ETHICAL REASONING IN ACTION
2017-2018 ASSESSMENT REPORT

Chi Au, M.A.
Jessica N. Jacovidis, Ph.D.
Allison Ames, Ph.D.
Madison Holzman, M.Ed.

Center for Assessment and Research Studies

James Madison University

Table of Contents

Executive Summary

The 2017 – 2018 school year marked five years of Madison Collaborative: 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. 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 is used to evaluate students' facility 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. 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 and second-year students ranked ethical reasoning as an important skill, with 48% of first-year and 40% of second-year students selecting ethical reasoning in their top three most important desirable skills.
- First and second-year students reported that ethical reasoning skills are important (4.53 and 4.46 out of 5 points, respectively). Additionally, first and second-year students were confident in their ability to apply the ethical reasoning process (4.12 and 4.21 out of 5 points, respectively).
- Students' ratings of the importance of ethical reasoning decreased significantly over time(Fall 2016 to Spring 2018); 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 recalled about 6 and explained about 3 of the 8KQs; second-year students accurately recalled and explained about 1 of the 8KQs.
- For both first and second-year students, Fairness, Outcomes, and Responsibilities were the easiest 8KQs to recall and Fairness and Outcomes were the easiest key questions to explain.
- Empathy and Rights were the most difficult key questions for first-year students to recall and Empathy and Authority were the most difficult for first-year students to explain. Authority, Liberty, and Rights were the most difficult key questions for second-year students to recall and Liberty and Authority were the most difficult for second-year students to explain.

- Students recalled and explained significantly more of the 8KQs as first-year students than as second-year students, recalling at least four fewer key questions and explaining at least two fewer key questions as second-year student 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, both first and second-year students scored about 68% correct.
- First-year and second-year students answered the most items correctly for Character, Fairness, and Outcomes, 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 did not change significantly from Fall 2016 to Spring 2018.

Ethical Reasoning - Writing (ER-WR) Essay

The ER-WR essay is a performance assessment instrument that asks students to respond to an essay prompt in which they are asked to describe an ethical dilemma from their own lives. The ER-WR2 and ER-WR3 essays are performance assessment instruments that provide students with a hypothetical scenario and asked them to apply the 8KQs.

- First-year students scored 1.10 points (e.g. just above marginal) on the ER-WR rubric and second-year students scored 0.97 points (e.g. marginal) on the ER-WR rubric.
- Both first and second-year students scored the highest on Element A and lowest on Elements C, D, and E.
- Students scored statistically significantly higher as first-year students (1.39 out of 4 possible points) than as second-year students (1.08 out of 4 possible points), by about 0.31 points.
- Given that both the ER-WR2 and ER-WR3 are new writing prompts, only first-year student data are available for analyses.
- The average score for the ER-WR2 (M = 1.15) is similar to the average score for the ER-WR (M = 1.10); however, the average score for the ER-WR3 (M = 0.71) is lower than the other two prompts.

Overall Summary

Generally, students' **attitudes** toward ethical reasoning appear to remain stable over time. Students' **ethical reasoning skills** appear to be either stable, or decreasing, over time. 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. Additionally 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. As its foundation, the Ethical Reasoning in Action team has set the 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 2017-2018 academic year was the fifth year of the Ethical Reasoning in Action intervention plan. A comprehensive assessment protocol is 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.
   Alternate assessment: From a list of ways of conceptualizing issues, students will correctly identify the eight Key Questions.
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.
   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 study with small groups of first-year students over 75-minutes. Students are introduced to the 8KQs, watch a video depicting the case study, and grapple with a tough decision to be made as part of the case study. *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 study (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, and/or jmUDESIGN 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 2017-2018 academic year, six Ethical Reasoning in Action assessment instruments were administered. All objectives are 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

Last year, a new reporting format was implemented. Similar to previous years, the 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. This

report also includes a discussion section that summarizes results by SLOs and describes current and future work. 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 be referred to the previous years' reports.

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 perceived 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 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*). 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*).[1] 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 2017; α = .87 for Spring 2018) as well as the confidence subscale scores (α = .84 for Fall 2017; α = .83 for Spring 2018).

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

In Fall 2017, 522 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 255 first-year students (48%) indicated that ethical reasoning was in their top three most important desirable skills, with 80 first-year students (15%) indicating that ethical reasoning was their most desirable skill. Ethical reasoning was most often ranked the 2nd most important skill. Figure 1 displays the frequency of first-year students who placed ethical reasoning at each importance ranking.

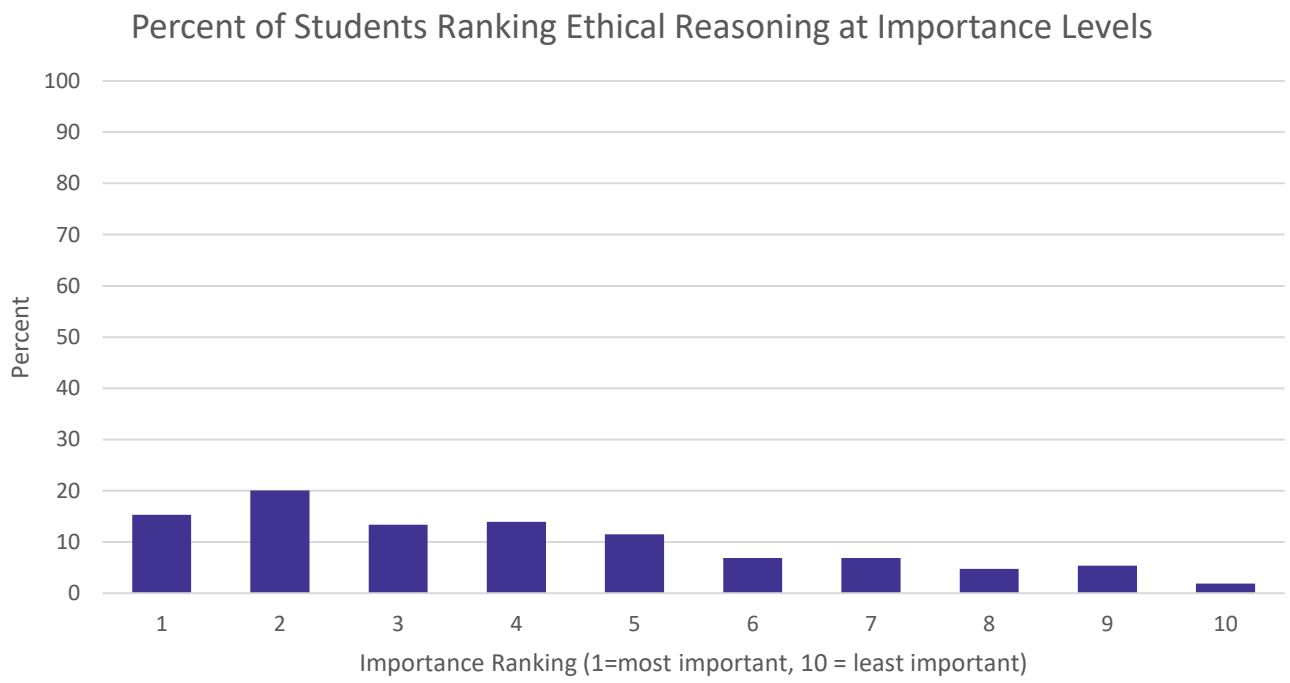

Figure 1. Percent of first-year students ranking ethical reasoning skills at importance levels

In addition to rank-ordering desirable skills, first-year students were explicitly asked about their perceived importance of ethical reasoning skills through 5 Likert-type items. The distribution of average scores is provided in Figure 2. On average, first-year students scored 4.53 out of 5 points on the importance subscale, suggesting that first-year students agree that ethical reasoning skills are important.

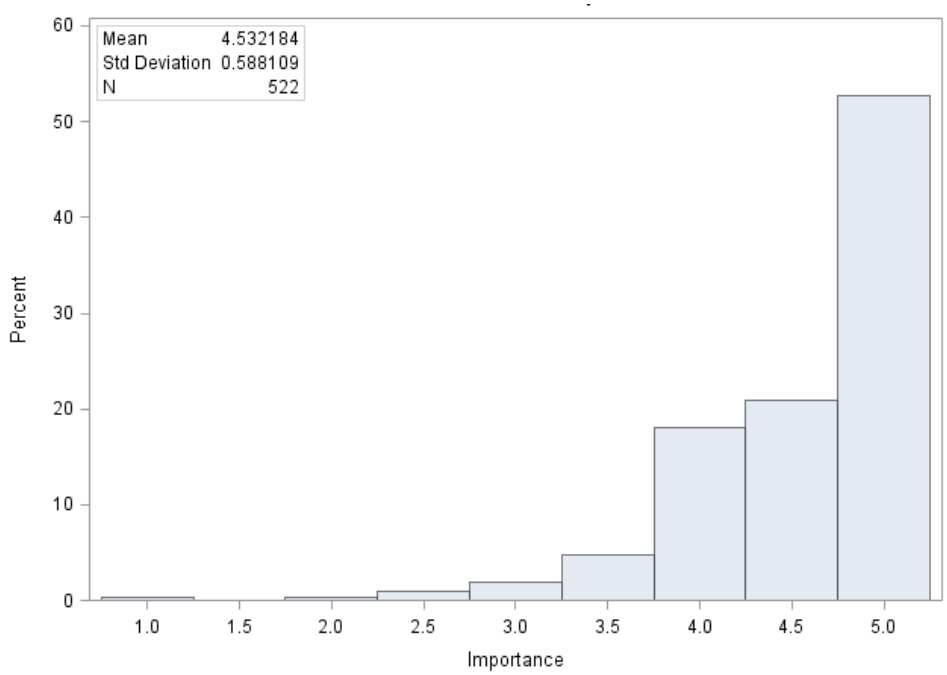

Figure 2. Distribution of average importance scores

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

In Fall 2017, 522 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 4.12 out of 5 points on the confidence subscale, suggesting that first-year students agree that that they have confidence to apply the ethical reasoning process.

