

EVALUATION REPORT

Innovative Instructional Technology Award Program (IITAP) Evaluation

June 30, 2006

Tomoko Traphagan PhD
tomoko.traphagan@austin.utexas.edu



Division of Instructional Innovation and Assessment

research, evaluation and assessment section

THE UNIVERSITY OF TEXAS AT AUSTIN

<http://www.utexas.edu/academic/diia>

P.O. Box 7246
2616 Wichita Street
Austin, TX 78713-7246

Contents

Executive Summary	4
Full Report	5
Background	6
Program Description	6
Program Facts	9
Evaluation Description	11
Research Questions and Evaluation Plan	11
Purpose of evaluation	11
Research questions	11
Data Collection and Analysis Method	13
Findings and Conclusions	14
Question 1.1: Is the First-Round Review Process Efficient?	14
Time spent for each entry	15
Coordination with work schedule	16
Efficiencies and strategies	17
Online review submission system	19
Time-consuming process and difficult process	20
Suggestions for improving the reviewing process	21
Summary of findings for question 1.1	21
Question 1.2: Is the First-Round Review Accurate?	23
Qualification	23
IITAP materials	24

Suggestions for improving the reviewing accuracy	26
Reviewer orientation	26
Consistency of reviewer scorings across reviewers	27
Consistency of the reviewer scorings across review questions	34
Consistency of the review scoring	39
with (external) judge scoring	39
Summary of findings for Question 1.2	40
Question 1.3: Is the Second-Round Judging Process Efficient?	42
Time spent for each entry	42
Efficiencies and strategies	42
IITAP materials	42
Online judge score submission system	43
Difficult process	44
Suggestions for improving the judging process	44
Summary of findings for Question 1.3	44
Recommendations	45
Appendices	49
Appendix A Teaching with Technology Detailed Review Instructions	49
Appendix B Resource Development Review: Detailed Instructions	52
Appendix C Program Staff Interview Protocol	56
Appendix D IITAP Reviewer Survey	57
Appendix E IITAP Judge Survey	62
Appendix F IITAP Reviewer Survey: Responses to the Open-Ended Questions	65

Executive Summary

The annual Innovative Instructional Technology Awards Program (IITAP) is conducted to encourage, recognize, and reward faculty members' efforts to incorporate technology in their teaching. In 2006, one gold, one silver, and one bronze winner were selected in each of two categories—Teaching with Technology and Resource Development—from 21 entrants, using a first-round selection process with eight internal reviewers and a second-round selection process with six external judges.

Key Findings

- ◆ The reviewers spent 2.5 hours or less and the judges spent 1 to 1.5 hours for each entry, on average. For both the reviewers and judges, examining the entries' products and supplementary materials was time consuming. (pp.15-16, 20, & 42)
- ◆ Finding enough time was the most difficult part of the task for the reviewers. (pp.16-17)
- ◆ Both reviewers and judges reported that instructions for their tasks were clear. (pp. 24-25, 42-43)
- ◆ Reviewers reported that the orientation was helpful and suggested providing hands-on rating training. (pp. 26-27)
- ◆ The scoring consistency among reviewers was poor. (pp.27-34)
- ◆ The outcome (assessment) questions were a locus of disagreement in reviewer scoring and received low scores. (pp.34-39)
- ◆ The internal reviewers' rankings were substantially different from the external judges' rankings. (pp.39-40)
- ◆ The judges mentioned that the most difficult part of the judging process was comparing entries from very different fields. (p. 44)
- ◆ The reviewers reported the online score submission system was easy to use but suggested some improvements. (pp.19-20)

Key Recommendations (pp.45-8)

- ◆ Maintain a recommendation of 2.5 hours to review each entry for reviewers. Recommend judges spend 1.5 hours to review each entry.
- ◆ Consider changes to help reviewers complete their task more easily, such as assigning fewer entries per reviewer by increasing the number of reviewers.
- ◆ Enrich the reviewer orientation with hands-on training, where reviewers practice scoring exemplar entries individually with follow-up group discussions until they come to agreement within 10 point differences.
- ◆ Establish the same review criteria using example entries for both reviewers and judges.
- ◆ Assist entrants more in preparing responses in the "outcome (assessment)" area.
- ◆ Consider ways for judges to communicate after they complete their scoring.
- ◆ Improve the online score submission system by (a) extending the periods before timeouts, (b) providing a page to input/change all scores across entries, (c) providing a page to view all entrants' answers to the same questions, and (d) allowing users to move from one rating page to another directly.

FULL REPORT

Running Head: IITAP EVALUATION

Innovative Instructional Technology Award

Program (IITAP) Evaluation

Full Report

Tomoko Traphagan, Ph.D.

Division of Instructional Innovation and Assessment

The University of Texas at Austin

Background

Program Description

The annual Innovative Instructional Technology Awards Program (IITAP) is conducted to encourage, recognize, and reward faculty members' efforts to incorporate technology in their teaching. The awards are offered by the Office of the Provost and administered by the Division of Instructional Innovation and Assessment (DIIA). In 2005-06, IITAP was in its ninth year.

Faculty members and faculty-student/staff teams at the University of Texas at Austin (UT Austin) are annually invited to enter for judging innovative instructional materials using emerging technologies that they have developed for a primary audience of UT Austin students. Both award categories, Teaching with Technology (TWT) and Resource Development (RD), celebrate development of innovative instructional materials that promote students' knowledge and skills and make teaching and learning more effective and efficient. The TWT awards focus on the development of instructional materials that can be directly used for instruction, while the RD awards recognize the development of digital content resources. (<http://www.utexas.edu/academic/cit/services/incentive/iitap/>)

After UT Austin faculty members submitted entry forms to DIIA in the fall of 2005, eight internal reviewers (DIIA staff members) selected five TWT and five RD entries to recommend for the second-round, final judging. The reviewers consisted of three specialists in instruction, three specialists in technology, and two specialists in evaluation. Each reviewed 10 to 11 entries, half for TWT entries and half for RD. Each

entry was evaluated by four reviewers, a group that included at least one specialist from each of the areas of technology, instruction, and evaluation. The TWT entries were evaluated based on 15 criteria and the RD entries were evaluated based on 13 criteria (see Tables 1 and 2). The entries received whole number ratings between -10 and 50 for each criterion:

Missing = -10
Not applicable = NA
Very poor = 10
Poor = 20
Satisfactory = 30
Very good = 40
Excellent = 50

Reviewers were instructed to plan on spending an average of 2.5 hours reviewing each entry, for a total of 25 hours. For details about the review procedures, including the scoring guide, see Appendices A and B. The reviewers submitted scores and comments via an online score submission system.

For second-round evaluation, six judges—four from within UT Austin, including an IITAP 2004 gold winner and two nationally-recognized judges from outside UT Austin—considered the five entries for each category. Program staff compiled the six judges' evaluations to determine the winners. The same set of judging criteria was used for both the first-round review process and the second-round judging process.

Table 1: Evaluation Criteria for the Teaching with Technology Category

INNOVATION, EFFECTIVENESS, RELEVANCE, AND EFFICIENCY AS A TEACHING AND/OR LEARNING PRACTICE, PROCESS, OR TECHNOLOGY	
2.	The teaching with technology practice provides an innovative solution to the stated teaching and/or learning problem and offers clear advantages over traditional techniques. New techniques, unique tools and methods are used to develop this teaching with technology practice. Refer to the Development Process question and Question 1 (What is the teaching and/or learning problem that you were trying to address?).
3.	Students gain new knowledge and skills using the teaching with technology practice and are now able to do something they could not do before.
4.	The teaching with technology practice makes the learning experience and/or subject matter relevant to the students.
5.	The teaching with technology practice increases teaching and learning efficiency for the learner, for the instructor, for the institution.
OUTCOMES FOR TEACHING WITH TECHNOLOGY	
6.	The assessment plan and design described in the entry form demonstrates that appropriate, valid, and thorough processes, instruments, and techniques were used to measure the effectiveness, relevance, and efficiency of the teaching with technology practice on teaching and/or learning.
7.	The evaluation outcomes confirm the answer to Question 3 (What are the knowledge and/or skills gained by the students using your practice, process, or technology, and what can your students now do that they could not do before?) in the entry form and demonstrate that students have gained knowledge and skills and are now able to do something they couldn't do before. Supporting qualitative or quantitative data is provided.
8.	The evaluation outcomes confirm the answer to Question 4 (How does your practice, process, or technology make the learning experience and/or subject matter relevant to your students?) in the entry form and demonstrate that the teaching with technology practice makes the learning experience and/or subject matter relevant to the students. Supporting qualitative or quantitative data is provided.
9.	The evaluation outcomes confirm the answer to Question 5 (How does your practice, process, or technology increase efficiency for the learner, for the instructor, for the institution?) in the entry form and demonstrate that the teaching with technology practice increases efficiency for the learner, for the instructor, for the institution. Supporting qualitative or quantitative data is provided.
FUNCTIONALITY, DESIGN, INTERACTIVITY, AND QUALITY	
10.	Technical operation (i.e. runs smoothly without crashes or hang-ups, easy to install)
11.	User interface design (i.e. clear navigation and usability; consistent, intuitive, clean, aesthetically pleasing)
12.	Engagement and interactivity (i.e. Are students now able to be more engaged with learning through interactivity? Are students engaged beyond page turning?)
13.	Content presentation (i.e. Is content clear, readable, well-organized, well-written, rich and deep—supporting linkages and the building of customized, relevant knowledge?)
14.	Media quality (high-quality graphics, video, sound, animations), if applicable
15.	Learner modalities are accommodated for (i.e. Are appropriate types of content and media included to create a rich practice, process, technology?)

Table 2: Evaluation Criteria for the Resource Development Category

INNOVATION, EFFECTIVENESS, AND EFFICIENCY AS A RESOURCE	
2.	The resource provides an innovative solution to the stated teaching and/or learning problem and offers clear advantages over traditional techniques. New techniques, unique tools and methods are used to develop this resource. Refer to the Development Process question and Question 1 (What is the teaching and/or learning problem that you were trying to address?).
3.	Students gain new knowledge and skills using the resource and are now able to do something they could not do before.
4.	The resource increases teaching and learning efficiency for the learner, for the instructor, for the institution.
OUTCOMES FOR RESOURCE DEVELOPMENT	
5.	The assessment plan and design described in the entry form demonstrates that appropriate, valid, and thorough processes, instruments, and techniques were used to measure the effectiveness and efficiency of the resource on teaching and/or learning.
6.	The evaluation outcomes confirm the answer to Question 3 (What are the knowledge and/or skill(s) gained by the students using this resource, and what can your students now do what they could not do before?) in the entry form and demonstrate that students have gained knowledge and/or skill(s) and are now able to do something they couldn't do before. Supporting qualitative or quantitative data is provided.
7.	The evaluation outcomes confirm the answer to Question 4 (How does this resource increase efficiency for the learner, for the instructor, for the institution?) in the entry form and demonstrate that the resource increases efficiency for the learner, for the instructor, for the institution. Supporting qualitative or quantitative data is provided.
FUNCTIONALITY, DESIGN, INTERACTIVITY, AND QUALITY	
8.	Technical operation (i.e. runs smoothly without crashes or hang-ups, easy to install)
9.	User interface design (i.e. clear navigation and usability; consistent, intuitive, clean, aesthetically pleasing)
10.	Engagement and interactivity (i.e. Are students now able to be more engaged with learning through interactivity? Are students engaged beyond page turning?)
11.	Content presentation (i.e. Is content clear, readable, well-organized, and well-written?)
12.	Media quality (high-quality graphics, video, sound, animations), if applicable
13.	Learner modalities are accommodated for (i.e. Are appropriate types of content and media included to create a rich resource?)

