

Salisbury University DIT-2 Ethical Reasoning Report, Spring 2019

This report, authored by SU office of University Analysis, Reporting & Assessment (UARA) staff, discusses Ethical Reasoning-related survey data collected during spring 2019 GULL Week sessions.

To request more information about the assessment, results, or additional analyses, please contact the Assessment Coordinator, Dr. Sarah Winger.

Executive Summary

Background and Findings

1. Faculty and UARA staff agreed that the Defining Issues Test, Version 2 (DIT-2) is aligned with General Education student learning outcome, Ethical Reasoning.
2. The DIT-2 instrument is comprised of five dilemmas, each of which has one multiple choice decision item about the dilemma, twelve Likert-type rating items, and four ranking items. There are several indices which are measured by the DIT-2, but for brevity in this report we report only on N2 – which can be used to summarize an individual’s overall moral judgment. There are several studies compiled across years of institutional data collection on the DIT-2 N2 score to which we can compare our students’ results and identify areas of strength and those that need improvement.
3. The results of our administration of the DIT-2 instrument supported its validity and reliability.
 - a. DIT-2 scores demonstrated validity:
 - i. Content Validity: response options based on authentic interview responses in an earlier, open-ended version of the assessment
 - ii. Criterion and Construct Validity: supported in many studies (e.g., experts achieve high scores on the test; scores correlate with Moral Comprehension and the original DIT-1 instrument; as well as group differences, particularly based on age and experience) – also, scores on this instrument have a positive correlation with the SU students’ related measure of SAT Verbal scores, $r = .302$ ($p < .001$)
 - b. DIT-2 scores demonstrated reliability in previous studies and approached levels of generally acceptable reliability criteria in our more limited SU sample of only undergraduate students ($\alpha = .656$)
4. Generally, the students that completed the DIT-2 instrument were only somewhat representative of the overall and non-test-taker populations at SU.
5. The overall SU N2 average score (30.0) was below the average score of freshmen in the two most recent National Norm reports.
6. There was a significant difference between N2 average score of transfer students and SU native, first time students; where the latter had a higher average than the transfer students.
7. As SU and other institutions’ students’ class level (i.e., freshmen, sophomores, juniors, seniors) increased, so too did the N2 average score ([Table 10](#)). SU students’ N2 average score increased significantly by class level; freshmen’s average score was significantly less than juniors’ and seniors’ average score; sophomores’ average score did not significantly differ from the other groups.

8. There was a significant difference between N2 average score by SU college/school (i.e., CHHS, Fulton, Henson, Perdue, and Seidel; based on students' primary major); Perdue and CHHS majors' average scores were significantly less than Fulton and Henson majors' average scores; no other college or school comparisons were significantly different.
9. Although the sample was small, a paired analysis supported a significant increase in N2 average score for SU students over time.

Suggested Action Items

1. The National Norms/benchmarks with which SU students' Ethical Reasoning are compared should be evaluated by objective faculty and/or staff with expertise in the discipline or assessment of those skills. Although SU data are lower than published National Norms, it is not possible to test significance between SU data and the National Norms.
2. Perform an area/course mapping of the current SU courses that align with the revised Ethical Reasoning student learning outcome.
3. Teaching faculty, General Education Steering Committee, and other relevant parties should consider whether or not the DIT-2 instrument is well aligned with revised (as of November 2018) General Education Ethical Reasoning student learning outcome. If the DIT-2 instrument is not aligned, then an alternative assessment that is aligned should be identified.
4. Consider results from the assessment to develop interventions or review and update curriculum to align with areas that need improvement.
5. SU stakeholders should request follow-up analyses on DIT-2 data that were not included in this report that will be relevant to other questions of students' abilities in these outcomes.
6. Determine a timeline to re-collect assessment data related to Ethical Reasoning, tentatively set for re-assessing using the DIT-2 in fall 2021 and then every 3 years.
7. Increase student participation in future GULL Weeks, to increase the likelihood of participant samples that are representative of the entire SU student population, via competitions and marketing to both students as well as faculty that might offer course-embedded incentives for their students that participate.

Contents

Executive Summary	1
Background and Findings	1
Suggested Action Items	2
DIT-2 Instrument	4
Methodology and Sample	5
Results	5
Demographic Comparison of Test-takers vs. Non-test-takers	5
Validity and Reliability of the DIT-2 Instrument Administration at SU	8
SU Student Scores on DIT-2 Instrument	9
DIT-2 and SOS Survey Student Responses	11
Longitudinal SU Student Scores on DIT-2 Instrument	11
Discussion	12
References	13
Salisbury University DIT-2 Reporting Documentation	13
Appendices	14
Appendix 1. Example dilemma and associated decision-making, rating, and ranking items from the DIT-2 Instrument (Rest & Narvaez 1998)	14
Appendix 2. Student Opinion Scale (SOS) Survey (Sundre & Thelk 2007)	15

Detailed Ethical Reasoning Report

DIT-2 Instrument

The assessment is an instrument comprised of five dilemmas, each of which has one multiple choice decision item about the dilemma, twelve Likert-type rating items, and four ranking items. There are also other demographic items following the five dilemmas. See a DIT-2 example item in [Appendix 1](#) and its alignment with SU's student learning goals, outcomes, and curricular area mapping in [Table 1](#). Details about the instrument can be found at the [Center for the Study of Ethical Development, The University of Alabama – About the DIT website](#) (2019).

