An Evaluation of the Critical Engineering Literacy Test Instrument through Item Analysis and Comparison to the Critical Assessment Test

Ruth E. H. Wertz, Austin Saragih, Michael J. Fosmire, Şenay Purzer, and Amy S. Van Epps; Purdue University

Abstract

This paper reports reliability and validity measures for a two-tiered multiple choice instrument developed by the authors to assess information literacy skills in an engineering context. Classical test theory was used to describe item difficulty and item discrimination. Internal reliability was determined using the Kuder-Richardson KR-20. Content validity was assessed with a correlational analysis that explored the relationships between the CELT instrument and the validated Critical Assessment Test (CAT). This study was conducted in three first year courses ($N = 188$) in the Fall 2012 semester at Purdue University: engineering ($N = 72$), aviation technology ($N = 91$) and nursing ($N = 25$). Preliminary results indicate that overall, the CELT instrument has KR-20 of 0.67. Individual item analysis shows that 12 of the 18 items have sufficient item discrimination with discrimination scores greater than 0.15. In addition, for a sub-set of the population who took both the CELT and CAT instruments, there was a moderately strong association between the total scores ($r = 0.45$, $p < 0.05$, $N = 44$). The preliminary results indicate the CELT has good internal reliability for a multiple-choice instrument and appropriate levels of item difficulty. However, item discrimination results indicate that some individual items still need revision.

Background

Engineering students must possess the critical thinking skills necessary for solving complex problems in our knowledge-based society. However, assessing students' critical thinking ability has been a long-standing challenge for engineering educators. The authors previously developed a two-tiered multiple choice instrument, the Critical Engineering Literacy Test (CELT), to assess information literacy and critical thinking skills in an engineering context. Overall, the purpose of the CELT instrument is to provide an easy-to-administer assessment of engineering students' information literacy skills, which rely heavily on their ability to think critically. This study builds on the authors' previous work in the development of this instrument[1-2]. Specifically, this paper reports the item analysis, reliability analysis, and validity measures for the beta version of the CELT instrument.

Information Literacy and Critical Thinking

Information literacy, as it is most commonly defined, represents a person's ability to recognize the need for, effectively locate, and use information[3]. Our ability to examine students' demonstrated behaviors in an authentic context is a key challenge in information literacy assessment. Several existing information literacy assessments are limited in that they measure

self-reported data, or they are not contextualized. Assessments that do measure demonstrated skills in authentic or contextualized scenarios, such as the commonly used iSkills[4] instrument, have drawbacks as well, particularly in the time it takes to implement and evaluate the assessment. The objective of the CELT instrument is to provide a semi-authentic, contextualized information task with selected response test items that are easily scored for rapid feedback. The alpha version of the assessment consisted of one engineering-related scenario: a student team composed a memo to a University representative containing recommendations on ways to save energy in the dormitories, and an argument to support their recommendations. Students were required to read the short (one page) memo and respond to a series of ten multiple choice items. The internal reliability of the alpha version of the CELT instrument was poor, with a KR-20 of 0.39. This was not surprising given the range of information literacy skills targeted and the small number of items. To address the poor reliability, a second scenario was added to the assessment in the form of a letter to the editor regarding the public health and environmental concerns with the use of genetically engineered salmon versus traditional farm-raised salmon. The new scenario was accompanied by eight new selected response items, which included six multiple choice items and two select all that apply items. A summary of the blueprint of the revised CELT as it was used in this study is shown in Table 1. The original items were also revised based on preliminary item analysis[2].

