

# Measuring Critical Thinking in *Reacting to the Past*

Patrick Ludolph

Georgia Gwinnett College

Teaching History 48(1)

DOI: 10.33043/TH.48.1.91-105

© 2023 Ludolph.

Distributed with CC BY-NC-ND 4.0 License.

The literature on *Reacting to the Past* has generally been laudatory.<sup>1</sup> For those unfamiliar with *Reacting*, it is a pedagogy originally developed by Mark Carnes at Barnard College in the 1990s, eventually spreading to college faculties across the US and beyond. It requires students to take on the role of historical figures. Students are expected to do extensive reading, preparation, writing, and speaking during these games, taking on the perspective of their characters during key historical events.

One of the central contentions scholars have made is that *Reacting* can help improve students' critical thinking. Unfortunately, there is little consensus on what critical thinking is or how to measure it. Most works on *Reacting* rely on student self-report (from class discussion, surveys, or focus groups) or faculty observations.<sup>2</sup> However, without both faculty and students having a clear idea of what critical thinking is, it is doubtful that either group can be trusted to report its development. In addition, even if we could trust the largely anecdotal evidence, how could we compare it? For instance, how could we know if *Reacting* is as effective in promoting critical thinking as writing a research paper? To begin to address these problems, I conducted a preliminary study in Spring 2023 measuring the effects of *Reacting* on critical thinking using a standardized assessment, the CAT. I assessed nine students in an upper division history course at Georgia Gwinnett College, a small, public, open enrollment college. Students took a critical thinking assessment before and after playing two role-playing games. Students also took two surveys gathering more data on their experiences in the course and skill development. This study finds that *Reacting to the Past* led to improved critical thinking, but not evenly across subskills. The data suggests that *Reacting* can be a potent intervention for developing creative thinking and effective communication. The study also finds that, while student self-report can be useful, it was not reliable in predicting critical thinking scores. These findings demonstrate the value of using a standardized assessment to measure critical thinking, but future studies will require greater participation.

## Assessing Critical Thinking

Critical thinking is supposed to be a core element of a college education. Unfortunately, as often as critical thinking appears in college pamphlets, there is still little consensus about how best to define, teach, or assess it. Research suggests that, while college does improve students' critical thinking, much remains to be done.

In *Our Underachieving Colleges* (2006), Derek Bok makes the case that American schools have lost their premier position among global academies; American schools should be doing more to improve writing, critical thinking, foreign language acquisition, quantitative reasoning, and citizenship.<sup>3</sup> Bok argues that students, faculty, and the changing world all bear some responsibility. Students are more distracted by technology and entertainment than ever—and studying less than ever as a result. Faculty are relying too much on lecture and both short answer and multiple-choice exams. In this, both faculty and students collude, because these methods are easier for both parties. Students will use their memory to cover over not understanding underlying concepts and faculty have less

---

1 For example, Jacqueline A. Gilbert, "Reacting to the Past Drives Business Students' Engagement," *Journal of Excellence in College Teaching* 32, 3 (2021): 161-91; Carl A. Anderson and T. Keith Dix, "'Reacting to the Past' and the Classics Curriculum: Rome in 44 BCE," *The Classical Journal* 103, 4 (2008): 449-55; Thomas C. Buchanan, Nicole Rarulevicz, and Edward Palmer, "Reacting to the Past in Australasia: From Early Adoption to COVID-19," *Australasian Journal of American Studies* 39, 1 (December 2020): 208-24; Kevin Burke, "Roleplaying Music History: Honing General Education Skills via 'Reacting to the Past,'" *Journal of Music History Pedagogy* 5, 1 (2014): 1-21. In addition, virtually every other article or book cited on *Reacting* in this article finds it to be a valuable pedagogy.

2 For example, see Christine L. Albright, "Harnessing Students' Competitive Spirit: Using *Reacting to the Past* to Structure the Introductory Greek Culture Class" *The Classical Journal* 112, 3 (2017): 364-79. For further examples, see footnote 11. Of course, student self-report and faculty observations are valuable, but an objective assessment offers different insights.

