

## Objectives & Criteria

### Communication

- Access and Use Information Ethically and Legally
- Audience-Appropriate Approach & Structure
- Central Message
- Clarity of Peer Review
- Command of Interpretive Strategies
- Comprehension
- Content Development
- Context of and Purpose for Communication
- Control of Language, Syntax, and Mechanics
- Data Interpretation
- Data Representation
- Delivery
- Empathy
- Explanation of Issues
- Genre and Disciplinary Conventions
- Integrated Communication
- Organization
- Other\*-Communication
- Relationship to Text
- Uses Information Purposefully

### Critical Thinking

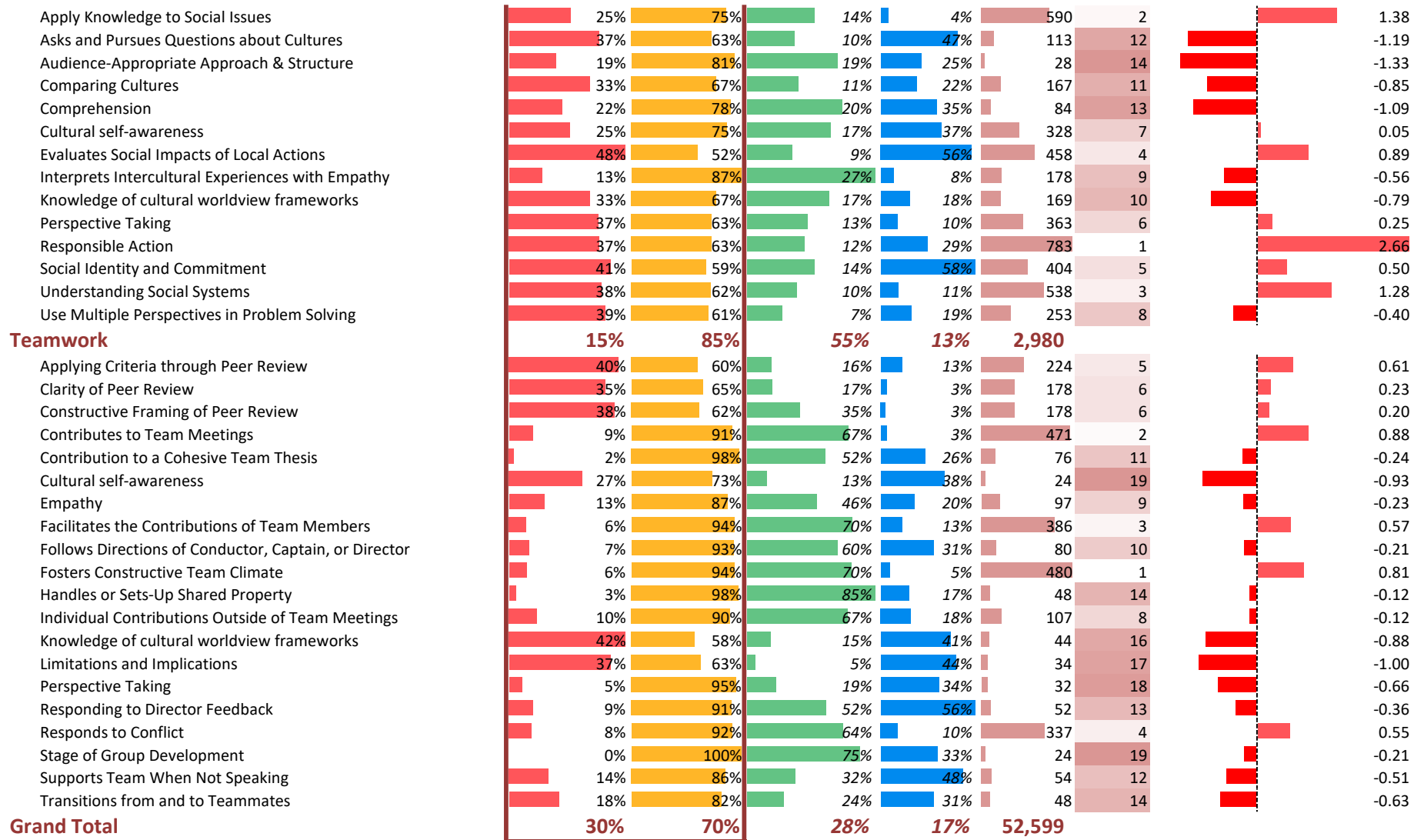
- Analysis of Knowledge
- Application / Analysis
- Application of Ethical Perspectives/Concepts
- Apply Disciplinary Knowledge
- Apply Knowledge to Social Issues
- Applying Disciplinary Methods
- Conclusions and Related Outcomes
- Define Problem
- Embracing Contradictions
- Ethical Issue Recognition
- Ethical Self-Awareness
- Evaluate Information and its Sources Critically
- Evaluate Outcomes
- Evaluate Potential Solutions
- Evidence Analysis
- Existing Knowledge, Research, and/or Views
- Identify Strategies