Table 1. *CELT instrument blueprint*

Objective (Students can…)	Item number	Item Stem
1. identify implicit and explicit assumptions	1	Which one of the following is an assumption made in this memo/report?
2. identify and resolve conflicts between presented information and prior knowledge	2	Which one of the following statements is incorrectly presented as factual information?
3. accurately interpret information	3	According to the authors, what does the $2,760 represent?
4. evaluate the reliability of information and use reliable information sources	4	Which of the following pieces of information provided in the memo is likely the least reliable?
	16_a,	As with most drug-related FDA reviews, much of the data being evaluated was collected by the same company that produced these fish. What would help the review panel validate the data presented?
	17	Where would you likely find authoritative information on typical omega-3 levels of salmon?
5. accurately document the sources referenced	5	Which one of the citations is incorrect or incomplete?
	6	Which reference included in the text (in-text citation) is incorrect?
	12_a	Which of the following passage(s) from the text should be referenced (i.e., include a citation in the text)?
	18	In the following citation, what required component is missing

Table 1. *CELT instrument blueprint (continued)*

Objective (Students can…)	Item number	Item Stem
6. evaluate overall quality of a written document	7	Which of the below is a strength of this memo/report?
	8	Which of the below is a weakness of this memo?
	13	Which of the following is a weakness of this letter?
7. determine what information is needed to make a strong argument	9	What information is missing that would strengthen the memo?
	10	What information is least relevant to the memo's argument?
	11	Which of the following is most relevant in determining whether the GE fish is equivalent to its non-GE counterpart?
	14	Which of the following additional pieces of information would allow you to make the most informed decision on whether the GE salmon is a healthy fat source?
8. determine key words to locate information relevant to a specific topic	15	If you wanted to find out more information about…, which of the following searches would likely yield the best results?

Note: [a]Select all that apply items.

Several of the objectives identified for assessment of information literacy require critical thinking skills, specifically evaluating reliability of information sources, accuracy of documentation, and overall quality of a written document. In this study, the Critical Assessment Test (CAT), a well-known and valid instrument developed by Tennessee Technological University[5], was used to assess critical thinking, and as a comparison measure for the CELT.

Methods

The CELT instrument was disseminated to a total of 188 first-year students, 72 engineering students, 91 aviation technology students, and 25 nursing students. The instrument was administered online as part of regular course activities. Students could access the internet while completing the assessment if they chose. The CAT instrument was disseminated as a paper and pencil instrument. The test was administered to the all the students in the current study, however only preliminary results for a subset of first-year engineering students ($N=44$) is available at this time.

Item analysis was performed as a function of item difficulty (p-value), expressed as the proportion of students to select the correct response, and item discrimination (r_{pb}), expressed at the point-biserial correlation of an individual item to the CELT total score. Internal reliability of the instrument was measured using the Kuder-Richardson KR-20. In addition, a correlational analysis was performed for the subset of first-year engineering students for which we have both CELT and CAT assessment scores.

Results

For the purpose of this study, items 12 and 16 were excluded from the preliminary analysis using classical test theory since these are both 'select all that apply' items. The quantitative analysis reported is for the 16 multiple choice items. The mean score and standard deviation for these 16 items is presented in Table 2 for the engineering, aviation technology, and nursing students, as well as the overall aggregated group. The KR-20 of the instrument was 0.67 ($N = 188$).

Table 2. *Descriptive statistics for the CELT assessment*

	N	M	SD
Engineering	72	10.14	2.31
Aviation Tech	91	7.19	2.71
Nursing	25	10.16	2.56
All Students	188	8.71	2.94

Overall the item difficulty (p-value) ranged from 0.13 to 0.84, and the item discrimination coefficients (r_{pb}) ranged from 0.05 to 0.45. A summary of the p-value and r_{pb} for each item is shown in Table 3.