3 Derek Bok, *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More* (Princeton: Princeton University Press, 2006), 1-10.

to grade. Developing critical thinking requires active learning, in which faculty pose problems, challenge student answers, encourage students to apply learned concepts in new situations, expose misconceptions, and regularly evaluate and provide feedback to students.<sup>4</sup>

Richard Arum and Josip Roksa's *Academically Adrift* makes a similar case to Bok. The researchers found that student critical thinking is only minimally improving over the course of two years of college education. However, they also considered many variables to try to explain and even predict improvement. They found that certain variables, to some extent controllable by faculty, could predict scores both within and between institutions. Namely, they found that students who reported taking classes that required more than 40 pages of reading per week and 20 or more pages of writing per semester were strongly correlated with greater improvement. They also found a strong positive correlation with the number of hours spent studying—alone. There was in fact a negative correlation with more hours students spent studying with peers.<sup>5</sup> The reading and writing findings bode well, but this last factor can call into question the extent to which *Reacting* can spur growth in critical thinking. It is important to note that these were correlations, not causations, so there may be other factors that encourage more studying alone or course selection for more rigorous reading and writing courses. Arum and Roksa relied on a standardized measure, the Collegiate Learning Assessment (CLA, now CLA+), that could be administered in the same way across college campuses twice for more than 2000 students. Their study demonstrates the value of objective assessments in bringing to light ways to target critical thinking development.

The study in this paper used the Critical thinking Assessment Test (CAT) developed by the Center for the Assessment and Improvement of Learning (CAIL) at Tennessee Technological University. The CAT has 15 questions, each exploring one or more skills, broken down into four major categories: evaluating information, creative thinking, learning and problem solving, and communication.<sup>6</sup> Each of these categories is further refined:

Evaluating and Interpreting Information	Separate factual information from inferences. Interpret numerical relationships in graphs. Understand the limitations of correlational data. Evaluate evidence and identify inappropriate conclusions.
Creative Thinking	Identify alternative interpretations for data or observations. Identify new information that might support or contradict a hypothesis. Explain how new information can change a problem.
Learning and Problem Solving	Separate relevant from irrelevant information. Integrate information to solve problems. Learn and apply new information, Use mathematical skills to solve real-world problems.
Communication	Communicate ideas effectively.

Figure 1: The CAT's definition of critical thinking

There is no single consensus definition of critical thinking. For instance, the above definition conflicts with

<sup>4</sup> Bok, *Our Underachieving Colleges*, 110-27.

<sup>5</sup> Richard Arum and Josipa Roksa, *Academically Adrift: Limited Learning on College Campuses* (Chicago: University of Chicago Press, 2011), 93-95, 100-101.

<sup>6</sup> It has been validated using multiple methods: Ada Haynes, Elizabeth Lisic, Kevin Harris, Katie Leming, Kyle Shanks, and Barry Stein, "Using the Critical Thinking Assessment Test (CAT) as a Model for Designing Within-Course Assessments: Changing How Faculty Assess Student Learning," *Inquiry: Critical Thinking Across the Disciplines* 30, 3 (2015): 38-40.

the Delphi Report, one of the more commonly accepted definitions. According to the Delphi Report, “not every valuable thinking skill is CT skill. CT is one among a family of closely related forms of higher-order thinking, along with, for example, problem-solving, decision making, and creative thinking.”<sup>7</sup> This would appear to contradict two of the four categories assessed by the CAT. However, closer examination of CAIL’s definition of creative thinking shows that it overlaps significantly with Delphi’s definition of inferences—querying evidence, conjecturing alternatives, and drawing conclusions.<sup>8</sup> Similarly, one could make a case that much of the CAT’s problem solving could be pieced together from skills identified in the Delphi Report. The Delphi Report also identifies areas not included in the CAT’s definition at all, such as a more elaborate apparatus for evaluating communication or a dispositional element: self-regulation. However, even without a complete consensus, the CAT is a valuable tool for assessing critical thinking.