There are several indices which are measured by the DIT-2 and are described below. The first index is to what degree each individual is accessing each of three moral judgment schemas for their decision-making: Personal Interest, Maintain Norms, and Post Conventional (Bebeau & Thoma 2003; Rest *et al.* 1999a; Rest *et al.* 1999b). An individual always uses each schema to some extent, but the degree of usage of each schema changes across moral judgment development. *Personal Interest* is described as considering things such as direct advantages to the actor, intentions of the parties, or on maintaining approval with parties. *Maintain Norms* is described as “maintaining the existing legal system...existing roles, or a formal organizational structure” (Bebeau & Thoma 2003). *Post Conventional* is described as considering an organized society through consensus-building procedures, safeguarding minimal basic rights, and “organizing social arrangements and relationships in terms of intuitively appealing ideals” (Bebeau & Thoma 2003). Other useful summary scores include the Type Indicator and N2. *Type indicator* helps to determine the overall extent to which an individual, at the time of assessment, is either consolidated in a single schema, or is transitioning between two schemas – there are seven types. The *N2* score indicates the degree to which Post Conventional items are prioritized as well as the degree to which Personal Interest items are rated lower than Post Conventional items – which can be used to summarize an individual's overall moral judgment. Although more details can be derived from the SU student responses of the DIT-2, for brevity in this report we will only report on N2.

Faculty and UARA staff agreed that the DIT-2 instrument is aligned with the General Education Ethical Reasoning student learning outcome ([Table 1](#)).

Table 1. The SU General Education student learning goal, outcome, and area mapping related to Ethical Reasoning.

Student Learning Goal	Outcome	Area Mapping
Personal, Social, and Cultural Responsibility	<p>Ethical Reasoning: Students will be able to...</p> <ul style="list-style-type: none"> • reason about right and wrong human conduct; • assess their own ethical values and the social context of problems; • recognize ethical issues in a variety of settings; • think about how different ethical perspectives might be applied; and • consider the ramifications of alternate actions. 	TBD*

Note. Revised SU General Education student learning goals and outcomes were approved by Faculty Senate November 20, 2018. Asterisk (*) denotes that, at this time, there has not been an official area mapping of current courses to the revised SU General Education student learning goals and outcomes.

Related to Ethical Reasoning, results from this instrument can: provide a benchmark of the student outcome at SU; inform instructional efficacy and possible interventions; evaluate curricular strengths and weaknesses; and continuously improve student outcomes if we use this instrument for future GULL Week administrations.

Methodology and Sample

Data were collected from volunteer students at SU that self-selected and signed up to participate in various Gaining Understanding as a Lifelong Learner (GULL) Week testing sessions during a week in February, 2019. GULL Week sessions were open to the entire SU undergraduate student population. The assessments were administered in a proctored computer lab setting and lasted approximately one hour, of which ~25 minutes was dedicated to the DIT-2 instrument administration, ~25 minutes was dedicated to a different assessment aligned with a separate General Education Area, and ~5 minutes for a Student Opinion Scale (SOS) Survey (Appendix 2; Sundre & Thelk 2007). The SOS Survey estimates the GULL Week participant's perceived importance of the assessment(s) and effort expended by the participant in completing the assessment(s) (i.e., DIT-2 instrument).

Some faculty offered incentives (such as extra credit) to participating students, some mentioned GULL Week and encouraged students to participate, and some did not interact with students about GULL Week. The office of University Analysis, Reporting & Assessment (UARA) publicized GULL Week across campus via many avenues. For example, competitions between both College/Schools and Greek life groups were set up to improve participation.

In all, n = 2537 undergraduates participated in spring 2019 GULL Week and of those, quality responses from n = 1769 students were collected on the DIT-2 instrument (34.7% and 24.2% of total final spring 2019 undergraduate enrollment [n = 7308] at SU, respectively). Demographic analyses of the non-DIT-2 test-takers (n = 5539; 75.8%) were compared to the test-takers that completed DIT-2 to evaluate the extent to which the sample of test-takers was representative of the entire SU undergraduate population during spring 2019. Further analyses within the test-takers were performed to evaluate the validity and reliability of the instrument administration at SU as well as to determine whether or not scores on the instrument varied by student characteristic(s). The students with data for both DIT-2 and the SOS Survey were analyzed to evaluate student responses on those scales.

