M R MULTIMEDIA RESEARCH
33 BROWNS LANE • BELLPORT, NY 11713 • (631) 286-8925

Impact of Museum of Science's Research Communication Laboratory Program on Research Presentation Skills of Graduate Students

Report to
Carol Lynn Alpert, Museum of Science

Submitted by
Barbara N. Flagg, Ed.D., Director, Multimedia Research
February 8, 2013

Abstract

Learning to design and deliver research information customized for particular audiences is one major goal of the Museum of Science's Research Communication Laboratory (RCL). Judging of short research presentations by an independent judge revealed that graduate students from MIT's Center for Excitonics who participated in RCL demonstrated significantly better spoken and graphic communication skills compared with graduate students who did not experience RCL instruction. The judge rated RCL students as significantly better than non-RCL students with respect to three criteria: 1) presentation overall; 2) efficacy of slides and graphics in enhancing understanding; and 3) poise, comportment and connection with audience. RCL students also scored better, but not significantly better, on two remaining criteria: 4) clarity of motivation for (and potential significance of) the research; and 5) clarity of what distinguished the chosen approach.

Introduction

The Research Communication Laboratory (RCL) program, implemented in the 2012-2013 academic year by the Museum of Science (MoS) at MIT's Center for Excitonics, is an intensive science communication skill-building program for a subset of the Center's graduate students. Learning to design and deliver information about research customized for particular audiences is one major goal of the RCL program. Progress toward achievement of this goal is reported here with data from judgments of students' research presentations.

Center for Excitonics is an Energy Frontier Research Center, directed by MIT Professor Marc Baldo and funded by the U.S. Dept. of Energy. On January 23, 2013, at an "All-Hands" retreat, graduate and post-doctoral students from the Center presented brief talks about their research to an audience of student and faculty colleagues. Among the presenters were two groups of graduate students who are the subject of this evaluation report: One group (n = 12) had been participating in the RCL program (Treatment group); the other group (n = 9) had not been participating in the RCL experience (Control group). This evaluation assesses the hypothesis:

- The treatment group compared with the control group will demonstrate better spoken and graphic communication skills as measured by an independent judge's assessment of their research presentations.

Procedure

Presentation Guidance. The Center's program administrator, Catherine Bourgeois, informed students via email to cover the following content for their presentations in a recommended five minutes with five or more slides. A PowerPoint template was supplied:

- Title of project, Authors/contributors;
- Motivation: Why is this project significant? E.g., does it address either of the two main challenges? 1) Understand & control the movement of excitons, 2) Understand & control the lifecycle of excitons. In addition, is it directly technologically relevant to solar energy conversion or solid state lighting? If so, explain.
- Approach: What is your experimental/theory approach to the problem? How does it differ from previous work?
- Key results: Present key results/progress to date
- Plans: Where is this project going? What do you expect to achieve by the next review in summer 2012?

The RCL instructor gave the treatment students further guidance: "Please use the outline for the talks provided by Catherine Bourgeois in her January 3rd email. However, remember that the template provided is just a guide: you may design your own slides and use any number of slides, as long as the talk stays within the 5 minute limit, and you hit all the points outlined in the template."[1]

---

[1] Jan. 15, 2013 email from Carol Lynn Alpert to RCL students

<u>Judging rubric.</u> RCL instructor Carol Lynn Alpert of MoS and evaluator Barbara Flagg of Multimedia Research developed an explicit set of criteria for judging the presentations – a rubric. The structure of the rubric reflects the RCL curriculum, the Center's presentation content outline described above, and criteria identified as important to Center faculty via a survey administered in May 2012. The rubric applied by judges to each presentation appears below:[2]

All-Hands Meeting: 5-Minute Presentation Assessment Sheet

Scorer Name _____

Presenter Name _____

Instructions:

- Assess the *presentation* of the research, not the research itself or the progress of the research.
- Your assessment of the clarity of each presentation should take into consideration that the presenters have only 5 minutes total to cover all the required segments.
- Circle one number on each 1 to 7 scale:

Needs considerable improvement = 1   Outstanding = 7

Clarity of motivation for (and potential significance of) the research	1 2 3 4 5 6 7
Clarity on what distinguishes the chosen approach	1 2 3 4 5 6 7
Efficacy of slides and graphics in enhancing understanding	1 2 3 4 5 6 7
Poise, comportment and connection with audience	1 2 3 4 5 6 7
Presentation overall	1 2 3 4 5 6 7

<u>Notes</u> (such as particular strengths/weaknesses, English as second language, etc.):

