0.55)
Element E	0.95 (0.70)	0.70 (0.58)	1.53 (0.84)	0.98 (0.64)	1.50 (0.85)	-	1.26 (0.79)	0.94 (0.54)	-	0.69 (0.57)
OVERALL AVERAGE	**1.22 (0.59)**	**1.00 (0.57)**	**1.57 (0.79)**	**1.00 (0.58)**	**1.56 (0.78)**	**-**	**1.54 (0.70)**	**1.11 (0.46)**	**-**	**0.96 (0.53)**

How do students who took the ER-WR during a makeup session compare to students who took the same version on Assessment Day?

As a reminder, the ER-WR was only provided in makeup sessions in Spring 2019 (not in Fall 2018). The makeup data available used the modified prompt. As displayed in Table 13 above, the average score for makeups was lower than the average score for the non-makeup modified prompt group. A t-test was conducted to test the difference in overall average scores between the modified group (using only data from Spring 2019) and the makeup group, excluding transfer students ($N_{modified}$ = 30, N_{makeup} = 250). The test showed no statistically significant differences between the average scores of these two groups (t[278] = 1.43, p = .15, M_{diff} = .14). However, we decided to exclude these students from the analyses in this report. Students taking make-up assessments often exhibit lower motivation than students taking assessments during Assessment Day, and adding these students into our analytic sample may cause interpretations to be affected by these differences. Information about the performance of students completing the makeup test is available upon request from CARS.

ER-WR: Original

Reliability

To evaluate the extent to which scores reflect students' abilities, rather than rater characteristics or other random error, we employed a generalizability analysis. There are two types of reliability estimates generated from a generalizability analysis: relative and absolute. Relative estimates are represented by the G-coefficient and are most useful when comparing students to one another. The G-coefficient is more appropriate for the desired ER-WR comparisons. Thus, those are the only ones reported here.

For both cohorts, G coefficients were lower than desirable (G = 0.59 for first-year scores, G = 0.63 for second-year scores). Ideally, G-coefficients should be above 0.70 to indicate adequate reliability. These lower than desirable results could be due to several factors. First, it could be the case that raters are introducing unwarranted factors into scores. Thus, additional rater training could be beneficial. It could also be the case that there is not enough variability in students' scores to produce stable reliability estimates. To compute reliability estimates, it is assumed that there is variability in students' scores. However, if students' scores are similar (i.e. limited variability), it may result in decreased reliability estimates. For first year students, the average score on the ER-WR was 1.22 with a standard deviation of 0.66. This means that about 68% of first-year students scored between 0.56 and 1.88 points. For second-year students, the average score on the ER-WR was 1.00 with a standard deviation of 0.57. This means that about 68% of second-year students scored between 0.43 and 1.57. It is possible that the relatively low variability in scores is contributing to decreased reliability. Given low reliability, the results below should be interpreted cautiously.

First year scores

In all, there were 140 "rate-able" essays from first-year students in Fall 2018. Essays were deemed non rate-able if they did not present an ethical dilemma. For example, some students may have presented a difficult decision and their thought processes to arrive at a decision, but the decision was not necessarily an ethical one, so the essay was not rated.

On average, first-year students scored 1.22 points (i.e. just above marginal) on the ER-WR rubric in Fall 2018. As can be seen in Figure 15, first-year students scored the highest on Element A, which requires students to identify an ethical situation. Students scored lowest on Elements C, D, and E, with scores below marginal for each of these elements. These results are similar to previous years. A t-test comparing the overall average scores for this version of the ER-WR in Fall 2017 and Fall 2018 revealed no statistically significant difference between the two groups of first-year students (t[456] = -1.85, p = .065).

Average Element Scores ER-WR Essays for First-Year Students

Element	Score
A. Ethical Situation	2.13
B. Key Question Reference	1.13
C. Key Question Applicability	0.96
D. Ethical Reasoning: Analyzing individual KQ	0.91
E. Ethical Reasoning: Weighing the relevant factors & deciding	0.95
OVERALL AVERAGE	1.22
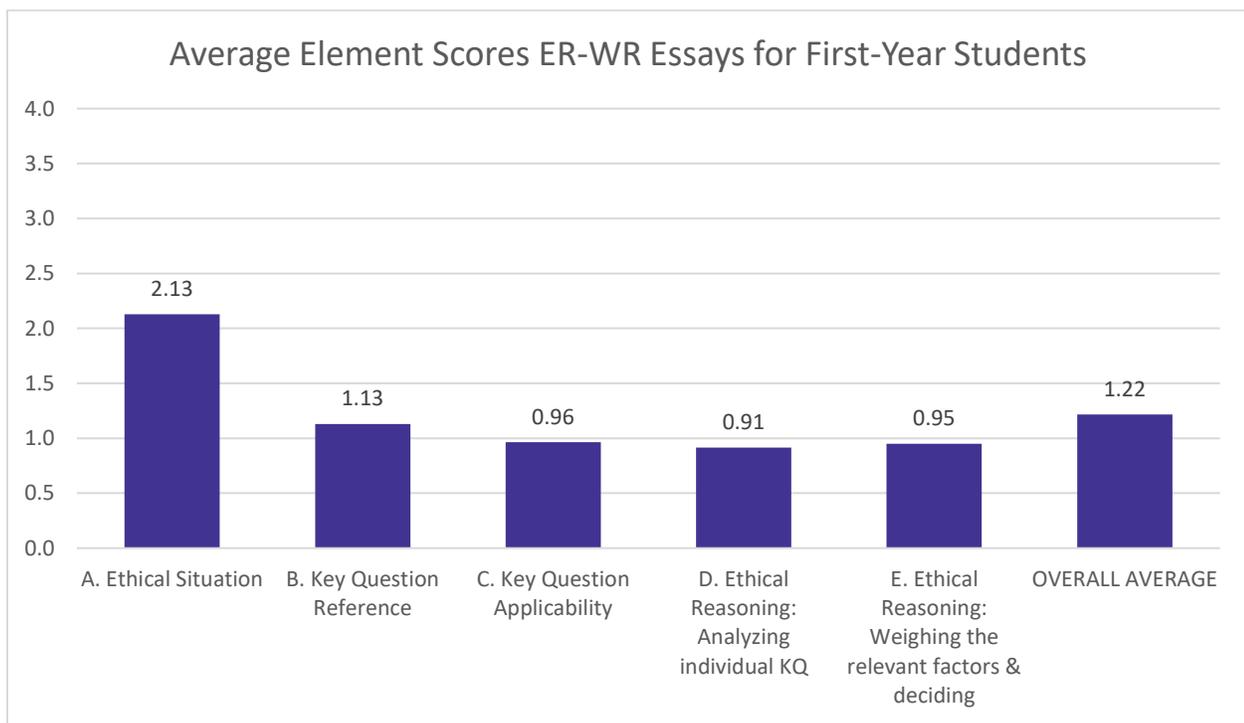
Figure 15. Average ER-WR element scores for first-year students.