Furthermore, there were two instances of using the DIT-2 instrument during a time frame when students could have participated in both GULL Week administrations (fall 2015 and spring 2019). There were 33 students that participated in both administrations of the DIT-2 with quality data. Therefore, a paired analysis was performed to determine if their scores changed significantly over their tenure at SU.

Results

Demographic Comparison of Test-takers vs. Non-test-takers

The demographics of the students that took the DIT-2 instrument were somewhat similar to the non-test-takers (Tables 2-7; not significant (ns) annotations), but due to the nature of the sampling method there were groups that were not well represented. The comparison of DIT-2 test-takers (vs. non-test-takers) showed that: Asian students were disproportionately high, whereas Caucasian students were disproportionately low ([Table 2](#)); females were disproportionately high, whereas males were disproportionately low ([Table 3](#)); SU native first time students were disproportionately high, whereas transfer students were disproportionately low ([Table 4](#)); freshmen and sophomores were disproportionately high, whereas seniors and unclassified non-degree undergrads were disproportionately low ([Table 5](#)); and students from Henson and Seidel were disproportionately high,

whereas undeclared students were disproportionately low (Table 6); and, for the student success metrics (i.e., High School GPA, SU Cumulative GPA, as well as SAT Mathematics and SAT Verbal scores), the test-takers of the DIT-2 instrument were significantly more successful than the non-test-takers (Table 7 and Table 8). Therefore, the sample of DIT-2 test-takers was only somewhat representative of the entire SU undergraduate population during spring 2019. In the future, efforts to publicize GULL Week should be targeted more directly to Caucasian students, males, transfer students, seniors and unclassified non-degree undergrads, students who have not yet declared a major, and students that represent the less successful students (in terms of GPAs and SAT scores) as well as continuing previous publicity efforts to ensure even further representative sampling.

Table 2. Student Race/Ethnicity Compared between the DIT-2 Test-takers, Non-test-takers and All SU Undergraduates

Race/Ethnicity	Test-taker	Non-test-taker	Total
African American	264 (14.9%)	770 (13.9%)	1034 (14.2%)
American Indian/ Alaska Native	15 (0.8%)	32 (0.6%)	47 (0.6%)
Asian	78 (4.4%)*	186 (3.4%)*	264 (3.6%)
Caucasian	1197 (67.7%)*	3888 (70.3%)*	5085 (69.7%)
Hispanic	89 (5.0%)	230 (4.2%)	319 (4.1%)
Native Hawaiian/ Pacific Islander	3 (0.2%)	5 (0.1%)	8 (0.1%)
Non-resident Alien	19 (1.1%)	78 (1.4%)	97 (1.3%)
Two or more races	45 (2.5%)	150 (2.7%)	195 (2.7%)
Unknown/ Not specified	58 (3.3%)	192 (3.5%)	250 (3.4%)
Total	1768 (100.0%)	5531 (100.0%)	7299 (100.0%)

Note. Cell values are counts with percentages reported parenthetically. Significant difference of participation categories between test-takers' and non-test-takers' proportions are indicated by an asterisk (*), $p \leq .05$.

Table 3. Student Gender Compared between the DIT-2 Test-takers, Non-test-takers and All SU Undergraduates

Gender (code)	Test-taker	Non-test-taker	Total
Male (1)	572 (32.4%)*	2620 (47.4%)*	3192 (43.7%)
Female (2)	1190 (67.3%)*	2880 (52.1%)*	4070 (55.8%)
Total	1768 (100.0%)	5531 (100.0%)	7299 (100.0%)

Note. Cell values are counts with percentages reported parenthetically. Significant difference of participation categories between test-takers' and non-test-takers' proportions are indicated by an asterisk (*), $p \leq .05$.

Table 4. Student Admit Type, to SU, Compared between the DIT-2 Test-takers, Non-test-takers and All SU Undergraduates

SU Admit Type (code)	Test-taker	Non-test-taker	Total
First time student (F)	1331 (75.6%)*	3438 (64.5%)*	4769 (67.2%)
Transfer (T + U)	430 (24.4%)*	1893 (35.5%)*	2323 (32.8%)
Total	1761 (100.0%)	5331 (100.0%)	7092 (100.0%)

Note. Cell values are counts with percentages reported parenthetically. Significant difference of participation categories between test-takers' and non-test-takers' proportions are indicated by an asterisk (*), $p \leq .05$.

Table 5. Student Undergraduate Class Level Compared between the DIT-2 Test-takers, Non-test-takers and All SU Undergraduates

Class Level (code)	Test-taker	Non-test-taker	Total
Freshmen (1)	372 (21.0%)*	934 (16.9%)*	1306 (17.9%)
Sophomores (2)	433 (24.5%)*	1132 (20.5%)*	1565 (21.4%)
Juniors (3)	485 (27.4%)	1398 (25.3%)	1883 (25.8%)
Seniors (and +) (4)	456 (25.8%)*	1798 (32.5%)*	2254 (30.9%)
Unclassified non-degree undergrads (7)	22 (1.2%)*	269 (4.9%)*	291 (4.0%)
Total	1768 (100.0%)	5531 (100.0%)	7299 (100.0%)

Note. Cell values are counts with percentages reported parenthetically. Significant difference of participation categories between test-takers' and non-test-takers' proportions are indicated by an asterisk (*), $p \leq .05$.

