

Refining a Rubric for Evaluating Lifelong Learning and Career Awareness in a First-Year Learning Community

**Edmund Tsang, Betsy Aller, Troy Place,
Andrew Kline, Tracey Moon, Frank Severance,
Cynthia Halderson**

College of Engineering & Applied Sciences, Western Michigan University

2007 Frontiers In Education Conference, Milwaukee, WI

Overview

- Project Background
- Methodology
- Findings
- Recommendations & Conclusion

Project Background

- An evaluation rubric has been created to determine the effects of co-curricular activities on student development in lifelong learning & career awareness.
- This effort is part of the CEAS program to improve retention of first-time first-year students – Session F4A.
- The rubric is used to holistically evaluate students' written responses to co-curricular activities.

Project Background

- Evaluation Rubric is based on Bloom's learning taxonomy:
 - Level 1 (Knowledge): dis-/uninterested; rote responses or superficial report
 - Level 2 (Comprehension): express interest or acknowledge its value; report activities in their own words
 - Level 3 (Application and Analysis): connect with own career; recognize lack of prior knowledge; willing to apply information/experience
 - Level 4 (Synthesis and Evaluation): will seek additional information or opportunities; become champions for others

Project Background

- Researchers have identified key words or phrases associated with each of the four levels.
- Generally there was a strong agreement among the three researchers who scored the students' written responses.
- Co-curricular activities (professional, social, cultural) can be effective in student development (Level 1 = 10%; Level 2 = 67%; Level 3 = 22%; Level 4 = 1%).

Methodology

- Objectives of current research
 - Further refinement of evaluation rubric
 - Investigate inter-rater reliability with 3 additional new researchers
 - Does providing guidelines for written responses influences student performance?

Methodology

- Training the 3 new evaluators
 - Familiarize the new evaluators with rubric and scoring procedure
 - New evaluators practiced scoring with 15 sample essays from previous year
 - Roundtable discussion of the evaluators' rationale for their scores

Methodology

- All six evaluators scored two new sets of essays.
- One set of 32 essays – students were not provided with any guidelines
- One set of 32 essays – students were given guidelines consisting of headings (introduction, details, conclusion) to organize the writing
- Evaluators' scores were tabulated and analyzed.
- After discussion, a second round of evaluation for essays where scores differed by 1 point or more

Methodology

- Data Analyses
 - Frequency and point difference among evaluators
 - Average correlation among ratings
 - Intra-class correlation
 - T-test of ratings of papers without vs. with guidelines

Findings

- Initial scores of six raters

FREQUENCIES AND POINT DIFFERENCE AMONG RATERS

Without Guidelines			With Guidelines			Total
% Agree	N	Pt. Diff.	% Agree	N	Pt. Diff.	
100	8	0	100	5	0	13
83 (5 of 6)	8	1	83 (5 of 6)	8	1	16
67 (4 of 6)	7	1	67 (4 of 6)	11	10 = 1 1 = 2	18
50 (3 of 6)	9	8 = 1 1 = 2	50 (3 of 6)	8	4 = 1 4 = 2	17

Findings

- Based on roundtable discussion, revise criteria:
 - Level 1: Uninterested/Disinterested; unaware of value or relevance of activity; rote response (what/when/where/who)
 - Level 2: Express interest; acknowledge some value or relevance of activity; able to report in own words or critique
 - Level 3: See connection to career development/lifelong learning; recognize lack of prior knowledge/awareness; express willingness to apply information or experience
 - Level 4: Demonstrate interest to seek additional information; express willingness to seek additional opportunity; become a champion for others

Findings

- Results of 2nd round of scores

Frequency and Point Difference among Raters after Revision

Without Guidelines			With Guidelines			Total
% Agree	N	Pt. Diff.	% Agree	N	Pt. Diff.	
100	10	0	100	6	0	16
83 (5 of 6)	10	1	83 (5 of 6)	15	1	25
67 (4 of 6)	11	10 = 1 1 = 2	67 (4 of 6)	10	1	21
50 (3 of 6)	1	1	50 (3 of 6)	1	1	2

Findings

- 2nd Round of Scores

- Number of papers with 100% agreement increased from 13 to 16
- 83% agreement increased from 16 to 25
- 67% agreement increased from 18 to 21
- 50% agreement decreased from 17 to 2
- Number of papers where point difference was 2 decreased from 6 to 1
- Discussion and revised criteria led to greater agreement among the 6 raters

Findings

- Inter-rater correlation: Initial Score; Papers without Guidelines

	E1	E2	E3	E4	E5	E6
E1	1.00					
E2	0.84	1.00				
E3	0.75	0.71	1.00			
E4	0.71	0.61	0.67	1.00		
E5	0.74	0.78	0.70	0.65	1.00	
E6	0.73	0.69	0.68	0.63	0.74	1.00

- 6 of 15 correlations have intra-class correlation (ICC) < 0.70 → need stronger ICC if intent is for individual rater to evaluate consistently using rubric

Findings

- Inter-rater correlation: Revised Score; Papers without Guidelines

	E1	E2	E3	E4	E5	E6
E1	1.00					
E2	0.92	1.00				
E3	0.82	0.80	1.00			
E4	0.80	0.72	0.70	1.00		
E5	0.74	0.76	0.74	0.51	1.00	
E6	0.76	0.82	0.88	0.67	0.74	1.00