Table 3. *Item analysis for the CELT assessment*

	Engineering	*Aviation Tech*	*Nursing*	*Overall*
Item 1	0.30	0.47	0.46	0.45
	(0.64)	(0.43)	(0.64)	(0.54)
Item 2	0.36	0.14	0.13	0.16
	(0.11)	(0.14)	(0.16)	(0.13)
Item 3	0.20	0.17	0.24	0.22
	(0.47)	(0.40)	(0.68)	(0.46)
Item 4	0.19	-0.07	0.25	0.12
	(0.44)	(0.31)	(0.40)	(0.37)
Item 5	0.19	0.33	0.12	0.34
	(0.75)	(0.45)	(0.56)	(0.58)
Item 6	0.25	0.29	0.14	0.37
	(0.85)	(0.56)	(0.88)	(0.71)
Item 7	0.08	0.28	0.24	0.21
	(0.40)	(0.34)	(0.44)	(0.38)
Item 8	0.11	0.06	0.56	0.18
	(0.21)	(0.11)	(0.16)	(0.15)
Item 9	0.18	0.30	0.33	0.41
	(0.82)	(0.41)	(0.76)	(0.61)
Item 10	0.30	0.28	0.00	0.38
	(0.96)	(0.70)	(1.00)	(0.84)

(continued)

	Engineering	*Aviation Tech*	*Nursing*	*Overall*
Item 11	0.06	0.22	0.35	0.22
	(0.58)	(0.41)	(0.44)	(0.48)
Item 13	0.40	0.20	0.19	0.33
	(0.83)	(0.66)	(0.88)	(0.76)
Item 14	0.19	0.13	0.00	0.27
	(0.92)	(0.67)	(1.00)	(0.81)
Item 15	0.01	0.00	0.32	0.05
	(0.78)	(0.74)	(0.68)	(0.74)
Item 17	0.16	0.30	0.00	0.34
	(0.88)	(0.65)	(1.00)	(0.78)
Item 18	0.15	0.24	0.45	0.33
	(0.50)	(0.22)	(0.48)	(0.36)

Note: p-value is shown in parentheses below the r_{pb} for each item.

The preliminary results of the correlational analysis between the CELT and the CAT indicate that there is a positive association between the scores of the two instruments ($r = 0.47$, $p < 0.01$). A more detailed item to item analysis has been performed by the authors as part of a separate, but related study.[6] For the scope of this paper, the relationships (i.e., Pearson's correlation, r) of the CELT items and total score to the CAT total score are presented in Table 4.

Table 4. *Correlation matrix of CELT items and CAT Total Score*

CELT Item	CAT Total	CELT Item	CAT Total	CELT Item	CAT Total
Item 1	0.20	Item 7	0.12	Item 14	0.20
Item 2	0.17	Item 8	0.05	Item 15	-0.19
Item 3	0.08	Item 9	0.24	Item 17	**0.35***
Item 4	0.14	Item 10	0.22	Item 18	0.05
Item 5	**0.51***	Item 11	**0.32***	CELT Total	**0.47***
Item 6	0.21	Item 13	0.08		

Note. $N = 44$. [a]Multiple binary (e.g., select all that apply) items. Bold values represent significant relationships. *$p < .05$. **$p < .01$.

Discussion

In general, an ideal multiple choice instrument will have a range of p-values with the mean falling close to 0.5. For this instrument, the range of p-values was from 0.13 to 0.84 with a mean of 0.54. These results indicate that, overall, the level of difficulty is acceptable. Item discrimination (r_{pb}), however, is a more appropriate indicator of item quality. Overall, items 1, 5, 6, 9, 10, 13, 14, 17 and 18 had item discrimination coefficients greater than 0.25, which indicates that five of the eight objectives are represented by strong test items, including:

- identify implicit and explicit assumptions (item 1);
- evaluate the reliability of information and use reliable information sources (item 17);
- accurately document the sources referenced (items 5, 6, 18);

- evaluate overall quality of a written document (item 13); and
- determine what information is needed to make a strong argument (items 9, 10, 14).