Table 6. Student School Enrollment Compared between the DIT-2 Test-takers, Non-test-takers and All SU Undergraduates

School	Test-taker	Non-test-taker	Total
CHHS	410 (23.2%)	1188 (21.4%)	1598 (21.9%)
Fulton	434 (24.5%)*	1719 (31.0%)*	2153 (29.5%)
Henson	352 (19.9%)*	727 (13.1%)*	1079 (14.8%)
Perdue	365 (20.6%)	1141 (20.6%)	1506 (20.6%)
Seidel	160 (9.0%)*	415 (7.5%)*	575 (7.9%)
Undeclared	48 (2.7%)*	349 (6.3%)*	397 (5.4%)
Total	1769 (100.0%)	5539 (100.0%)	7308 (100.0%)

Note. Cell values are counts with percentages reported parenthetically. Significant difference of participation categories between test-takers' and non-test-takers' proportions are indicated by an asterisk (*), $p \leq .05$.

Table 7. Student SAT Scores Compared between the ILT Test-takers, Non-test-takers and All SU Undergraduates

SAT Score Range	SAT Mathematics			SAT Verbal		
	Test-taker	Non-test-taker	Total	Test-taker	Non-test-taker	Total
< 500	295 (23.7%)*	931 (27.1%)*	1226 (26.2%)	303 (24.4%)*	939 (27.3%)*	1242 (26.5%)
500-599	619 (49.8%)	1766 (51.3%)	2385 (50.9%)	612 (49.2%)	1700 (49.4%)	2312 (49.4%)
600-699	302 (24.3%)*	681 (19.8%)*	983 (21.0%)	301 (24.2%)	741 (21.5%)	1042 (22.3%)
700-800	27 (2.2%)	62 (1.8%)	89 (1.9%)	27 (2.2%)	60 (1.7%)	87 (1.9%)
Total	1243 (100.0%)	3440 (100.0%)	4683 (100.0%)	1243 (100.0%)	3440 (100.0%)	4683 (100.0%)

Notes. Cell values are counts with percentages reported parenthetically. Within each SAT subject, significant difference of participation categories between test-takers' and non-test-takers' proportions are indicated by an asterisk (*), $p \leq .05$. The SAT score ranges were used so that both the student scores on the old and 2016 SAT versions could be included.

Table 8. Student GPA Scores Compared between ILT Test-takers and Non-test-takers

Success Metric	Test-taker		Non-test-taker	
	n	Avg (SD)	n	Avg (SD)
High School GPA	1378	3.68 (.45)**	3732	3.55 (.47)**
SU Cumulative GPA	1766	3.15 (.58)**	5460	2.96 (.68)**

Notes. Cell values are sample sizes (n) or averages with standard deviation reported parenthetically. Significant difference of participation categories between test-takers' and non-test-takers' average values are indicated by two asterisks (**), $p \leq .001$.

Validity and Reliability of the DIT-2 Instrument Administration at SU

The results of our administration of the DIT-2 instrument supported its validity and reliability. Much of the validity of the DIT-2 instrument is described in the published documentation of the development of the instrument (Bebeau & Thoma 2003; Rest *et al.* 1998; Rest *et al.* 1999a; Rest *et al.* 1999b). A summary of details about the instrument's validity and reliability can be found at the [Center for the Study of Ethical Development, The University of Alabama – About the DIT website](#) (2019; Indicators of Validity section). Content validity was achieved through development of the response options based on authentic interview responses in an open-ended earlier version of the assessment (Rest *et al.* 1999b). Criterion and construct validity were supported in many studies, including those whose results demonstrated experts achieve high scores on the test; scores correlate with Moral Comprehension and the original DIT-1 instrument; as well as discriminant validity (e.g., group differences), particularly based on age and experience. Based on the SU student scores in spring 2019, criterion and construct validity were supported because students' N2 scores on this instrument had a moderate positive correlation with the SU students' related measure of SAT Verbal score range categories, $r = .302$ ($p < .001$). The score range categories were from 1 - 4 where: 1 = < 500; 2 = 500-599; 3 = 600-699; and 4 = 700-800). Correlation coefficients $\geq .3$, but $\leq .5$ are evidence of medium effect sizes (Field 2013).

Similarly, past studies have supported the reliability of the dilemmas in DIT-2, where Cronbach's alpha (α) value, which is a measure of reliability, or consistency, of these data has been $> .7$ over the five dilemmas for the N2 score (Bebeau & Thoma 2003; Rest *et al.* 1998a, Rest *et al.* 1998b). Typically, an α score $\geq .7$ is considered indicative of a reliable scale (DeVellis 2012). However, Bebeau & Thoma (2003)

warn that “if your sample does not contain the entire range of educational levels (from Junior high to Graduate School), your Cronbach alpha is likely to be lower.” Therefore, since the SU sample is only comprised of undergraduates, the SU spring 2019 DIT-2 instrument’s value of $\alpha = .656$ is therefore quite close to acceptable instrument reliability for this group.