Program Facts

- 21 entrants—12 TWT and 9 RD
- Entrants from 16 areas: Anthropology, Architecture, Biomedical Engineering, Communication Studies, Educational Psychology, English, Germanic Studies, Government, Kinesiology and Health Education, Liberal Arts ITS, Marketing,

McCombs School of Business, Mechanical Engineering, Nuclear Engineering
Teaching Lab, Pharmacy, and Social Work

- Eight internal reviewers (DIIA staff members)
- Six judges (four UT Austin faculty members and two external judges)
- One gold winner, one silver winner, and one bronze winner for each category, TWT and RD
- A total of \$8,500 awarded (\$2,000 for gold winners, \$1,500 for silver winners, and \$750 for bronze winners)
- Program timeline

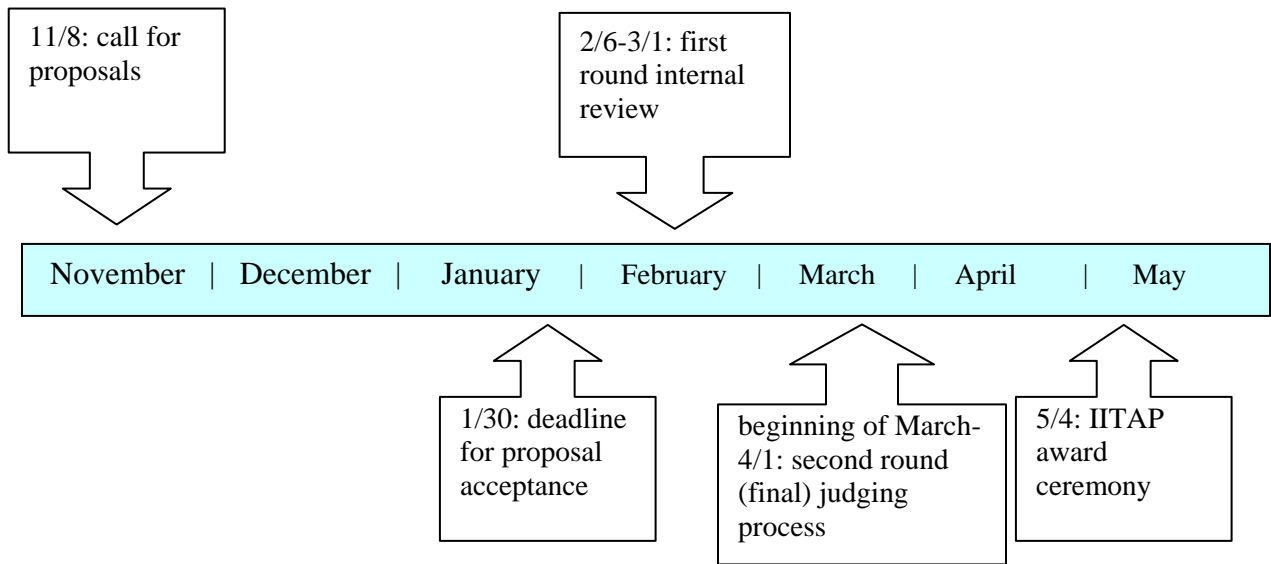


Figure 1: Program Timeline

Evaluation Description

Research Questions and Evaluation Plan

Purpose of Evaluation. The purpose of the IITAP evaluation was to identify strengths and areas for improvement to inform future implementation. The program was not systematically evaluated in its first eight years, so this was the first attempt to do so.

Research Questions. The IITAP program evaluation addressed four research questions:

1. Is the program efficient and accurate? What requires greater efficiency and accuracy?

1.1) Is the first-round review process efficient? Based on estimates of time spent by prior reviewers, it is recommended that each first-round reviewer spend an average of 2.5 hours per entry, which means spending 25 to 28 hours to complete the rating of all assigned entries. How can the internal review process be made more efficient?

1.2) Is the first-round review accurate? In order to review entries accurately, reviewers must be qualified, trained to follow the recommended review process, and willing to spend sufficient time on each entry. Therefore, there were five sub-questions: What are the necessary qualifications for reviewers? Do the reviewers have these qualifications? Do they understand the review process? Are they

spending the time required for accuracy? Are their ratings consistent across reviewers and review criteria?

1.3) Is the second-round judging process efficient? It is estimated that each second-round judge spends an average of 25 hours to complete the assessment. How can the second-round judging process be made more efficient?

2. How can the consistencies between the IITAP and ~FAST Tex programs be improved?

While the IITAP and ~FAST Tex programs were developed independently, they share the goal of supporting effective use of instructional technology, so some ~FAST Tex projects invariably are entries for IITAP awards. Therefore, it is desirable to align the two programs in terms of proposal submission protocols, proposal review protocols, and judging criteria. The focus of this analysis was major IITAP program documents used by staff, entrants, reviewers, and judges—such as proposal forms, protocols for proposal reviews, and criteria for judging and quality review—to determine whether they are consistent with ~FAST Tex program documents and streamlined.

The document analysis was conducted before the IITAP 2006 program cycle began, and the report of its outcome was prepared separately from this report.

Data Collection and Analysis Methods

The methods used to address the central questions are summarized in Table 3. The project components across the columns indicate evaluation activities used to address the central questions at the head of each row.

Table 3: Overview of Data Collection Methods to Address Central Questions

Components/ Central Questions	Program Staff Interview	Reviewer Survey	Judge Survey	Review Analysis	Document Analysis
1.1 Is the first round review process efficient?	X	X			
1.2 Is the first-round review accurate?	X	X		X	
1.3 Is the second-round judging process efficient?	X		X		
2 How can consistency between the IITAP and ~FAST Tex programs be improved?					X

- **Program staff interview:** The purpose of a program staff interview was to understand expectations about program operations. The interview was conducted with one program manager at the beginning of the evaluation project. For the interview questions, see Appendix C.
- **Reviewer survey:** The survey was conducted online using SurveyMonkey (<http://www.surveymonkey.com>), with participation solicited by e-mail to the eight reviewers immediately after the review deadline. By the last day of the

survey period from March 1, 2006, to March 10, 2006, five of eight reviewers had responded, so another solicitation was sent to the reviewers with the deadline extended by three days, improving the response rate to seven out of eight.

Descriptive statistics were obtained from SurveyMonkey and qualitative data were analyzed for recurring themes. For the survey instrument, see Appendix D.

- **Judge survey:** The purpose of the IITAP judge survey was to assess the efficiency of the final (second-round) judging. The survey was conducted online using SurveyMonkey (<http://www.surveymonkey.com>), with participation solicited by e-mail immediately after the review deadline, for the period May 1 through 12th. Two of the six judges responded. For the survey instrument, see Appendix E.
- **Review analysis:** Review analysis was conducted to determine the accuracy of the IITAP first-round (internal) review process as indicated by the consistency of ratings across reviewers and review questions. Ratings were retrieved from the online review submission site (<https://www.utexas.edu/academic/cit/test/iitap/iitaprr.php>) and analyzed for consistency using several descriptive statistics.

Findings and Conclusions

Question 1.1: Is the First-Round Review Process Efficient? (Based on reviewer survey)

Time spent on each entry

The reviewers spent **2.5 hours or less**, on average, to review each entry. The recommendation of 2.5 hours for each entry appears to be adequate, although some reviewers reported that they spent a shorter time—1.5 to 2 hours.

- Three reviewers spent about the recommended time of 2.5 hours, on average, to consider each entry, while four spent less time (Figure 2).

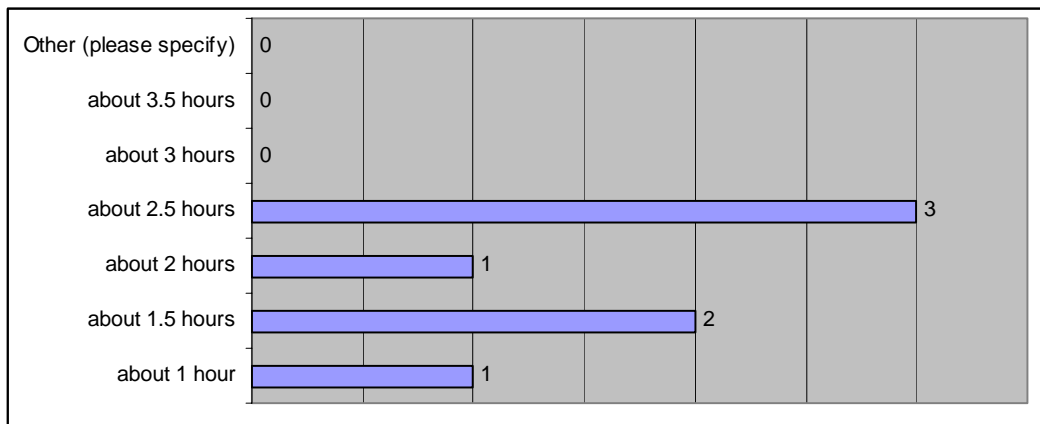


Figure 2: Responses to “The time I spent for each entry on average was:”

- **Most reviewers (five out of seven) reported that the time they spent was adequate**, while two reported it was somewhat short. None reported spending “somewhat long” or “too long” a time (Figure 3).
- **The perception of what was an appropriate amount time to spend for considering each entry differed significantly among reviewers.** Those who reported spending the right amount of time spent 1.5 to 2.5 hours, and the two reviewers who reported spending somewhat short time spent 1 hour and 2.5 hours.

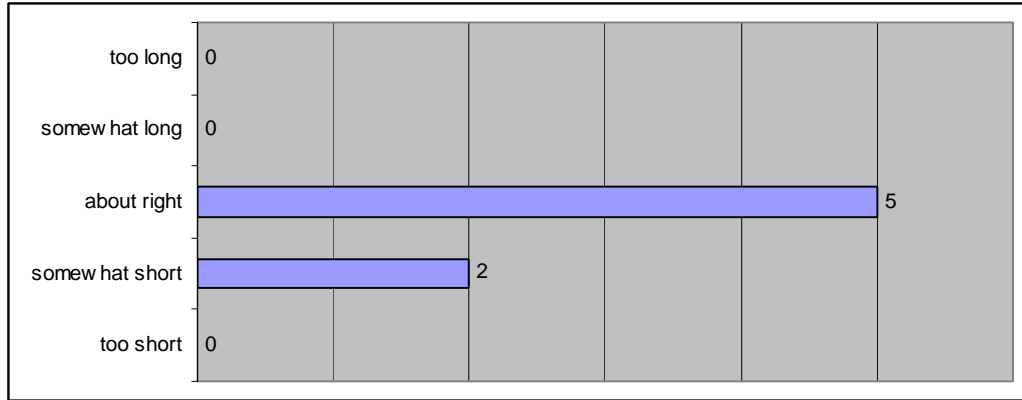


Figure 3: Responses to “The time I spent for reviewing each entry was:”

Coordination with work schedule

Allocating enough time to complete the task was challenging for some.

- Three reviewers **disagreed with the statement that they had enough time** for the review within their work schedule, while the others agreed or were neutral (Figure 4).

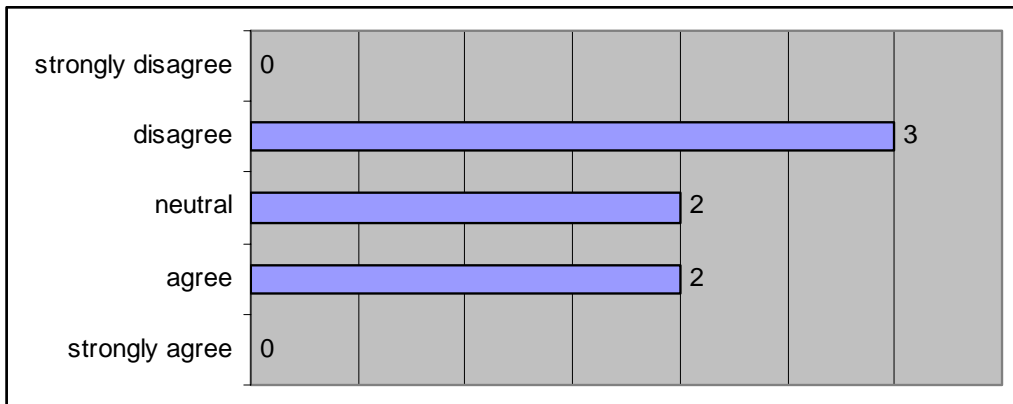


Figure 4: Responses to “I had enough time for the IITAP review within my work schedule.”

- The three reviewers who did not have enough time **worked overtime** (including weekends), **reviewed entries more quickly** than they would have liked, and/or **sacrificed other work** (Figure 5).

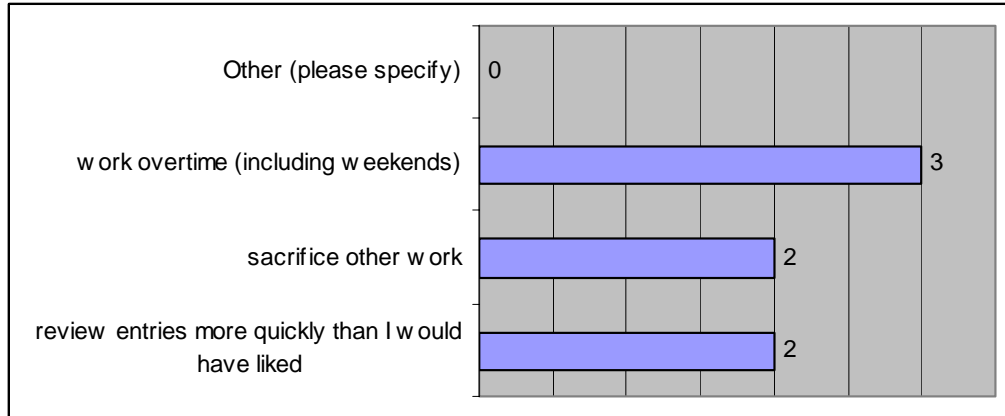


Figure 5: Responses to “In order to complete the IITAP reviews I had to (Check all that apply).”