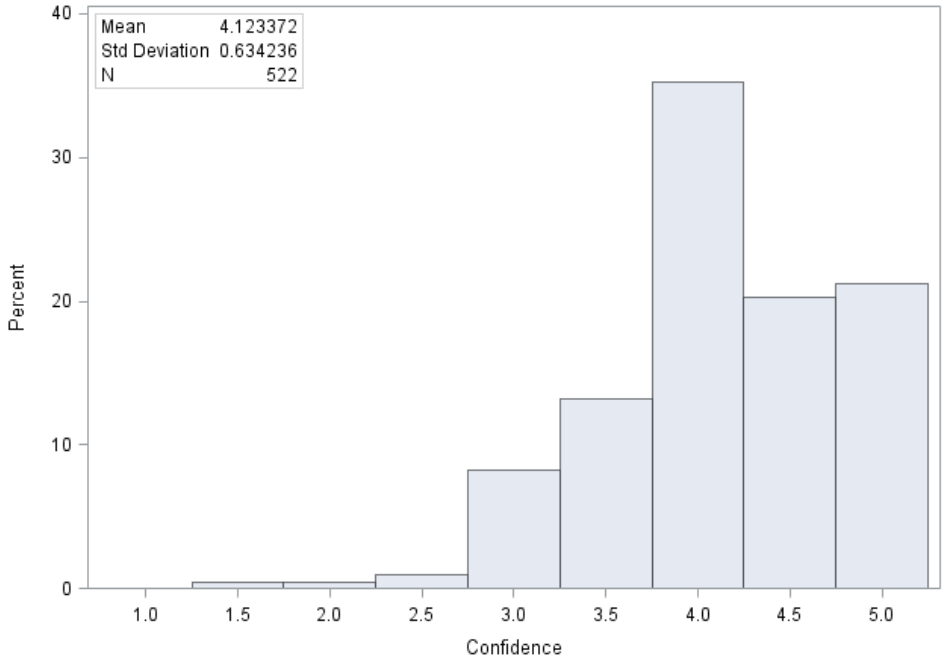

Figure 3. Distribution of average confidence scores

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

In Spring 2018, 274 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 111 second-year students (40%) indicated that ethical reasoning was in their top three most important desirable skills, with 28 second-year students (10%) indicating that ethical reasoning was their most desirable skill. Ethical reasoning was most often ranked the 2nd, 3rd, or 4th most important skill. Figure 4 displays the frequency of second-year students who placed ethical reasoning at each importance ranking.

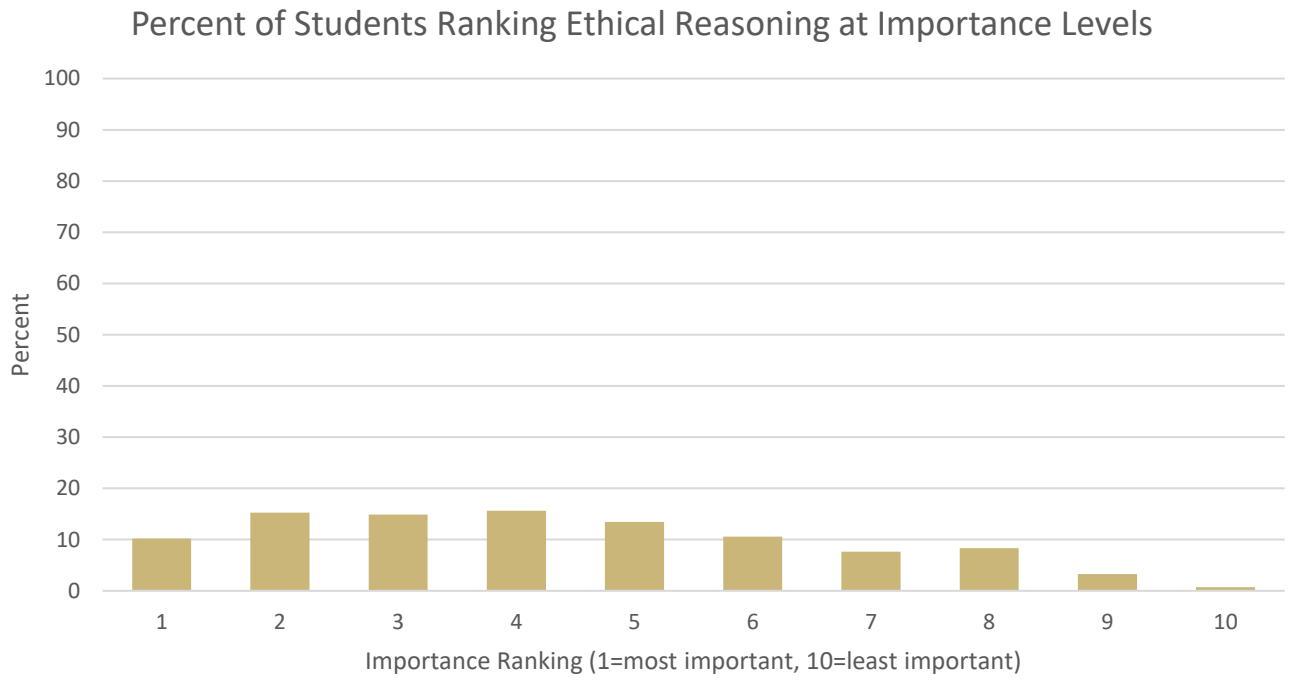

Figure 4. Percent of second-year students ranking ethical reasoning skills at importance levels

In addition to rank-ordering desirable skills, second-year students were explicitly asked about their perceived importance of ethical reasoning skills through 5 Likert-type items. The distribution of average scores is provided in Figure 5. On average, second-year students tend to score 4.46 out of 5 points on the importance subscale, suggesting that second-year students agree that ethical reasoning skills are important.

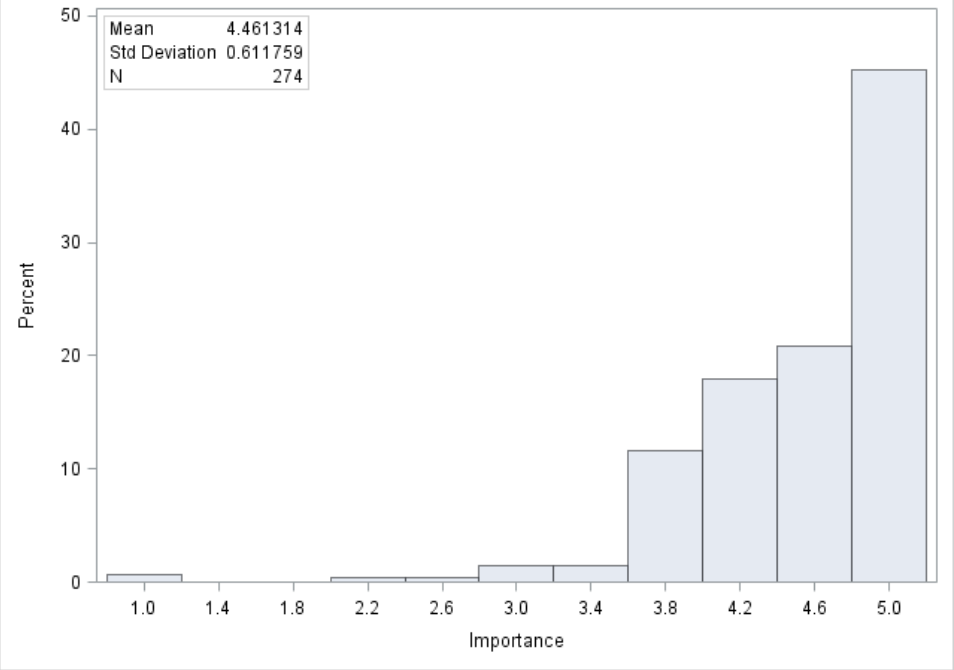

Figure 5. Distribution of average importance scores

How confident are second-year students in their ethical reasoning

In Spring 2018, 272 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 4.21 out of 5 points on the confidence subscale, suggesting that second-year students agree that that they have confidence to apply the ethical reasoning process.

