

Assessing Collaborative Online Inquiry and Social Deliberation in Digital Environments

Jesse R. Sparks, Educational Testing Service, jsparks@ets.org
Julie Coiro, University of Rhode Island, jcoiro@snet.net
Jill M. Castek, University of Arizona, jcastek@email.arizona.edu
Carita Kiili, University of Oslo, Norway, c.p.s.kiili@iped.uio.no
Beth R. Holland, University of Rhode Island, beth@brholland.com
Changhee Lee, University of Rhode Island, lechhe86@gmail.com

Abstract: We describe efforts to design and validate a digitally-based assessment of collaborative online inquiry and social deliberation using a digital virtual world platform with embedded supports for real-time collaboration. Cognitive validity studies were conducted to examine the collaborative prompts and the overall task, with 21 dyads participating in either Face-to-Face (FTF; n=5) or Computer-Mediated (CM; n=16) conditions. Quantitative results suggest the task captured variation in dyads' inquiry performances and processes.

Assessment of 21st century skills: Online inquiry and collaboration

Critical thinking, complex problem solving, and collaboration are required 21st century skills for success in college and the workforce. Current large-scale assessments in the United States do not fully represent the range of 21st century skills associated with the complexities of digital literacy and information-based problem solving in collaborative, networked environments; thus, there has been increasing interest in the development of assessments that capture valid evidence of these constructs. This project aimed to develop and validate an assessment of students' collaborative online inquiry and social deliberation skills, using a digital environment for collaboration.

Small-scale cognitive validity study

We developed an assessment of online inquiry skills in a collaborative context by adapting an existing individual scenario-based virtual world assessment task for use with dyads. Evidence-centered design (ECD) principles (Mislevy, Almond, & Lukas, 2003) were applied to expand our online inquiry construct definitions, evidence collection, and task designs to incorporate collaborative work (Coiro, Sparks & Kulikowich, 2018). This included adding explicit prompts to collaborate, delivered through a digital environment designed to support real-time remote collaboration (Hao et al., 2017). We conducted several small-scale studies to examine the utility and validity of the collaborative prompts and the collaborative task, with 21 dyads participating in either Face-to-Face (FTF) or Computer-Mediated (CM) conditions. We collected multiple sources of evidence of students' inquiry proficiency, including item responses, moment-to-moment actions (recorded in log files), and students' real-time conversational dialogue (audio/video, text-based chat) as they worked collaboratively to complete the online inquiry task. Here, we present preliminary quantitative evidence of task performance.

Methods

Participants

Participants were rising or current 9th and 10th grade students (mean age=14.8 years, range: 14-16) in the Northeastern U.S. Altogether 21 dyads (42 individuals; 27 females, 15 males) participated across four rounds of data collection, in FTF (n=5) and CM (n=16) conditions, with CM groups being divided across data collection phases, including play testing (n=5), in-school tryouts (n=6), and laboratory-based tryouts (n=4).

Collaborative virtual world task

The collaborative scenario-based virtual world research task engaged dyads in locating, evaluating, reading, and synthesizing information from multiple sources by incorporating evidence from those sources into an overall response to an inquiry question (i.e., whether or not an artifact should be placed in a museum based on its historical accuracy). The task has three phases (see Coiro et al., 2018): *Setup* (scenario and task introduction), *Free Roam* (exploring, gathering and evaluating available resources), and *Conclusion* (apply information from collected sources to construct an overall response). Dyads also completed an *Oral Presentation Task* summarizing their overall conclusions and using information from key resources to support their reasoning. Responses and actions in the task were scored based on ECD documentation (e.g., credit for correct answers or for actions that move students closer to a correct solution to the inquiry task), yielding a total of 87 points, with subscores computed by

task phase (Setup = 12 points, Free Roam = 51 points, Conclusion = 24 points) and by inquiry construct (Planning = 6 points, Locating = 22 points, Evaluating = 35 points, Synthesizing = 24 points).