<u>Judges</u>. Four judges scored each presentation. Judges included two Center for Excitonics faculty (Marc Baldo, Karl Berggren) whose graduate students were among those presenting; the RCL instructor (Carol Lynn Alpert) who had worked with the treatment graduate students; and an independent judge (Janet Rankin) who did not know which students were participating in RCL and which were not. Dr. Rankin, Senior Associate Director for Teaching Initiatives at MIT, has an MIT Ph.D. in Materials Science and Engineering and as a member of MIT's Teaching and Learning Laboratory "works with faculty and departments to integrate efforts to promote better learning at MIT with departmental needs and constraints. Her interests include: active learning methods, improving learning in large-classes, interdisciplinary learning and teaching, and working with departments to support the professional development of TAs."[3]

<u>Additional data</u>. In addition to the scoring and notes of the judges, RCL assistant instructor Karine Thate recorded the presentation duration; the presence or absence of coverage of motivation, approach, results, and plans; and the number of slides presented for each of these four content categories.

---

[2] See Appendix for analysis of rubric inter-rater agreement.
[3] http://web.mit.edu/tll/about-tll/rankin.html

Results: Did presentations of treatment and control groups differ significantly?

- **Did content coverage differ between treatment and control groups?**

All 21 presenters covered each of the four expected content areas: motivation, approach, results, plans.

- **Did presentation length differ between treatment and control groups?**

Of the 21 graduate student presenters, only four (19%) managed to keep their presentation at the recommended five minutes; three of the four were in the treatment group. Table 1 presents the duration data for the treatment and control groups. There was no statistically significant difference between mean durations for the two groups.

Table 1. Presentation Duration

	Treatment (n = 12)	Control (n = 9)
Mean presentation duration	6.5 mins	7.8 mins
Range of duration	5 to 9 mins	5 to 10 mins

- **Did number of content slides differ between treatment and control groups?**

The Center's initial directions to the graduate students suggested a title slide and one slide for each of the four major content areas; however, RCL's instruction encouraged as many slides as necessary to enhance visual clarity and understanding within the five minutes allotted. Of the 21 graduate students, only one (5%) - from the control group - followed the recommendation of four content slides. Total number of content slides did not correlate with overall duration of presentation.

Table 2 presents the content slide data for the treatment and control groups. There was a statistically significant difference between mean number of total content slides for the two groups, with the treatment group utilizing a significantly larger mean number of slides in their presentations compared with the control group.[4] The treatment group also presented a significantly larger number of slides on average per minute of presentation.[5]

Table 2. Presentation Content Slides

	Treatment (n = 12)	Control (n = 9)
Mean number of total content slides	10.4 slides	6 slides
Range of total content slides	7 to 17 slides	4 to 8 slides
Mean content slides per minute of presentation	1.7 slides/min	0.8 slides/min

---

[4] Two-tailed t-test of means: $t(19)$ = -3.94, p = .0009. This is a parametric test to assess a null hypotheses that the means of two normally distributed independent populations are equal.
[5] Two-tailed t-test of means: $t(19)$ = -4.08, p = .0006.

4

Differences in slide numbers were analyzed for each of the four content areas. As indicated in Table 3, the treatment group presented a significantly larger mean number of slides for Motivation,[6] Approach,[7] and Results[8] but not for Plans.

Table 3. Mean Presentation Slides by Content Area

Content Area	Treatment (n = 12)	Control (n = 9)
Motivation	2.5 slides	1.6 slides
Approach	2.7 slides	1.6 slides
Results	3.9 slides	1.9 slides
Plans	1.3 slides	1.0 slides

---

[6] Two-tailed t-test of means: $t(19) = -2.39$, $p = .0274$
[7] Two-tailed t-test of means: $t(19) = -3.06$, $p = .0064$
[8] Two-tailed t-test of means: $t(19) = -3.03$, $p = .0069$

Results: Did rubric scores of treatment and control group presentations differ significantly?

- **Did presentation scores of the <u>independent</u> judge differ between groups?**

The independent judge scored each presentation on the five criteria of the rubric. Table 4 presents median scores rather than mean scores because the rubric response scale is an ordinal scale and scores are not normally distributed.

Table 4. Independent Judge's Median and Range of Presentation Scores

Rubric Criteria		Median and Range of Scores on Scale of 1 to 7	
		Treatment (n = 12)	Control (n = 9)
Clarity of motivation for (and potential significance of) the research	Median	6	5
	Range	2 to 7	4 to 6
Clarity on what distinguished the chosen approach	Median	4.4	3.4
	Range	1 to 6	1 to 6
Efficacy of slides and graphics in enhancing understanding	Median	6	5
	Range	5 to 7	4 to 6
Poise, comportment and connection with audience	Median	6	4
	Range	5 to 7	3 to 6
Presentation overall	Median	6	4
	Range	3 to 6	3 to 6

Statistical analyses of the five criteria reveal that the independent judge rated the treatment group as significantly better than the control group on three criteria:

Efficacy of slides and graphics in enhancing understanding,[9]
Poise, comportment and connection with audience, [10] and
Presentation overall.[11]

Clarity of motivation and clarity of approach were scored higher for RCL students but differences were not beyond chance. Presentation overall scores were highly correlated with the other four criteria. [12] The presentation overall scores of the independent judge were not correlated with presentation duration, total number of content slides or slides per minute. This judge's scores for efficacy of slides did not correlate with slides per minute of presentation; number of slides appeared to be less relevant than slide content.