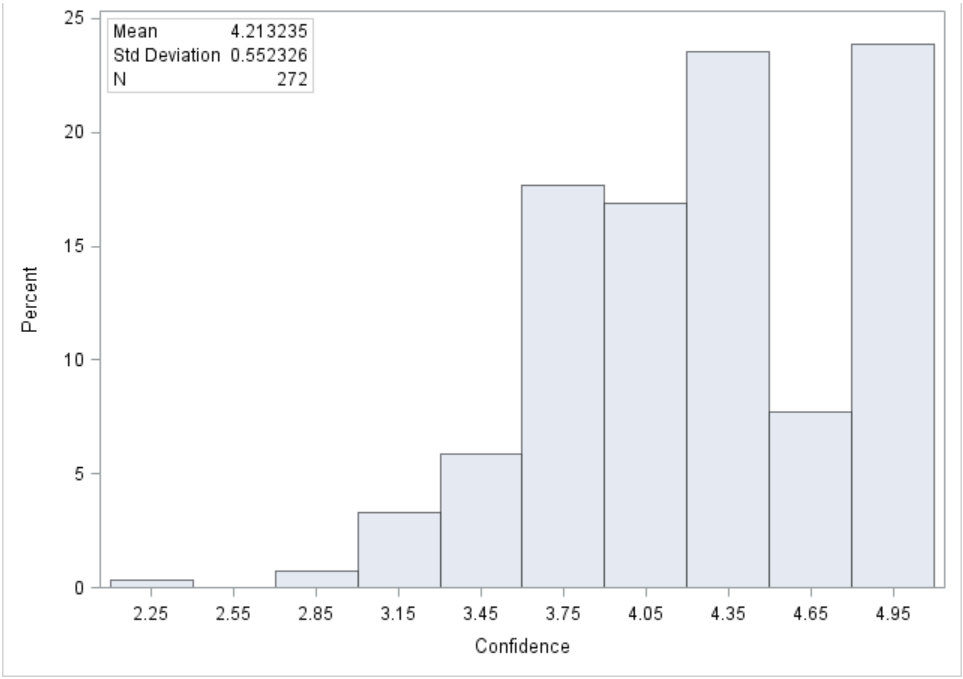

Figure 6. Distribution of average confidence scores

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

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 2016) scores were compared to their SER scores as second-year students (Spring 2018). Only students who completed the SER in Fall 2016 **and** Spring 2018 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 2018 than in Fall 2016, [$t(269) = -0.23$, $p = .004$]. 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 2016 to Spring 2018, [$t(269) = -0.22$, $p = .825$].

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 2016	Importance	1.80	5.00	4.60	0.48
	Confidence	1.80	5.00	4.20	0.56
Spring 2018	Importance	1.00	5.00	4.46	0.61
	Confidence	2.20	5.00	4.21	0.55

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 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. Thus, those are the only ones reported here. The KQ recall scores for first-year students had lower reliability ($G = 0.78$) than second-year students ($G = 0.95$). For KQ explanation scores, first-year students had lower reliability as well ($G = 0.61$) than the second-year students ($G = 0.86$). 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.

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 6 out of the 8KQs; these results are similar to the results from previous years.

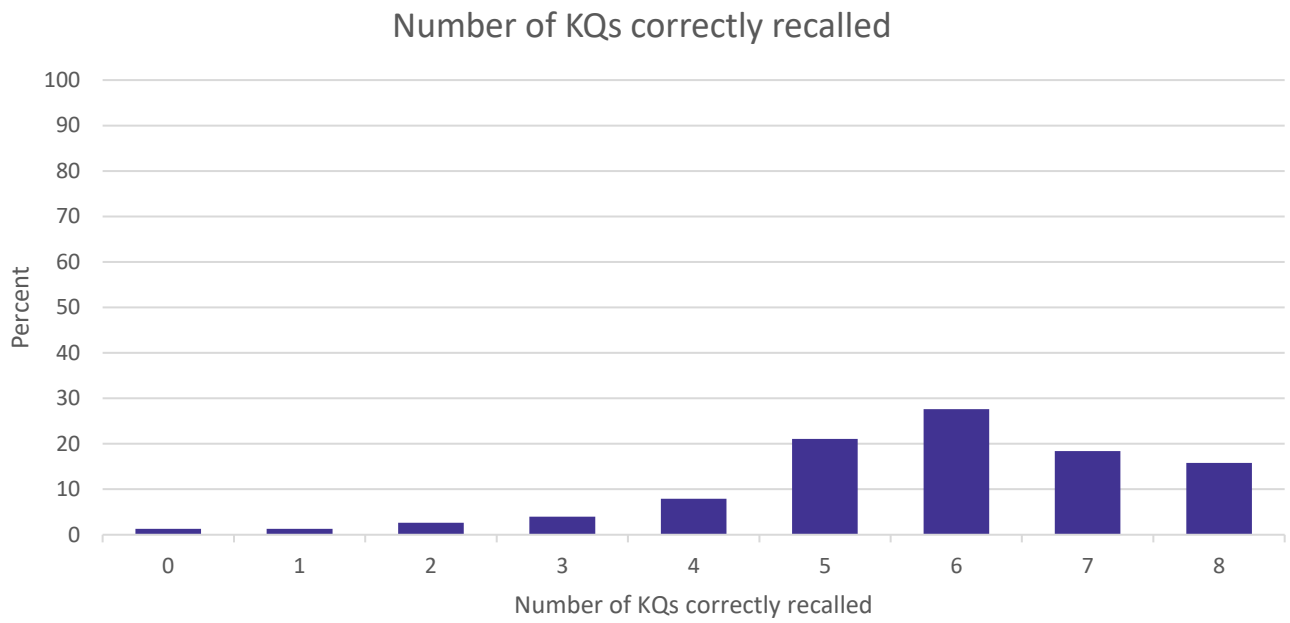

Figure 7. Percentage of students correctly recalling KQs on the ERRT.

Table 2 provides the percentage of first-year students recalling each of the 8KQs. It appears that Fairness, Outcomes, and Responsibilities were the easiest 8KQs to recall, as at least 80% of first-year students correctly recalled these 8KQs. It appears that Empathy and Rights were the most difficult key questions to recall, as only approximately 71% of first-year students were able to recall each of these 8KQs.

Table 2. ***Difficulty of Key Question Recall (Fall 2017)***

Key Question	Percent of Students Recalling KQ
Fairness	97.37%
Outcomes	92.10%
Responsibilities	82.89%
Character	72.37%
Liberty	73.68%
Empathy	71.05%
Authority	78.59%
Rights	70.79%

Note. N = 76

How many 8KQs do first-year students explain?

The distribution of first-year students' scores is provided in Figure 8. On average, first-year students accurately explained about 3 out of the 8KQs; these results are similar to the results from previous years.

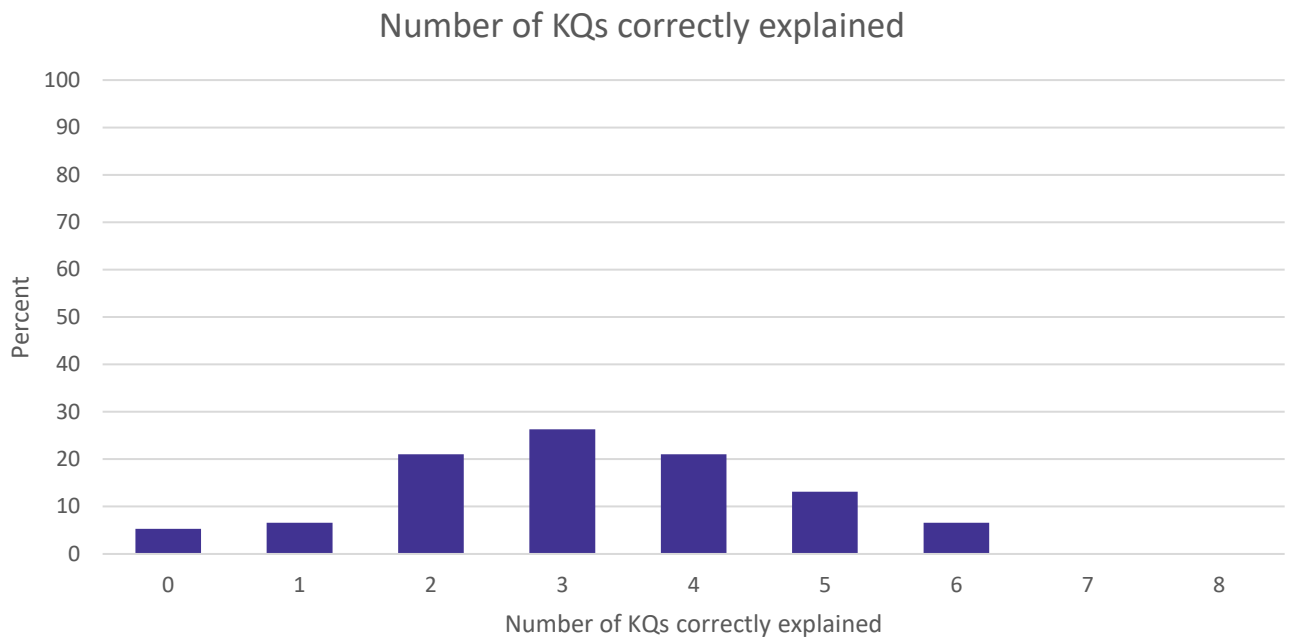

Figure 8. Percentage of students correctly explaining KQs on the ERRT.