Preliminary results

Quantitative analysis

Mean scores for the total task, task phase (Setup, Free Roam, and Conclusion), as well as proportion correct for each inquiry subskill ($P+$, i.e., raw score divided by maximum points per subskill of Plan, Locate, Evaluate, and Synthesize) and total score are presented (see Table 1). Performance varied across the dyads, with scores in the Free Roam phase (i.e., locating and evaluating tasks) showing greatest variability. Conclusion Phase/Synthesis subscores also varied, in part because some dyads did not finish. Altogether 17 of 21 dyads completed the main task, and 12 of 21 completed the culminating oral presentation task. The task captured variability in inquiry scores.

Table 1: Mean Scores (Standard Deviations in Parentheses) and Subscores by Task Phase and Inquiry Subskill

Condition	N	Total Scores (max=87)	Setup (max=12)	Free Roam (max=51)	Conclusion (max=24)	Total Score ($P+$)	Plan ($P+$)	Locate ($P+$)	Evaluate ($P+$)	Synthesize ($P+$)
FTF-Playtest	5	54.30 (7.18)	9.00 (0.79)	26.00 (7.72)	19.30 (1.89)	.62 (.08)	.83 (.12)	.59 (.17)	.49 (.14)	.80 (.08)
CM-Playtest	6	49.67 (9.30)	9.58 (0.80)	24.92 (5.48)	15.17 (6.64)	.57 (.11)	.86 (.13)	.54 (.18)	.50 (.08)	.63 (.28)
CM-School	6	43.83 (11.47)	9.00 (1.48)	23.67 (4.63)	13.40 (4.60)	.50 (.13)	.86 (.07)	.58 (.11)	.42 (.09)	.56 (.19)
CM-Lab	4	57.25 (4.99)	9.63 (1.11)	27.50 (4.14)	20.13 (2.43)	.66 (.06)	.79 (.21)	.60 (.10)	.55 (.09)	.84 (.10)
Total	21	50.55 (9.74)	9.29 (1.06)	25.31 (5.41)	16.75 (5.06)	.58 (.11)	.84 (.12)	.58 (.14)	.48 (.10)	.70 (.21)

Note: $P+$: proportion correct, or the dyads' earned score divided by the maximum possible points for each skill.

Discussion

Preliminary analysis of quantitative scores, capturing dyad's performance with collaborative online inquiry and social deliberation, revealed that the scenario-based virtual world task seems to elicit a range of performances from student dyads. Performance was best on planning and synthesis tasks, with moderate performance on evaluate and locate tasks. All dyads experienced more difficulty with locating and evaluating tasks, and several lost points due to inefficient or ineffective allocation of time across multiple resources in the task, such that some failed to finish in the allotted 2.5 hour timeframe. Overall, however, the preliminary results indicate that the inquiry task was feasible for students to use and presented an appropriate level of challenge. Relationships among collaborative processes, dyad characteristics, and the quantitative scores reported above will also be discussed during the poster presentation, with implications for future research on collaborative assessments of online inquiry and social deliberation, as relevant to the CSCL community and to the design of collaborative assessments.

References

- Coiro, J., Sparks, J.R., & Kulikowich, J.M. (2018). Assessing online reading comprehension, collaborative inquiry and social deliberation across multiple sources and perspectives. In J.L.G. Braasch, I. Bråten, & M.T. McCrudden (Eds.), *Handbook of Multiple Source Use* (pp. 485-517). Routledge.
- Hao, J., Liu, L., von Davier, A.A., Lederer, N., Zapata-Rivera, D., Jakl, P., & Bakkenson, M. (2017). EPCAL: ETS Platform for Collaborative Assessment and Learning. *ETS Research Report Series* (RR-17-49), 1-14. doi:10.1002/ets2.12181
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *Research Report 16-2003*. Educational Testing Service: Princeton, NJ.

Acknowledgments

The authors express gratitude to Colleen Appel, Eowyn Winchester, Zhitong (Lin) Yang, and Wen Wen for their contributions to this research, and to the reviewers who provided feedback on earlier versions of this paper.