Critical thinking is still critical
An institutional approach to assessing an enduring competency

Jennifer Hill, EdD
Evan Widney, MA
Alessandra Dinin, PhD

Office of Assessment
Trinity College at Duke University

SACSCOC Annual Meeting
December 2019

Image credit: Nasher Museum of Art
Background

Align general education and programmatic assessment of critical thinking
(Standard 8.2.a)

Describe our study of Critical Thinking in the general education
Administration • Instrumentation • Analysis • Understanding
(Standard 8.2.b)

Strategies to evaluate and deploy an instrument
Duke belongs to SACSCOC reaffirmation class of 2019.


Current 2019-2029 QEP focuses on excellence in undergraduate education in students’ first contacts with the field of study, especially those occurring in the first two years of college.

To understand how critical thinking is represented in undergraduate education at Duke, we need to look at the curriculum.
Bloom: Levels of cognitive development

Piaget: Stages of cognitive development

King & Kitchener: Theories of reflective judgment and reflective practice

AAC&U: Critical thinking VALUE rubric


## Crowd-sourcing a definition of critical thinking

<table>
<thead>
<tr>
<th>Option 1 (on the web)</th>
<th>Option 2 (by text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In a web browser, open</td>
<td>On your phone, text:</td>
</tr>
<tr>
<td>Type in your definition of critical thinking.</td>
<td>to number <strong>22333</strong></td>
</tr>
<tr>
<td></td>
<td>Then text your definition of critical thinking.</td>
</tr>
</tbody>
</table>

Admin page: [https://www.polleverywhere.com/free_text_polls/XhprPhvJex55heomPm7DJ](https://www.polleverywhere.com/free_text_polls/XhprPhvJex55heomPm7DJ)
Our operational definition of critical thinking largely is based on the VALUE definition and capstone levels. https://www.aacu.org/value/rubrics/critical-thinking

Critical thinking learning outcomes:

A. Issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.

B. Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis. Viewpoints of experts are questioned thoroughly.

C. Thoroughly (systematically and methodically) analyzes own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.

D. Specific position (perspective, thesis/hypothesis) takes into account the complexities of an issue. Others' points of view are acknowledged within position (perspective, thesis/hypothesis).

E. Conclusions and related outcomes (consequences and implications) are logical and reflect student’s informed evaluation and ability to place evidence and perspectives discussed in priority order.
# Critical Thinking VALUE Rubric

Critical thinking is a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.

**Definition**

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

<table>
<thead>
<tr>
<th>Capstone</th>
<th>Milestones</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

## Explanation of issues

- **Issue/problem to be considered critically is stated clearly and described comprehensively,** delivering all relevant information necessary for full understanding.
- **Issue/problem to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions.**
- **Issue/problem to be considered critically is stated but description leaves some terms undefined, ambiguities unexplored, boundaries undetermined, and/or backgrounds unknown.**
- **Issue/problem to be considered critically is stated without clarification or description.**

## Evidence

- **Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis.** Viewpoints of experts are questioned thoroughly.
- **Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis.** Viewpoints of experts are subject to questioning.
- **Information is taken from source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis.** Viewpoints of experts are taken as mostly fact, with little questioning.
- **Information is taken from source(s) without any interpretation/evaluation. Viewpoints of experts are taken as fact, without question.**

## Influence of context and assumptions

- **Thoroughly (systematically and methodically) analyzes own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.**
- **Identifies own and others' assumptions and several relevant contexts when presenting a position.**
- **Questions some assumptions. Identifies several relevant contexts when presenting a position. May be more aware of others' assumptions than one's own (or vice versa).**
- **Shows an emerging awareness of present assumptions (sometimes labels assertions as assumptions). Begins to identify some contexts when presenting a position.**

## Student's position (perspective, thesis/hypothesis)

- **Specific position (perspective, thesis/hypothesis) is imaginative, taking into account the complexities of an issue. Limits of position (perspective, thesis/hypothesis) are acknowledged. Others' points of view are acknowledged within position (perspective, thesis/hypothesis).**
- **Specific position (perspective, thesis/hypothesis) takes into account the complexities of an issue.**
- **Specific position (perspective, thesis/hypothesis) acknowledges different sides of an issue.**
- **Specific position (perspective, thesis/hypothesis) is stated, but is simplistic and obvious.**

## Conclusions and related outcomes (implications and consequences)

- **Conclusions and related outcomes (consequences and implications) are logical and reflect student's informed evaluation and ability to place evidence and perspectives discussed in priority order.**
- **Conclusion is logically tied to a range of information, including opposing viewpoints; related outcomes (consequences and implications) are identified clearly.**
- **Conclusion is logically tied to information (because information is chosen to fit the desired conclusion); some related outcomes (consequences and implications) are identified clearly.**
- **Conclusion is inconsistently tied to some of the information discussed; related outcomes (consequences and implications) are oversimplified.**
Arts, Literature, or Performance (2)

Civilizations (2)

Natural Sciences (2)

Quantitative Studies (2)

Social Sciences (2)

Cross-cultural Inquiry (2)

Ethical Inquiry (2)

Foreign Language (1-3)

Research (2)

Science, Technology, Society (2)

Writing (3)
Background

There’s great freedom for students to craft an authentic and purposive pathway.