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 90% of first-year students receiving partial or full credit for their explanations. It appears that Empathy and Authority were the most difficult key questions to explain, as less than 65% of first-year students received partial or full credit for their explanations of these 8KQs.

Table 3. ***Difficulty of Key Question Explanation (Fall 2017)***

Key Question	Percentage of Students Receiving Partial Credit	Percentage of Students Receiving Full Credit	Percentage of Students Receiving ANY Credit
Fairness	80.25%	14.47%	94.74%
Outcomes	72.37%	18.42%	90.79%
Responsibilities	65.79%	14.47%	80.26%
Character	59.21%	19.74%	78.95%
Liberty	57.90%	7.89%	65.79%

Key Question	Percentage of Students Receiving Partial Credit	Percentage of Students Receiving Full Credit	Percentage of Students Receiving ANY Credit
Empathy	61.84%	7.32%	63.16%
Authority	55.27%	5.26%	60.53%
Rights	63.16%	2.63%	65.79%

Note. N=76

How many 8KQs do second-year students recall?

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; nearly 20% of second-year students recalled only one of the 8KQs.

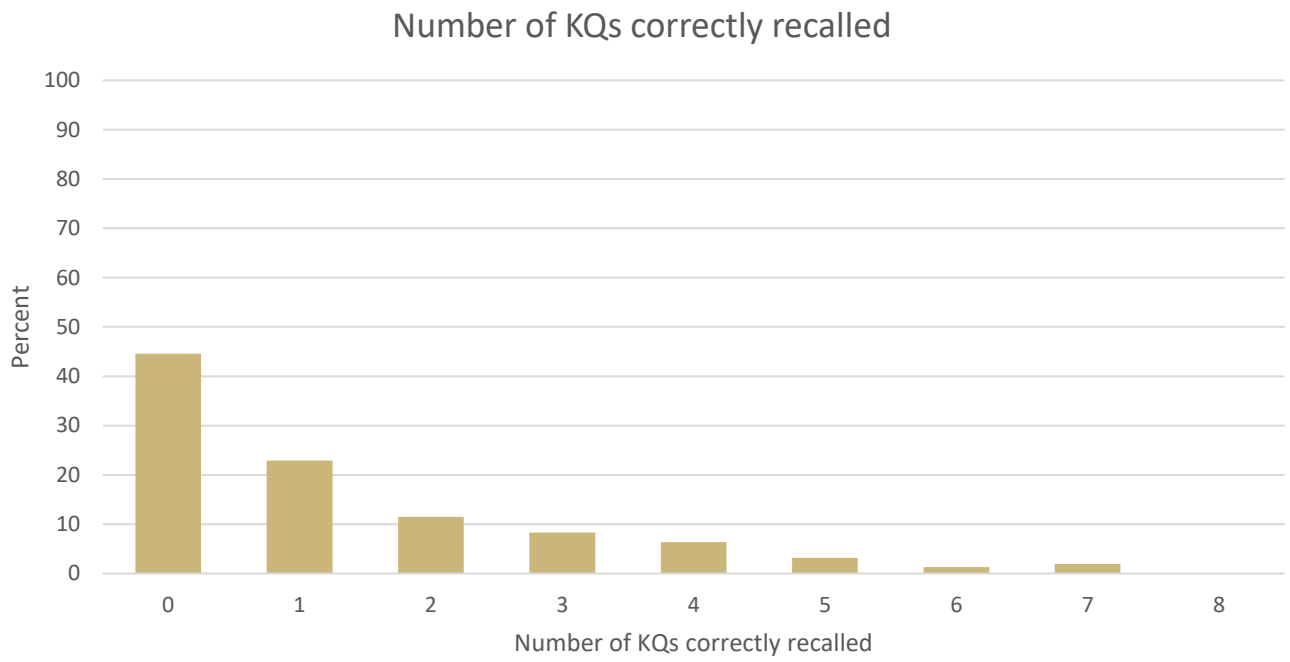

Figure 9. Percentage of students recalling KQs correctly on the ERRT.

Table 4 provides the percentage of second-year students recalling each of the 8KQs. It appears that Fairness, Outcomes, and Responsibilities were the easiest 8KQs to recall, with more than 30% of second-year students correctly recalling these 8KQs. It appears that Authority, Liberty, and Rights were the most difficult key questions to recall, as less than 20% of second-year students were able to recall each of these 8KQs.

Table 4. ***Difficulty of Key Question Recall (Spring 2018)***

Key Question	Percent of Students Recalling KQ
Fairness	36.94%
Outcomes	32.48%
Responsibilities	33.76%
Character	20.39%
Liberty	19.10%
Empathy	24.20%
Authority	15.92%
Rights	19.11%

Note. N = 157

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 about 1 out of the 8KQs. About 70% of second-year students did not explain any of the 8KQs correctly.

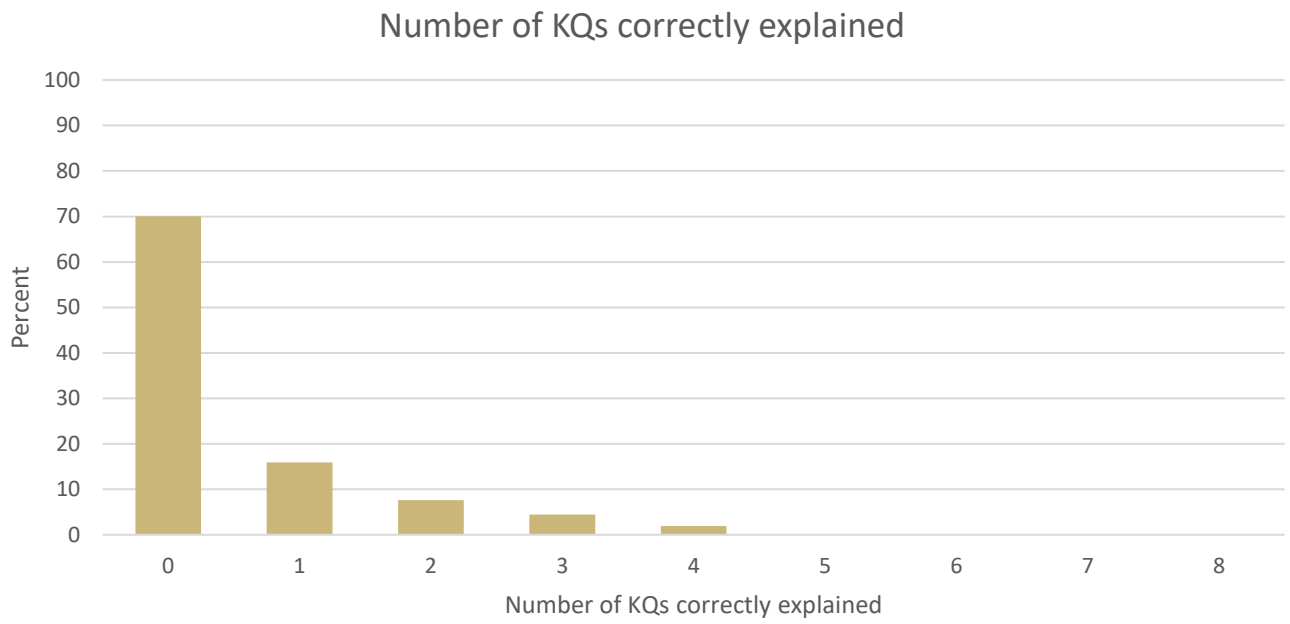

Figure 10. Percentage of students correctly recalling KQs on the ERRT.

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 Outcomes were the easiest key questions to explain, with more than 25% 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 8KQs.

Table 5. ***Difficulty of Key Question Explanation (Spring 2018)***

Key Question	Percentage of Students Receiving Partial Credit	Percentage of Students Receiving Full Credit	Percentage of Students Receiving ANY Credit
Fairness	31.84%	0.64%	32.48%
Outcomes	26.11%	0.64%	26.75%
Responsibilities	22.93%	1.91%	24.84%
Character	22.29%	2.55%	24.84%
Liberty	11.46%	0.64%	12.10%
Empathy	22.93%	1.91%	24.84%
Authority	12.74%	0.64%	13.38%
Rights	15.28%	0.64%	15.92%

Note. N=157

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

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 2016) scores were compared to their ERRT scores as second-year students (Spring 2018). Only students who completed the ERRT in Fall 2016 **and** Spring 2018 were included in the analyses.

On average, students' ERRT recall scores were statistically significantly higher as first-year students than as second-year students, [$t(97) = 18.19$, $p < .0001$; $d = 2.09$]. Specifically, students are recalling at least four fewer key questions as second-year student ($M = 5.52$) than as first-year students ($M = 1.63$). Figure 11 presents the percentage of students correctly recalling each of the 8KQs as first-year students and second-year students.