SU Student Scores on DIT-2 Instrument

On average, the students that participated ($n = 1769$) had a N2 score of 30.0 (standard deviation, $SD = 14.0$) with a range of -4.3 to 73.6 on the DIT-2 instrument. Although there are no benchmark values for the N2 score, there are several studies compiled across years of institutional data collection on the DIT-2 (Bebeau & Thoma 2003; Dong 2014; Saculla *et al.* 2016) to which we can compare our students’ results. The overall SU N2 average score (30.0) was below the average score of freshmen in the two most recent National Norm reports (Dong 2014; Saculla *et al.* 2016) and equal to the average score of freshmen in the original National Norm report (Bebeau & Thoma 2003). However, since we do not have the raw data to compare the National Norm data sets to our own, we cannot evaluate statistically significant differences between them. Also, overlapping standard deviation of the National Norm studies’ groups’ scores with our groups’ scores indicates high variance within samples that may result in inability to identify statistical differences between groups. The Saculla *et al.* (2016) study is the most comparable to the data collected at SU as it was only comprised of data collected in an online format with the most recent testing dates (2010-14) of the various norm studies. Therefore, the Saculla *et al.* (2016) study will hereafter be referred to as the 2010-14 National Norm(s) or simply by National Norm(s), unless otherwise specified.

On average, SU native, first time students were significantly higher than transfer students in terms of N2 score from the DIT-2 instrument ([Table 9](#)). The difference, 2.3 was significant $t(1759) = 3.04, p < .05$.

Table 9. Student Admit Type, to SU, Average Scores on the DIT-2 Instrument.

SU Admit Type (code)	n	N2 Score	SD
First time student (F)	1331	30.6	14.1
Transfer (T + U)	430	28.3	13.7

In general, based on the National Norm data reports from DIT-2 researchers (Bebeau & Thoma 2003; Dong 2014), as SU and other institutions’ students’ class level (i.e., freshmen, sophomores, juniors, seniors) increased, so too did the average score on the instrument ([Table 10](#)). Although not shown in [Table 10](#) below, there was a general increase in N2 score as class level increased in the National Norm studies from both Bebeau & Thoma (2003) and Dong (2014), except that, in the former, juniors scored only slightly higher than freshmen and less than sophomores. Specifically at SU, after removing the unclassified students (whose sample size was less than 30), juniors and seniors scored significantly higher than freshmen on the DIT-2 instrument. However, the difference in average scores between class level groups was very small based on effect size value interpretation ($F(3, 955) = 5.4, p = .001, r = .09$). Post hoc comparisons, via the Tukey HSD test, were used to identify which class levels’ average scores were significantly different. Tests revealed significant pairwise differences between the average scores of freshmen as compared to juniors, $p < .05$, and seniors, $p < .05$. Sophomores did not significantly differ from the other groups, $p > .05$.

In a separate analysis, the covariate, student admit type, was found to be significantly related to the test-taker’s class level. However, the effect size was very small $F(1, 1606) = 12.3, p < .001, r = .09$. There was also a significant effect of the “dose” of class level (1 = freshmen; 2 = sophomores; 3 = juniors; 4 =

seniors) on N2 score after controlling for the effect of the student admit type, with a similarly small effect size, $F(3, 1606) = 8.3$, $p < .001$, partial $\eta^2 = .02$.

Also, a simple linear regression was calculated to predict DIT N2 score based on class level. A significant regression equation was found ($F(1,1767) = 11.969$, $p = .001$), with an R^2 of .007. Participants' predicted N2 score is equal to $27.516 + .965$ (N2 score) class level.

Table 10. SU (white columns) and National Norm (gray columns) Student Undergraduate Class Level Average Scores on the DIT-2 Instrument.

Class Level (code)	2010-14 National Norms			SU Fall 2015			SU Spring 2019		
	n	N2 Score	SD	n	N2 Score	SD	n	N2 Score	SD
Freshmen (1)	3482	35.8	15.05	154	28.5	13.6	372	27.7	13.05
Sophomores (2)	1308	33.5	15.85	155	29.3	14.1	433	29.9	13.84
Juniors (3)	1189	35.3	14.94	213	28.8	13.9	485	31.1	13.66
Seniors (and +) (4)	1807	37.0	15.39	205	30.4	15.1	456	30.9	15.10
Unclassified non-degree undergrads (7)	n/a	n/a	n/a	6	20.7	19.1	22	32.3	15.82

Note. The 2010-14 National Norms are reported in Saculla *et al.* (2016).