Second year scores

In all, there were 98 "rate-able" essays from second-year students. On average, second-year students scored 1.00 points (e.g. precisely marginal) on the ER-WR rubric in Spring 2019. As can be seen in Figure 16, second-year students scored the highest on Element A, which requires students to identify an ethical situation. Students scored lowest on Elements B C, D, and E, with scores below marginal for each of these elements. These results are similar to previous years. A t-test comparing the overall average scores for this version of the ER-WR in Spring 2018 and Spring 2019 revealed no statistically significant difference between the two groups of second-year students (t[238] = -0.38, p = .70)


Average Element Scores ER-WR Essays for Second-Year Students

1.99	0.91	0.74	0.67	0.70	1.00
A. Ethical Situation	B. Key Question Reference	C. Key Question Applicability	D. Ethical Reasoning: Analyzing individual KQ	E. Ethical Reasoning: Weighing the relevant factors & deciding	OVERALL AVERAGE
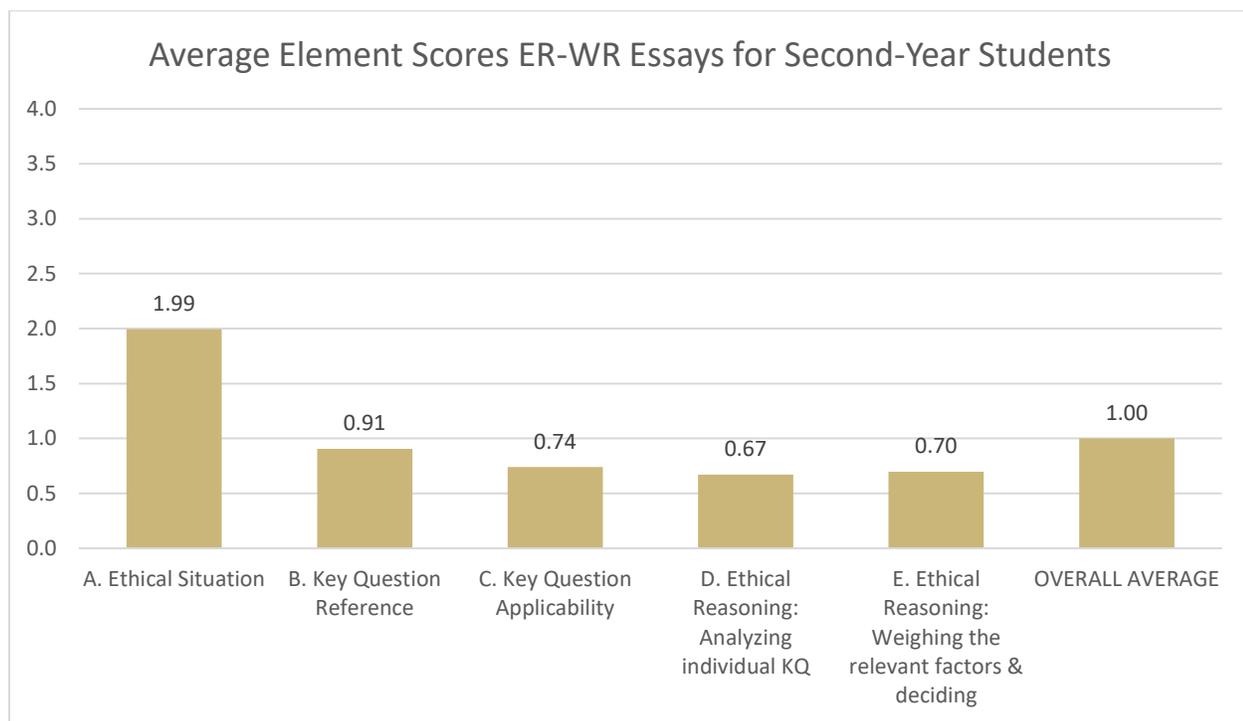
Figure 16. Average ER-WR element scores for second-year students.

Percent meeting benchmark

During the 2012-2013 academic year, university stakeholders set a tentative standard or expectation for student performance on the ER-WR rubric. This tentative standard represented an overall average score of 2 ("Good") on the rubric. As shown in Figure 17, about 9% (e.g., 22 out of 238 students) of the students who responded to the ER-WR during Fall 2018 and Spring 2019 met this standard. Additionally, the percentage of students who met the standard for the ER-WR rubric and also experienced a mandatory

Ethical Reasoning in Action intervention at JMU (i.e., It's Complicated) was noticeably greater each year compared to the baseline (e.g., students who did not experience any ethical reasoning interventions).
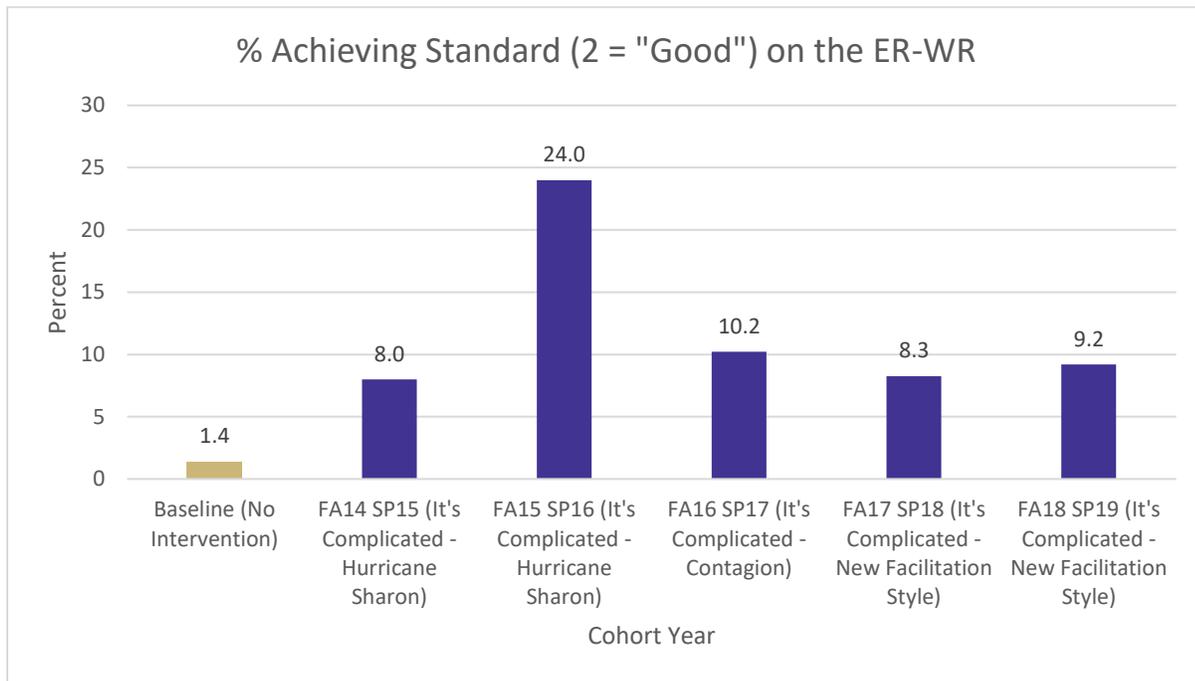
Figure 17. Percentage of students achieving the standard each year on the original ER-WR.