Beyond being a validated assessment, the CAT was used in this study for three reasons. First, my home institution, Georgia Gwinnett College (GGC), uses the CAT in assessment, making it possible to draw broader comparisons between experimental groups and the larger student population. Second, CAIL has developed a pedagogy for improving critical thinking based on the CAT: CAT Applications or CAT-apps. Third, as a result of the first two reasons, I was part of a research group that tested the effectiveness of CAT-apps. We carried out a semester-length longitudinal study of 208 GGC students in General Education courses and one upper division business course. A control group (98 students) was given no extra critical thinking training while an experimental group (110 students) completed and scored two CAT-apps. Both groups were assessed with the CAT at the start and end of the semester.<sup>9</sup>

The study found that the experimental group benefitted significantly from the exercises. The control group improved their average score by .98 points. We considered this to be the practice effect, a consequence of taking the exam twice in the same semester. The experimental group that completed two of the CAT-apps improved their scores by an average of 2.38 points, or 1.4 points beyond the control group, a 10.6% increase. The gap between freshmen and seniors at GGC from 2015-2019 was only 1.97 points. Thus in one semester, the effect of two CAT-Apps was equivalent to more than two-thirds of the critical thinking gains of a college career.<sup>10</sup>

The above study was a roadmap for the one reported in this paper with some important caveats. First, as the current one is a small, preliminary study, I will not be able to do more than discuss what the evidence suggests and provide direction for a much larger study in the future. Second, the CAT and *Reacting* are foundationally untethered. Whereas the CAT-apps were developed with the express purpose of focusing on skills addressed by the CAT, *Reacting* requires a wide range of skills, many of which may not be captured by the CAT and may not even be properly considered critical thinking. It is difficult to see, for instance, where the ability to cope with distress would fit into a definition of critical thinking. Alternatively, while perspective-taking is an important skill in historical thinking and valuable to critical thinking, learning how to think like a sans-culottes mob would not seem to fit into CAIL’s definition of critical thinking. *Reacting* games are not made with the CAT in mind, and the CAT will only measure a piece of what *Reacting* offers. It is a test of particular (rather than the total) premises of both using the others’ measure: that the CAT can assess critical thinking across disciplines and that *Reacting* helps students develop their critical thinking.

---

7 Peter Facione, *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction, Executive Summary: “The Delphi Report”* (Millbrae, CA: California Academic Press, 1990), 5. <https://eric.ed.gov/?id=ED315423>.

8 Haynes, et al., “Using the Critical Thinking Assessment Test,” 39; Facione, “The Delphi Report,” 9.

9 Unfortunately, my own courses were not able to be included because of the arrival of the pandemic.

10 Tom Lilly, Pratima Darr, Matthew Schmolesky, Todd Lindley, Patrick Ludolph, Marieke Schilpzand, Young Shim, Rebecca Higgins, Aurelie Weinstein, Lior M. Burko, and Daniel von Deutsch, “Intending to Teach Critical Thinking: A Study of the Learning Impacts over One Semester of Embedded Critical Thinking Learning Objects,” *Journal of the Scholarship of Teaching and Learning* 22, 3 (September 2022): 34-54. <https://scholarworks.iu.edu/journals/index.php/josotl/article/view/31801>.

## Reacting and Critical Thinking

Much of the literature on *Reacting* indicates that it is helpful for critical thinking. However, few studies give a clear, working definition of critical thinking. It may be an entry on a rubric, or students may be asked to rate how much they had to use critical thinking, but students do not appear to have been given a definition to consider.<sup>11</sup> Students' understanding of critical thinking tends to be very vague; a recent survey of students at Georgia Gwinnett College about critical thinking suggests that students have a very limited idea of what it is.<sup>12</sup> It seems doubtful that students would be able to reliably identify which of their thinking was "critical" without further prompting.

Generally, *Reacting* researchers have tended to discuss critical thinking as being associated with three areas: primary source analysis, writing or otherwise forming arguments, and perspective-taking.<sup>13</sup> Paula Lazrus and Gretchen McKay, in discussing *Reacting*, argue that "students learn to make reasoned arguments and support their ideas with textual and factual information, and that is the heart of critical thinking."<sup>14</sup> Richard Powers, John Burney, and Mark Carnes have argued that *Reacting* "educates students in critical thinking by requiring them to advance claims and reasons, and to conduct research to develop supporting evidence while also accounting for alternative points of view." Students must do so in a "dramatic context" while trying to solve "messy" problems.<sup>15</sup> One of the more interesting definitions has come from Mark Carnes. He argued that the core of critical thinking was leaving the self, at least temporarily. For Carnes, the Socratic method was a mainstay of education because it could be used to break down the self through relentless critical examination. This process is very unpleasant. Instead, Carnes argues that *Reacting* students can develop their critical thinking by "adding" new selves; not destroying their old self, but temporarily taking on an alternate self.<sup>16</sup> Frequently, though, studies do not provide a clear definition of critical thinking, and virtually none have used a definition founded in critical thinking pedagogical research.