Student performance by SU College or School is listed in [Table 11](#). There was a significant difference in the DIT-2 N2 score based on enrollment in College or School at SU, but the difference in average scores between groups was small based on effect size value interpretation ($F(5, 369) = 6.4$, $p < .001$, $r = .13$). Post hoc comparisons, via the Tukey HSD test, were used to identify which College or Schools' average scores were significantly different. Tests revealed pairwise differences between the average score of students from Perdue, which was significantly lower, as compared to the average scores of students from Fulton ($p < .001$) and Henson ($p < .001$). Similarly, the College of Health and Human Services (CHHS) students' average score was significantly lower, as compared to the average scores of students from Fulton ($p < .01$) and Henson ($p < .05$). The average scores of Seidel majors and students whose majors are undeclared do not significantly differ from the other groups' average scores, $p > .05$.

Table 11. Student School Enrollment Average Scores on the DIT-2 Instrument.

College or School	SU Fall 2015			SU Spring 2019		
	n	N2 Score	SD	n	N2 Score	SD
CHHS	n/a	n/a	n/a	410	28.6 ^{b*}	13.0
Fulton	191	29.2 ^{a*}	14.0	434	31.9 ^{a*}	14.8
Henson	159	31.9 ^{a*}	15.0	352	32.0 ^{a*}	14.3
Perdue	183	28.1	14.5	365	27.6 ^{b*}	13.6
Seidel	176	28.4 ^b	13.6	160	29.4 ^b	13.5
Undeclared	24	26.7	13.9	48	33.0 ^a	13.9

Note. Significant difference, $p < .05$, of categories' average scores are indicated by group letters ^a and ^b, where the group ^a categories differ significantly compared to group ^b category are indicated by an asterisk (*), $p \leq .05$.

Although not presented here, student performance by primary major is available [upon request](#) to programs or Departments when at least 30 students in that major participated in this instrument's administration. These data can be used for informal review and improvement efforts, or for more formal

program review and improvement efforts such as Academic Program Review required reporting related to assessment of program student learning outcomes aligned with this instrument, when applicable.

DIT-2 and SOS Survey Student Responses

The DIT-2 test-takers also took the SOS Survey (n = 1769; [Table 12](#)). We were able to evaluate the reliability of both subscales within the SOS Survey. The *Importance* subscale, which addresses the extent to which the student thought it was important to do well on the DIT-2 instrument, demonstrated reliability ($\alpha = .775$). Similarly, the *Effort* subscale, which addresses the extent to which the student fully engaged in effortful behavior on the DIT-2 instrument, demonstrated reliability ($\alpha = .813$). The validity of the instrument is discussed in the SOS Survey Manual (Sundre & Theik 2007). The 10 items, five in each subscale, are measured in a 1 to 5 scale, where 1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; and 5 = Strongly Agree. There are four items that are negatively worded, and their scores were reverse coded prior to analysis.

In general, students' average response was "Agree" for the *Importance* subscale, whereas their average response tended to be within the "Agree" to "Strongly Agree" response range for the *Effort* subscale. For *Importance*, this indicates that students thought that their scores on the DIT-2 instrument would affect them somewhat in either a negative or positive way. For *Effort*, it indicates that students put in a moderate to high effort towards completing the DIT-2 instrument. The two subscales had a positive correlation with one another, $r = .358$ ($p < .001$; medium effect size) and both subscales were also minimally positively correlated with the DIT-2 N2 score (*Importance*, $r = .108$ ($p < .001$; small effect size); *Effort*, $r = .232$ ($p < .001$; small effect size). The latter two correlations of SOS subscales with the N2 scores seem to indicate that the students that self-reported that the DIT-2 was an important test and exerted more effort on performance on the test also scored higher than those who did not self-report the high importance of the test or exerting as much effort on the test, respectively, although the effect sizes were small.

Table 12. Student Opinion Scale (SOS) Survey subscales' administrative results for the students that also participated in the DIT-2 instrument administration.

SOS Subscale	Number of Items	Reliability (α)	n	Average Score (out of 25)	SD
Importance	5	.775	1769	16.2	3.8
Effort	5	.813	1769	20.6	3.1

Longitudinal SU Student Scores on DIT-2 Instrument

There were 33 students that took the DIT-2 instrument, with quality data, during both the fall 2015 and spring 2019 GULL Week administrations. Initial analyses (kurtosis and normal distribution) indicated that the samples were not normally distributed; therefore, the non-parametric form of the paired samples T test, the Wilcoxon signed-rank test, was performed. N2 scores were significantly higher for the spring 2019 administration (median = 39.1) as compared to fall 2015 administration (median = 25.3), $T = 102$, $p = .001$, $r = -.56$ (large effect size).

Discussion

Based on the results presented here it seems that there is room for improvement in student learning outcomes related to Ethical Reasoning at SU. Several action items are suggested below towards this end.