BUT

There are challenges for advising and planning.

Students can face uncertainty.

It’s hard to study the impact of the curriculum.
Initial model of general education test administration

Web-based tests of
Ethical reasoning
Quant. Reasoning
Global Perspectives

Proctored test of critical thinking

Start year 1
Pre-matriculation

Start year 2
Major declaration

Start year 3

Start year 4

Graduation

Alumni
Instruments administered by the institution to assess learning across the curriculum

Nat. Science dept.
Lab reports, surveys

Humanities dept.
Recorded performances, capstone papers

Social Sci. dept.
Exams, portfolios, exit interviews

Other (e.g., Writing)
Reflection papers, writing samples

CAT = Critical thinking Assessment Test
DIT-2 = Defining Issues Test
GPI = Global Perspectives Inventory
QLRA = Quantitative Literacy and Reasoning Assessment

Helping general education assessment serve academic units
Using program assessment to illustrate Gen Ed learning outcomes

Trinity College student learning outcomes

Critical thinking
- Biology Program: Problem-solving exercise
- Literature Program: Honors thesis oral defense
- Neuroscience Program: Course evaluations
- Dance: Technique and performance rubric
- Computer Science: Regular online programming quizzes

Research, Inquiry, Analysis
- Cultural Anthropology: Senior capstone evaluation
- Documentary Studies: Student free-write exercise

Written communication
SECTION 8: Student Achievement

2. The institution identifies expected outcomes, assesses the extent to which it achieves these outcomes, and provides evidence of seeking improvement based on analysis of the results in the areas below:

a. Student learning outcomes for each of its educational programs.

b. Student learning outcomes for collegiate-level general education competencies of its undergraduate degree programs.

c. Academic and student services that support student success. [Not addressed in this session.]

Measures and data are used and shared between program-level and general education assessment.

Provide a superior liberal education to undergraduate students, attending not only to their intellectual growth but also to their development as adults committed to high ethical standards and full participation as leaders in their communities.

Advance the frontiers of knowledge and contribute boldly to the international community of scholarship.

Promote an intellectual environment built on a commitment to free and open inquiry.

Promote a deep appreciation for the range of human difference and potential, a sense of the obligations and rewards of citizenship, and a commitment to learning, freedom and truth.

| Percentage of Trinity College Programs manifesting this domain in at least one SLO |
|-------------------------------------------------|---------------------|---------------------|
| 2016-17                                        | 2017-18             |
| Civic engagement                               | 26%                 | 18%                 |
| Collaboration                                  | 24%                 | 24%                 |
| Content knowledge                              | 39%                 | 49%                 |
| Creativity                                     | 24%                 | 37%                 |
| Critical thinking                              | 80%                 | 82%                 |
| Engaging difference                            | 0%                  | 20%                 |
| Ethical reasoning                              | 28%                 | 33%                 |
| Foundations for lifelong learning              | 35%                 | 41%                 |
| Global learning                                | 31%                 | 35%                 |
| Information literacy                           | 26%                 | 14%                 |
| Integrative learning                           | 45%                 | 39%                 |
| Oral communication                             | 29%                 | 22%                 |
| Problem solving                                | 43%                 | 39%                 |
| Quantitative reasoning                         | 18%                 | 20%                 |
| Reading and text analysis                      | 35%                 | 51%                 |
| Research inquiry, and analysis                 | 73%                 | 69%                 |
| Visual analysis                                | 15%                 | 16%                 |
| Written communication                          | 47%                 | 55%                 |
Background

**Initial** model of general education test administration

- Pre-matriculation
- Web-based tests of
  - Ethical reasoning
  - Quant. Reasoning
  - Global Perspectives
- Proctored test of critical thinking
- Start year 1
- Start year 2
- Major declaration
- Start year 3
- Start year 4
- Graduation
- Alumni
Revised model of general education test administration

Web-based tests of
Ethical reasoning
Quant. Reasoning
Global Perspectives
Critical thinking

Pre-matriculation
Start year 1
Start year 2
Start year 3
Start year 4
Graduation
Alumni

Major declaration
• Which methodology serves us best?