Longitudinal analysis

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

Assessment Day data collection methodologies allow for longitudinal comparisons (e.g., comparing a student's pretest score to that same student's posttest score). Thus, students' ER-WR scores as entering first-year students (Fall 2017) scores were compared to their ER-WR scores as second-year students (Spring 2019). Only students who completed the ER-WR in Fall 2017 and Spring 2019 were included in the analyses (N = 73). On average, Fall 2017 first-year students scored statistically significantly higher than they did when they were assessed as second-year students in Spring 2019 [$t(72)=3.00$, p = .004)]. Specifically, as first-year students, the average score was 1.24 out of 5 possible points and as second-year students, the average scores was 1.00 out of 5 possible points. That is, students' average scores decreased by 0.24 points from their first to second years. Further, as shown in Figure 18, students scored lower on each element as second-year students than as first-year students. Ethical Reasoning in Action stakeholders may consider whether this is a meaningful difference between first-year and second-year student scores.
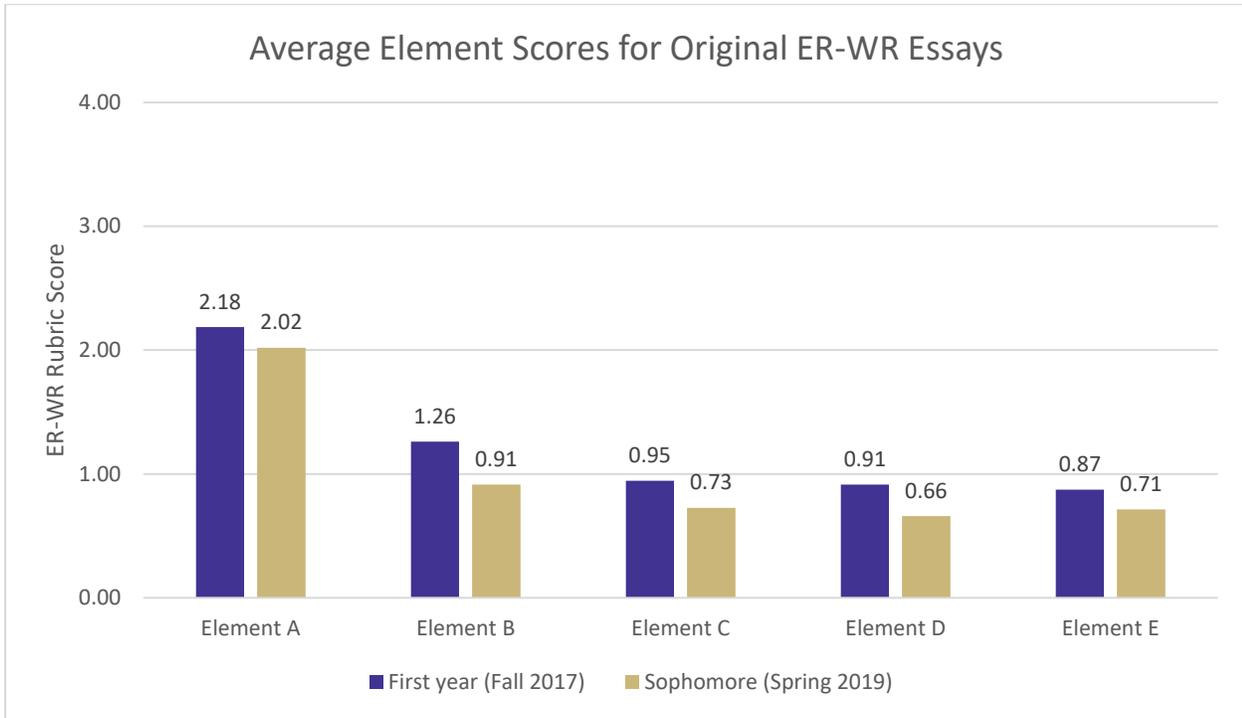
Figure 18. Average element scores for original ER-WR essays.

As shown in Table 14, the students who responded to the ER-WR during Fall 2018 and Spring 2019, on average, scored about "Marginal." Similar to other years, the easiest element was A (Ethical Situation). Elements C, D, and E were the most difficult elements. Although scores had been steadily decreasing for the past two assessment cycles, scores from Fall 2018 and Spring 2019 appear to represent a leveling off.

Table 14. *Average Scores by ER-WR Rubric Element*

Rubric Element	FA12-SP13 *N* = 110		FA13-SP14 *N* = 180		FA14-SP15 *N* = 284		FA15-SP16 *N* = 293		FA16-SP17 *N* = 450		FA17-SP18 *N* = 460		FA18-SP19 N = 238	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Element A	1.56	0.90	1.94	1.16	1.98	0.89	2.18	0.72	2.16	0.76	2.03	0.87	**2.07**	**0.80**
Element B	0.76	0.58	1.13	0.94	1.01	0.74	1.43	0.79	1.05	0.71	1.03	0.69	**1.04**	**0.66**
Element C	0.44	0.48	0.82	0.78	0.71	0.60	1.14	0.72	0.84	0.65	0.76	0.61	**0.87**	**0.61**
Element D	0.48	0.54	0.86	0.82	0.74	0.68	1.14	0.71	0.82	0.68	0.74	0.63	**0.81**	**0.68**
Element E	0.50	0.55	0.90	0.83	0.81	0.67	1.07	0.69	0.89	0.67	0.75	0.65	**0.85**	**0.66**

| OVERALL AVERAGE	0.75	0.61	1.13	0.79	1.05	0.87	1.39	0.63	1.15	0.60	1.06	0.60	1.13	0.59

ER-WR2: Blood Bank

Reliability

The G coefficient was acceptable for first-year scores (G = .79) but were lower than desirable for second-year scores (G = 0.64). Ideally, G-coefficients should be above 0.70 to indicate adequate reliability. Therefore, these values imply adequate reliability for first-year scores and somewhat lower reliability for second-year scores. Results for second-year ER-WR2 scores should be interpreted cautiously.