Efficiencies and strategies

The reviewers reported in response to an open-ended question that they employed strategies to **regulate their work procedure** and to obtain an **overview of scores across the entries**.

- Most strategies the reviewers reported were concerned with their procedure for approaching the review task. For example, the most elaborated procedure reported contained seven steps: (1) read proposal, (2) enter initial impression scores, (3) thoroughly examine all materials, (4) re-read proposal more carefully again, (5) enter comments, re-examine materials, (6) give/reconsider scores and comments, and (7) go over all scores to check for consistency across proposals. Two other procedures reported included steps one (read proposal), three (examine materials),

and six (give/reconsider scores and comments) above. Two reviewers said they chose to read all the entries first and then rate them all.

- One reviewer created a table in which he/she recorded the scores in all categories of all entries. Another one agreed that such a table was very helpful.
- One reviewer described a procedure involving printing out all the entry descriptions and writing notes on them.
- One reviewer reported having to copy responses for questions when the system was about to logout the reviewer.
- Most reviewers (five of seven) reported that they considered their review process to be efficient (Figure 6). The one who disagreed that his/her process was efficient took an approach of reading/viewing all entries first and then rating them all.

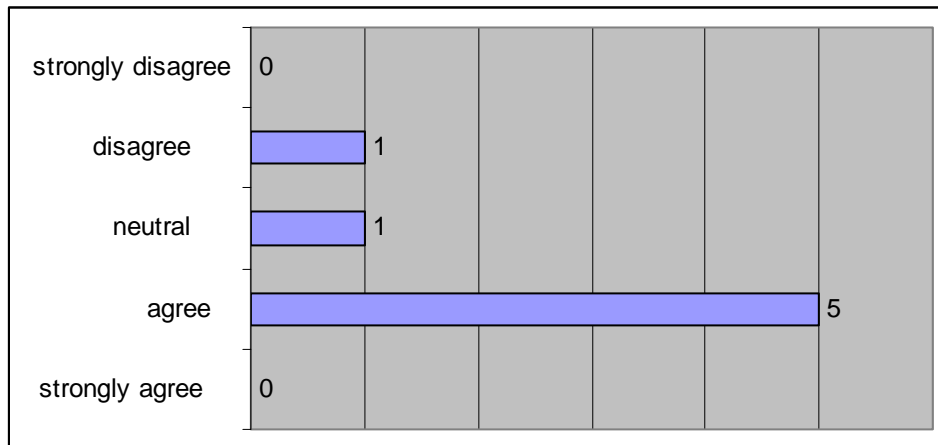


Figure 6: Responses to “In general, the review process was efficient.”

See Appendix F for the reviewers’ complete responses to open-ended questions.

Online review submission system

Most reviewers reported that the online review submission system was **easy to use**, but they **suggested improvements**, such as providing a longer timeout setting and incorporating a way to obtain an overview of all scores.

- Most reviewers (five out of seven) agreed or strongly agreed that the online review submission system was **easy to use** (Figure 7).

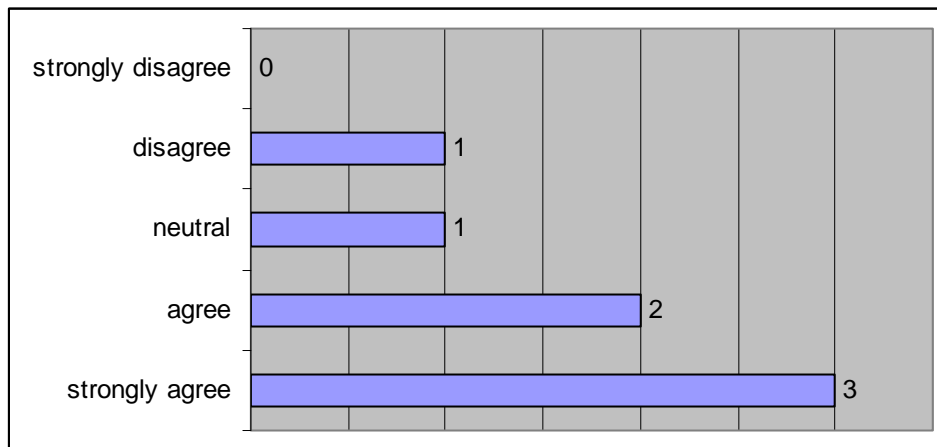


Figure 7: Responses to “The online review submission system was easy to use.”

- Reviewers suggested **improvements** in two areas, in an open-ended question. See Appendix F for the reviewers’ complete responses to the open-ended questions.
 - **Lengthen the period before timeout** (3 reviewers). Reviewers lost the comments that they had written when the timeout period ended before they clicked the “submit” button.
 - **Create an easy way to view and change scores across questions and entries** (3 reviewers). Reviewers found that viewing and changing scores they

had assigned previously was cumbersome, because they had to go through all the pages in between. One reviewer created a table to provide an overview of all the scores, and reported that it was helpful to be able to refer to scores easily and quickly. One stated that he/she would like to have had a hard copy of such a table.

Other suggestions included allowing users to **move from one rating page to another directly** and giving users an **area for general comments**.

Time-Consuming process and difficult process

Examining the entries, in particular the products, was the most time-consuming part of the task, and **finding enough time and achieving consistent scoring** were the most difficult parts.

- Most reviewers (four out of seven) said that the most time-consuming part of the reviewing process was **examining and understanding the entries' products**. The three others said that **reviewing all materials in the submissions** was most time-consuming.
- Three reviewers said that **finding time** to complete the review was the most difficult part of the task. Three also mentioned **making sure to provide adequate (consistent and fair) ratings**, and others mentioned **scoring the technical aspects** and **running the products on a given platform**.

See Appendix F for the reviewers' complete responses to the open-ended questions.

Suggestions for improving the reviewing process

Reviewers provided suggestions for improvement in **time management**.

- To help reviewers allocate enough time to complete the review tasks, several strategies were suggested:
 - **Assign more reviewers, each with fewer entries to review** (mentioned by 2 reviewers). The DIIA faculty advisory committee or the student advisory committee may be good sources for more reviewers.
 - Help reviewers commit themselves to their review tasks by sending them **weekly reminders** and asking about their progress, as the program staff did in this program cycle.

See Appendix F for the reviewers' complete responses to the open-ended questions.

Summary of findings for Question 1.1: Is the first round review process efficient?

- Program staff recommended reviewers to spend 2.5 hours to review each entry, and the reviewers spent **2.5 hours or less**, on average. Most reviewers reported that the time they spent was adequate.
- **Allocating enough time to complete the task was challenging** for some reviewers, who worked overtime, reviewed entries more quickly, or sacrificed other work to complete the review task.
- Reviewers employed strategies to **regulate their work procedure** and to obtain an **overview of scores across the entries**.

- Most reviewers reported that they considered their review process to be efficient. The one who did not took an approach of reading/viewing all entries first and then rating them all.
- Most reviewers reported that the online review submission system was **easy to use**.

Suggested improvements include

- lengthen the period before timeout
 - create an easy way to obtain an overview of all category scores across entries
 - allow users to move from one rating page to another directly
 - give users an area for general comments
- **Examining and understanding the entries' products and submitted materials** was the most time consuming part of the task.
 - **Finding enough time and achieving consistent scoring** were the most difficult parts of the task.
 - Suggestions reviewers provided for improvement include assigning more reviewers, each with fewer entries to review, and sending weekly reminders to help reviewers complete the tasks in time.

Question 1.2: Is the First-Round Review Accurate? (Based on the reviewer survey and review analysis)

Qualifications (Based on the reviewer survey)

Reviewers considered as necessary qualifications having **knowledge, skills, and/or experiences in technology, learning/instruction and/or evaluation/assessment**, and they considered themselves to be **well qualified**.

- All agreed or strongly agreed that they were well qualified for the review task (Figure 8).

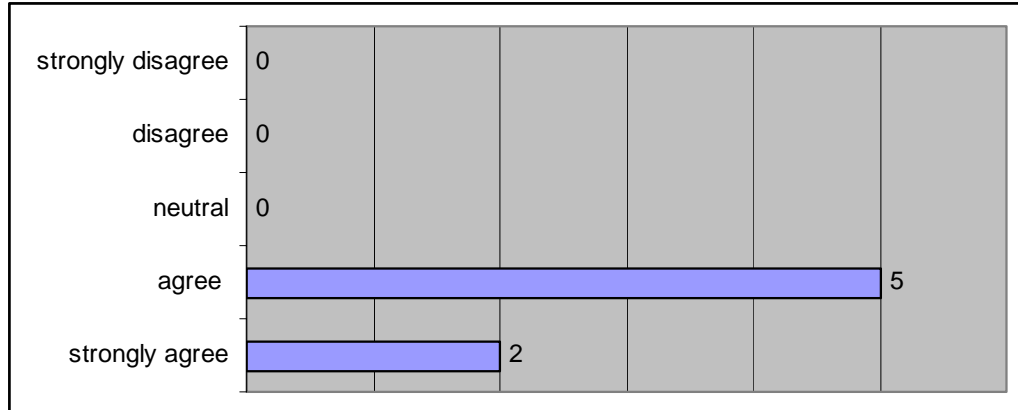


Figure 8: Responses to “I was well qualified to do this review task.”

- Reviewers reported in an open-ended question that they considered **knowledge, skills, and/or experiences in technology, instruction/learning, and/or evaluation/assessment** as essential for IITAP reviewers:
 - Technology competencies of various kinds were mentioned by seven reviewers: website design, technical knowledge/skills, technology use/innovation. Two reviewers’ responses concerned having an appropriate

disposition toward technological knowledge and skills: one said that it is important to possess “a sense for when to give the ‘benefit of the doubt’ for technology’s sake,” and the other said that it is important to have the “ability to recognize what is ‘innovative.’”

- Instruction/learning experience of various kinds was mentioned by six reviewers: teaching, pedagogy, instructional/pedagogical design, teaching principles, instructional experiences, learning, the ability to imagine the student’s perspective.
- Evaluation/assessment competence was mentioned by two reviewers.

Other skills and knowledge the reviewers mentioned included “an appreciation for visual design,” “the ability to write critiques that are helpful, true, and as kind as possible,” “experience reviewing IT proposals,” and “familiarity with the content (i.e., get someone with a science background to review science content, etc.).”

See Appendix F for the reviewers’ complete responses to the open-ended questions.

IITAP materials (Based on the reviewer survey)

Most reviewers reported that **instructions for the review task were clear**. Reviewers **read the materials given and/or sent** by e-mail directly to them, but they less frequently read the online materials.

- Most reviewers (six out of seven) agreed or strongly agreed that the **instructions for the review tasks were clear** (Figure 9).

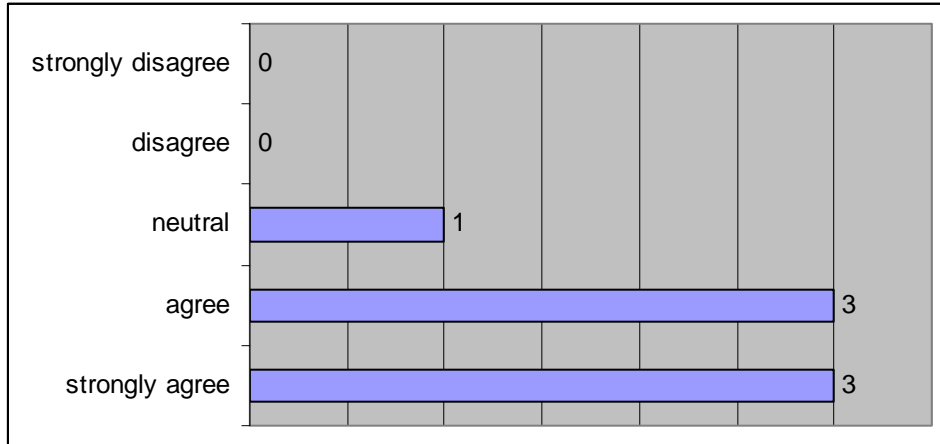


Figure 9: Responses to “Instructions for the review task were clear to me.”

- Most reviewers (at least six out of seven) **read the materials given at the orientation and/or sent via e-mail by the program staff**, but only about half of the reviewers (three or four) read materials on the DIIA website that the program staff asked them to read (Figure 10).