## Standard

## Count

	Not Met	Met	Exceeded	N/A*	N	Rank	DISAPPOINTMENT INDEX**
	<b>28%</b>	<b>72%</b>	<b>29%</b>	<b>13%</b>	<b>17,175</b>		
Access and Use Information Ethically and Legally	38%	62%	20%	4%	343	11	-0.65
Audience-Appropriate Approach & Structure	30%	70%	15%	2%	377	10	-0.59
Central Message	18%	82%	29%	11%	1529	5	0.64
Clarity of Peer Review	0%	100%	0%	95%	21	19	-0.90
Command of Interpretive Strategies	23%	77%	22%	33%	122	17	-0.80
Comprehension	28%	72%	30%	27%	2625	2	1.87
Content Development	38%	62%	19%	5%	1951	4	1.39
Context of and Purpose for Communication	26%	74%	29%	10%	387	9	-0.49
Control of Language, Syntax, and Mechanics	24%	76%	32%	4%	1439	6	0.57
Data Interpretation	31%	69%	40%	26%	1162	7	0.30
Data Representation	28%	72%	38%	11%	107	18	-0.73
Delivery	44%	56%	11%	96%	210	14	-0.93
Empathy	41%	59%	25%	65%	194	15	-0.83
Explanation of Issues	26%	74%	30%	7%	2674	1	1.87
Genre and Disciplinary Conventions	24%	76%	26%	0%	140	16	-0.75
Integrated Communication	12%	88%	40%	51%	255	13	-0.47
Organization	22%	78%	36%	3%	2495	3	1.50
Other*-Communication	0%	100%	50%	0%	2	20	-0.46
Relationship to Text	37%	63%	24%	1%	308	12	-0.67
Uses Information Purposefully	41%	59%	20%	12%	820	8	-0.05
	<b>34%</b>	<b>66%</b>	<b>26%</b>	<b>18%</b>	<b>19,148</b>		
Analysis of Knowledge	47%	53%	14%	55%	330	20	-0.72
Application / Analysis	36%	64%	28%	34%	481	16	-0.20
Application of Ethical Perspectives/Concepts	34%	66%	28%	21%	180	29	-0.94
Apply Disciplinary Knowledge	30%	70%	29%	8%	1345	3	1.82
Apply Knowledge to Social Issues	36%	64%	16%	15%	293	23	-0.74
Applying Disciplinary Methods	36%	64%	23%	25%	322	22	-0.62
Conclusions and Related Outcomes	38%	62%	32%	7%	1450	2	2.17
Define Problem	26%	74%	24%	14%	1097	6	1.26
Embracing Contradictions	64%	36%	9%	11%	279	25	-1.02
Ethical Issue Recognition	36%	64%	17%	5%	186	28	-1.03
Ethical Self-Awareness	49%	51%	14%	67%	105	31	-1.42
Evaluate Information and its Sources Critically	43%	57%	15%	31%	634	12	0.21
Evaluate Outcomes	27%	73%	31%	48%	597	14	0.08
Evaluate Potential Solutions	30%	70%	33%	31%	703	9	0.31
Evidence Analysis	30%	70%	35%	10%	1646	1	2.36
Existing Knowledge, Research, and/or Views	25%	75%	18%	17%	1318	4	1.85
Identify Strategies	33%	67%	28%	28%	609	13	0.11