• How could you begin to evaluate a methodology? What evaluative criteria could you take home today?

• Since this assessment work happens outside classrooms and programs, how do we align gen. ed. and program assessment? (Standards 8.2.a. and 8.2.b.)
**Brief critical thinking study timeline**

- **2008**: Initial CAT Training
- **2009**: First CAT “pilot” use with small program
- **2010**: First full CAT admin. to FY students
- **2011**: First CAT admin. to FY students continues annually through fall 2018
- **2012**: Regular reliability and validity checks
- **2013**: First CAT admin. to graduating seniors
- **2014**: First CAT reports to acad. depts.
- **2015**: First full CCTST admin. to FY students
- **2016**: CAT data reviews
- **2017**: CAT admin. to graduates continues annually through spr. 2018
- **2018**: Initial CCTST pilot, Full analysis of scores
- **2019**:
Paper & paper written test (not computer-mediated)

In-person proctored

15 questions spanning four constructs

Requires 45-90 minutes per student

We incentivized participation; different incentives for FYs and seniors

Scored by a team of faculty and/or graduate students

Required rater training and ongoing recalibration

Our capacity was approx. 300 students annually, splitting FY and senior
About the CAT

The Critical thinking Assessment Test (CAT)

www.tntech.edu / cat

Develops faculty and future faculty interested in assessment

We have a direct role in the evaluation of critical thinking

Questions represent real situations

Coursework and/or tests can be modeled after CAT questions

We have limited scoring capacity

High labor demand for a small office

We had trouble seeing movement from FY to senior year (Motivation?)

Faculty were unsure how to interpret and use the findings (Small Ns)
<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment</td>
<td>Voluntary participation</td>
</tr>
<tr>
<td></td>
<td>Mass email invitation to full cohort</td>
</tr>
<tr>
<td>Administration</td>
<td>Sign-up via web form (Qualtrics), with waitlist</td>
</tr>
<tr>
<td>Scoring</td>
<td>Small award to each participant</td>
</tr>
<tr>
<td>Findings</td>
<td>Larger drawing across all participants</td>
</tr>
<tr>
<td>Reporting</td>
<td>Approx. 150 FY students in the fall</td>
</tr>
<tr>
<td></td>
<td>Approx. 150 senior students in the spring</td>
</tr>
</tbody>
</table>
Recruitment

- On-campus room reservations
- Approx. 15 testing sessions, with 10-30 students each

Administration

- Mostly Sundays, with some evenings
- Test packets provided by TnTech; we provide consent and release forms

Scoring

Findings

- Electronic record keeping, including maintaining test and subject IDs

Reporting
Recruitment

Faculty raters preferred; we hired and paid graduate students

Administration

Weekend scoring sessions

Scoring

At our peak, a team of 10 veteran raters could score 150 tests in 8 hours (with breaks)

Findings

Regular review of rater reliability metrics

Reporting

Ongoing training, especially onboarding new raters
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment</td>
<td>Cohort average exceed the national mean for Research Extensive institutions</td>
</tr>
<tr>
<td>Administration</td>
<td>Changes from year 1 to year 4 are variable, ambiguous</td>
</tr>
<tr>
<td>Scoring</td>
<td>Low Ns due to our limited scoring capacity complicate analysis</td>
</tr>
<tr>
<td>Findings</td>
<td>Causation is impossible to determine. Students have highly variable academic pathways $\rightarrow$ many confounding factors</td>
</tr>
<tr>
<td>Reporting</td>
<td></td>
</tr>
</tbody>
</table>
Recruitment
Aggregate results reported to Trinity College leadership annually (narrative)

Administration
Aggregate results for students in each major/minor reported to the academic department (Tableau data dashboards)

Scoring
Results letters sent to student via email, with explanatory context and group benchmarks

Findings
We use pre-test results to support recruitment of seniors in year 4

Reporting
Aggregating graduating classes, it looks like there’s an increase!

But Ns for repeating students often are small due to our scoring capacity.