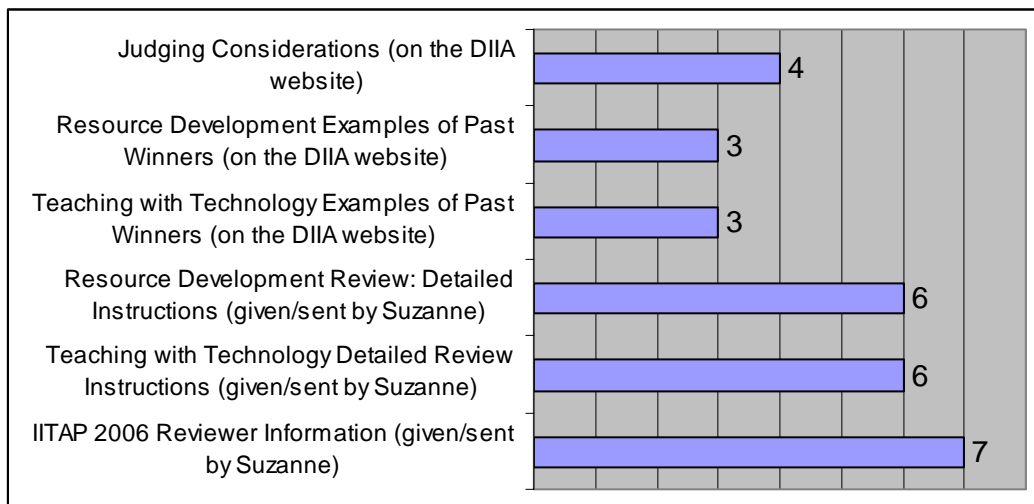


Figure 10: Responses to “I looked at the following materials before I started (Check all that apply).”

Suggestions for improving the reviewing accuracy (Based on the reviewer survey)

- To help reviewers maintain accuracy in completing their review tasks, several strategies were suggested:
 - Hold a **training session** for reviewers.
 - Provide **examples of ratings**.
 - Clarify the approach one should take in reviewing. Indicate whether one should (1) review a proposal, (2) review supporting content, (3) review content in the context of a whole lesson plan or class, or (4) make suppositions about learning that likely occurred, but that may not have been adequately addressed in a proposal.
 - Allow reviewers to **discuss their ratings when they share the same entries or teams to review**.
 - Provide online resource to view all the ratings across entries.

See Appendix F for the reviewers' complete responses to the open-ended questions.

Reviewer orientation (Based on the reviewer survey)

Reviewers reported that the orientation was **helpful**, and they suggested providing **hands-on rating training**.

- Those five reviewers who attended the reviewer orientation session agreed that it was **helpful** (Figure 11).

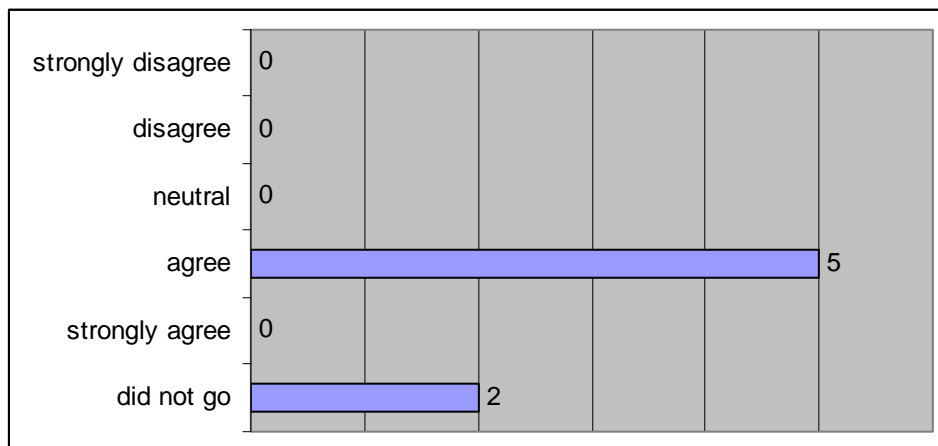


Figure 11: Responses to “The reviewer orientation session was helpful.”

- All three reviewers who offered ideas for improving the reviewer orientation session suggested conducting **hands-on training** using entries from previous years or simulated entries as exemplars for individual consideration, followed by group discussion of ratings criteria. One reviewer said that such training sessions would help reviewers develop common understanding of what to look for, what to watch out for, and what to expect.

See Appendix F for the reviewers’ complete responses to the open-ended questions.

Consistency of reviewer scorings across reviewers (Based on the review analysis)

Inter-rater reliability was computed for each entry (see Table 4), which revealed several tendencies about scoring consistency across reviewers:

- The inter-rater reliability averaged 0.49.

- There were several entries for which extremely low (less than 0.3) inter-rater reliabilities (yellow and red cells in Table 4) were obtained, including two with negative reliabilities (red cells in Table 4).
- Extremely low inter-rater reliabilities were not associated with score values; reviewers disagreed about both high-score and low-score entries.
- Excluding the negative inter-rater reliabilities, average inter-rater reliability improves to 0.57, and excluding inter-rater reliabilities less than 0.3, average inter-rater reliability improves to .65.

In order to understand where disagreements between the reviewers occurred, the cases in which inter-rater reliability increased by 0.10 or more when a specific rater was excluded are summarized in Table 5. While all but two raters disagreed with other raters significantly (to the extent their scores lowered the inter-rater reliability by 0.10 or more), three raters (raters A, C, and H) in particular disagreed with others, lowering the inter-rater reliability by 0.10 or more for two or more entries and by 0.20 or 0.30 or more in some cases.

Table 4: Inter-Rater Reliability by Entries

Entries	Inter-rater reliability	Mean score
Paradise Lost, Book 9	.34	43.75
Vietnam: A Movie	.73	38.69
Flash Animations in Nuclear and Radiation Engineering Courses	.63	25.69
Integrating Creativity and Ethics into the Undergraduate Educational Experience	.35	18.33
Operations Research Models and Methods	.44	43.33
Sámi Culture	.65	39.00
Texas Politics	.84	44.65
The Sister Arts: British Gardening, Painting, and Poetry, 1700-1832	.01	44.60
Transforming Lives Through Resilience Education	.65	42.38
An Instructional Technology Scaffold for Reinforcing Learning of Probability and Statistics	-.12	35.09
Applying Counseling Theories Website	.27	45.98
Does Product Costing Matter When Setting Prices in a Competitive Product Market?	.69	43.83
I-Movie Demonstration of Teaching Skills	.80	33.44
iESP Online: The Experience of Time at Work	.68	43.15
Interactive Laboratory Module in Primate Anatomy: Understanding Primate Quadrupedal Locomotion	.59	40.91
Kinematic Analysis of Human Movement	.61	40.70
Making Structural Issues More Accessible to Architecture Students: A New Methodology for Teaching Construction	.66	39.96
Memories of Older Adults	.76	31.79
Not Just for Theaters Anymore: Documentary Approaches to Inquiry and Education	.17	46.19
Office Production of Lectures	.91	22.18
Technology Meets Business Education	-.36	43.75
Average	.49	38.44

* Red and yellow cells indicate inter-rater reliabilities less than 0 and between 0 and .3, respectively.

Table 5: Effects on Rater Reliabilities of Outlier Scores from One Reviewer

Reviewers	Entries for which inter-rater reliability increased without this reviewer's ratings by 0.10 (black), 0.20 (blue), 0.30 or more (red)
A	<ul style="list-style-type: none"> Integrating Creativity and Ethics into the Undergraduate Educational Experience The Sister Arts: British Gardening, Painting, and Poetry, 1700-1832 Technology Meets Business Education Transforming Lives Through Resilience Education
B	
C	<ul style="list-style-type: none"> An Instructional Technology Scaffold for Reinforcing Learning of Probability and Statistics Vietnam: A Movie
D	<ul style="list-style-type: none"> Technology Meets Business Education
E	
F	<ul style="list-style-type: none"> Paradise Lost, Book 9
G	<ul style="list-style-type: none"> Making Structural Issues More Accessible to Architecture Students: A New Methodology for Teaching Construction
H	<ul style="list-style-type: none"> An Instructional Technology Scaffold for Reinforcing Learning of Probability and Statistics Applying Counseling Theories Website Flash Animations in Nuclear and Radiation Engineering Courses

In order to understand whether the score variability was related to the categories of reviewers (technology, instruction/learning, and evaluation/assessment), the average scores are summarized by reviewer categories in Table 6 and Figures 12 and 13.

- Instruction/learning specialists in general tended to award higher scores than other specialist groups (Table 5).

Table 6: Comparison of Average Scores by Reviewer Categories

Categories of Reviewers	Mean	N (number of scores)	Std. Deviation
Instruction	40.64	436	11.23
Technology	36.47	425	12.94
Evaluation	37.64	273	11.63
Total	38.36	1134	12.13

- In the TWT category, technology specialists awarded lower scores for questions 3 (new skills and knowledge), 10 (technical operation), and 15 (learner modalities) than other specialists, whereas evaluation/assessment specialists awarded lower scores for all outcome (assessment) related questions (6, 7, 8 and 9).

Instruction/learning specialists tended to give higher scores than others for many questions (3, 7, 8, 9, and 15). Differences of 5 points or more in average scores were found among the specialist groups in the following areas (Figure 12):

- Question 3: Instruction > Technology
- Question 6: Instruction and Technology > Evaluation
- Question 7: Instruction > Technology > Evaluation
- Question 8: Instruction > Technology > Evaluation
- Question 9: Instruction > Technology > Evaluation
- Question 10: Evaluation > Technology
- Question 15: Instruction > Evaluation > Technology

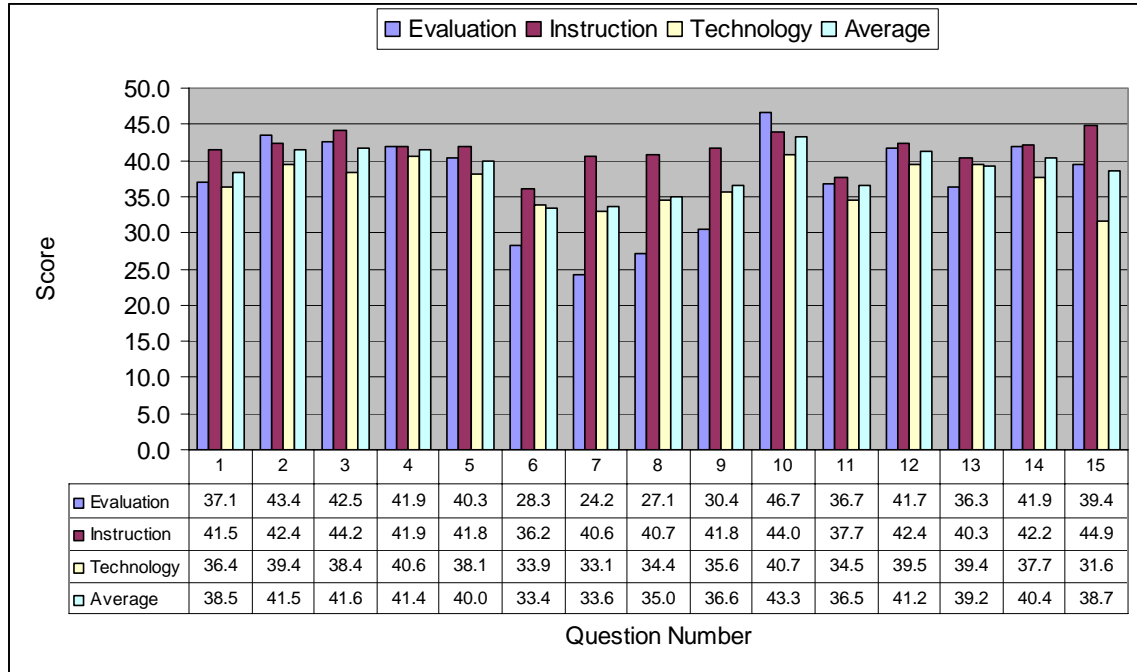


Figure 12: Scores by Reviewer Category in Teaching with Technology Category

- In the RD category, technology specialists awarded lower scores than other specialists for questions 3 (new knowledge and skills), 4 (efficiency), 8 (technical operation), and 13 (media quality), whereas evaluation/assessment specialists awarded lower scores for all outcome (assessment) related questions (5, 6, and 7) and higher scores for questions 8 (technical operation), 9 (user interface design), 11 (content presentation), and 13 (learner modalities). Instruction/learning specialists tended to give higher scores for outcome (assessment) related questions (5, 6, and 7) and question 4 (efficiency). Differences of 5 points or more were found among the specialist groups in the following areas (Figure 13):
 - Question 3: Evaluation & Instruction > Technology