First year scores

Fall 2017 〉 Spring 2018 〉 **Fall 2018** 〉 Spring 2019

In all, there were 69 "rate-able" ER-WR2 "Blood Bank" essays from first-year students. Because this prompt provided students with an ethical dilemma to which they were asked to respond, these essays were not rated on rubric Element A (identifying the ethical situation). On average, first-year students scored 1.57 points (between "marginal" and "good") on the ER-WR rubric in Fall 2018. As can be seen in Figure 19, first-year students scored the highest on Element B, which requires students to reference the 8 Key Questions. Average scores were slightly lower on elements C, D, and E. These results demonstrate similar performance to Fall 2017. A t-test comparing the overall average scores for this version of the ER-WR2 in Fall 2017 and Fall 2018 revealed no statistically significant difference between the two groups of first-year students (t[181] = -1.38, p = .17).
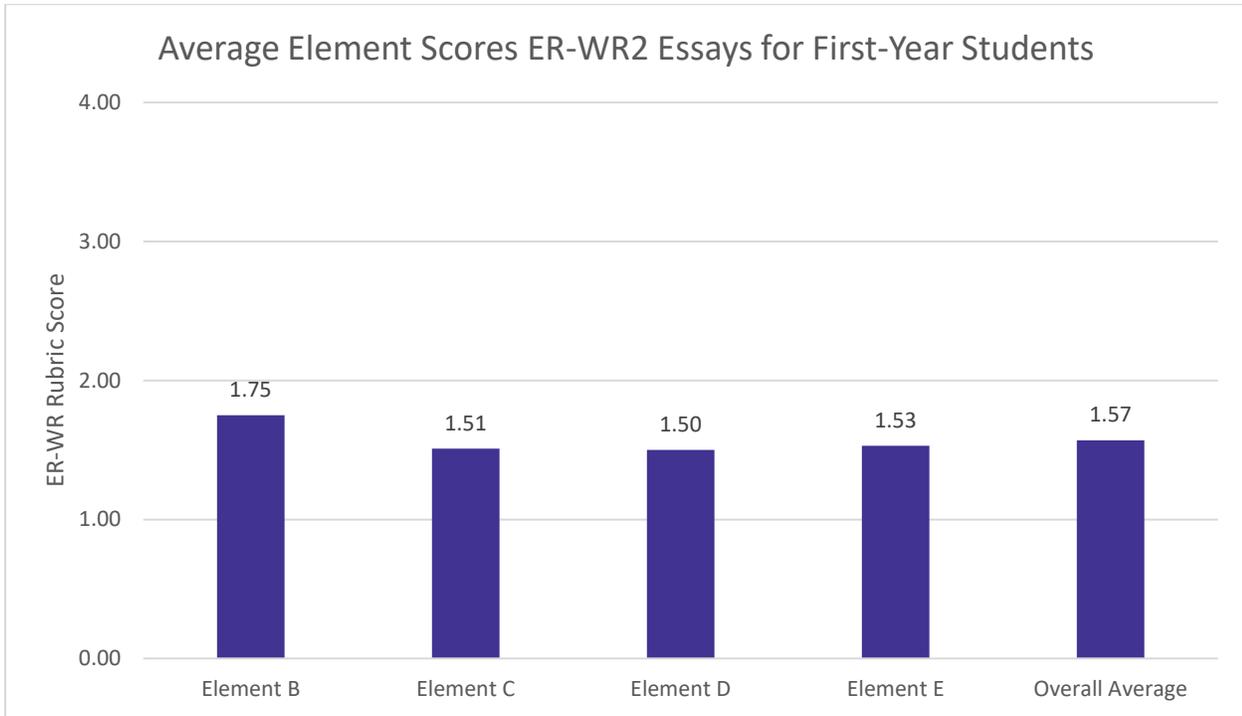
Figure 19. Average element scores for ER-WR2 essays for first-year students.

Second year scores

Fall 2017 > Spring 2018 > Fall 2018 > **Spring 2019**

In all, there were 93 "rate-able" essays from second-year students. On average, second-year students scored 1.06 points (i.e., just above marginal) on the ER-WR rubric in Spring 2019. As can be seen in Figure 20, second-year students scored the highest on Element A, which requires students to identify an ethical situation. Students scored slightly lower on elements C, D, and E, with scores below marginal for each of these elements. A t-test comparing the overall average scores for this version of the ER-WR2 in Spring 2018 and Spring 2019 revealed a statistically significant difference between the two groups of first-year students (*t*[151] = -2.89, *p* = .005). On average, students in Spring 2019 scored .25 points higher than students in Spring 2018 on the ER-WR2.
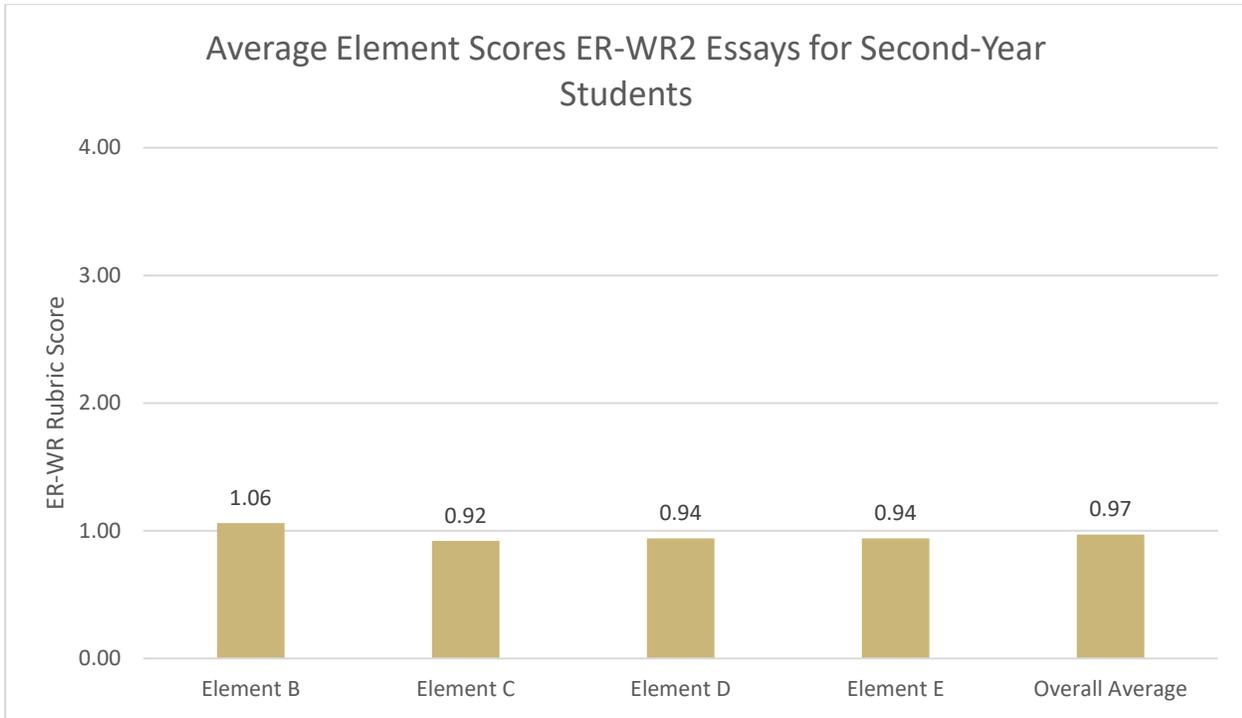
Figure 19. Average element scores for ER-WR2 essays for second-year students.