---

[9] Mann-Whitney U = 20.5, p = .019. Because we have an ordinal scale, small samples, and non-normal score distributions, we use Mann-Whitney's U test to compare ordinal values of the two independent samples to determine if they reflect the presence of a real difference in the larger populations they represent.

[10] Mann-Whitney U = 13.5, p = .005

[11] Mann-Whitney U = 20, p = .017

[12] Clarity of Motivation (r_s = .78); Clarity of Approach (r_s = .77); Efficacy of Slides (r_s = .78); and Poise (r_s = .64). The nonparametric Spearman rank-order correlation coefficient, r_s , assesses the strength of association between two ranked variables.

6

- **Did presentation scores of the faculty judges differ between groups?**

Two Center for Excitonics faculty judged most of the presentations using the rubric. One judge was mentor for three treatment students and one control student; the other judge mentored one treatment student. Eliminating their own graduate students from their scoring data, we can look at the same issues as for the independent judge. Table 5 presents median scores for the two faculty judges. Like the independent judge, the presentation overall scores of the two faculty judges were highly correlated with the four other rubric criteria.[13]

Table 5. Center Faculty's Median and Range of Presentation Scores

Rubric Criteria		Median and Range of Scores on Scale of 1 to 7			
		Faculty Judge 1		Faculty Judge 2	
		Treatment (n = 9)	Control (n = 8)	Treatment (n = 9)	Control (n = 9)
Clarity of motivation for (and potential significance of) the research	Median	6	5	6	5
	Range	4 to 6	3 to 7	5 to 7	4 to 6
Clarity on what distinguished the chosen approach	Median	6	4.5	6	5
	Range	5 to 7	3 to 6	5 to 7	4 to 6
Efficacy of slides and graphics in enhancing understanding	Median	6	4.5	7	5
	Range	5 to 7	3 to 6	6 to 7	2 to 6
Poise, comportment and connection with audience	Median	6	4	6	3
	Range	5 to 6	4 to 6	5 to 7	2 to 7
Presentation overall	Median	6	4	6	5
	Range	4 to 7	3 to 6	5 to 7	4 to 6

Faculty judge #1 rated the treatment group as significantly better than the control group on four of the five criteria,[14] the exception being clarity of motivation. Faculty judge #2 rated the treatment group as significantly better than the control group on all five criteria.[15]

The overall presentation scores of the two faculty judges were not correlated with duration of presentation but were highly correlated with total number of slides.[16] Higher presentation scores tended to accompany higher slide numbers but note that high slide numbers do not guarantee or cause high overall presentation scores.

---

[13] Clarity of Motivation (Judge 1 r_s = .87; Judge 2 r_s = .75); Clarity of Approach (r_s = .85; r_s = .76); Efficacy of Slides (r_s = .85; r_s = .68); and Poise (r_s = .74; r_s = .84).
[14] Clarity on Approach: Mann-Whitney U = 20, p = .009; Efficacy of Slides: Mann-Whitney U = 10, p = .009; Poise: Mann-Whitney U = 7.5, p = .003; Presentation Overall: Mann-Whitney U = 10, p = .031.
[15] Clarity of Motivation: Mann-Whitney U = 8.5, p = .003; Clarity on Approach: Mann-Whitney U = 9.5, p = .004; Efficacy of Slides: Mann-Whitney U = 1.5, p = .0007; Poise: Mann-Whitney U = 10.5, p = .007; Presentation Overall: Mann-Whitney U = 5.5, p = .001.
[16] Judge 1: r_s = .62; Judge 2: r_s = .81

- **Did presentation scores of the <u>RCL instructor</u> judge differ between groups?**

Table 6 presents median scores for the RCL judge who was instructor for students in the treatment group. The RCL judge rated the treatment group as significantly better than the control group on all five criteria.[17] As with the other judges, the presentation overall scores of the RCL instructor were highly correlated with the four other rubric criteria.[18] The overall presentation scores of the RCL instructor judge were not correlated with duration of presentation but were highly correlated with total number of content slides.[19]