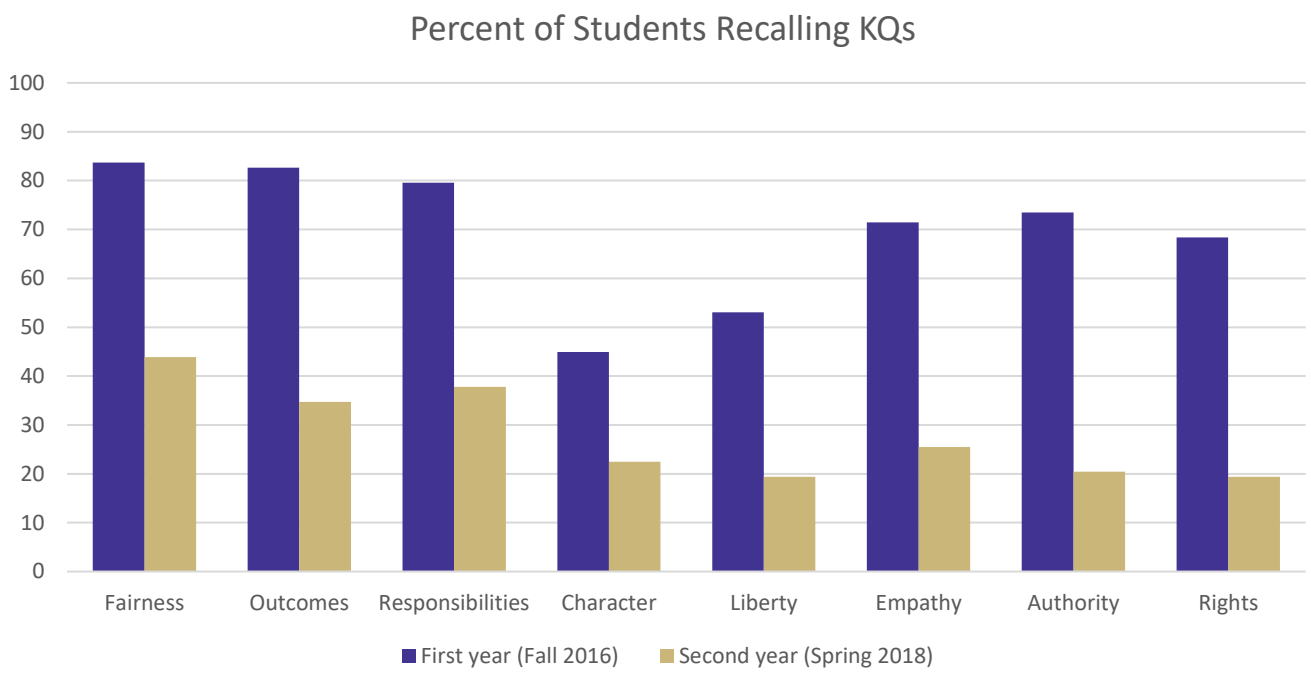

Figure 11. Percent of students recalling each KQ.

Additionally, students' ERRT explanation scores were statistically significantly higher as first-year students than as second-year students, [$t(97) = 11.00$, $p < .0001$; $d = 1.41$]. Students are explaining at least two fewer key questions as second-year student ($M = 0.83$) than as first-year students ($M = 2.74$). Figure 12 presents the percentage of students correctly explaining each of the 8KQs as first-year students and second-year students.

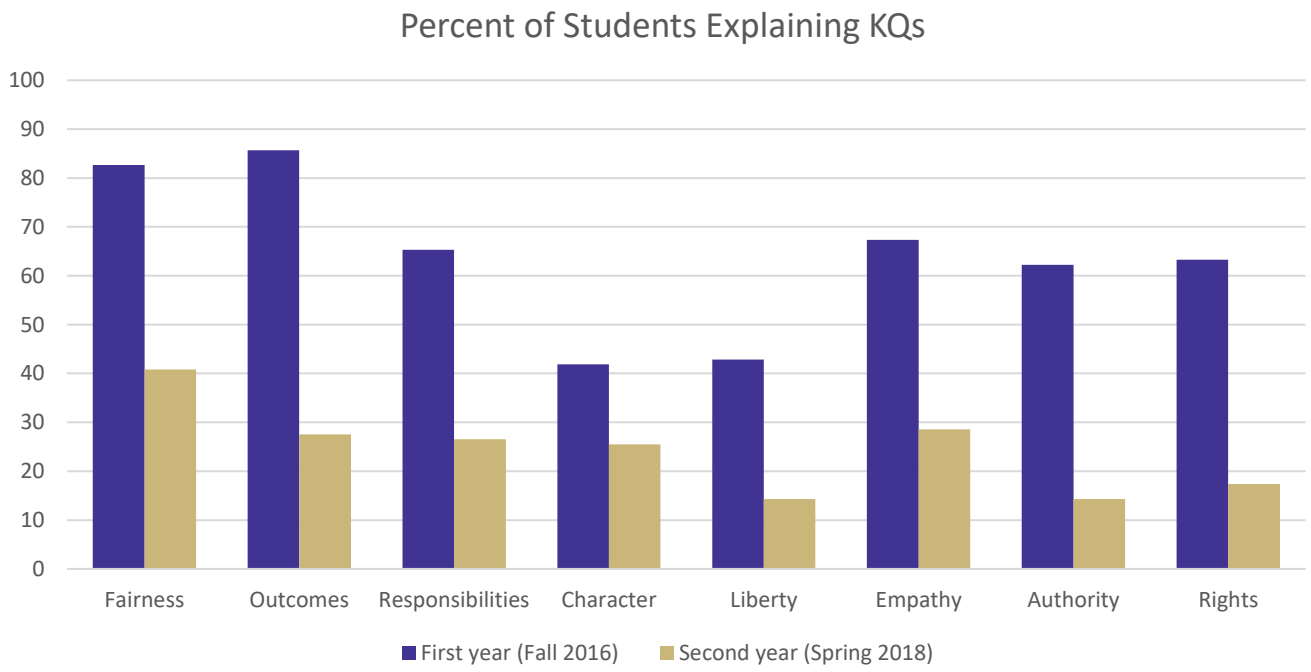

Figure 12. Percent of students receiving credit for explaining each KQ.

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. Additionally 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 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 = .86 for Fall 2017; α = .91 for Spring 2018) 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 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 547 first-year students responded to the ERIT. Figure 13 shows the distribution of Fall 2017 ERIT scores. On average, first-year students scored about 68% correct. The majority of students scored between 54% and 82% correct on the ERIT.

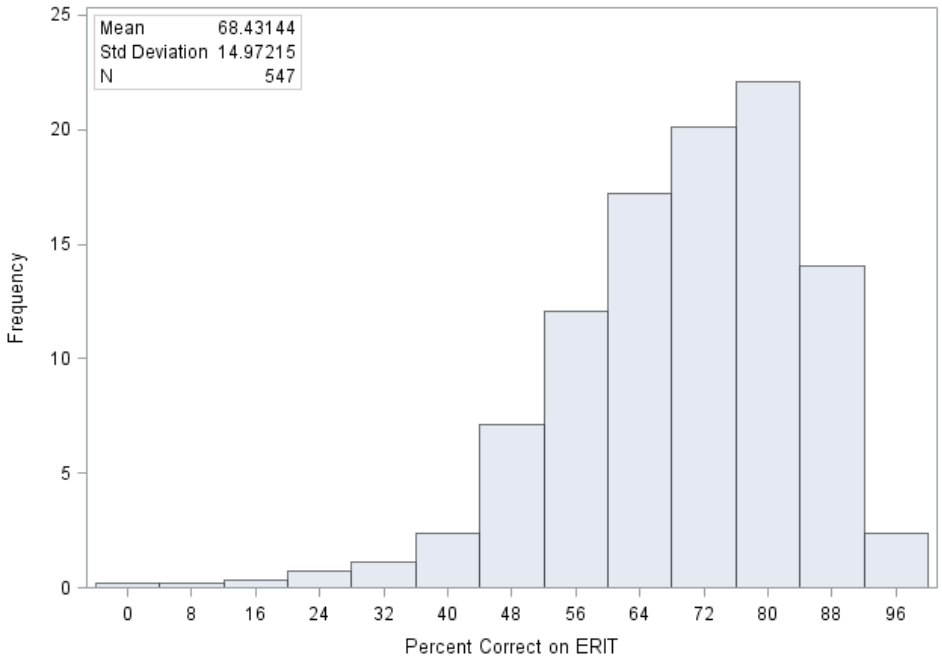

Figure 13. Distribution of ERIT Percent Correct Scores.

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 Character, Fairness, Empathy, and Outcomes, 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	76.14%	20.23
Outcomes	73.28%	19.96
Responsibilities	62.39%	23.18
Character	83.03%	19.71
Liberty	56.12%	27.21
Empathy	74.22%	18.35
Authority	71.33%	23.98
Rights	54.06%	23.92
TOTAL	**68.43%**	**14.97**

Note. N = 547

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

A totl of 291 second-year students completed the ERIT. Figure 14 shows the distribution of Spring 2018 ERIT scores. On average, second-year students scored about 68% correct. The majority of students scored between 50% and 86% correct on the ERIT.