Usefulness declines when we try to sort results for individual departments.
## CAT scores, first-year to fourth-year

**Students affiliated with your major, minor, and/or certificate**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time 1, Q1:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarize figure</td>
<td>1.00</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>0.00</td>
<td>0.70</td>
</tr>
<tr>
<td>without making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inappropriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inferences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q1:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarize figure</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>without making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inappropriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inferences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 1, Q2:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determine how well</td>
<td>0.50</td>
<td>0.33</td>
<td>0.50</td>
<td>0.67</td>
<td>0.33</td>
<td>0.47</td>
</tr>
<tr>
<td>evidence supports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q2:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determine how well</td>
<td>0.78</td>
<td>0.73</td>
<td>0.67</td>
<td></td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td>evidence supports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 1, Q3:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide alternative</td>
<td>0.33</td>
<td>0.33</td>
<td>0.17</td>
<td>0.33</td>
<td>0.33</td>
<td>0.30</td>
</tr>
<tr>
<td>explanations for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q3:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide alternative</td>
<td>0.33</td>
<td>0.53</td>
<td>0.78</td>
<td></td>
<td></td>
<td>0.55</td>
</tr>
<tr>
<td>explanations for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 1, Q4:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify types of</td>
<td>0.25</td>
<td>0.25</td>
<td>0.13</td>
<td></td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>information to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evaluate hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q4:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify types of</td>
<td>0.25</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>information to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evaluate hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 1, Q5:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge whether data</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>support a hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q5:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge whether data</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>support a hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 1, Q6:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide alternative</td>
<td>0.50</td>
<td>0.83</td>
<td>0.33</td>
<td></td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>explanations for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q6:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide alternative</td>
<td>0.78</td>
<td>0.73</td>
<td>0.89</td>
<td></td>
<td></td>
<td>0.80</td>
</tr>
<tr>
<td>explanations for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 1, Q7:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify types of</td>
<td>0.25</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.20</td>
</tr>
<tr>
<td>information to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evaluate hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q7:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify types of</td>
<td>0.17</td>
<td>0.30</td>
<td>0.00</td>
<td></td>
<td></td>
<td>0.16</td>
</tr>
<tr>
<td>information to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evaluate hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 1, Q8:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge whether data</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>support a hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time 2, Q8:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge whether data</td>
<td>1.00</td>
<td>0.80</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.93</td>
</tr>
<tr>
<td>support a hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Graduating class 2016-17**

N: 5
Mean: 0.53
St. Dev.: 0.1826
We reached a point where we had to think seriously about trade-offs and limitations.

We have a professional responsibility to continue learning about and discussing ways to study core learning outcomes.
Exploring Alternative Instruments

Considerations

- Administrative time and labor
- Student engagement
- Reporting scores to departments
- Representative nature of questions across disciplines
- Maximum capacity
- Financial cost
- Connections to general education
- Reporting scores to students
- Involving faculty and graduate students
- Transparency
- Ease of use of results
- Load on students
- Consistency and reliability of scores
Exploring the CCTST

The California Critical Thinking Skills Test (CCTST)

Characteristics

Computer-mediated

Forty multiple choice questions spanning seven constructs

Requires 45-50 minutes per student

Scored automatically, electronically, without participation from faculty and instructors

We incentivized participation; different incentives for FYs and seniors
Recruitment

Voluntary participation

Administration

Email invitation to designated subset of cohort

Small award to each participant

Scoring

Larger drawing across all participants

Findings

Approx. 150 FY students in the pre-matriculation pilot. Sent summer before arriving to campus.

Reporting

Approx. 100 senior students in the spring pilot.
<table>
<thead>
<tr>
<th>Recruitment</th>
<th>Instruction for completion included in invitation email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administration</td>
<td>Administered online during timed session</td>
</tr>
<tr>
<td>Scoring</td>
<td>Non-proctored session – completed at students convenience</td>
</tr>
<tr>
<td>Findings</td>
<td>Electronic record keeping, including maintaining test and subject IDs</td>
</tr>
<tr>
<td>Reporting</td>
<td></td>
</tr>
</tbody>
</table>
Recruitment

- Scored electronically

Administration

- Results immediately available to students and test administrators
- Multiple choice questions remove test administrators from scoring process

Scoring

- Scores are calculated by construct – not by individual question

Findings

Reporting
<table>
<thead>
<tr>
<th>About the CCTST</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recruitment</strong></td>
</tr>
<tr>
<td><strong>Administration</strong></td>
</tr>
<tr>
<td><strong>Scoring</strong></td>
</tr>
<tr>
<td><strong>Findings</strong></td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
</tr>
</tbody>
</table>
Individual participant reports are provided to students at the completion of the test

Internal reporting tools for CCTST data are currently under development
We are interested in movement between first-year and senior year, including ceiling effects.