- Question 4: Instruction > Technology
- Question 5: Instruction > Technology & Evaluation
- Question 6: Instruction > Technology & Evaluation
- Question 7: Instruction > Technology > Evaluation
- Question 8: Evaluation > Technology
- Question 9: Evaluation > Instruction & Technology
- Question 11: Evaluation > Instruction
- Question 13: Evaluation > Technology

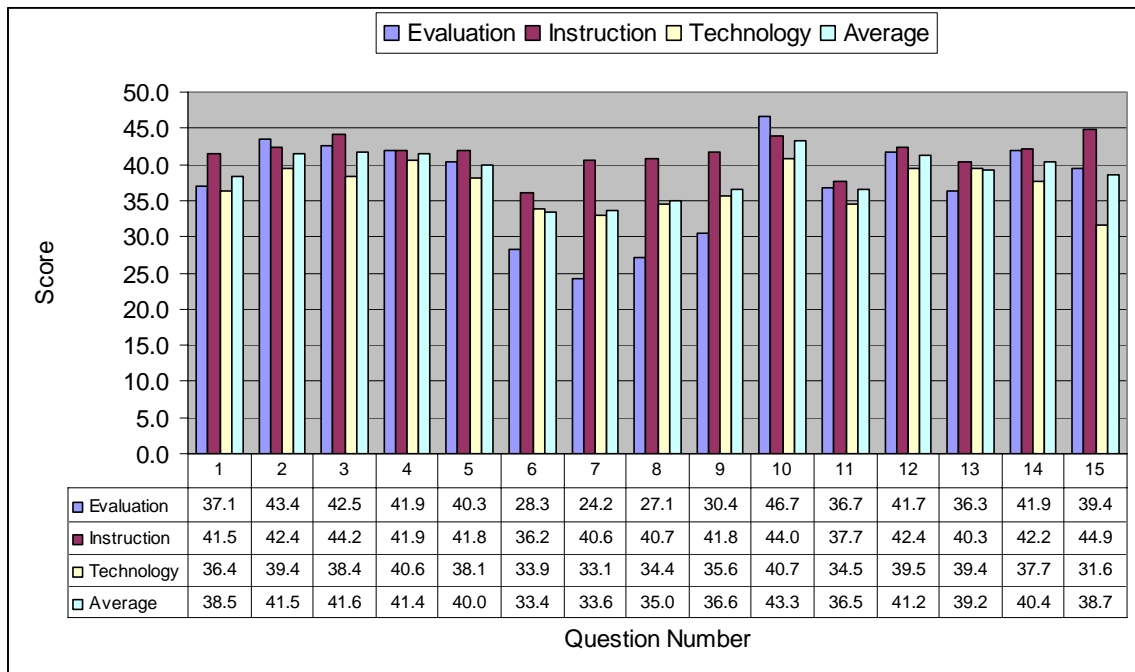


Figure 13: Scores by Reviewer Category in Resource Development Category

In summary, the consistency among reviewers in scorings was not great, as reflected in the inter-rater reliability of 0.49. Certain entries had substantially lower

inter-rater reliability than others. Also, a few reviewers diverged substantially from the other reviewers' scorings. Furthermore, the scoring seemed to differ significantly among specialist groups of reviewers. Hence, aligning reviewers' scoring individually as well as by specialist group would appear to contribute to the improvement of inter-rater reliability.

Consistency of the reviewer scorings across review questions

(Based on the review analysis)

Figures 14 and 15 and Tables 7 and 8 summarize score variability by question, using the ranges and standard deviations.

- In the TWT category, scoring variability was great (the reviewers tended to disagree more substantially) for questions 5, 6, 7, 8, and 9 (in particular 7). Questions 6, 7, 8, and 9 concerned the “outcomes (assessment)” area, and Question 5 concerned teaching and learning efficiency.
- In the RD category, scoring variability was great for questions 5, 6, and 7, concerning the “outcomes (assessment)” area.

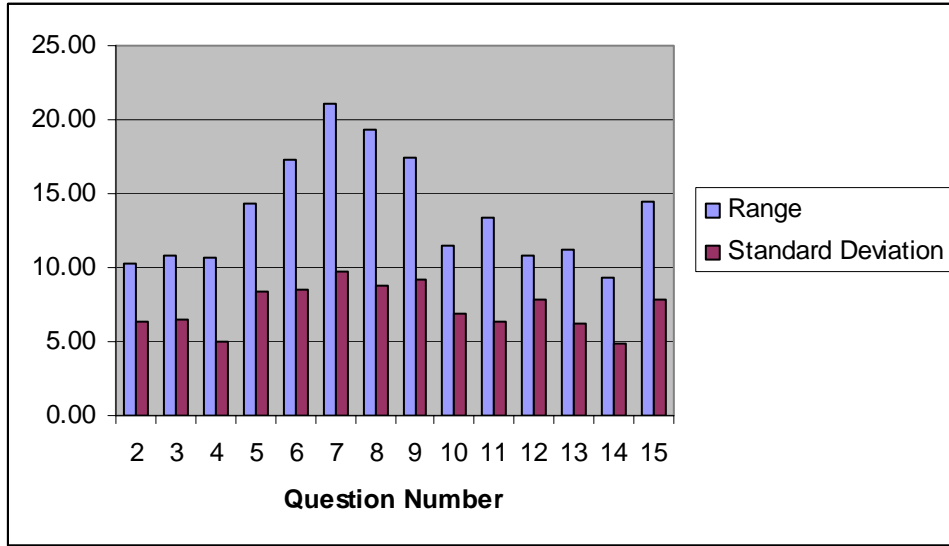


Figure 14: Score Variability in the Teaching with Technology Category

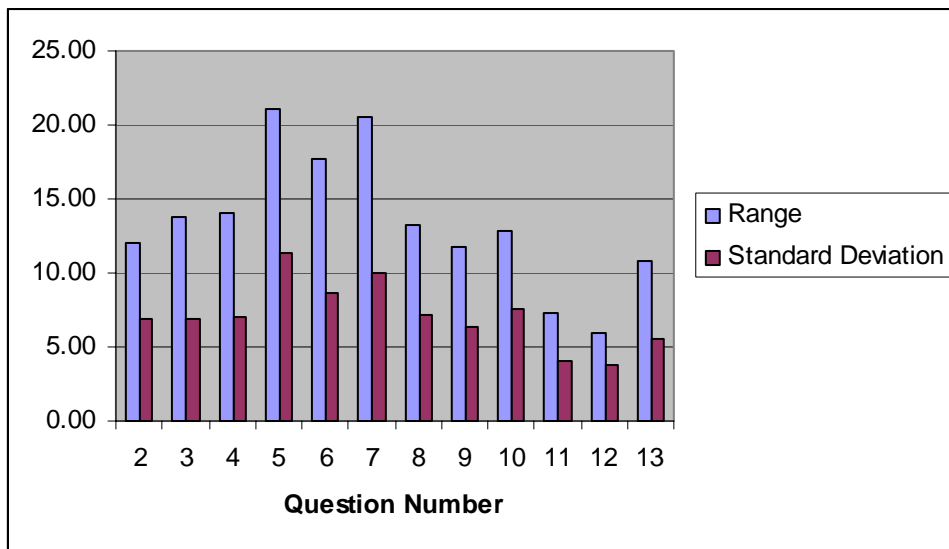


Figure 15: Score Variability in the Resource Development Category

Table 7: Ranges and Standard Deviations of Ratings for Each Entry in the Teaching with Technology Category

Entries		Question Numbers														Average
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	
An Instructional Technology...	Range	10.00	10.00	30.00	15.00	25.00	10.00	25.00	20.00	15.00	20.00	25.00	25.00	35.00	25.00	21.54
	Std. Dev.	4.79	4.79	8.54	6.45	12.50	4.79	11.09	8.16	6.29	8.66	10.80	12.50	14.72	11.09	9.26
Applying Counselin ...	Range	5.00	5.00	10.00	10.00	25.00	20.00	15.00	15.00	10.00	10.00	5.00	5.00	5.00	2.00	10.54
	Std. Dev.	2.89	2.50	4.76	5.77	11.35	11.00	7.23	6.45	4.86	4.79	2.89	2.50	2.06	1.00	5.17
Does Product Costing ...	Range	5.00	5.00	5.00	5.00	25.00	22.00	25.00	15.00	5.00	20.00	0.00	10.00	2.00	15.00	11.85
	Std. Dev.	2.89	2.36	2.50	2.89	10.80	9.88	10.80	6.45	2.89	8.54	0.00	5.00	1.41	7.50	5.46
I-movie demonstration...	Range	20.00	20.00	15.00	15.00	5.00	35.00	20.00	5.00	0.00	0.00	30.00	0.00	0.00	10.00	11.92
	Std. Dev.	8.54	9.46	7.07	6.45	2.89	15.54	8.54	2.50	0.00	0.00	15.28	0.00	0.00	5.77	5.65
iESP Online	Range	10.00	20.00	10.00	20.00	15.00	20.00	10.00	20.00	11.00	0.00	10.00	15.00	0.00	0.00	11.62
	Std. Dev.	5.77	9.57	5.00	9.57	6.29	9.57	5.77	8.16	7.78	0.00	5.00	10.61	0.00	0.00	5.95
Interactive Laboratory ...	Range	5.00	10.00	20.00	22.00	25.00	30.00	30.00	20.00	10.00	20.00	10.00	10.00	10.00	12.00	17.62
	Std. Dev.	2.89	4.79	9.57	9.43	12.25	14.36	13.54	8.54	4.79	10.00	4.79	4.79	5.77	5.41	8.31
Kinematic Analysis...	Range	5.00	5.00	5.00	10.00	10.00	30.00	30.00	30.00	20.00	20.00	10.00	20.00	10.00	10.00	16.15
	Std. Dev.	2.89	2.50	2.45	4.35	4.79	13.15	13.15	14.36	10.31	8.66	4.86	8.54	5.00	5.77	7.53
Making Structural ...	Range	10.00	10.00	10.00	10.00	40.00	40.00	15.00	25.00	5.00	10.00	15.00	10.00	2.00	20.00	16.31
	Std. Dev.	5.00	4.79	5.77	5.77	17.50	16.52	6.45	11.09	2.50	5.77	6.29	4.79	1.00	9.46	7.52
Memories of Older Adults	Range	23.00	15.00	8.00	10.00	15.00	10.00	25.00	20.00	5.00	20.00	0.00	30.00	15.00	40.00	16.38
	Std. Dev.	10.28	7.50	4.00	4.79	6.29	5.00	10.41	8.54	2.50	14.14	0.00	12.58	7.07	16.33	7.63
Not Just for Theaters...	Range	12.00	10.00	10.00	20.00	15.00	25.00	25.00	5.00	20.00	16.00	0.00	0.00	5.00	5.00	12.00
	Std. Dev.	5.68	5.00	5.00	10.00	7.07	12.50	11.90	2.50	11.55	8.50	0.00	0.00	2.89	3.54	6.19
Office Production ...	Range	20.00	20.00	5.00	25.00	10.00	5.00	5.00	20.00	20.00	0.00	31.00	10.00	5.00	15.00	13.15
	Std. Dev.	9.57	8.66	2.50	11.09	4.79	2.89	2.89	9.43	10.00	0.00	17.90	5.77	2.50	7.64	6.62
Technology meets business..	Range	2.00	5.00	5.00	15.00	8.00	15.00	15.00	30.00	16.00	30.00	5.00	10.00	30.00	30.00	16.46
	Std. Dev.	1.00	2.50	2.50	7.50	3.32	7.07	7.07	14.14	6.16	13.54	2.50	4.08	16.07	13.99	7.73
Average	Range	10.21	10.77	10.62	14.38	17.31*	21.12*	19.38*	17.50*	11.43	13.42	10.85	11.15	9.38	14.43	13.98
	Std. Dev.	6.32	6.49	4.97	8.39*	8.45*	9.79*	8.76*	9.26*	6.89	6.35	7.79	6.24	4.88	7.88	7.40

* The yellow highlights indicate ranges of 15 or more and standard deviations of 8 or more.