1. To determine whether or not our students are meeting SU expectations for Ethical Reasoning, the benchmarks with which SU students' Ethical Reasoning are compared should be evaluated by objective faculty and/or staff with expertise in the discipline or assessment of it. Since the DIT-2 instrument lacks any definition of a proficiency level and only provides National Norm values for certain groups for which we can only assume and not test whether or not our data are statistically less than those National Norms, it is difficult to evaluate actual student ability with the current instrument. Also, since for some schemas there are not clear trends for scoring across groups for which we would expect to see increases with age, schooling, and experience – this instrument may have deficiencies in how we can use it to evaluate and inform curricular interventions for improvements in student learning outcomes.
2. Perform an area/course mapping of the current SU courses that align with the revised Ethical Reasoning student learning outcome.
3. Based on discussions and decisions related to Action Items #1-2, relevant parties such as faculty teaching courses aligned with this student learning outcome and the General Education Steering Committee should consider whether or not the DIT-2 instrument is aligned well with the revised (as of November 2018) SU General Education Ethical Reasoning student learning outcome. If it is not aligned, then an alternative assessment that is aligned should be identified.
4. Relevant stakeholders at SU should consider the results from the DIT-2 assessment to develop interventions or review and update curricula to align with areas that need improvement. In particular, and possibly in conjunction with Action Item #2, these data can be re-evaluated to help identify particular courses that students with high N2 scores have completed at SU to investigate potentially successful Ethical Reasoning-related interventions on campus. Successful projects at other institutions may be considered to guide instructional interventions at SU.
5. Relevant stakeholders at SU should request further analyses of the DIT-2 data to address additional questions of interest that relate to other variables measured that were not described here (e.g., schema and type index scores). Some areas of potential research questions as well as literature review are provided by King & Mayhew (2002).
6. Based on discussions and decisions related to Action Items #1-5, a timeline for re-assessment of the SU General Education Ethical Reasoning student learning outcome should be finalized. At this time, the DIT-2 is planned to be re-assessed in fall 2021 and every three years after. This will allow an analysis of whether or not there is change in student learning outcomes based upon either a change in assessment or instructional or curricular interventions.
7. Attempt to increase student participation in future GULL Weeks, particularly in disproportionately low groups (i.e., Caucasians, [Table 2](#); males, [Table 3](#); transfer students, [Table 4](#); seniors and unclassified non-degree undergrads, [Table 5](#); Fulton and undeclared majors, [Table 6](#); students with SAT mathematics or verbal scores less than 500, [Table 7](#); students with lower High School GPA or SU Cumulative GPA, [Table 8](#)), to increase the likelihood of participant samples that are representative of the entire SU student population. This can be done via efforts that have occurred in the past, such as competitions and marketing to both students as well as faculty that might offer course-embedded incentives for their students that participate. However, new ways to incentivize participation of disproportionately low groups should also be identified and implemented.

References

- Bebeau, M.J., & Thoma, S.J. 2003. Guide for DIT-2: A guide for using the Defining Issues Test, Version 2 (DIT-2) and the scoring service of the Center for the Study of Ethical. Tuscaloosa, AL: Center for the Study of Ethical Development.
- Center for the Study of Ethical Development, The University of Alabama. 2019. About the DIT. Retrieved March 5, 2020, from <http://ethicaldevelopment.ua.edu/about-the-dit.html>
- DeVellis, R.F. 2012. Scale development: Theory and applications (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc., pp. 109.
- Dong, Y. 2014. Norms for DIT2: From 2005-2009. Tuscaloosa, AL: Center for the Study of Ethical Development.
- Field, A.P. 2013. Discovering statistics using IBM SPSS Statistics (4th ed.). Thousand Oaks, CA: Sage Publications, Inc., pp. 267.
- King, P.M., & Mayhew, M.J. 2002. Moral Judgement Development in Higher Education: Insights from the Defining Issues Test. *Journal of Moral Education*, 31(3), 247–270.
- Rest, J. R., & Narvaez, D. 1998. Defining Issues Test 2 Version 3.0. Minneapolis, MN: University of Minnesota.
- Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. J. 1999a. A Neo-Kohlbergian Approach: The DIT and Schema Theory. *Educational Psychology Review*, 11(4), 291–324.
<https://doi.org/10.1023/A:1022053215271>
- Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. J. 1999b. Postconventional Moral Thinking: A Neo-kohlbergian Approach. Lawrence Erlbaum Associates, Inc. Mahwah, NJ.
- Saculla, M. M., O'Connor, E., & Thoma, S. J. 2016. Online DIT-2 Normative Trends From 2010 to 2014. Presented at the American Educational Research Association (AERA) Annual Meeting, Washington, D.C.
- Sundre, D.L., & Thelk, A.D. 2007. [“The Student Opinion Scale \(SOS\): A Measure of Examinee Motivation: Test Manual.”](#) James Madison University, The Center for Assessment & Research Studies.
- Salisbury University DIT-2 Reporting Documentation
1. Fall 2015 Defining Issues Test Social Responsibility & Humane Values Assessment Report: <http://www.salisbury.edu/fall-2015-humane-values/>
 2. Fall 2015 Defining Issues Test Social Responsibility & Humane Values ONESHEET: <http://www.salisbury.edu/socialResponsibilityHumanValues/>