Table 6. RCL Instructor's Median and Range of Presentation Scores

Rubric Criteria		Median and Range of Scores on Scale of 1 to 7	
		Treatment (n = 12)	Control (n = 9)
Clarity of motivation for (and potential significance of) the research	Median	7	5
	Range	5 to 7	3 to 6
Clarity on what distinguished the chosen approach	Median	7	5
	Range	5 to 7	2 to 6
Efficacy of slides and graphics in enhancing understanding	Median	6	2
	Range	4 to 7	2 to 3
Poise, comportment and connection with audience	Median	6	3
	Range	3 to 7	2 to 6
Presentation overall	Median	6.5	4
	Range	5 to 7	2 to 5

---

[17] Clarity of Motivation: Mann-Whitney U = 19.5, p = .009; Clarity on approach: Mann-Whitney U = 10, p = .001; Efficacy of slides: Mann-Whitney U = 0, $p \leq$ .0001; Poise: Mann-Whitney U = 9, p = .001; Presentation overall: Mann-Whitney U = 3, p = .0002.
[18] Clarity of Motivation (r_s = .76); Clarity of Approach (r_s = .88); Efficacy of Slides (r_s = .90); and Poise (r_s = .93).
[19] r_s = .73

8

Results: How did judges' freehand notes reflect the scoring criteria?

The rubric permitted judges to comment on the presentations beyond the 7-point scoring system. These notes were coded according to mention of the four individual criteria in order to explore which criteria were most salient for the different judges. The randomly chosen illustrative comment that follows gives an example of the coding procedure:

Motivation for work is that all models require a spectral density…a very theoretical motivation. [clarity of motivation] *To distinguish current work - stated that they wanted to be able to account for many degrees of freedom - was clear about how the work was different.* [clarity of approach] *Would have been beneficial to have this articulated on a slide.* [efficacy of slides]

Table 7 presents the frequency with which judges' optional notes made mention of each of the four scoring criteria. The criteria of "poise, comportment and connection with audience" appeared most salient across all four judges. The independent judge focused comments mostly on clarity of content (i.e., motivation and approach); whereas the RCL judge focused comments mostly on slides and poise.

Table 7. Frequency of Criteria Referenced in Judges' Notes

Criteria	Judges			
	Independent (N =21)	Faculty 1 (N =21)	Faculty 2 (N = 19)	RCL (N = 21)
Clarity of motivation for (and potential significance of) the research	95%	33%	16%	10%
Clarity on what distinguished the chosen approach	90%	5%	0%	0%
Efficacy of slides and graphics in enhancing understanding	52%	5%	37%	71%
Poise, comportment and connection with audience	81%	43%	53%	76%

Conclusion

In the opinion of four different judges, the graduate students who participated in the Research Communication Laboratory experience demonstrated significantly better spoken and graphic communication skills in their short research presentations compared with graduate students who were not enrolled in the RCL. With respect to the judging criteria, RCL students were strongest in their use of slides and graphics and in their poise and connection with the audience. Further clarifying motivation and approach for their research will add to the effectiveness of future presentations.

APPENDIX

Rubric Inter-Rater Agreement

The scoring rubric was developed specifically for the purpose of this evaluation, and implementation of the rubric did not entail any training or discussion of criteria among the raters, possibly resulting in low inter-rater agreement. Rater agreement in the case of this rubric has two components: (1) judges may differ in their interpretation of the five criteria; and/or (2) judges may differ in their application of the scoring levels of 1 to 7.

To estimate rater agreement for the first component, we look at each individual judge's scores as they correlate with the group score average for each criteria (rater-group rank-order correlations). Table 8 presents a matrix of correlations for the four judges for the five criteria for 14 students whom all judges rated. A value of 1.0 is a perfect correlation; all correlations in Table 8 are statistically significant at $p \leq .01$. The four judges appear to interpret the five criteria with high consistency.

Table 8. Judge vs. Group Correlations (r_s) for Five Criteria

Criteria	Judges			
	Independent	Faculty 1	Faculty 2	RCL
Clarity of motivation for (and potential significance of) the research	.59	.61	.83	.70
Clarity on what distinguished the chosen approach	.75	.58	.89	.77
Efficacy of slides and graphics in enhancing understanding	.76	.79	.95	.87
Poise, comportment and connection with audience	.94	.93	.91	.86
Presentation overall	.68	.81	.86	.87

For the second component of usage of the scoring levels, we can look at Figure 1, which shows the proportion of each judge's scores distributed across the seven scoring levels for 14 students whom all judges rated. The independent judge and faculty judge 1 show similar positively skewed distributions, with Faculty Judge 2 showing more variation but still a positive skewness. In contrast, the RCL instructor displays almost a bimodal distribution of scores.

Fig. 1. Distribution of Judges' Rating Levels