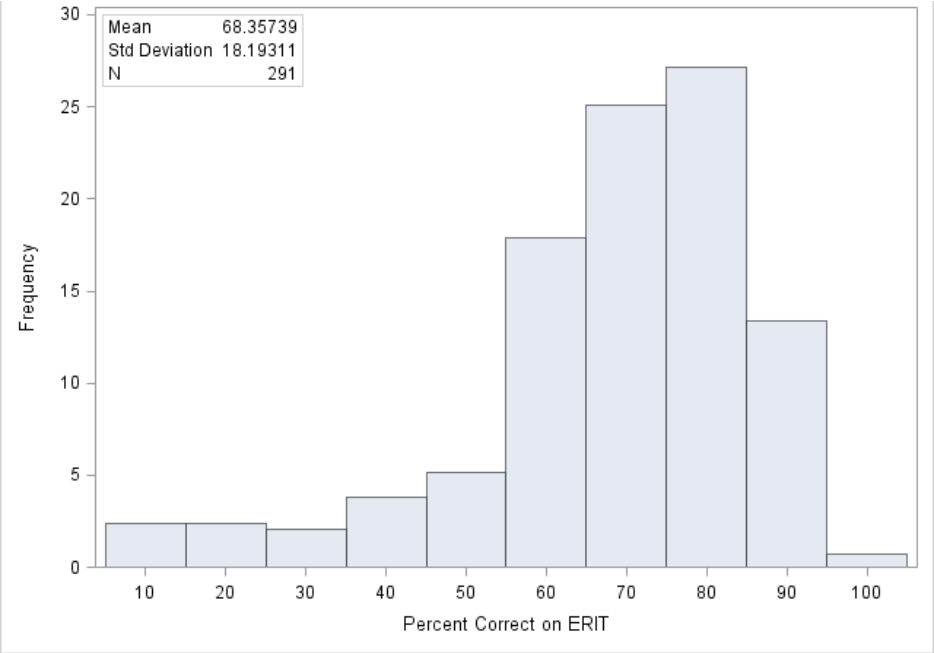

Figure 14. Distribution of ERIT percent correct scores.

Table 7 provides the average percent of items answered correctly for each key question. On average, second-year students answer the most items correctly for Character, Fairness, Authority, and Outcomes, suggesting these key questions may be easier for students to grasp. 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	78.41%	22.80
Outcomes	72.05%	22.61
Responsibilities	62.35%	25.24
Character	82.47%	22.32
Liberty	55.67%	29.05

Key Question	Average % of items answered correctly	*Standard Deviation (SD)*
Empathy	71.19%	20.84
Authority	72.79%	26.47
Rights	55.04%	23.96
TOTAL	**68.36%**	**18.19**

Note. N = 291

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

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 2016) scores were compared to their ERIT scores as second-year students (Spring 2018). Only students who completed the ERIT in Fall 2016 **and** Spring 2018 were included in the analyses. Students' ERIT total scores did not change significantly from Fall 2016 to Spring 2018, [$t(183) = -.52$, $p = .606$]. In fact, averages were nearly identical with students scoring 35.41 points on average as a first-year student and 35.67 points as a second-year student. These results are similar to prior years' results and suggest that, on average, from their first to second-year year, students are not meaningfully changing in their ethical reasoning abilities, as measured by the ERIT. 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
35.00	19.00	0.26	6.84

Recall that *The One Book* and *It's Complicated* are the only required intervention for 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' first and second-year years. Thus, it may be expected that students' ERIT scores would not improve from Fall 2016 to Spring 2018.

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 describe an ethical dilemma from their own lives. Students are expected to describe the dilemma, provide the considerations from which they analyzed the dilemma, and explain the decision they came to and why. Student responses are scored by trained raters using the ER-WR rubric (see Appendix B). Scores are assigned to five rubric elements on a five-point scale (0 = *Insufficient*, 1 = *Marginal*, 2 = *Good*, 3 = *Excellent*, and 4 = *Extraordinary*).

A total of 460 essays (318 Fall 2017 and 142) were rated. Both veteran and new MC raters attended the rating session. We provided an hour and half rater training to raters to ensure that all raters were adequately calibrated to the ER-WR rubric. This training session included two practice essays that were rated as group and individually to ensure rubric calibration. Eleven raters participated in the rating session, one of which was a graduate student 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 11 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.

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. The KQ recall scores for first-year students had lower reliability (G = 0.78) than second-year students (G = 0.95). For KQ explanation scores, first-year students had lower reliability as well (G = 0.61) than the second-year students (G = 0.86). 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.

For both cohorts, G coefficients were lower than desirable (G = 0.66 for first-year scores, G = 0.54 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 1.33 with a standard deviation of 0.66. This means that about 68% of first-year students scored between 0.67 and 1.99 points. For second-year students, the average score on the ER-WR was 0.99 with a standard deviation of 0.50. This means that about 68% of second-year students scored between 0.49 and 1.49. For the second-year students in particular, there is minimal variability in scores, which could be a reason for decreased reliability of second-year students' scores. Given low reliability, results below should be interpreted cautiously.

How do first-year students score on the ER-WR?

In all, there were 318 "rate-able" essays from first-year students. 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 1.10 points (e.g. just above marginal) on the ER-WR rubric in Fall 2017. As can be seen in Figure 15, first-year students scored the highest on Element A, which requires students to identify an ethical decision and describe the context surrounding the decision. 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.

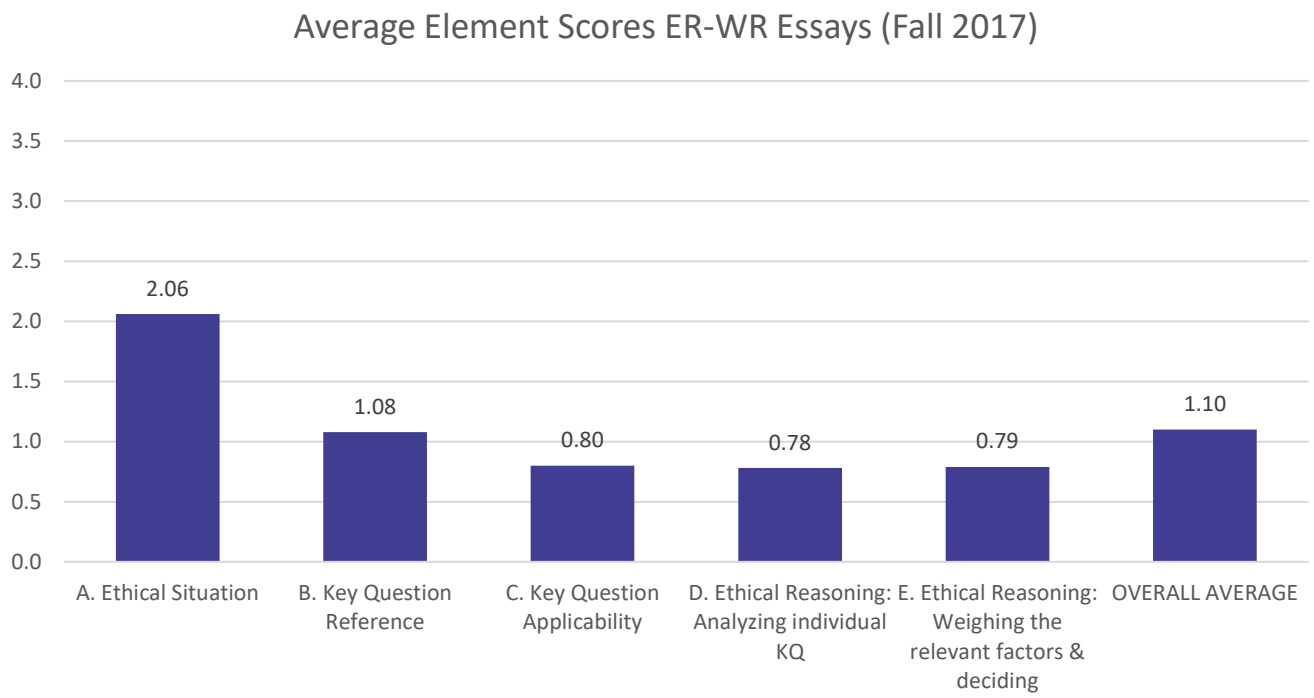

Figure 15. Average scores by element for first-year students (Fall 2017).

How do second-year students score on the ER-WR?

In all, there were 318 "rate-able" essays from first-year students. 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, second-year students scored 0.97 points (e.g. marginal) on the ER-WR rubric in Spring 2018. As can be seen in Figure 16, second-year students scored the highest on Element A, which requires students to identify an ethical decision and describe the context surrounding the decision. Students scored the lowest on Elements C, D, and E, with scores at below marginal for each of these elements. These results are similar to previous years.