Implement Solution	31%	69%	39%	10%	466	17	-0.21
Influence of context and assumptions	34%	66%	23%	20%	511	15	-0.13
Limitations and Implications	25%	75%	24%	50%	102	32	-1.08
Other*-CriticalThinking	44%	56%	26%	24%	82	33	-1.30
Propose Solutions/Hypotheses	25%	75%	45%	18%	890	7	0.60
Reflection and Self-Assessment	23%	77%	36%	13%	242	26	-0.65
Research Design	50%	50%	0%	89%	56	34	-1.75
Social Self-Awareness	28%	72%	24%	36%	132	30	-1.04
Source Use & Evaluation	30%	70%	20%	20%	1271	5	1.79
Student's position (perspective, thesis/hypothesis)	58%	42%	14%	10%	640	11	0.26
Textual Analysis	58%	42%	10%	18%	643	10	0.27
Topic selection	9%	91%	59%	11%	370	19	-0.22
Understanding Social Systems	41%	59%	19%	7%	288	24	-0.77
Use of Evidence	34%	66%	17%	11%	866	8	0.82
Uses Information Purposefully	29%	71%	26%	12%	326	21	-0.56
Student's Position	41%	59%	15%	13%	447	18	-0.34
Evaluate Outcomes of Attempted Solutions	25%	75%	50%	26%	234	27	-0.57
<b>Empirical/Quantitative</b>	<b>32%</b>	<b>68%</b>	<b>38%</b>	<b>14%</b>	<b>3,414</b>		
Application / Analysis	31%	69%	36%	12%	585	1	1.46
Assumptions	51%	49%	23%	8%	132	9	-1.00
Calculation	23%	77%	55%	6%	489	3	0.71
Define Problem	35%	65%	29%	52%	162	8	-0.67
Evaluate Outcomes	27%	73%	36%	47%	267	6	-0.08
Evaluate Potential Solutions	27%	73%	41%	8%	358	5	0.32
Evidence Analysis	24%	76%	37%	9%	407	4	0.54
Interpretation	40%	60%	33%	4%	585	1	1.65
Propose Solutions/Hypotheses	50%	50%	42%	14%	188	7	-0.53
Representation	24%	76%	35%	19%	120	10	-0.75
Research Design	63%	37%	11%	10%	51	12	-1.82
Textual Analysis	20%	80%	26%	34%	70	11	-1.03
<b>Personal Responsibility</b>	<b>31%</b>	<b>69%</b>	<b>23%</b>	<b>21%</b>	<b>5,426</b>		
Access and Use Information Ethically and Legally	35%	65%	28%	16%	643	3	0.91
Application of Ethical Perspectives/Concepts	42%	58%	20%	17%	634	4	0.99
Connect Concepts to Experience	28%	72%	15%	52%	659	2	1.05
Content Development	35%	65%	17%	3%	216	10	-1.25
Ethical Issue Recognition	26%	74%	17%	19%	429	7	-0.11
Ethical Self-Awareness	33%	67%	22%	53%	305	9	-0.74
Evaluate Information and its Sources Critically	32%	68%	30%	6%	478	6	0.12
Reflection and Self- Assessment	24%	76%	20%	26%	592	5	0.65
Transfer	31%	69%	13%	21%	211	11	-1.28
Understanding Different Ethical Perspectives/Concepts	42%	58%	25%	24%	341	8	-0.58
Use of Evidence	33%	67%	20%	10%	96	12	-1.81
Uses Information Purposefully	25%	75%	27%	4%	822	1	1.62
<b>Social Responsibility</b>	<b>33%</b>	<b>67%</b>	<b>13%</b>	<b>26%</b>	<b>4,456</b>		



\* N/A percentage represents share of raw totals. All success-oriented percentages are based on remaining artifacts not classified as N/A.

\*\* Disappointment Index takes the sum of Not Met and those not Exceeded for a criterion and multiplies it by the standardized score for its N, relative to the rest of its Core Objective.

In a nutshell, the Disappointment Index asks: Given how many faculty expect students to do this, how disappointed should we be by the results?

# EXPLANATORY NOTES ON THE FIRST FOUR YEARS TABLE

## BY THE NUMBERS

The table above reflects assessment of four years' worth of artifacts, 4,793 artifacts in all, representing 57,304 total ratings across all criteria, artifacts, and raters.

The ratings were carried out by 145 volunteer raters, a mix of tenure-track faculty, adjunct faculty, graduate students, and staff members, with participation also by a few alumni, administrators, and educators from nearby campuses who were curious how our approach worked. We even had two undergraduate students go through the training and help out. All artifacts were rated twice on each criterion to enable us to measure rater reliability (that is, how much we can expect two raters to give a piece of student work the same rating).

Student artifacts are recruited from faculty based on a stratified random sample of students across core classes, with the sample selecting roughly 8% of the students in those courses. During the period covered by our table above, 317 faculty submitted artifacts representing the work of 3,591 students.

When we ask faculty to provide artifacts for assessment, we also ask them what the assignment specifics were, and we ask faculty to identify which criteria, of those listed here, best pertain to the assignment in question. For this reason, some criteria are more frequently assessed than others. *Explanation of Issues* is very frequently selected (2,674 ratings in our data set) while relatively few faculty have asked us to rate *Stage of Group Development* (24 ratings in our data set).

As a general rule, the more frequently a criterion is assessed, the more robust our data for it is: rater reliability tends to be higher for frequently-assessed criteria than for rarely-assessed criteria, and those criteria more often match well with what students were asked to do, so the validity is often better as well. It should be noted that reliability on assessments will be lower than one sees for studies carried out under research conditions because teaching doesn't take place under research conditions. We're training a revolving door of raters to rate a wide variety of assignments across dozens of disciplines with six rubrics each encompassing ten to twenty criteria. The only way to ensure high reliability would be to rigidly control what and how faculty taught students, which runs us headlong into what we call the Schrödinger Cat problem: the danger that by trying to measure learning, we end up interfering with it. We don't want to do that, so we accept that low reliability is part-and-parcel of the assessment process and tread cautiously in our interpretations and responses. Boiled down, this means we look for longstanding patterns involving criteria that are frequently rated and which seem to have better-than-usual reliability and validity.