Appendices

[Appendix 1](#). Example dilemma and associated decision-making, rating, and ranking items from the DIT-2 Instrument (Rest & Narvaez 1998)

[Appendix 2](#). Student Opinion Scale (SOS) Survey (Sundre & Thelk 2007)

Appendix 1. Example dilemma and associated decision-making, rating, and ranking items from the DIT-2 Instrument (Rest & Narvaez 1998)

Famine Dilemma. The small village in northern India has experienced shortages of food before, but this year's famine is worse than ever. Some families are even trying to feed themselves by making soup from tree bark. Mustaq Singh's family is near starvation. He has heard that a rich man in his village has supplies of food stored away and is hoarding food while its price goes higher so that he can sell the food later at a huge profit. Mustaq is desperate and thinks about stealing some food from the rich man's warehouse. The small amount of food that he needs for his family probably wouldn't even be missed.

Item 1. What should Mustaq Singh do? Do you favor the action of taking food?

- A. Should take the food
- B. Can't decide
- C. Should not take the food

Item 2. Rate the following issues in terms of importance.

1. Is Mustaq Singh courageous enough to risk getting caught for stealing?	Great	Much	Some	Little	No
2. Isn't it only natural for a loving father to care so much for his family that he would steal?	Great	Much	Some	Little	No
3. Shouldn't the community's laws be upheld?	Great	Much	Some	Little	No
4. Does Mustaq Singh know a good recipe for preparing soup from tree bark?	Great	Much	Some	Little	No
5. Does the rich man have any legal right to store food when other people are starving?	Great	Much	Some	Little	No
6. Is the motive of Mustaq Singh to steal for himself or to steal for his family?	Great	Much	Some	Little	No
7. What values are going to be the basis for social cooperation?	Great	Much	Some	Little	No
8. Is the epitome of eating reconcilable with the culpability of stealing?	Great	Much	Some	Little	No
9. Does the rich man deserve to be robbed for being so greedy?	Great	Much	Some	Little	No
10. Isn't private property an institution to enable the rich to exploit the poor?	Great	Much	Some	Little	No
11. Would stealing bring about more total good for everybody concerned or wouldn't it?	Great	Much	Some	Little	No
12. Are laws getting in the way of the most basic claim of any member of a society?	Great	Much	Some	Little	No

Item 3. Consider the 12 issues above and rank which issues are the most important.

Most important item	1	2	3	4	5	6	7	8	9	10	11	12
Second most important	1	2	3	4	5	6	7	8	9	10	11	12
Third most important	1	2	3	4	5	6	7	8	9	10	11	12
Fourth most important	1	2	3	4	5	6	7	8	9	10	11	12

Descriptions of the Neo-Kohlbergian developmental schemas of moral judgment aligned with some issues from the example dilemma above (Bebeau & Thoma 2003; Rest et al. 1999b):

Schema	Kohlberg Stage(s)	Schema Foci	Aligned Issues
Personal Interest	Stages 2 and 3	“the direct advantages to the actor and on the fairness of simple exchanges of favor for favor” or “the good or evil intentions of the parties, on the party’s concern for maintaining friendships and good relationships, and maintaining approval.”	1. Is Mustaq Singh courageous enough to risk getting caught for stealing? 2. Isn’t it only natural for a loving father to care so much for his family that he would steal?
Maintaining Norms	Stage 4	“maintaining the existing legal system, maintaining existing roles and formal organizational structure.”	3. Shouldn't the community's laws be upheld? 5. Does the rich man have any legal right to store food when other people are starving?
Post Conventional	Stages 5 and 6	“organizing a society by appealing to consensus-producing procedures...insisting on due process...and safeguarding minimal basic rights” or “organizing social arrangements and relationships in terms of intuitively appealing ideals”	7. What values are going to be the basis for social cooperation? 11. Would stealing bring about more total good for everybody concerned or wouldn’t it? 12. Are laws getting in the way of the most basic claim of any member of a society?

Appendix 2. Student Opinion Scale (SOS) Survey (Sundre & Theik 2007)

Item	Item Text	Subscale
1	Doing well on these tests was important to me.	Importance
2	I engaged in good effort throughout these tests.	Effort
3*	I am not curious about how I did on these tests.	Importance
4*	I am not concerned about the scores I receive on these tests.	Importance
5	These were important tests to me.	Importance
6	I gave my best effort on these tests.	Effort
7*	While taking these tests, I could have worked harder on them.	Effort
8	I would like to know how well I did on these tests.	Importance
9*	I did not give these tests my full attention while completing them.	Effort
10	While taking these tests, I was able to persist to completion of the tasks.	Effort

* Denotes items that are reversed prior to scoring.