Scores are not the only consideration in instrument selection.
Exploring Alternative Instruments

General Education Assessment
Planning and Implementation

Ongoing Study of General Education - Outcome 1
Current Instrument

General Education - Outcome 1
Alternative Instruments Identified Through Other Sources

Develop Criterion For Evaluation of Instruments and Development of Decision Matrix

Give Initial Weight To Criterion

Analyze Pilot Results
Pilot Leading Alternative

Solicit Stakeholder Input to Gauge Criterion Weight and Score Instruments

Review Decision Matrix Results
Exploring Alternative Instruments

Review process

General Education Assessment
Planning and Implementation

Ongoing Study of General Education - Outcome 1
Current Instrument

General Education - Outcome 1
Alternative Instruments Identified Through Other Sources

Develop Criterion For Evaluation of Instruments and Development of Decision Matrix

Give Initial Weight To Criterion

Analyze Pilot Results

Pilot Leading Alternative

Solicit Stakeholder Input to Gauge Criterion Weight and Score Instruments

Review Decision Matrix Results
Exploring Alternative Instruments

Review process
Weighted Decision Matrix

Useful quantitative technique to help guide decision making

Helpful in evaluating a set of choices against a set of important criteria

Most helpful when faced with:
- Multiple options
- Multiple decision criteria
- Varying degrees importance among criteria

Helps remove emotion and guesswork from the decision making process
Exploring Alternative Instruments

General Education Assessment
Planning and Implementation

Ongoing Study of General Education - Outcome 1
Current Instrument

Develop Criterion For Evaluation of Instruments and Development of Decision Matrix

Give Initial Weight To Criterion

Analyze Pilot Results

Pilot Leading Alternative

Solicit Stakeholder Input to Gauge Criterion Weight and Score Instruments

Review Decision Matrix Results

General Education - Outcome 1
Alternative Instruments Identified Through Other Sources
Before continuing to the next page please refer to the CAT-CCTST comparison primer for information related to both instruments. This primer includes relevant information for each of the elements being judged. You may wish to leave the primer open for reference when responding.

Paper copies of the CAT are available for review and screenshots of the CCTST are available for review by clicking the following link: CCTST screen shots.
Exploring Alternative Instruments

Decision makers were asked to indicate degree of importance for each of the relevant criteria.

This guided the weighting process for the various criteria during the instrument evaluation period.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial cost</td>
<td></td>
</tr>
<tr>
<td>Maximum capacity</td>
<td></td>
</tr>
<tr>
<td>Reporting scores to departments</td>
<td></td>
</tr>
<tr>
<td>Consistency and reliability of test scores</td>
<td></td>
</tr>
<tr>
<td>Representative nature of questions across disciplines</td>
<td></td>
</tr>
<tr>
<td>Reporting scores to students</td>
<td></td>
</tr>
<tr>
<td>Student engagement</td>
<td></td>
</tr>
<tr>
<td>Load on student</td>
<td></td>
</tr>
<tr>
<td>Connection to general education and Duke curriculum</td>
<td></td>
</tr>
</tbody>
</table>
Exploring Alternative Instruments

No instrument is perfect.

Different institutional factors will influence the ideal solution.

We answer: How close is each option to the ideal solution for our institution?
Exploring Alternative Instruments

Review process

General Education Assessment
Planning and Implementation

Ongoing Study of General Education - Outcome 1
Current Instrument

General Education - Outcome 1
Alternative Instruments Identified Through Other Sources

Develop Criterion For Evaluation of Instruments and Development of Decision Matrix

Give Initial Weight To Criterion

Solicit Stakeholder Input to Gauge Criterion Weight and Score Instruments

Analyze Pilot Results

Pilot Leading Alternative

Review Decision Matrix Results
Exploring Alternative Instruments

Weighted Criterion Results

- Score - Post Survey Weighting
- Score - Original weighting
- Score - Unweighted

CCTST | CAT
We are interested in:

- Formalizing the decision making process
- Making sense of numerous decision making consideration
- Ongoing review of assessment instruments and institutional needs
Some conversation starters to bring home:

• Do we have any information on critical thinking? What does it tell us?

• How we are sharing evidence with academic and co-curricular partners? Is it working for them?

• What are the essential characteristics of an effective assessment strategy?

• How are we coming to consensus about the factors by which we create a strategy?

• Do we have a roadmap for the evaluation of measures?
What’s Next For Us?

- Review of other general education outcomes and instruments
  - Critical Thinking
  - Quantitative Literacy
  - Ethical Reasoning and Moral Development
  - Global Perspectives and Intercultural Competency
Critical thinking is still critical
An institutional approach to assessing an enduring competency

Jennifer Hill, EdD
Evan Widney, MA
Alessandra Dinin, PhD

Office of Assessment
Trinity College at Duke University

SACSCOC Annual Meeting
December 2019

Image credit: Nasher Museum of Art