In contrast, items 4 and 15 had overall item discrimination coefficients of less than 0.15, which indicates that these are poor items that do not discriminate between students who score well overall and those who do not. Item 4 was a relatively difficult item with 37 percent of the students selecting the correct response. The stem for this item is negative, using the words *least likely* instead of *most likely*, which may be the reason for its poor functionality. Item 15 was an easier item, with 74 percent of students selecting the correct answer. There were five possible choices for this item instead of four, with distracters attracting 7 to 26 percent of the responses. In addition to the poor item discrimination, item 15 was the only item to show a negative correlation with the CAT total score. The relationship was not signification at the 0.05 level, but the trend is enough to support that Item 15 is a poor item and should be revised or removed from future implementations of the CELT. Currenlty, item 15 is the only item satisfying the objective of determining key words to locate information relevant to a specific topic. Therefore, future implementation of the CELT should include additional items in this category to determine if the problem is localized to a specific item, or if this objective cannot be effectively measured with this type of assessment.

Our analysis also included the comparison between the different student groups within our sample. The population size, particularly of the nursing students (N = 25) is a limiting factor for strong statistical analyses. Two patterns emerged from the data, however, that are of particular interest. The first is the pattern of item difficulty and item discrimination for Item 2 across the three groups. Item 2 is what the authors called a *scientific literacy* question where students were required to use prior knowledge to identify data presented in the memo that was incorrect or unreasonable. This item was difficult for all three populations with 16 to 32 percent answering the item correctly. According to the item discrimination coefficients, however, the item was high functioning for engineering students (r_{pb} = 0.36) but low functioning for the aviation technology and nursing students (r_{pb} = 0.14 and r_{pb} = 0.13, respectively). This implies there may be an important contextual element that should be considered if this instrument is used outside of a general engineering context.

A similar phenomenon occurred for item 8, which also a difficult item (*p*-values range from 0.11 to 0.21). In this case the item was high functioning for nursing students (r_{pb} = 0.56) but low functioning for engineering and aviation technology students (r_{pb} = 0.11 and r_{pb} = 0.06, respectively). Item 8 requires students to identify a weakness of the associated written memo. There was a computational element to this item where the correct response was that the student authors made computational errors when presenting the data in the memo, however the numbers shown in the data table where arithmetically correct. The fact that the arithmetic was correct, even if the numbers being added and multiplied were misinterpreted appeared to be a stronger distracter for engineering and technology students than for the nursing students.

Implications for Practice and Future Work

Overall, the increase in the internal reliability KR-20 = 0.67 is good for a multiple choice instrument, particularly for one with less than 20 items. In addition, there were high functioning

items distributed throughout all but three of the objectives, and all but one objective had at least one acceptable item. Finally, the moderate association of the CELT total score with the CAT total score implies that while the instruments are measuring different constructs, there is a significant relationship between information literacy and critical thinking. This is particularly relevant to librarians and educators who are in a position to design and/or implement information literacy interventions. Future work will include more sophisticated item analysis and validity studies for further development of the CELT instrument, as well as more in-depth item to item analysis of the CELT and CAT assessments.

References

1.   Purzer, S., Fosmire, M.J., Wertz, R.E.H. & Yoon, S.Y. Development of the Critical Engineering Literacy Test. in *NARST Annual International Conference* (Indianapolis, IN, 2012).
2.   Wertz, R.E.H., Ross, M.C., Purzer, S., Fosmire, M. & Cardella, M.E. Assessing engineering students' information literacy skills: an alpha version of a multiple-choice instrument. in *2011 Annual American Society for Engineering Education Conference & Exposition* (Vancouver, BC, 2011).
3.   American Library Association. Information literacy competency standars for higher education (2000).
4.   Katz, I.R. Testing Information literacy in Digital Environments: ETS's iSkills Assessment. *Information Technology and Libraries* **26**, 3-12 (2007).
5.   Stein, B. & Haynes, A. Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test. *Change* **43**, 44-49 (2011).
6.   Wertz, R.E.H., Saragih, A., Van Epps, A.S., Sapp Nelson, M., Purzer, S., Fosmire, M.J., & Dillman, B. Work in Progress: Critical thinking and information literacy: Assessing student performance. in *2013 Annual Americal Society for Engineering Education Conference & Exposition* (Atlanta, GA, in review).