Table 8: Ranges and Standard Deviations of Scorings for Each Entry in the Resource Development Category

Entries	Question Numbers													Average
		2	3	4	5	6	7	8	9	10	11	12	13	
"Paradise Lost," Book 9	Range	10.00	15.00	10.00	30.00	30.00	10.00	5.00	10.00	20.00	10.00	5.00	15.00	14.55
	Std. Dev.	4.79	7.50	5.00	12.50	12.50	5.77	2.89	4.79	10.31	5.00	2.50	7.50	6.93
"Vietnam: A Movie"	Range	20.00	20.00	20.00	30.00	20.00	20.00	30.00	25.00	0.00	20.00	5.00	20.00	19.09
	Std. Dev.	8.54	9.46	10.00	13.77	8.54	11.55	14.36	12.58	0.00	10.31	2.36	11.55	9.50
Flash Animations in Nuclear ...	Range	15.00	25.00	25.00	30.00	25.00	25.00	30.00	20.00	10.00	10.00	15.00	15.00	20.91
	Std. Dev.	7.07	10.31	12.25	14.36	11.90	11.90	12.58	10.00	5.77	5.77	7.07	6.45	9.85
Integrating Creativity and Ethics ...	Range	10.00	15.00	20.00	25.00	10.00	10.00	0.00	0.00	21.00	0.00	0.00	0.00	9.18
	Std. Dev.	4.79	7.07	9.57	12.25	5.00	4.79	0.00	0.00	12.12	0.00	0.00	0.00	4.62
Operations Research Models ...	Range	10.00	10.00	10.00	10.00	10.00	15.00	5.00	15.00	10.00	5.00	10.00	15.00	10.45
	Std. Dev.	4.79	4.79	4.79	4.79	4.79	6.45	2.89	7.07	5.77	2.50	4.08	6.45	4.94
Sámi Culture	Range	10.00	10.00	10.00	15.00	25.00	28.00	40.00	35.00	20.00	10.00	10.00	20.00	20.27
	Std. Dev.	5.00	4.79	4.79	7.50	10.41	11.79	17.97	14.43	8.54	5.00	5.00	8.54	8.98
Texas Politics	Range	10.00	2.00	10.00	10.00	15.00	40.00	5.00	0.00	10.00	5.00	0.00	5.00	9.27
	Std. Dev.	5.00	1.00	5.00	5.77	7.50	17.50	2.50	0.00	4.08	2.50	0.00	2.50	4.40
The Sister Arts	Range	15.00	5.00	5.00	20.00	10.00	10.00	10.00	8.00	13.00	5.00	10.00	5.00	9.18
	Std. Dev.	7.07	2.50	2.50	8.54	4.08	4.08	5.26	3.95	5.38	2.06	4.24	2.50	4.10
Transforming Lives Through Resilience ...	Range	15.00	35.00	25.00	35.00	25.00	30.00	5.00	5.00	20.00	5.00	5.00	10.00	18.18
	Std. Dev.	7.07	16.52	11.90	14.43	11.81	16.01	2.50	2.36	10.31	2.50	2.38	4.79	8.68
Average	Range	12.00	13.80	14.00	21.08*	17.75*	20.55*	13.25	11.80	12.81	7.25	6.00	10.75	12.00
	Std. Dev.	6.91	6.89	7.08	11.39*	8.65*	9.98*	7.10	6.32	7.53	4.06	3.76	5.53	6.91

* The yellow highlights indicate the ranges of 15 or more and standard deviations of 8 or more.

In order to understand the scoring tendency by question, the average scores by questions are summarized in Figures 16 and 17.

- In the TWT category, the average scores were particularly low for questions 6, 7, 8, and 9 in the “outcomes (assessment)” area . The average score was also low for question 11, about the user interface design.
- In the RD category, the average scores were particularly low for Questions 5, 6, and 7 in the “outcomes (assessment)” area. The average scores were also low for questions 9 and 10, about the user interface design and engagement/interactivity.

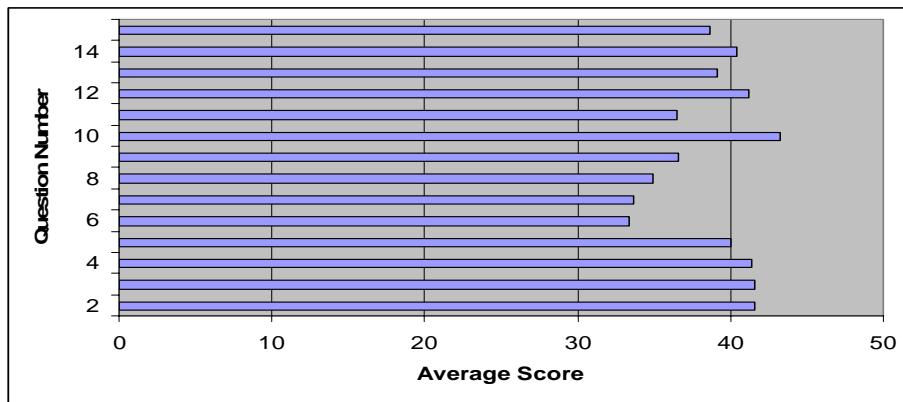


Figure 16: Average Scores by Questions in the Teaching with Technology Category

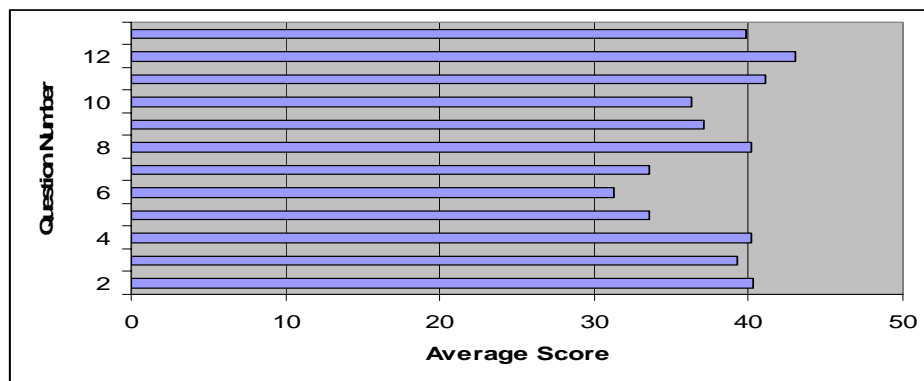


Figure 17: Average Scores by Questions in the Resource Development Category

In summary, reviewers disagreed more substantially in their scorings for certain questions. Reviewer training should focus on these questions. Also, the entries received low scores for certain questions, and entrants may need more guidance in these areas in preparing their proposals. The outcome (assessment) questions were problematic in the both regards: more substantial disagreement and lower scores.

Consistency of reviewer scoring with (external) judge scoring

There were **substantial differences** between the rankings of internal reviewers and those of the external judges, as shown in Tables 9 and 10.

- The gold award winner in the TWT category, Kinematic Analysis of Human Movement, was ranked 6th according to the reviewers' scorings, with low scores for questions 6 (assessment plan), 9 (assessment of efficiency), and 11 (user interface design).
- The entries ranked second and third by the internal reviewers in the TWT category were not selected for an award by the external judges.
- The gold award winner in the RD category, Operations Research Models and Methods, was ranked 4th according to reviewers' scorings, with low scores for questions 9 (user interface design) and 12 (media quality).
- The entries ranked second and third by the internal reviewers in the Resource Development category were not selected for an award by the external judges.

Table 9: Comparison of Rankings by Judges and Reviewers in the Teaching with Technology Category

Award (determined by the judges)	Entries	Ranking based on Reviewers' Scoring	Average Reviewer Score
Gold	Kinematic Analysis of Human Movement ("Biomech")	6th	40.70
Silver	Applying Counseling Theories Website	1st	45.98
Bronze	Technology Meets Business Education—Teaching Students Marketing Research Techniques through Online Interactive Interfaces	4 th	43.53
No award	Not Just for Theaters Anymore: Documentary Approaches to Inquiry and Education	2nd	45.26
No Award	Does Product Costing Matter When Setting Prices in a Competitive Product Market?	3rd	43.90

Table 10: Comparison of Rankings by Judges and Reviewers in the Resource Development Category

Award (determined by the judges)	Entries	Ranking based on Reviewers' Scoring	Average Reviewer Score
Gold	Operations Research Models and Methods	4th	43.33
Silver	Paradise Lost, Book 9	3rd	43.75
Bronze	Transforming Lives Through Resilience Education	5th	42.38
No Award	Texas Politics	1st	44.90
No Award	The Sister Arts: British Gardening, Painting, and Poetry, 1700-1832	2nd	44.61

Summary of findings for Question 1.2: Is the first-round review process accurate?

- Reviewers considered as necessary qualifications having **knowledge, skills, and/or experiences in technology, learning/instruction and/or evaluation/assessment**, and they considered themselves to be **well qualified**.

- Most reviewers reported that **instructions for the review task were clear**. Reviewers **read the materials given and/or sent** by e-mail directly to them, but they less frequently read the online materials.
- Suggestions for improving the reviewing accuracy that reviewers suggested include
 - Hold a **training session** for reviewers.
 - Provide **examples of ratings**.
 - **Clarify the approach** one should take in reviewing. Allow reviewers to **discuss their ratings when they share the same entries or teams to review**
 - Provide online resource to **view all the ratings across entries**.
- Reviewers reported that the orientation was **helpful**, and they suggested providing **hands-on rating training**.
- Consistency among reviewers in scorings was not great, as reflected in the **inter-rater reliability of 0.49**. Certain entries had substantially lower inter-rater reliability than others. Also, a few reviewers diverted substantially from the other reviewers' scorings. Furthermore, the scoring differed significantly among specialist groups of reviewers.
- Reviewers disagreed more substantially in their scorings for certain questions. Also, the entries received low scores for certain questions. **The outcome (assessment) questions were problematic in both regards: more substantial disagreement and lower scores.**

- **The internal reviewers' rankings were substantially different from those of the external judges.**

Question 1.3: Is the Second-Round Judging Process Efficient?

Time spent for each entry

The judges spent **1 to 1.5 hours** to review each entry.

- One judge spent about one hour and the other spent about 1.5 hours on each entry.
- Both judges reported that the time they spent was adequate.
- Concerning the most time-consuming part of the process, the judges responded as follows:
 - "...going back and checking for consistency in the judging of the same question from entry to entry"
 - "looking through the supplementary material, and, in some cases, navigating some pretty detailed Web sites"

Efficiencies and strategies

Both judges agreed that they considered their judging process to be efficient.

IITAP materials

Both judges looked at most materials and reported that instructions for judging were clear to them.

- Both judges looked at the following materials:

- IITAP 2006 Judging Information
- Teaching with Technology Scoring Instructions
- Resource Development Scoring Instructions
- The Entry Guidelines/2006 Call for Entries
- One judge looked at the examples as well:
 - Teaching with Technology Examples of Past Winners
 - Resource Development Examples of Past Winners
- Both judges agreed that instructions for the judging task were clear to them.

Online judge score submission system

The judges' reactions to the online judge score submission system were mixed.

- One judge agreed that the system was easy to use, while the other was neutral about it.
- One judge contributed a suggestion for improvement: “For consistency's sake, it would have been useful to be able to access my scoring and comments on the same question from one entry to another, i.e., to have a spreadsheet of all info available from the very beginning. The linear, lockstep format of the judging submission system made it hard to work in small chunks of time. I would have liked to have compared all entrants' answers to one question. A printed hard copy would be a good solution.”

Difficult process

Concerning the most difficult part of the judging process, both judges mentioned comparing entries from very different fields:

- “...judging such diverse entries. It was sometimes difficult to judge how much of a difference the entry actually made to the students' learning.”
- “Trying to compare/contrast a few, excellent contributions, in vastly different fields of study.”

Suggestions for Improving the Judging Process

The judges offered two suggestions for improvement.

- “I think a meeting of the judges would have been useful, to talk through the quality of some of the entries.”
- “Pruning the field to four excellent entries in each category would allow for a more detailed analysis/assessment.”

Summary of findings for Question 1.3: Is the second-round judging process efficient?

- The judges spent **1 to 1.5 hours to review each entry**, which they thought was adequate. They reported that checking scoring consistency and understanding supplementary materials were time consuming.
- Both judges agreed that they considered their judging process to be efficient.
- Both judges looked at most materials and reported that **instructions for judging were clear to them.**

- The judges' reactions to the online score submission system were mixed. One judge suggested **allowing judges to access all of their scorings and comments in a spreadsheet and to view all entrants' answers to the same questions at once.**
- Both judges said that **the most difficult part of the process was comparing entries from very different fields.**
- The judges suggested having a meeting of judges and reducing the number of entries that advance to the final evaluation round.

Recommendations

It is recommended that DIIA leadership and IITAP program staff consider the following recommendations for the improvement of the program.

Recommendation concerning entrants

- Provide entrants examples and assistance in preparing responses to the questions in the “outcomes” area.

Recommendations for improving the online review/judge submission system

- Provide a longer period before timeouts.
- Provide a printable spreadsheet where reviewers/judges may input and change all category scores across entries.