Percent meeting benchmark

Fall 2017 ⟩ Spring 2018 ⟩ **Fall 2018** ⟩ **Spring 2019**

During the 2012-2013 academic year, university stakeholders set a tentative standard or expectation for student performance on the ER-WR rubric. This tentative standard represented an overall average score of 2 ("Good") on the rubric. About 18% (e.g., 27 out of 151) of the students who responded to the ER-WR2 during Fall 2018 and Spring 2019 met this standard. This is similar to the percentage of students who met the standard in the 2017-2018 academic year. Figure 20 displays the percentage of students achieving the standard on the ER-WR2 for both years it has been administered.

% Achieving Standard (2 = "Good")
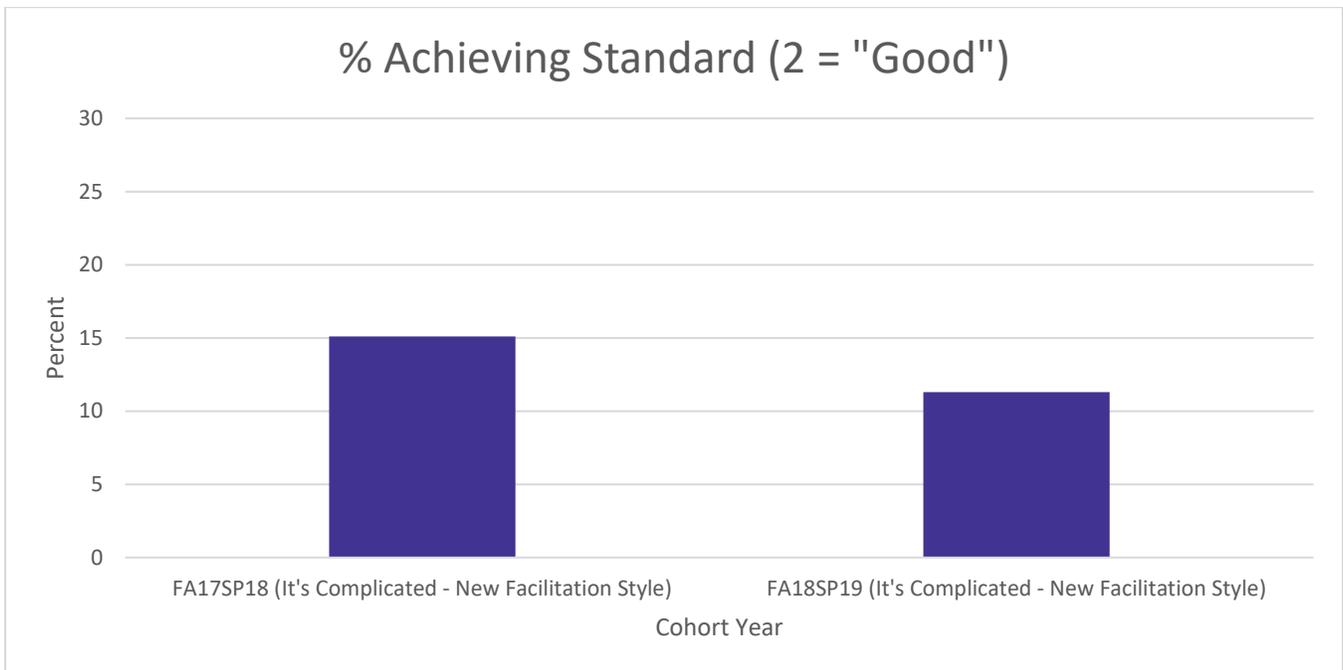
Figure 20. Percentage of students achieving the standard each year on the ER-WR2.

Longitudinal analysis

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

Assessment Day data collection methodologies allow for longitudinal comparisons (e.g., comparing a student's pretest score to that same student's posttest score). Thus, students' ER-WR2 scores as entering first-year students (Fall 2017) scores were compared to their ER-WR2 scores as second-year students (Spring 2019). Only students who completed the ER-WR2 in Fall 2016 and Spring 2018 were included in the analyses (N = 50). On average, first-year students in assessed in Fall 2017 scored statistically significantly higher than they did when they were assessed as second-year students in Spring 2018, [$t(49)$=3.91, p = .0003)]. Specifically, as first-year students, the average score was 1.50 out of 5 possible points and as second-year students the average score was 1.06 out of 5 possible points. That is, students' average scores decreased by 0.44 points from their first to second years. Further, as shown in Figure 21, students scored lower on each element as second-year students than as first-year students. Ethical Reasoning in Action stakeholders may consider whether this is a meaningful difference between first-year and second-year student scores.