There have been few studies that have used standardized measures to quantify the effects of *Reacting* on students in areas outside of critical thinking, but they show the value of using standardized measures. Steven

---

11 Marie Gasper-Hulvat, David M. Dees, and Anthony V. Shreffler, "Eliciting Meaningful Engagement in an Art History Survey Course: Reacting to the Past and Active Learning," in *Playing to Learn with Reacting to the Past: Research on High Impact, Active Learning Practices*, ed. C. Edward Watson and Thomas Chase Hagood (Cham, Switzerland: Palgrave Macmillan, 2018), 119-20; Thomas Chase Hagood, Naomi J. Norman, Hyeri Park, and Brittany M. Williams, "Playing and Learning and Teaching in Higher Education: How Does *Reacting to the Past* Empower Students and Faculty?" in *ibid.*, 170; Tracy Lightcap, "Creating Political Order: Maintaining Student Engagement through *Reacting to the Past*," *PS: Political Science and Politics* 42, 1 (Jan. 2019): 175-79.

12 Marieke Schilpzand, Matthew Schmolesky, and Patrick Ludolph, "Student and faculty perspectives on critical thinking: a qualitative and quantitative analysis" (Georgia Gwinnett College, Lawrenceville, GA: *Teaching, Learning, and Research Symposium*, 2022). Another study found that students and academics in Education have largely similar concepts of critical thinking, though students tend to emphasize the product and academics tend to emphasize the process and disposition: Margaret Lloyd, "Thinking Critically about Critical Thinking in Higher Education," in *International Journal for the Scholarship of Teaching and Learning* 4, 2 (2010). <https://eric.ed.gov/?id=EJ1136134>.

13 Among historians, this has frequently taken the form of "historical thinking" rather than critical thinking more broadly: Russell Olwell and Azibo Stevens, "I had to double check my thoughts': How the Reacting to the Past Methodology Impacts First-Year College Student Engagement, Retention, and Historical Thinking," *The History Teacher* 48, 3 (2015): 561-72, [http://www.societyforhistoryeducation.org/pdfs/M15\\_Olwell\\_and\\_Stevens.pdf](http://www.societyforhistoryeducation.org/pdfs/M15_Olwell_and_Stevens.pdf); Gretchen Galbraith, "I Had Almost Forgotten I Was in a Classroom Setting': Reacting to the Past and Engagement with Historical Thinking" in *The Role of Agency and Memory in Historical Understanding: Revolution, Reform, and Rebellion*, ed. Gordon Andrews and Yosay Wagdi (Cambridge Scholars Publishing, 2017), 394.

14 Paula Kay Lazrus and Gretchen Krehling McKay, "The Reacting to the Past Pedagogy and Engaging the First-Year Student," *To Improve the Academy: A Journal of Educational Development* 32 (2013): 358. <https://quod.lib.umich.edu/cgi/p/pod/dod-idx/21-the-reacting-to-the-past-pedagogy-and-engaging.pdf?c=tia;idno=17063888.0032.025;format=pdf>.

15 Richard Gid Powers, John M. Burney, and Mark C. Carnes, "Reacting to the Past: A New Approach to Student Engagement and to Enhancing General Education," A White Paper Report for the Teagle Foundation (2010), 28. [https://reacting.barnard.edu/sites/default/files/inline-files/reacting\\_white\\_paper\\_teaglefoundation\\_0.pdf](https://reacting.barnard.edu/sites/default/files/inline-files/reacting_white_paper_teaglefoundation_0.pdf).

16 Mark Carnes, *Minds on Fire: How Role-Immersion Games Transform College* (Cambridge, MA: Harvard University Press, 2014), 87-123.