- 13 of 15 correlations have intra-class correlation (ICC) > 0.70 → discussion improves inter-rater reliability

Findings

- Comparison of aggregate scores by all 6 raters for papers without vs. with guidelines

<u>Without</u> guideline	Aggregate Mean	Standard Deviation	t	p-value
Before Discussion	1.97	0.64	-2.24	0.032*
After Discussion	2.05	0.67		

<u>With</u> guideline	Aggregate Mean	Standard Deviation	t	p-value
Before Discussion	1.98	0.56	-3.43	0.003*
After Discussion	2.08	0.61		

* Alpha is significant at $\alpha < 0.5$

Findings

- Comparison of revised scores for papers without vs. with guidelines

	Aggregate Mean	Standard Deviation	t	p-value
Without Guidelines	2.05	0.67	-0.23	0.82
With Guidelines	2.08	0.61		

- No statistical difference → guidelines for organization of paper do not improve a student's appreciation of co-curricular activity

Qualitative Findings

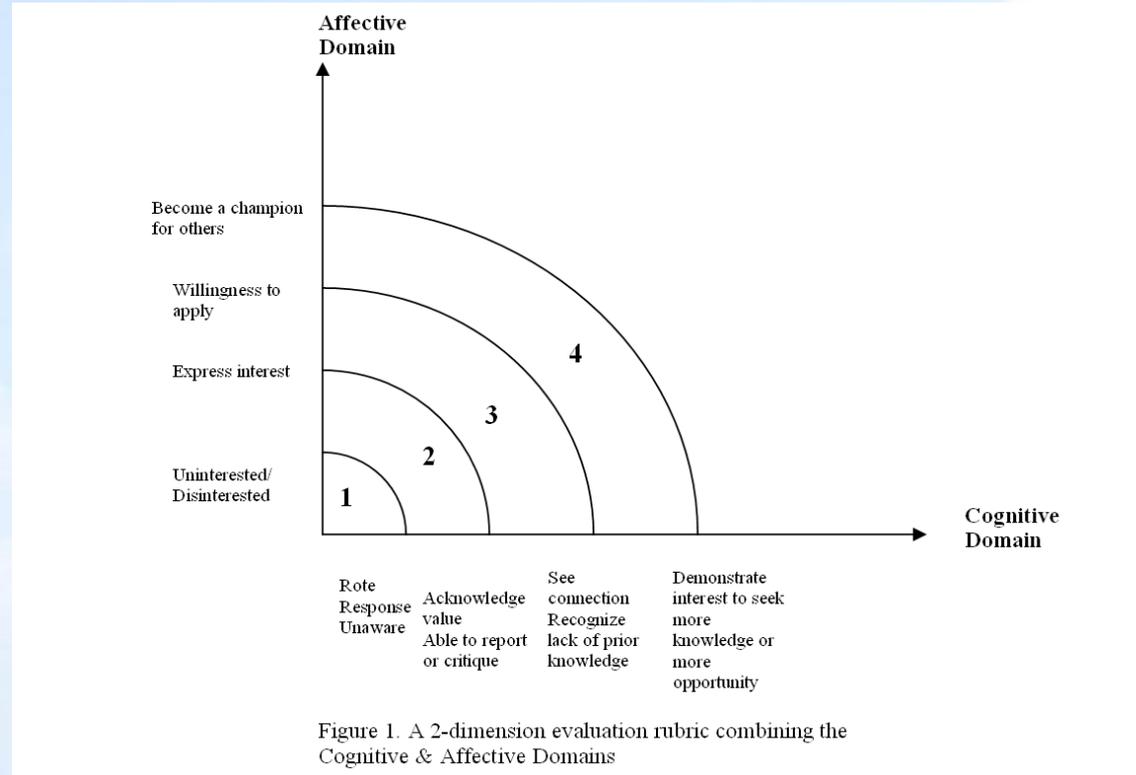
- Evaluator's familiarity with co-curricular activity may bias rating → interpret information as “listing of facts” rather than “express interest” or “apply information”
- Some of the evaluators were involved in teaching technical communication → need to look beyond writing mechanics or style to examine the information student tried to convey

Qualitative Findings

- The “Conclusion” section could be falsely interpreted as “comprehension,” “application of knowledge,” or “synthesis” when in reality students were wrapping up paper in a positive manner → research has not explored how to minimize this effect
- Artistic or cultural activities tended to result in Level 2 rating because students do not make connection → instructors may need to show how activities can enrich lives and social relations

Recommendation/Conclusion

- Revise rubric to combine the cognitive and affective domains



Recommendation/Conclusion

- Discussion among raters led to a better understanding of the rubric, improvement of the rubric, and a more consistent application of the rubric.
- The evaluator must be cognizant of his/her own bias when using the rubric for evaluation.
- Isolate an individual student's papers over the course of a semester to track his/her development.
- Add new researchers from other institutions

Acknowledgment

Partial support was provided by the National Science Foundation STEM Talent Expansion Program (STEP) under grant #0336581.