Average Element Scores for ER-WR2 Essays

	Element B	Element C	Element D	Element E	Overall Average
First year (Fall 2017)	1.68	1.49	1.39	1.43	1.50
Sophomore (Spring 2019)	1.15	1.01	1.05	1.04	1.06
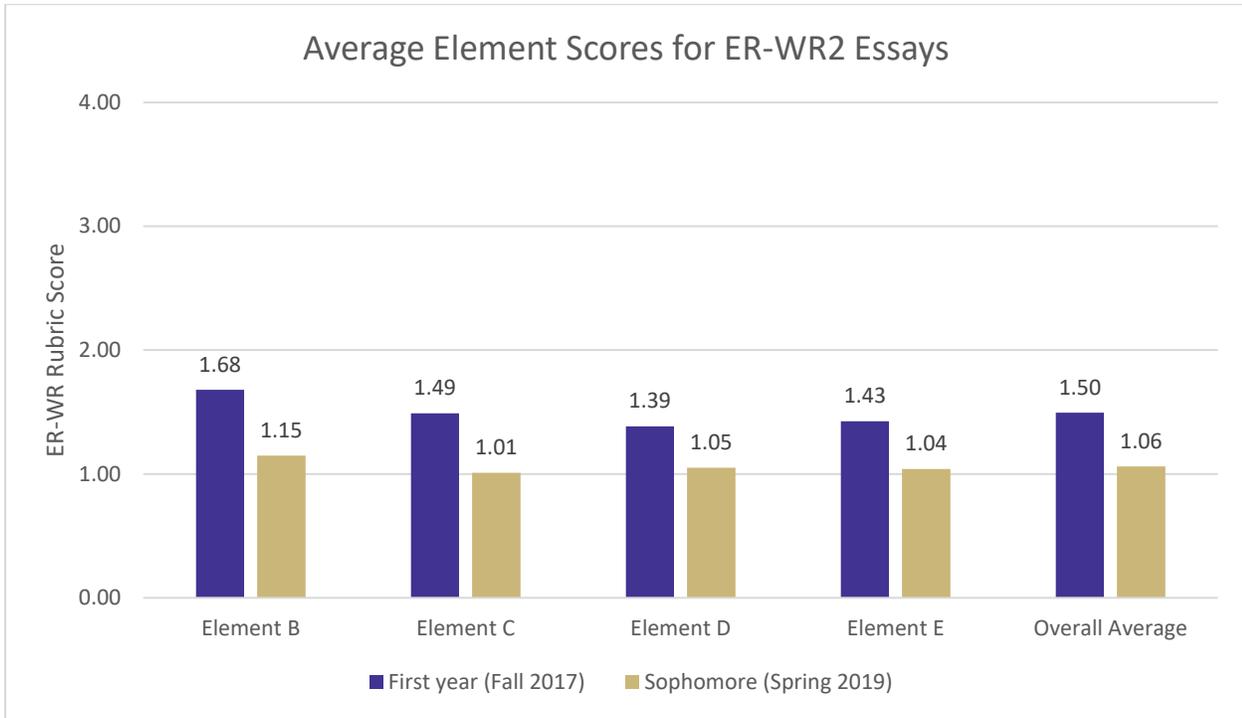
Figure 21. Average element scores for ER-WR2 essays.

ER-WR3: Tip Jar

Reliability

For this prompt, data were only collected during the fall semester (i.e., for first-year students). Therefore, a reliability estimate is only available for first-year scores. The G coefficient was relatively high (G = .76 for first-year students), indicating adequate reliability of the scores.

First year scores

Fall 2017 > Spring 2018 > **Fall 2018** > Spring 2019

In all, there were 91 "rate-able" essays from first-year students. Because the ER-WR3 "Tip Jar" prompt provided students with an ethical dilemma to which they were asked to respond, these essays were not rated on rubric Element A (identifying the ethical situation). On average, first-year students scored 1.56 points (between "marginal" and "good") on the ER-WR rubric in Fall 2018. As can be seen in Figure 22, first-year students scored the highest on Element B, which requires students to reference the 8 Key Questions. Average scores were slightly lower on elements C, D, and E. This score pattern is almost identical to that observed for the Blood Bank scenario for first-year students in Fall 2018. A t-test

comparing the overall average scores for this version of the ER-WR3 in Fall 2017 and Fall 2018 revealed a statistically significant difference between the two groups of first-year students ($t[151] = -5.05$, $p < .0001$), with overall average scores in Fall 2018 being .60 points higher than students in Fall 2017.

Average Element Scores ER-WR3 Essays for First-Year Students

	Element B	Element C	Element D	Element E	Overall Average
ER-WR Rubric Score	1.73	1.50	1.51	1.50	1.56
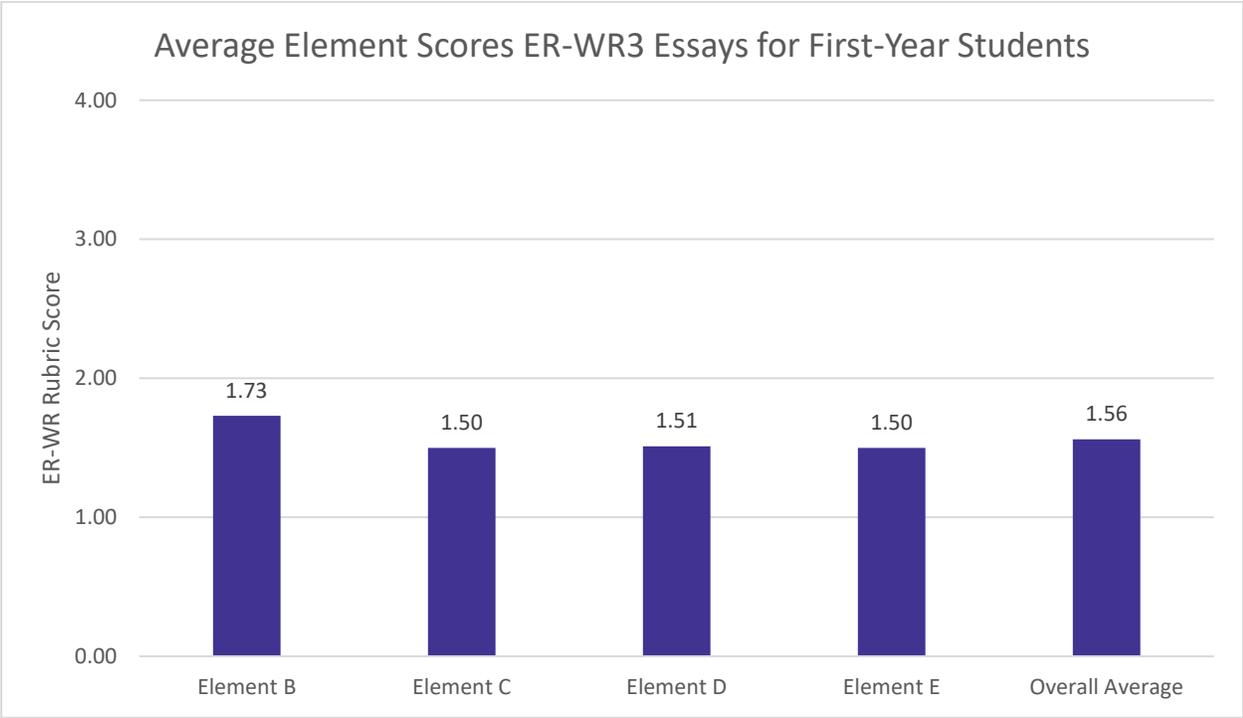
Figure 22. Average element scores for ER-WR2 essays for first-year students.

Percent meeting benchmark

Fall 2017 > Spring 2018 > Fall 2018 > Spring 2019

During the 2012-2013 academic year, university stakeholders set a tentative standard or expectation for student performance on the ER-WR rubric. This tentative standard represented an overall average score of 2 ("Good") on the rubric. About 29% (e.g., 26 out of 91) of the students who responded to the ER-WR3 during Fall 2018 met this standard. Comparisons to previous cohorts are not yet possible for the number of students meeting baseline because the ER-WR3 was administered only to first-year students in Fall 2017. The lack of data from Spring 2018 for the ER-WR3 would likely result in inflated rates of benchmark attainment for that academic year, as scores are typically higher for first-year students than for second-year students. However, in future years, these comparisons should be examined.