Stroessner carried out an assessment on *Reacting* at Barnard College and later expanded it with Laurie Susser Beckerman and Alexis Whittaker. As they noted, “the vast majority of assessments [of *Reacting*] to date have relied on descriptive or anecdotal accounts, and empirical evaluations have generally lacked credible measures and appropriate controls, undermining causal claims about the effects of role playing.”<sup>17</sup> Such measures and controls are important because “people tend to be quite inaccurate in judging both the nature and the extent of an impact of the experience.”<sup>18</sup> In other words, self-report is unreliable because students and faculty may be unable to accurately gauge the effects of the pedagogy. Unfortunately, their study did not measure critical thinking. They used standardized assessments to measure psychosocial effects: empathy, locus of control, mastery, self-esteem, implicit person theory, confidence as speakers, Machiavellianism, social avoidance and distress, and optimism. They found that *Reacting* generally leads to increases in self-esteem, emotional empathy, and the belief in human malleability. However, they also found, conversely, that *Reacting* appears to make students feel less in control of their lives, but without the negative consequences that tend to follow that externalized locus of control: drops in self-esteem and the belief in human malleability. Their explanation was that, considering how frequently sudden and seemingly random events take place in *Reacting* games, it is not surprising that students feel less in control.<sup>19</sup> In Phase 3 of their study, Stroessner, Beckerman, and Whitaker teased out how individual differences could affect enjoyment of *Reacting*. They found that “students who have a high degree of self-confidence . . . tend to like the pedagogy. In contrast, students who do not like receiving attention as a result of disagreement and students who are highly emotionally empathetic tend to enjoy the pedagogy to a lesser degree.”<sup>20</sup>

A study by Robert Bledsoe and Deborah Richardson uncovered a curious effect of role selection on students. The authors used standardized measures for self-efficacy and perspective-taking, along with surveys, to measure the effects of *Reacting* on students, noting that anecdotal evidence has not yet been backed up by empirical data.<sup>21</sup> Their most provocative finding was that students who were given indeterminate roles did not see an increase in their self-efficacy, whereas students in factionalist roles did.<sup>22</sup> This finding largely agrees with Matthew Weldenfeld and Kenneth Fernandez, who also lamented the lack of systematic studies of *Reacting* and used surveys and focus groups to gather data on student engagement.<sup>23</sup> They found that students in “moderate” or “crowd” roles did not feel the same need to prepare for class. Moderates believed it was the job of the factionalists to persuade them, and the crowd students felt they had no power over the outcome of the game. They also found that supposedly negative emotions, such as nervousness and anger, led to increased preparation and participation.<sup>24</sup>

All of this is to suggest that there is a great deal left to learn about the effects of *Reacting* on student learning, and that empirical studies using standardized assessments can yield fruitful data. This study is an attempt to do something similar for *Reacting* and critical thinking.

---

17 Steven J. Stroessner, Laurie Susser Beckerman, and Alexis Whittaker, “All the World’s a Stage? Consequences of a Role-Playing Pedagogy on Psychological Factors and Writing and Rhetorical Skill in College,” *Journal of Educational Psychology* 101, 3 (2009): 605.

18 Stroessner, Beckerman, and Whittaker, “All the World’s a Stage?” 608.

19 Stroessner, Beckerman, and Whittaker, “All the World’s a Stage?” 612, 614. April Lidinsky argues that *Reacting* builds on feminist pedagogy that encouraged, among other things, collaboration and empathy: April Lidinsky, “‘Reacting to the Past’ to Be Proactive in the Present: Feminist Roots of High-Impact Practices,” *Feminist Teacher* 24, 3 (2014): 208-12.

20 Stroessner, Beckerman, and Whittaker, “All the World’s a Stage?” 617.

21 Robert S. Bledsoe and Deborah S. Richardson, “Impact of Reacting to the Past and Effect of Role on Student Attributes and Academic Outcomes,” *International Journal of Teaching and Learning in Higher Education* 33, 3 (2022): 362-64. 365, 370-71. <https://eric.ed.gov/?id=EJ1366192>.

22 Bledsoe and Richardson, “Impact of Reacting,” 365, 370-71.

23 Matthew C. Weldenfeld and Kenneth E. Fernandez, “Does *Reacting to the Past* Increase Student Engagement? An Empirical Evaluation of the Use of Historical Simulations in Teaching Political Theory,” *Journal of Political Science Education* 13, 1 (2017): 47.