- Provide a page where all entrants' answers to the same questions are listed together to allow comparison.
- Enable navigation that allows users to move from one rating page to another directly.
- Create an area for general comments.

Recommendation concerning both the first-round review process and the second-round judging process

- Establish review criteria using example entries: for example, "Entry A is an example of a score of 40 for question 2." Provide the criteria to both reviewers and judges to resolve disparities between reviewers' and judges' ratings.

Recommendations for the first-round review process

Recommendations 1-2 are the most important and require substantial work, while recommendations 3-6 require very little work.

1. Consider changes in procedure and policy to help reviewers more easily complete their task on time:
 - a. Assign fewer entries per reviewer by increasing the number of reviewers.
 - b. Ask reviewers to clear their schedule for the review period. For example, if a reviewer is expected to spend 25 hours in total, ask that three entire days or six half days be kept completely clear of other commitments.

Reviewers might also be encouraged to work from home during that period.

- c. Send weekly reminders, as was done this year.
2. Enrich the reviewer orientation with hands-on rating training during which program staff help reviewers understand how to rate consistently.
 - Clarify priorities for focus: the proposal, the products, or some combination of them.
 - Make reviewers aware of the different scoring tendencies of specialist groups.
 - Show reviewers examples of previously scored entries, then have them evaluate exemplar entries individually with follow-up group discussion of their ratings. Program staff should focus on the questions for which scoring tends to vary substantially, making sure that all reviewers come to agreement within a 10-point differential by the end of the practice session. Aim to improve average inter-rater reliability to .60 initially and .70 ultimately, by focusing on eliminating extremely low inter-rater reliabilities.
 - Experiment with collaborative, group review of the products in the future, depending on how useful reviewers find the group discussions.
 3. Maintain a recommendation of spending 2.5 hours to review each entry, so reviewers will be able to allocate enough time to complete their task.
 4. Develop for reviewers' use a review strategy inventory drawn from the self-reported strategies identified in this study. While all strategies may not be equally useful to all reviewers, making the strategies available will present more options

for reviewers to choose from. Future reviewer surveys should include inquiries about the effectiveness of the strategies, so that only selected strategies may be recommended.

5. Give/send reviewers important reading materials directly rather than expecting them to read online materials.
6. Create a profile of the ideal reviewer as a basis for selection of reviewers.

Because the current selection process appears to address qualifications that reviewers consider important—knowledge, skills and/or experiences in technology, learning/instruction, and evaluation/assessment—no change in the reviewer selection process is necessary. However, documenting the ideal reviewer profile will help inform future selections.

Recommendations for the second-round judging process

1. Consider ways for judges to communicate after they complete their individual scorings, so that they can reach agreement on the final decision of awards.
2. Recommend that judges spend 1.5 hours to review each entry.

Appendix A: Teaching with Technology Detailed Review Instructions

This year, IITAP judges are looking at how applicants answer specific questions about an instructional problem or challenge that they are working on solving, and which specific features of their practice, process, technology and provide students with an innovative or efficient way of gaining knowledge and the skills pertinent to the instructional goals.

INNOVATION, EFFECTIVENESS, RELEVANCE, AND EFFICIENCY AS A TEACHING AND/OR LEARNING PRACTICE, PROCESS, OR TECHNOLOGY

2. The teaching with technology practice provides an innovative solution to the stated teaching and/or learning problem and offers clear advantages over traditional techniques. New techniques, unique tools and methods are used to develop this teaching with technology practice. Refer to the Development Process question and Question 1.
3. Students gain new knowledge and skills using the teaching with technology practice and are now able to do something they could not do before.
4. The teaching with technology practice makes the learning experience and/or subject matter relevant to the students.
5. The teaching with technology practice increases teaching and learning efficiency for the learner, for the instructor, for the institution.

Scoring: Any whole number between 10–50 may be entered for each item, using the following as a guide:

Very poor = 10
 Poor = 20
 Satisfactory = 30
 Very good = 40
 Excellent = 50

Suggested definitions:

Very poor: Question is answered, but the answer is unclear and ineffective in describing what is requested; the solution to the problem is traditional or ineffective

Poor: Question is answered, but is only partially effective in communicating the essence of what is being asked; information is incomplete; solution to problem is quite traditional

Satisfactory: Question is answered clearly and succinctly; acceptable description; the technology or use of it is of acceptable quality, but is only somewhat effective, barely innovative

Very good: Question is answered clearly, succinctly; and the technology or use of it is of effective quality and performance; meets an instructional need and is somewhat innovative.

Excellent: Question is answered clearly and succinctly and the technology or use of it is of outstanding quality and performance; unique, very innovative, very efficient and very effective, an exemplar for others.

OUTCOMES FOR TEACHING WITH TECHNOLOGY

6. The assessment plan and design described in the entry form demonstrates that appropriate, valid, and thorough processes, instruments, and techniques were used to measure the effectiveness, relevance, and efficiency of the teaching with technology practice on teaching and/or learning.
7. The evaluation outcomes confirm the answer to Question 3 in the entry form and demonstrate that students have gained knowledge and skills and are now able to do something they couldn't do before. Supporting qualitative or quantitative data is provided.
8. The evaluation outcomes confirm the answer to Question 4 in the entry form and demonstrate that the teaching with technology practice makes the learning experience and/or subject matter relevant to the students. Supporting qualitative or quantitative data is provided.
9. The evaluation outcomes confirm the answer to Question 5 in the entry form and demonstrate that the teaching with technology practice increases efficiency for the learner, for the instructor, for the institution. Supporting qualitative or quantitative data is provided.

Scoring: Any whole number between 10–50 may be entered for each item, using the following as a guide:

Very poor = 10

Poor = 20

Satisfactory = 30

Very good = 40

Excellent = 50

Suggested definitions:

Very poor: Question is answered, but the answer is unclear and ineffective in describing what is requested; the solution to the problem is traditional or ineffective

Poor: Question is answered, but is only partially effective in communicating the essence of what is being asked; information is incomplete; solution to problem is quite traditional

Satisfactory: Question is answered clearly and succinctly; acceptable description; the technology or use of it is of acceptable quality, but is only somewhat effective, barely innovative

Very good: Question is answered clearly, succinctly; and the technology or use of it is of effective quality and performance; meets an instructional need and is somewhat innovative.

Excellent: Question is answered clearly and succinctly and the technology or use of it is of outstanding quality and performance; unique, very innovative, very efficient and very effective, an exemplar for others.

FUNCTIONALITY, DESIGN, INTERACTIVITY, AND QUALITY

- Technical operation (i.e. runs smoothly without crashes or hang-ups, easy to install)
- User interface design (i.e. clear navigation and usability; consistent, intuitive, clean, aesthetically pleasing)
- Engagement and interactivity (i.e. Are students now able to be more engaged with learning through interactivity? Are students engaged beyond page turning?)
- Content presentation (i.e. Is content clear, readable, well-organized, well-written, rich and deep—supporting linkages and the building of customized, relevant knowledge?)
- Media quality (high-quality graphics, video, sound, animations), if applicable
- Learner modalities are accommodated for (i.e. Are appropriate types of content and media included to create a rich practice, process, technology?)

Note: If applicants are integrating a custom-developed technology or tool, focus on how well applicants execute and use their interface, content design, and media quality. If applicants are integrating a commercially developed technology or tool, focus on how well their choice of tool or customization of it meets their goal, and how effectively the tool's performance, interface, and features engage students.

Scoring: Any whole number between -10–50 may be entered for each item, using the following as a guide:

Missing = -10

Not applicable = NA

Very poor = 10

Poor = 20

Satisfactory = 30

Very good = 40

Excellent = 50

Suggested definitions:

Missing: When a feature should be present as stated in the entry form but is not, for example critical navigational aids, missing content, media, text, etc.

Not applicable: When a feature is missing because it is not applicable or necessary in this particular context. This should not negatively affect scoring.

Very poor: Feature is present, but practically non-functional, ineffective; unacceptable quality and performance; needs great improvement; very traditional approach

Poor: Feature is not very effective, somewhat functional, needs improvement; traditional approach

Satisfactory: Feature is functional, acceptable quality, somewhat effective; barely innovative

Very good: Feature is of good solid quality and performance, effective and somewhat innovative

Excellent: Feature is of outstanding quality and performance; unique, very innovative, very efficient and very effective, an exemplar for others.

Appendix B: Resource Development Review: Detailed Instructions

This year, IITAP judges are looking at how applicants answer specific questions about an instructional problem or challenge that they are working on solving, and which specific features of their resource provide students with an innovative or efficient way of gaining knowledge and the skills pertinent to the instructional goals.

INNOVATION, EFFECTIVENESS, AND EFFICIENCY AS A RESOURCE

1. The resource provides an innovative solution to the stated teaching and/or learning problem and offers clear advantages over traditional techniques. New techniques, unique tools and methods are used to develop this resource. Refer to the Development Process question and Question 1.
2. Students gain new knowledge and skills using the resource and are now able to do something they could not do before.
3. The resource increases teaching and learning efficiency for the learner, for the instructor, for the institution.

Scoring: Any whole number between 10–50 may be entered for each item, using the following as a guide:

Very poor = 10
 Poor = 20
 Satisfactory = 30
 Very good = 40
 Excellent = 50

Suggested definitions:

Very poor: Question is answered, but the answer is unclear and ineffective in describing what is requested; traditional or ineffective solution to problem

Poor: Question is answered, but is only partially effective in communicating the essence of what is being asked; information is incomplete; solution to problem is quite traditional

Satisfactory: Question is answered clearly and succinctly; acceptable description; resource is of acceptable quality, but is only somewhat effective, barely innovative

Very good: Question is answered clearly, succinctly, and the resource is of effective quality and performance; meets an instructional need and is and somewhat innovative

Excellent: Question is answered clearly and succinctly and the resource is of outstanding quality and performance; unique, very innovative, very efficient and very effective, an exemplar for others.

OUTCOMES FOR RESOURCE DEVELOPMENT

4. The assessment plan and design described in the entry form demonstrates that appropriate, valid, and thorough processes, instruments, and techniques were used to measure the effectiveness and efficiency of the resource on teaching and/or learning.
5. The evaluation outcomes confirm the answer to Question 3 in the entry form and demonstrate that students have gained knowledge and/or skill(s) and are now able to do something they couldn't do before. Supporting qualitative or quantitative data is provided.
6. The evaluation outcomes confirm the answer to Question 4 in the entry form and demonstrate that the resource increases efficiency for the learner, for the instructor, for the institution. Supporting qualitative or quantitative data is provided.

Scoring: Any whole number between 10–50 may be entered for each item, using the following as a guide:

Very poor = 10
Poor = 20
Satisfactory = 30
Very good = 40
Excellent = 50

Suggested definitions:

Very poor: Question is answered, but the answer is unclear and ineffective in describing what is requested; traditional or ineffective solution to problem

Poor: Question is answered, but is only partially effective in communicating the essence of what is being asked; information is incomplete; solution to problem is quite traditional

Satisfactory: Question is answered clearly and succinctly; acceptable description; resource is of acceptable quality, but is only somewhat effective, barely innovative

Very good: Question is answered clearly, succinctly, and the resource is of effective quality and performance; meets an instructional need and is and somewhat innovative

Excellent: Question is answered clearly and succinctly and the resource is of outstanding quality and performance; unique, very innovative, very efficient and very effective, an exemplar for others.

FUNCTIONALITY, DESIGN, INTERACTIVITY, AND QUALITY

- Technical operation (i.e. runs smoothly without crashes or hang-ups, easy to install)
- User interface design (i.e. clear navigation and usability; consistent, intuitive, clean, aesthetically pleasing)
- Engagement and interactivity (i.e. Are students now able to be more engaged with learning through interactivity? Are students engaged beyond page turning?)
- Content presentation (i.e. Is content clear, readable, well-organized, and well-written?)
- Media quality (high-quality graphics, video, sound, animations), if applicable
- Learner modalities are accommodated for (i.e. Are appropriate types of content and media included to create a rich resource?)

Note: If applicants are integrating a custom-developed technology or tool, focus on how well applicants execute and use their interface, content design, and media quality. If applicants are integrating a commercially developed technology or tool, focus on how well their choice of tool or customization of it meets their goal, and how effectively the tool's performance, interface, and features engage students.

Scoring: Any whole number between -10–50 or N/A may be entered for each item, using the following as a guide:

Missing = -10

Not applicable = NA

Very poor = 10

Poor = 20

Satisfactory = 30

Very good = 40

Excellent = 50

Suggested definitions:

Missing: When a feature should be present as stated in the entry form but is not, for example critical navigational aids, missing content, media, text, etc.