For a look at some typical reliability figures, as well as a look at how well our student demographics match up with the demographics of our sampling process, see the 2015 Pilot Report. We hope to follow up with such data for the more recent data, but the data that we're seeing now are pretty consistent with what we saw in that pilot report.

The remaining subsections each address elements of the table.

## OBJECTIVES

We have six core objectives, assigned by the state of Texas. These are **COMMUNICATION, CRITICAL THINKING, EMPIRICAL & QUANTITATIVE SKILLS, PERSONAL RESPONSIBILITY, SOCIAL RESPONSIBILITY, and TEAMWORK.**

Under each of the objective headings, we have listed a range of **CRITERIA** like *Comprehension*, *Organization*, and *Evidence Analysis*. Some criteria are aligned with more than one core objective. For instance, *Evidence Analysis* appears under both **CRITICAL THINKING** and **EMPIRICAL & QUANTITATIVE SKILLS**.

## NOT MET, MET, & EXCEEDED

Raters assign each artifact a score of 1, 2, or 3 for each criterion selected by the faculty member. In general, a 1 means the student didn't meet the standard on the rubric, a 3 means the student clearly met the standard, and a 2 represents some mixed or uneven result in between. "Not Met" indicates the percentage of rated students who received a 1 for a given criterion. "Met" indicates the percentage who received a 2 *or* 3. "Exceeded," a subset of "Met," identifies only the percentage who received a 3 rating.

## N/A

Sometimes the criterion selected by the faculty member doesn't really fit the assignment very well. When this happens, raters assign a rating of "N/A." Note that there is an important difference between two common scenarios:

Scenario 1: The faculty member assigned the student to do something, but the student didn't even attempt it. If this happens, the student receives a 1.

Scenario 2: Based on the assignment materials, there was no reasonable way the student could have anticipated that we would be rating their ability to perform the criterion that the faculty member ended up selecting, and so the lack of that performance doesn't mean that the student cannot perform such a task. (Example: Sometimes faculty selected the criterion *Delivery* for a written assignment, apparently missing that *Delivery* applies only to oral presentations. Therefore, we have no way to know whether the student could have performed this paper orally; they weren't asked to do so. In this sort of case, we assign a rating of N/A.

## COUNT AND RANK

"Count" refers to how many times we rated student artifacts on each criterion, including the N/A ratings. "Rank" refers to how popular<sup>1</sup> a criterion was within its objective, with a 1 indicating the criterion most often selected within that objective.

---

<sup>1</sup> By the term *popular* we only mean *frequently selected*, nothing more.

## DISAPPOINTMENT INDEX

The Disappointment Index is the least intuitive but probably most useful part of the table. One of the challenges with assessing learning across something as large as the core curriculum is that it's difficult to say *what to do* about the results. So we created a way to prioritize among criteria and identify patterns that might warrant more attention.

In a nutshell, the Disappointment Index answers the question, *Given how frequently faculty are selecting this criterion, how disappointed should we be in the results?*

Weak results on a criterion that only a few faculty have selected may not warrant as much attention as similar results on a more frequently selected criterion.

It should be noted that the index figures are always *relative to the other criteria under the same objective*.

That is, don't compare a criterion under one objective to a criterion under another objective.

Because Communication and Critical Thinking apply to every course in the core, while the other objectives only apply to slices of the core, all of the criteria under those two objectives have much higher Ns than you'll find under other objectives. So we compare Communication criteria only to other Communication criteria, not to Teamwork criteria.

How is the index calculated? We add together Not Met and those not Exceeded for a criterion and multiply that sum by the standardized score for the criterion's N, relative to the rest of its Core Objective. This formula weights 1s more heavily than 2s, but counts both as disappointments.

Among the criteria that seem to most warrant attention are

*Comprehension*

*Explanation of Issues*

*Conclusions & Related Outcomes*

*Evidence Analysis*

*Application/Analysis*

*Interpretation* (now called *Data Interpretation* to distinguish it better from *Comprehension*)

*Responsible Action*

Behind those seven criteria lies a consistent pattern, implicit here but more easily seen if you participate in the rating sessions: students are struggling a lot with comprehension, in several forms, and often with serious secondary effects. Comprehension issues often undermine attempts at analysis. Students struggle to explain things that they haven't fully understood. Students sometimes struggle to interpret data, tables, and graphs. Even *conclusions and related outcomes* often suffers from misunderstandings of the material that the student was attempting to draw conclusions *from*.

For this reason, our office plans to start a series of workshops and other small initiatives to see if we can improve the ways in which we're preparing students to understand the materials with which they have to work. ♦