Longitudinal analysis

Longitudinal analyses were not possible for the ER-WR3 because this version was not administered during Spring 2019. However, beginning in the 2019-2020 report, longitudinal responses linking participants from Fall 2018 to Spring 2020 will be possible and should be examined.

ER-WR Modified

Reliability

Fall 2017 〉 Spring 2018 〉 Fall 2018 〉 Spring 2019

The G coefficient was reasonable for first year scores (G = 0.74) but much lower than desirable for second-year scores (G = .36). Ideally, G-coefficients should be above 0.70 to indicate adequate reliability. Therefore, these values imply adequate reliability for first-year scores but very low reliability for second-year scores. It is possible that the relatively low sample size for second-year students on the modified prompt (N = 31) and the restricted variability of this group's scores (SD = 0.45) contributed to the low reliability estimate. Results for second-year ER-WR modified scores should be interpreted cautiously, and further research should be conducted to determine why the reliabilities differ for the two sets of scores.

First year scores

Fall 2017 〉 Spring 2018 〉 Fall 2018 〉 Spring 2019

In all, there were 92 "rate-able" essays from first-year students. On average, first-year students scored 1.54 points (between "marginal" and "good") on the ER-WR rubric in Fall 2018. As can be seen in Figure 23, first-year students scored the highest on Element A, which requires students to identify an ethical situation. Average scores were slightly lower on elements C, D, and E. These results demonstrate somewhat higher performance than the data from Fall 2017, which is the only previous data available for this prompt. A t-test comparing the overall average scores for this version of the ER-WR2 in Fall 2017 and Fall 2018 revealed a statistically significant difference between the two groups of first-year students (t[252] = -3.81, p = .0002), with overall average scores in Fall 2018 being .42 points higher than students in Fall 2017, on average.
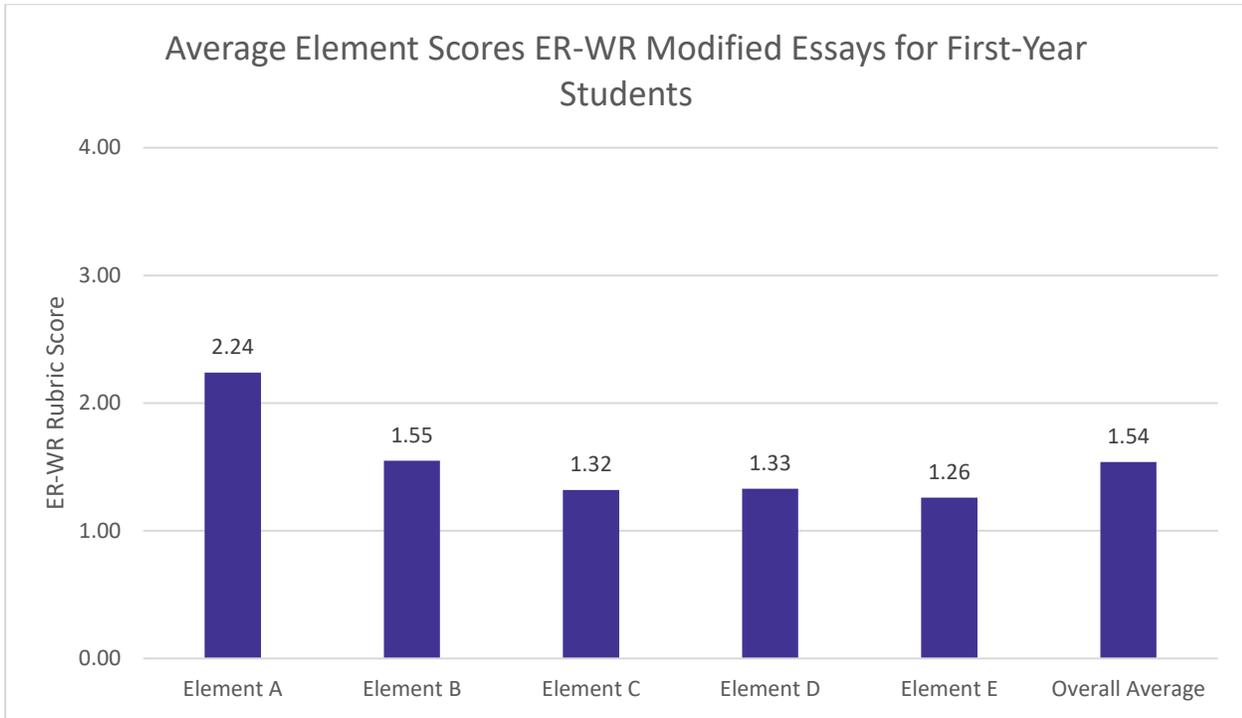
Figure 23. Average element scores for ER-WR2 essays for first-year students.

Second year scores

Fall 2017 > Spring 2018 > Fall 2018 > **Spring 2019**

In all, there were 30 "rate-able" essays from second-year students. On average, second-year students scored 1.11 points (i.e., just above marginal) on the ER-WR rubric in Spring 2019. As can be seen in Figure 24, second-year students scored the highest on Element A, which requires students to identify an ethical situation. Students scored slightly lower on elements C, D, and E, with scores below marginal for each of these elements. As a reminder, because this represents the first time the modified ER-WR was administered to second-year students, no comparisons to previous cohorts are possible.
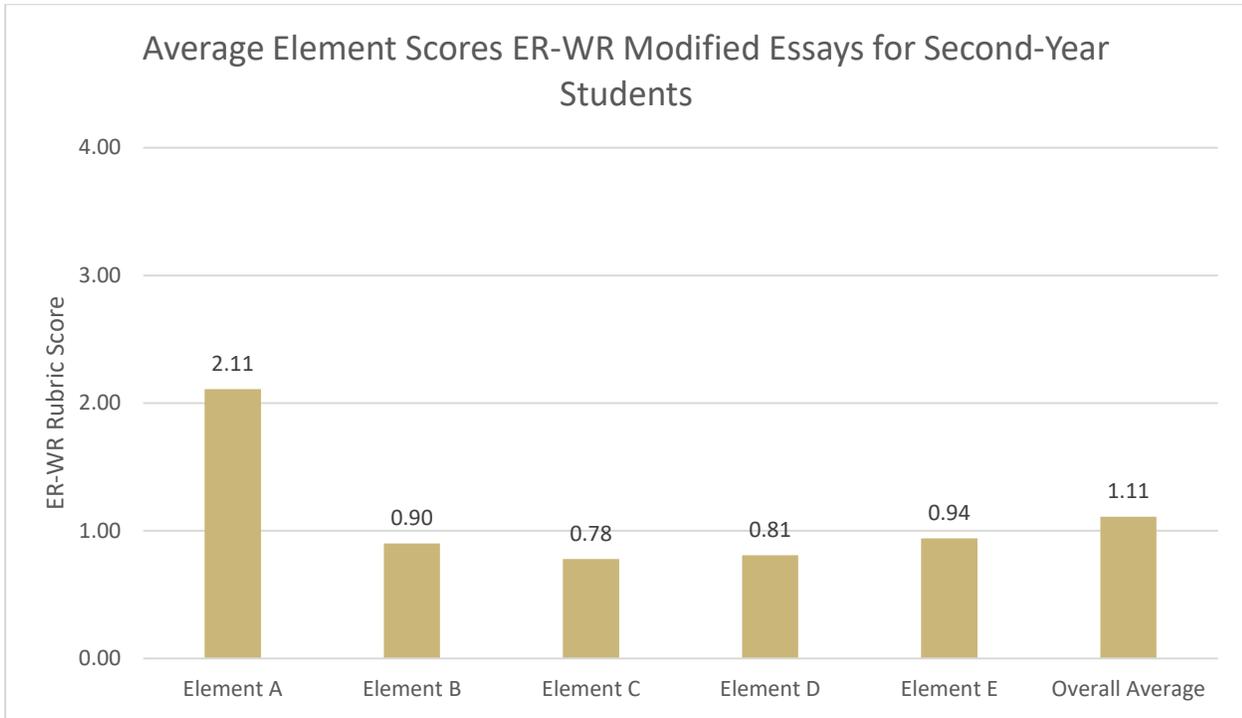
Figure 24. Average element scores for ER-WR2 essays for second-year students.