24 Weldenfeld and Fernandez, “Does *Reacting to the Past* Increase Student Engagement,” 57-58.

## Measuring the Effects of *Reacting* on Critical Thinking

In Spring 2023, I had the opportunity to teach an upper division course using three role-playing games: the Valladolid Debate; *Japan, 1941*; and *The Needs of Others*.<sup>25</sup> The Valladolid Debate game is under development; in keeping with *Reacting* standards, there were individualized roles for each student and multiple game mechanics for playing and winning the game.<sup>26</sup> In part because of the course's offering as an upper division Special Topics course, the students in the spring were all juniors or seniors, and mostly history majors. The course had been advertised as a role-playing game course, so students who signed up already had some idea of what it would be like. About half had played similar games in prior courses. Ten students enrolled, with nine completing the course. Unsurprisingly, these students were more committed to actively participating in the games than in a non-*Reacting* class. I graded student papers using a critical thinking writing rubric, but it was based on a non-CAT definition of critical thinking. It was intended to encourage students to inhabit their roles more fully and had categories for identifying the problem/question, perspectives, assumptions, and context. I did not provide any critical thinking training beyond explaining the terms in the rubric and related comments on their papers.

With assistance from the Office of Academic Assessment at GGC, I carried out a preliminary study using the CAT to measure changes in students' critical thinking as a result of playing *Reacting* games.<sup>27</sup> Students were given the CAT in the second week of the semester before playing any of the games and then again in the tenth week after finishing the second game (Valladolid Debate, *Japan, 1941*). Students were also given a two-part survey in the course meeting following the second CAT assessment. The first had open-ended questions about their experiences of the games and the second asked them to provide five-point Likert scale ratings about those experiences. Students were asked to fill out the second survey without returning to the first to avoid shaping their open-ended answers.<sup>28</sup> With only nine students having completed both CAT assessments and the surveys, it is not possible to draw broad conclusions about the effectiveness of *Reacting* on student's critical thinking. However, the data does give reason to be sanguine.

The CAT-app study at GGC established that there is practice effect. Administering the CAT to students twice in a semester with no critical thinking intervention led to a gain of .98 points, an increase of 7.42%.<sup>29</sup> For *Reacting* to show promise as a means of fostering critical thinking development, the study's subjects needed to improve by more than that.

	Average CAT Score	Evaluate and Interpret Information	Problem Solving	Creative Thinking	Effective Communication
GGC Freshmen (2016-23)	10.93	7.92	7.02	2.43	5.27
GGC Seniors (2016-23)	12.94	9.03	7.90	3.13	6.41
Point change	2.01	1.11	.88	.7	1.14
Percent change	18.39	14.02	12.54	28.81	21.63

Figure 2: Freshmen and Senior Performance on the CAT at GGC

25 John Moser, *Japan, 1941: Between Pan-Asianism and the West* (Chapel Hill, NC: University of North Carolina Press, 2022); Kelly McFall, *The Needs of Others* (Chapel Hill, NC: University of North Carolina Press, 2022).

26 I am working with Dr. Rebecca Carte at Cuyahoga Community College to turn it into a *Reacting* game.

27 My thanks to the director, Dr. Thomas Lilly.

28 Students were offered extra credit for completing both CATs and the surveys.

29 Lilly et al., "Intending to Teach Critical Thinking", 41.

	Average CAT Score	Evaluate and Interpret Information	Problem Solving	Creative Thinking	Effective Communication
National Seniors	17.64	10.73	10.82	6.11	10.5
Pretest	13.8	9.56	8.00	3.56	7.00
Posttest	15.56	10.56	8.89	4.22	8.56
Point change	1.76	1	.89	.67	1.56
Percent change	12.75	10.47	11.11	18.75	22.22