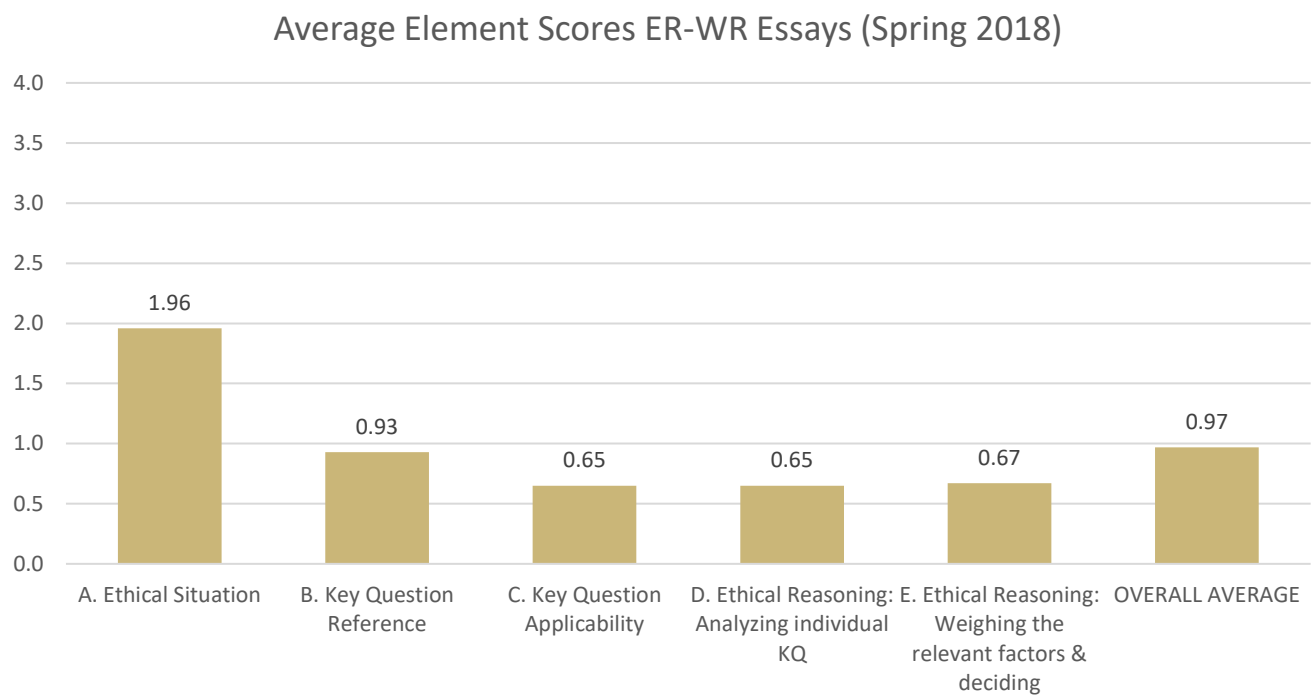

Figure 16. Average scores by element for second-year students (Spring 2018).

How do students' ER-WR scores change over time?

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 2016) scores were compared to their ER-WR scores as second-year students (Spring 2018). Only students who completed the ER-WR in Fall 2016 **and** Spring 2018 were included in the analyses (N = 100). On average, first-year students in assessed in Fall 2016 scored statistically significantly higher than they did when they were assessed as second-year students in Spring 2018, [$t(99)=3.90$, $p < .001$)]. Specifically, as first-year students, the average score was 1.39 out of 4 possible points and as second-year students, the average scores was 1.08 out of 4 possible points. That is, students' average scores decreased by 0.31 points from their first to second-year years. Further, as shown in Figure 17, students scored lower on each element as second-years 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.

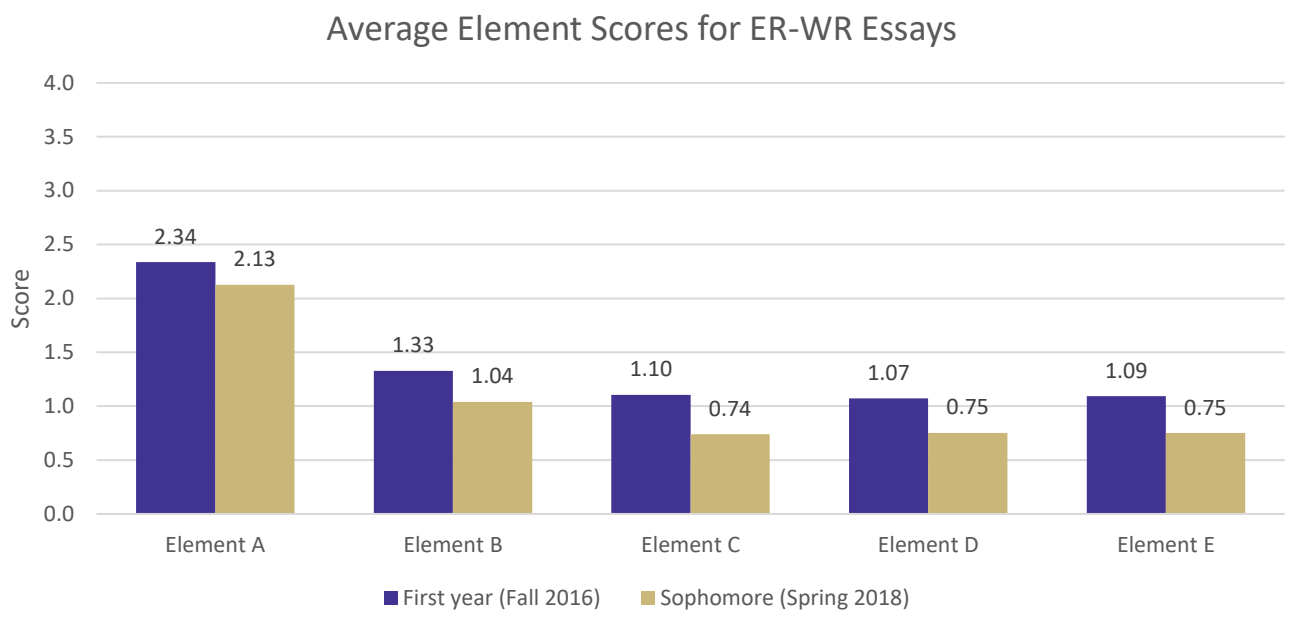

Figure 17. Longitudinal ER-WR scores by element.

Are students meeting the university strategic goal?

During the academic year (2012-2013), 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 18, about 8% (e.g., 38 out of 460 students) of the students who responded to the ER-WR during Fall 2017 and Spring 2018 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., student who did not experience any ethical reasoning interventions.

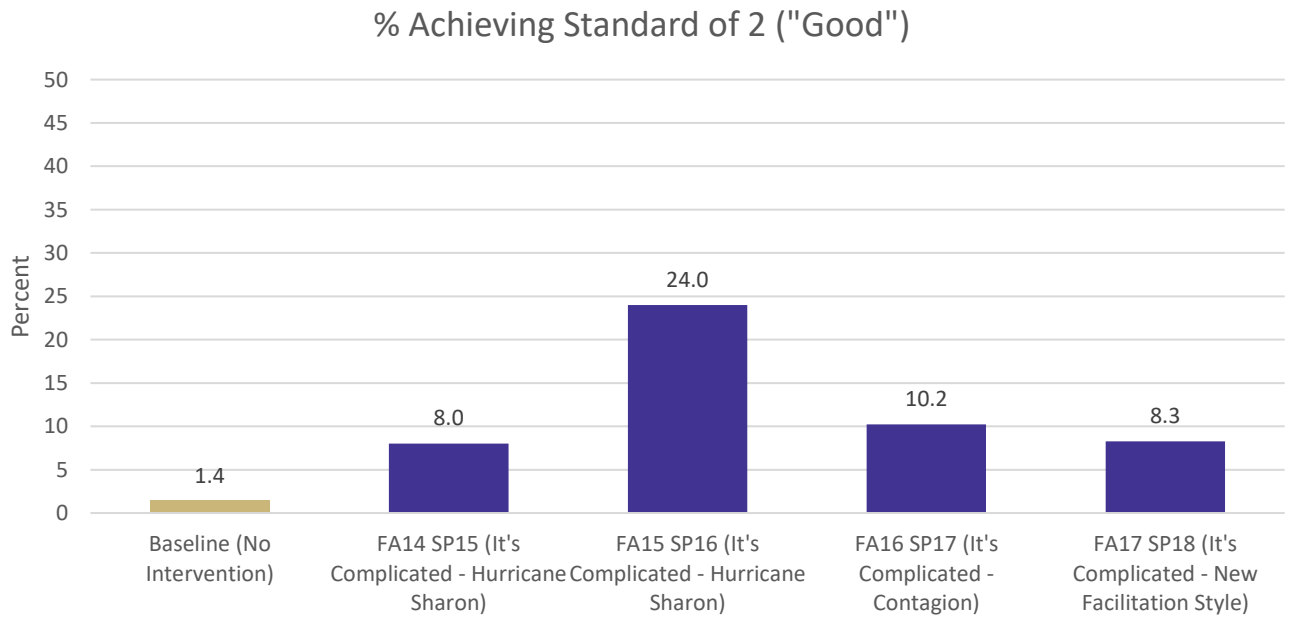

Figure 18. Percent of students meeting university-determined benchmark.