Percent meeting benchmark

Fall 2017 〉 Spring 2018 〉 **Fall 2018** 〉 **Spring 2019**

During the 2012-2013 academic year, university stakeholders set a tentative standard or expectation for student performance on the ER-WR rubric. This tentative standard represented an overall average score of 2 ("Good") on the rubric. About 22% (e.g., 27 out of 122) of the students who responded to the modified ER-WR during Fall 2018 and Spring 2019 met this standard. Comparisons to previous cohorts are not yet possible for the number of students meeting baseline because the modified ER-WR was administered for the first time in Fall 2018. In future years, these comparisons should be examined.

Longitudinal analysis

Longitudinal analyses were not possible for the modified ER-WR because this version was not administered during Spring 2019. However, beginning in the 2019-2020 report, longitudinal responses linking participants from Fall 2018 to Spring 2020 will be possible and should be examined.

References

Good, M. R. (2015). *Improving student learning in higher education: A mixed methods study* (Doctoral dissertation, James Madison University). Retrieved from https://commons.lib.jmu.edu/diss201019/18

Holzman, M. A. (2018). *Evaluating rater effects in the context of ethical reasoning essay assessment: An application of the Many-Facets Rasch Measurement Model* (Doctoral dissertation, James Madison University). Retrieved from https://commons.lib.jmu.edu/diss201019/192/

Smith, K. L. (2017). *Integrating implementation fidelity and learning improvement to enhance students' ethical reasoning abilities* (Doctoral dissertation, James Madison University). Retrieved from https://commons.lib.jmu.edu/diss201019/153

THE EIGHT KEY QUESTIONS

Fairness

How can I act equitably and balance legitimate interests?

Outcomes

What achieves the best short- and long-term outcomes for me and all others?

Responsibilities

What duties and/or obligations apply?

Character

What action best reflects who I am and the person I want to become?

Liberty

How does respect for freedom, personal autonomy, or consent apply?

Empathy

What would I do if I cared deeply about those involved?

Authority

What do legitimate authorities (e.g. experts, law, my religion/god) expect of me?

Rights

What rights (e.g. innate, legal, social) apply?

Appendix B

	Insufficient 0	Marginal 1	Good 2	Excellent 3	Extraordinary 4	Score
A. Ethical Situation: Identifying ethical issue in its context						
	No reference to decision option(s).	Implicit reference to decision options AND/OR little context given regarding decision option(s).	Explicit but unorganized reference to decision option(s) and context.	Clear description of decision option(s) and context.	Meets criteria for *Excellent* AND… • Context treated with nuance • Builds tension with organization and word choice.	
B. Key Question Reference: Mentioning the 8 KQs or equivalent terms						
	Reference to zero or only one key question.	Vague references to key questions OR only <u>two</u> key questions referenced.	References <u>four</u> key questions.	References <u>six</u> key questions.	References all <u>eight</u> key questions.	
C. Key Question Applicability: Describing which of the 8 KQs are applicable or not applicable to the situation and why						
	No rationale provided for the applicability or inapplicability of any KQs to the ethical situation.	Provides a rationale for the applicability or inapplicability of <u>two</u> key questions to the ethical situation.	Provides a rationale for the applicability or inapplicability of <u>four</u> key questions to the ethical situation.	Provides a rationale for the applicability or inapplicability of <u>six</u> key questions to the ethical situation.	For all <u>eight</u> questions provides a rationale for its applicability or inapplicability to the ethical situation.	
SPECIAL NOTE: If author identifies fewer than three applicable KQs, then Criteria "D" and "E" can be scored no higher than (1) "Marginal"						
D. Ethical Reasoning: Analyzing individual KQs						
	No attempt to analyze any <u>of the referenced</u> key questions.	Analysis attempted using two or more key questions. Typically <u>incorrect</u> ascription of the key questions to the ethical situation. Account is <u>unclear, disorganized, or inaccurate.</u>	Analysis attempted using three or more key questions. <u>Basically accurate</u> ascription of the key questions to the ethical situation. Account is <u>unclear</u> or <u>disorganized.</u>	Analysis attempted using three or more key questions. <u>Accurate</u> ascription of the key questions to the ethical situation. Account is <u>clear</u> and <u>organized</u>.	Meets criteria for *Excellent* AND… Nuanced treatment of key questions, for example: • elucidates subtle distinctions	

48

Insufficient 0	Marginal 1	Good 2	Excellent 3	Extraordinary 4	Score
				• uses analogies or metaphors • considers different issues within same key question.	
SPECIAL NOTE: If Criterion "D" is scored a 0 or 1 then Criterion "E" can be scored no higher than (1) "Marginal"					
E. Ethical Reasoning: Weighing the relevant factors and deciding					
No judgment is presented OR judgment presented with no rationale.	Uses products of the analysis and provides some weighing to make a decision. Account is unclear, disorganized, or inaccurate.	Conveys weighing approach using analysis products. Provides an intelligible basis for judgment.	Meets criteria for *Good* AND…. Logically terminates in decision that will be reached.	Meets criteria for *Excellent* AND… Products of analysis weighed to make judgment compelling.	