Not applicable: When a feature is missing because it is not applicable or necessary in this particular context. This should not negatively affect scoring.

Very poor: Feature is present, but practically nonfunctional, ineffective, unacceptable quality and performance; needs great improvement, very traditional approach

Poor: Feature is not very effective, only somewhat functional, needs improvement; traditional approach

Satisfactory: Feature is functional, acceptable quality, somewhat effective; barely innovative

Very good: Feature is of good solid quality and performance, effective and somewhat innovative

Excellent: Feature is of outstanding quality and performance; unique, very innovative, very efficient and very effective; an exemplar for others.

Appendix C

Program Staff Interview Protocol

The first-round review process:

1. What is involved in the first round review process?
2. How much time is expected of reviewers for them to complete the procedure?
3. What are the necessary qualifications for the first-round reviewers?
4. How do you find the first-round reviewers?
5. Did you ever have any difficulties or problems in the past with the first-round reviewers?
If so, what kinds of problems did you have?
6. What process should the reviewers follow?

The second-round judging process:

1. What steps are involved in the second-round judging process?
2. How much time is expected of judges for them to complete their evaluations?
3. How much time are they expected to spend for each step?
4. How do you find the second-round judges?
5. Did you ever have any difficulties or problems in the past with the second-round judges?

Appendix D:

IITAP Reviewer Survey

IITAP Reviewer Survey	Exit this survey >>
<p>1. Welcome!</p> <p>IITAP program staff and evaluation staff would like to improve the efficiency and accuracy of IITAP entry review process. To that end, we would like to get your input.</p> <p>The survey is anonymous and should take you about 5-15 minutes.</p> <p>If you are willing to participate, just click the "next" button below. Please submit your survey by March 10.</p> <p style="text-align: center;">Next >></p>	

IITAP Reviewer Survey
<p>1. The time I spent for reviewing each entry, on average, was:</p> <p><input type="radio"/> about 1 hour</p> <p><input type="radio"/> about 1.5 hours</p> <p><input type="radio"/> about 2 hours</p> <p><input type="radio"/> about 2.5 hours</p> <p><input type="radio"/> about 3 hours</p> <p><input type="radio"/> about 3.5 hours</p> <p><input type="radio"/> Other (please specify)</p> <div style="border: 1px solid black; width: 150px; height: 15px; margin-left: 20px;"></div>
<p>2. The time I spent for reviewing each entry was:</p> <p><input type="radio"/> too short</p> <p><input type="radio"/> somewhat short</p> <p><input type="radio"/> about right</p> <p><input type="radio"/> somewhat long</p> <p><input type="radio"/> too long</p>
<p>3. I had enough time for the IITAP review within my work schedule.</p> <p><input type="radio"/> strongly agree</p> <p><input type="radio"/> agree</p> <p><input type="radio"/> neutral</p> <p><input type="radio"/> disagree</p> <p><input type="radio"/> strongly disagree</p>

IITAP Reviewer Survey

4. In order to complete the IITAP reviews, I had to (check all that apply):

- review entries more quickly than I would have liked
- sacrifice other work
- work overtime (including weekends)
- Other (please specify)

IITAP Reviewer Survey

5. I was well qualified to do this review task.

- strongly agree
- agree
- neutral
- disagree
- strongly disagree

6. In your opinion, what knowledge and skills must IITAP reviewers have?

7. What process or strategies did you use to improve efficiency and/or accuracy?

IITAP Reviewer Survey

8. In general, the review process was efficient.

- strongly agree
- agree
- neutral
- disagree
- strongly disagree

9. I looked at the following materials before I started (check all that apply):

- IITAP 2006 Reviewer Information (given/sent by Suzanne)
- Teaching with Technology Detailed Review Instructions (given/sent by Suzanne)
- Resource Development Review: Detailed Instructions (given/sent by Suzanne)
- Teaching with Technology Examples of Past Winners (on the DIIA website)
- Resource Development Examples of Past Winners (on the DIIA website)
- Judging Considerations (on the DIIA website)

10. Instructions for the review task were clear to me.

- strongly agree
- agree
- neutral
- disagree
- strongly disagree

IITAP Reviewer Survey

11. The online review submission system was easy to use.

- strongly agree
- agree
- neutral
- disagree
- strongly disagree

12. Please give us suggestions for improving the online review submission system.

13. The reviewer orientation session was helpful.

- did not go
- strongly agree
- agree
- neutral
- disagree
- strongly disagree

14. Please give us suggestions for improving the reviewer orientation session.

IITAP Reviewer Survey

15. The most time consuming part of the reviewing process was:

16. The most difficult part of the reviewing process was:

17. Please give us suggestions for improving the efficiency and/or accuracy of the reviewing process.

Appendix E: IITAP Judge Survey

[Exit this survey >>](#)

IITAP Judge Survey

1. Welcome!

Thank you very much for serving as a judge for IITAP.

IITAP program staff and evaluation staff would like to improve the efficiency and accuracy of IITAP entry judging process. To that end, we would like to get your input.

The survey is anonymous and should take you about 5-15 minutes.

If you are willing to participate, just click the "next" button below. Please submit your survey by May 12.

[Next >>](#)

[Exit this survey >>](#)

IITAP Judge Survey

1. The time I spent for judging each entry, on average, was:

- less than 1 hour
- about 1 hour
- about 1.5 hours
- about 2 hours
- about 2.5 hours
- about 3 hours
- Other (please specify)

2. The time I spent for judging each entry was:

- too short
- somewhat short
- about right
- somewhat long
- too long

[<< Prev](#) [Next >>](#)

IITAP Judge Survey [Exit this survey >>](#)

3. In general, the judging process was efficient.

strongly agree
 agree
 neutral
 disagree
 strongly disagree

4. I looked at the following materials before I started (check all that apply):

IITAP 2006 Judging Information (document given/sent by Suzanne)
 Teaching with Technology Scoring Instructions (document given/sent by Suzanne)
 Resource Development Scoring Instructions (document given/sent by Suzanne)
 The Entry Guidelines/2006 Call for Entries (on the DIIA website)
 Teaching with Technology Examples of Past Winners (on the DIIA website)
 Resource Development Examples of Past Winners (on the DIIA website)

5. Instructions for the judging task were clear to me.

strongly agree
 agree
 neutral
 disagree
 strongly disagree

[<< Prev](#) [Next >>](#)

IITAP Judge Survey [Exit this survey >>](#)

6. The online judge score submission system was easy to use.

strongly agree
 agree
 neutral
 disagree
 strongly disagree

7. Please give us suggestions for improving the online judge score submission system.

[<< Prev](#) [Next >>](#)

IITAP Judge Survey Exit this survey

8. The most time consuming part of the judging process was:

9. The most difficult part of the judging process was:

10. Please give us suggestions for improving the efficiency and/or accuracy of the judging process.

[<< Prev](#) [Next >>](#)

Appendix F:

IITAP Reviewer Survey: Responses to Open-Ended Questions

Results (n = 7)

This appendix lists all reviewers' unedited responses to open-ended questions.

Question 6: In your opinion, what knowledge and skills must IITAP reviewers have?

- Knowledge on website design, instructional design, and assessment
- Technical knowledge, assessment knowledge pedagogical knowledge
- Learning, technology use/innovation, and assessment practices
- Optimally, a reviewer would be very comfortable with various types of technology. In addition, a reviewer must have some knowledge of effective teaching principles.
- Instructional and technological experience, the ability to imagine the student's perspective, a sense for when to give the 'benefit of the doubt' for technology's sake, an appreciation for visual and pedagogical design, the ability to write critiques that are helpful, true, and as kind as possible.
- Experience reviewing IT proposals helps a lot - Ability to recognize what is 'innovative' - Ability to read past faculty BS and call them on it - Familiarity with the content (i.e., get someone with a science background to review science content, etc),
- Technical skills and some information about teaching and learning.

Question 7: What process or strategies did you use to improve efficiency and/or accuracy?

- I created a table to keep track of the scores that I gave to each category of each entry. This way, when I wanted to check the rating I gave in order to confirm my consistency in rating, I could refer to it very easily and quickly.
- Viewed ALL of the ten entries first, then scored them all at the end.
- I copied responses for questions when I knew the system was about to force me to log back in. This enabled me to not have to rewrite my comments. In my attempt to be thorough, I typically reviewed a write-up, looked over the submission (often in great detail), and then provided feedback to the various questions (which sometimes necessitated reviewing the materials again).
- I printed out all the entry descriptions, then looked over the technology of the entries themselves, making notes on the printouts where I had questions or comments. Once I had done that, I walked through the online entry form, entering the comments I had written on the printout and fleshing them out enough so they made sense to someone besides me. I did one entry at a time.
- Setting aside enough time to do it!! Get started early and not wait until the last week!!!! (1) read proposal and enter initial impression scores (2) thoroughly examine all materials (3) re-read proposal more carefully again (4) enter comments, re-examine materials (5) reconsider scores (6) go over all scores to check for consistency across proposals. Sending out the Excel sheet of all scores was very, very helpful!!

- I read through all the information and accessed all the technology first and took notes. I think went back and did the evaluations.

Question 12: Please give us suggestions for improving the online review submission system.

- It was great. One thing to improve is perhaps to allow users to move to certain rating page of other entries directly. For example, when I am reviewing the question 2 of entry 5, I may want to look at the rating I gave to question 2 of entry 3. This function will make the rating much easier and accurate (consistent).
- I think it was useful for entering scores but I really prefer to read the entries on paper and would have preferred a paper scoring sheet so I could more readily compare projects and go back and make changes more easily. Next year, I'd like to see a hybrid system--paper scoring sheets that we then enter into the electronic system.
- I like the online review system, but often have problems with the restrictions of having to be online and timing out of the system.
- Lengthen the time period before time-outs.
- The EID timeout killed me. Several times, I had written somewhat lengthy comments, clicked submit, and poof they were gone. I got in the habit of copying my comments before clicking submit, so I could paste them back in when the EID thing timed out on me, but that is awkward and distracting.

- It was very cumbersome to have to page thru all the questions to change a score in the middle or end. It would help to be able to see all your scores for each and all proposals and be able to tweak scores just like you could in an excel sheet!!
There is no feedback to make sure you entered or submitted a score and it was saved. Going backwards does not save the score, you always have to hit submit and next, so that was awkward.
- Allow an area for general comments. Provide an online view all results for each entry so that you can look across the projects.

Question 14: Please give us suggestions for improving the reviewer orientation session.

- Hands-on Training in rating, different raters will give similar ratings.
- I think it would be helpful for the group to review a project together or in a small group. Perhaps you could send a previous years' entry to everyone beforehand, have us “review” it, and then we could discuss it as a group. An alternative would be to do a mini review by briefly demonstrating a previous years' project and then having us score a pedagogical question, an assessment question, and a technical question and then discuss our scores as a group.
- Maybe go over a fake proposal and have everyone score it or discuss how they should score the various questions/scenarios so everyone is on the same page with what to look for, what to watch out for, and what you expect?

Question 15: The most time consuming part of the reviewing process was:

- going over the products, especially when websites have many pages, took a lot of time.
- testing the products
- perusing the materials
- reviewing submissions in detail.
- looking through each entry and comparing my perception with what was written about it
- examining all the materials per submission
- reviewing the entries and figuring out how the entries worked.

Question 16: The most difficult part of the reviewing process was

- being consistent in rating
- scoring the technical aspects
- making the time to do this
- getting certain projects/examples to properly run on a given platform
- finding the time
- making sure nothing else got put on my plate that got in the way of reviewing!
Wondering if I was thorough enough, too critical, or too easily impressed, writing comments, trying to remember what I scored for each project. Getting the Excel sheet of all scores was very, very helpful!!
- Ensuring that your results were adequate.

Question 17: Please give us suggestions for improving the efficiency and/or accuracy of the reviewing process.

- Again, giving some sort of training for reviewers will be helpful. Also providing examples of rating (based on last year's rating) will be helpful.
- Often it was confusing to me whether I should be (1) reviewing a proposal, (2) reviewing supporting content, (3) reviewing content in the context of a whole lesson plan or class (the overall teaching strategy), (4) whether I should make suppositions about learning that likely occurred, but not have been adequately addressed in a proposal.
- Finding the time was the hardest part, since it really requires a dedicated chunk of attention. If possible, more reviewers, each with fewer entries to review. What about the DIIA faculty advisory committee or the student advisory committee?
- Get people's commitments for time and make them stick to it. Getting weekly reminders or asking for progress was very good, we need that. Allowing reviewers to discuss their same proposals? Maybe that would be cheating though.
- Provide online resource to view all results for your reviews. Have fewer items to review. Maybe team review to enable discussion of various aspects of the systems.