As shown in Table 9, the students who responded to the ER-WR during Fall 2017 and Spring 2018, 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. Further, scores have been steadily decreasing since FA15-SP16. The Ethical Reasoning in Action team should explore why scores are decreasing and potentially develop additional interventions to mitigate this issue.

Table 9. ***Average Scores by ER-WR 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	
Rubric Element	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
A. Ethical Situation	1.56	0.90	1.94	1.16	1.98	0.89	2.18	0.72	2.16	0.76	**2.03**	**0.87**
B. Key Question Reference	0.76	0.58	1.13	0.94	1.01	0.74	1.43	0.79	1.05	0.71	**1.03**	**0.69**
C. Key Question Applicability	0.44	0.48	0.82	0.78	0.71	0.60	1.14	0.72	0.84	0.65	**0.76**	**0.61**
D. Ethical Reasoning: Analyzing individual KQ	0.48	0.54	0.86	0.82	0.74	0.68	1.14	0.71	0.82	0.68	**0.74**	**0.63**
E. Ethical Reasoning: Weighing the relevant factors & deciding	0.50	0.55	0.90	0.83	0.81	0.67	1.07	0.69	0.89	0.67	**0.75**	**0.65**
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**

Note. *SD* = standard deviation and indicates the spread of scores around the mean. For example, a SD of about 1 on Element A with a mean of about 1 indicates that 68% of the essays (1 SD below and above the mean, assuming a normal distribution) received scores between 0 and 2. The scale is: 0 = Insufficient; 1 = Marginal; 2 = Good; 3 = Excellent; and 4 = Extraordinary.

Ethical Reasoning – Writing 2 (ER-WR2) and Ethical Reasoning – Writing 3 (ER-WR3) Essay

The ER-WR2 and ER-WR3 essays are performance assessment instruments designed to address SLO 5. Rather than asking students to think of their own ethical dilemmas to write about, the ER-WR2 and ER-WR3 pilot prompts provided students with a hypothetical scenario and asked them to apply the 8KQs to come to a decision about what they would do in that situation. The new prompts were reviewed and vetted by Ethical Reasoning in Action content experts. As with the original ER-WR, *JMU's Ethical Reasoning Rubric* (ER-WR rubric) is used to score the ER-WR2 and ER-WR3 essays; however, only the last four elements are used. That is, the first element is used to rate students on their abilities to identify an ethical situation and describe the situation surrounding the situation. However, because the ethical situation is provided for the students in the ER-WR2 and ER-WR3 prompt, students are not rated on this element. Scores are assigned to the four applicable rubric elements on a five-point scale (0 = *Insufficient*, 1 = *Marginal*, 2 = *Good*, 3 = *Excellent*, and 4 = *Extraordinary*).

A total of 185 ER-WR2 and 62 ER-WR3 essays were rated. One rating session was conducted with veteran raters who had participated in ER-WR rating in previous years. We provided an hour and half refresher rater training to these veteran raters to ensure that all raters were adequately calibrated to the ER-WR rubric, and understood the distinction between the new (i.e., ER-WR2 and ER-WR3) and the original (i.e., ER-WR) prompts. Six faculty member raters participated. Raters were assigned to anonymous pairs, and each rater pair evaluated their essays in reverse order to counteract fatigue effects.

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-WR2 and ER-WR3 comparisons. Thus, those are the only ones reported here. Reliabilities for the ER-WR2 (G = 0.757) and ER-WR3 (G = 0.750) scores were adequate

How do first-year students score on the ER-WR2 and the ER-WR3?

Given that both the ER-WR2 and ER-WR3 are new writing prompts, only first-year student data are available for analyses. For the Fall 2017 administration, the average score for the ER-WR2 (M = 1.15) is similar to the average score for the ER-WR (M = 1.10); however, the average score for the ER-WR3 (M = 0.71) is substanitally lower than the other two prompts (see Table 10). The suggests that the ER-WR3 prompt may be more difficult for students than the other prompts. The Ethical Reasoning in Action team should consider exploring what aspects of this prompt make it more difficult for students.

Table 10. *Average Scores by Element*

	ER-WR2 *N* = 185		ER-WR3 *N* = 62	
Rubric Element	**Mean**	***SD***	**Mean**	***SD***
B. Key Question Reference	1.31	0.88	1.15	0.75
C. Key Question Applicability	1.12	0.79	0.93	0.67
D. Ethical Reasoning: Analyzing individual KQ	1.07	0.80	0.86	0.65
E. Ethical Reasoning: Weighing the relevant factors & deciding	1.08	0.82	0.91	0.55
OVERALL AVERAGE	**1.15**	**0.78**	**0.71**	**0.63**

Note. *SD* = standard deviation. The scale is: 0 = Insufficient; 1 = Marginal; 2 = Good; 3 = Excellent; and 4 = Extraordinary.

Figure 19 depicts the percentage of students who met the university benchmark of "2" on the ER-WR2 and ER-WR3 writing prompt. For the Fall 2017 administration, about 15% of students met the standard for the ER-WR2 and about 11% of students met the standard for the ER-WR3. A higher percentage of students met the standard when they were provided an ethical dilemma (ER-WR2 and ER-WR3) compared to when students were asked to identify and justify their own ethical dilemma (ER-WR).

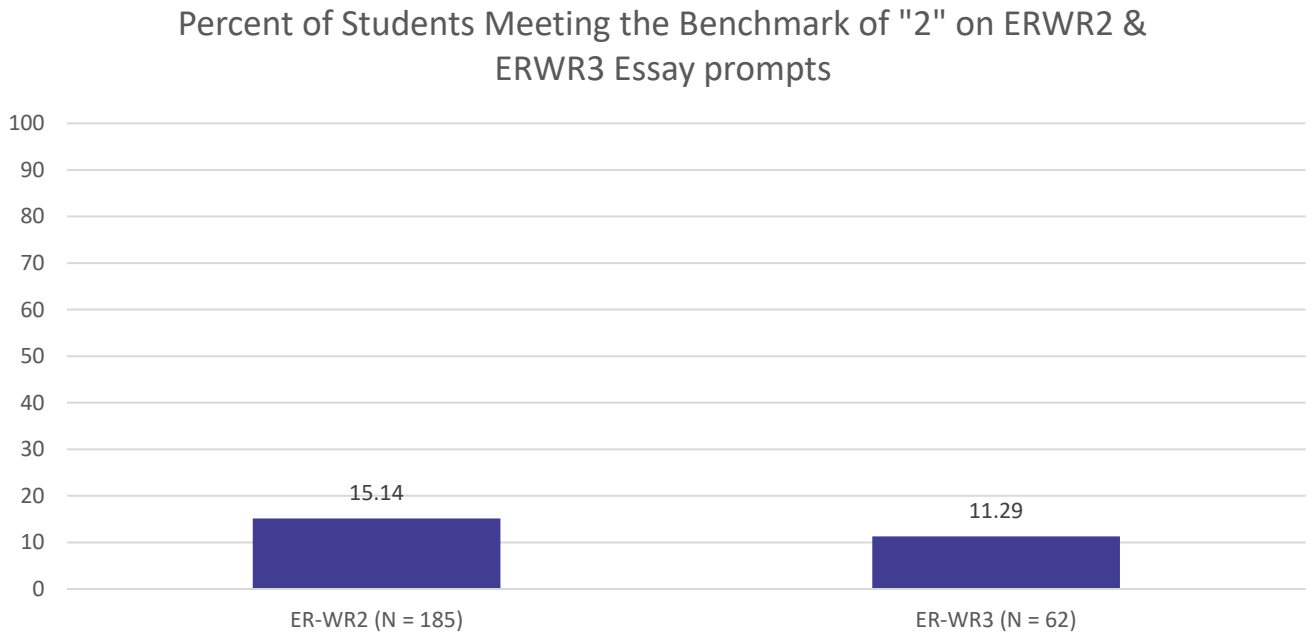

Figure 19. Percentage of students meeting university-determined benchmark

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).

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

Appendix A

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 two key questions referenced.	References four key questions.	References six key questions.	References all eight 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 two key questions to the ethical situation.	Provides a rationale for the applicability or inapplicability of four key questions to the ethical situation.	Provides a rationale for the applicability or inapplicability of six key questions to the ethical situation.	For all eight 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 of the referenced key questions.	Analysis attempted using two or more key questions. Typically incorrect ascription of the key questions to the ethical situation. Account is unclear, disorganized, or inaccurate.	Analysis attempted using three or more key questions. Basically accurate ascription of the key questions to the ethical situation. Account is unclear or disorganized.	Analysis attempted using three or more key questions. Accurate ascription of the key questions to the ethical situation. Account is clear and organized.	Meets criteria for *Excellent* AND… Nuanced treatment of key questions, for example: • elucidates subtle distinctions	

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.	