Figure 3: National Senior and *Reacting* Student Performance on the CAT

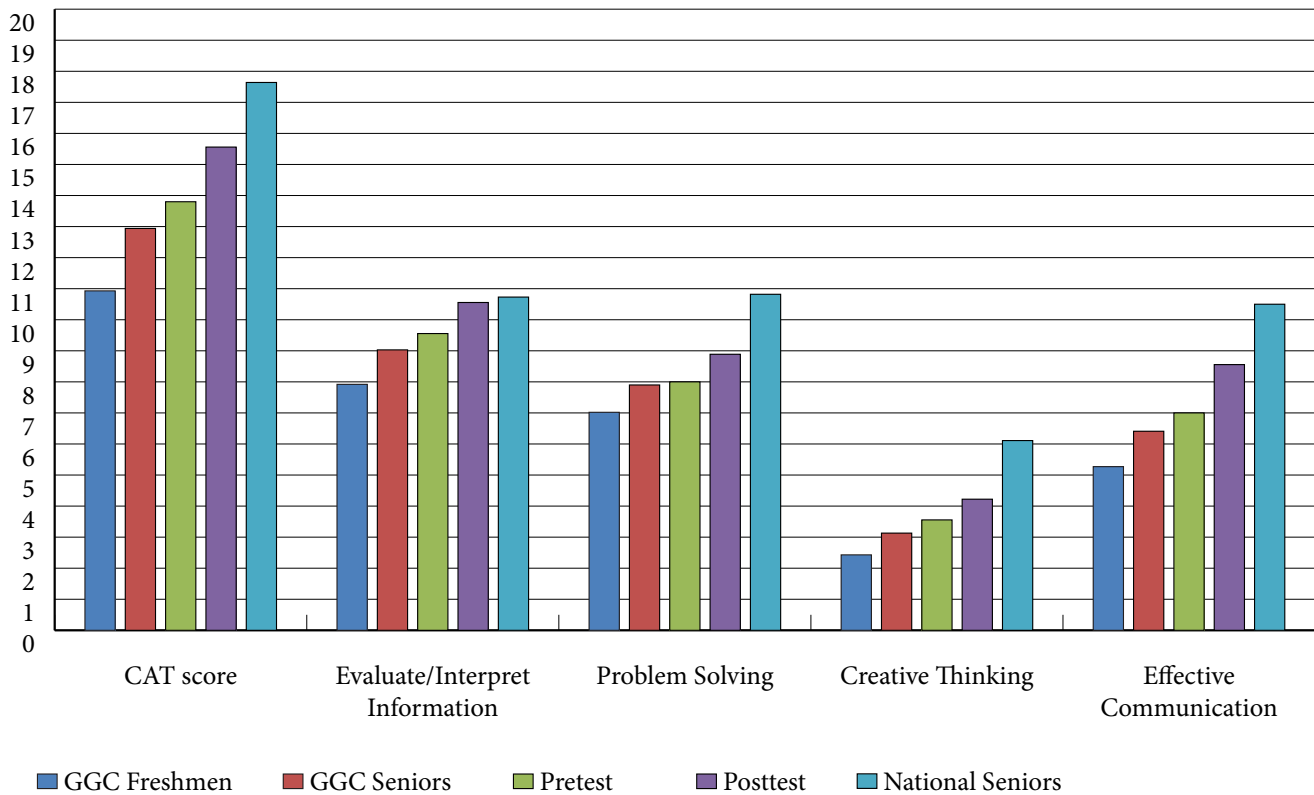


Figure 4: Mean total CAT and skill set scores

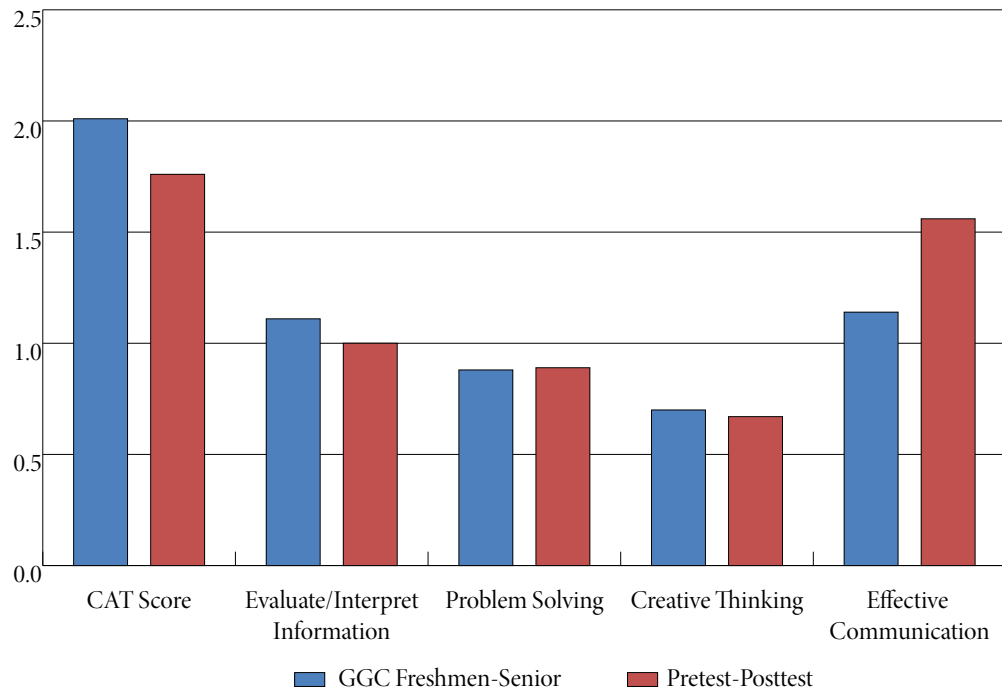


Figure 5: Change in total CAT and skill set scores by total points

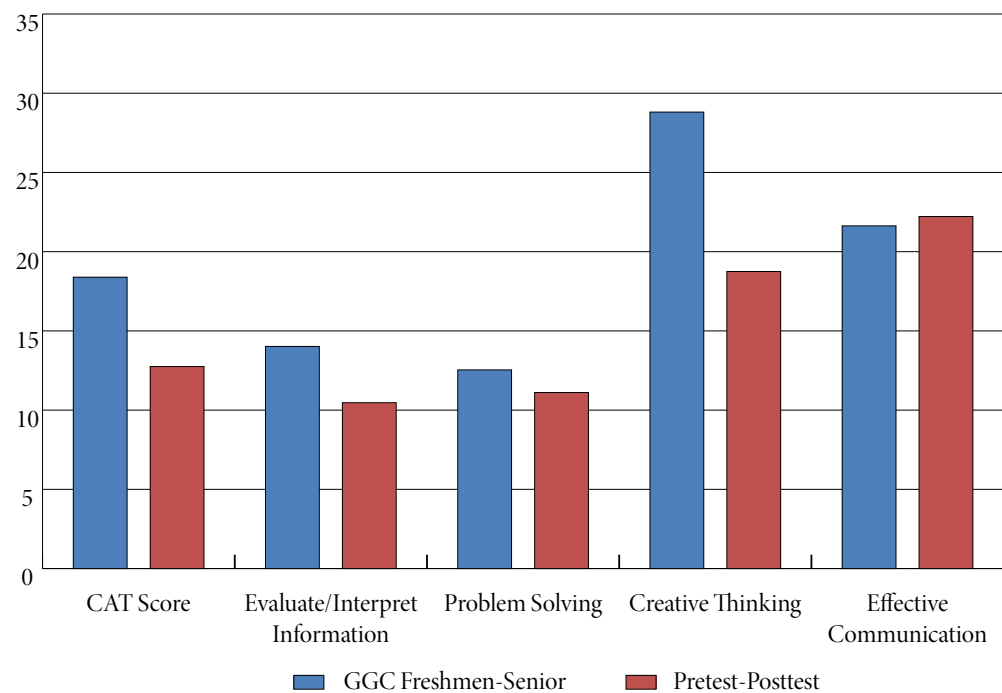


Figure 6: Change in total CAT and skill set scores by percentage

As the above figures show, before beginning to play any of the games, *Reacting* students (13.8) were a little above GGC seniors’ average (12.94), but not exceptionally so. After playing two games, their total scores rose by 1.76 points (12.75% increase), close to the 2.01 total difference between GGC freshmen and seniors, and a .78 point increase —5.65%—beyond the .98 point practice effect. In the four skill categories, *Reacting* students gained almost as many points as the difference between freshmen and seniors in evaluating and interpreting information



and creative thinking, slightly more in problem solving, and substantially more in effective communication. Because *Reacting* students started with higher scores, this was a smaller percentage gain, but even there, *Reacting* students' effective communication increase (22.22%) was greater than the difference between GGC freshmen and seniors (21.63%). *Reacting* students' chief gains, in terms of percentage increase, came in creative thinking and effective communication. Given GGC students' persistently low performance in these categories relative to national averages, the possibility of a pedagogy that targets those areas for improvement is particularly heartening.

**Survey data**

In the class session following the posttest CAT, students were asked to take two surveys. One had open-ended questions and the other asked them to rank the course relative to other courses on a five-point Likert scale. Students were asked to answer the open-ended questions first and not return to them after taking the second survey to avoid contaminating their answers with wording and ideas from the second survey. The surveys indicate that the role-playing games made a substantial emotional impact on the students, that they primarily focused on the social and speaking aspects of the games, and that, when compared with their CAT scores, they did not have a strong sense of which skills they were developing. These results reinforce the need for using objective assessments in addition to self-report and anecdotal evidence.

- |    |   |
|----|---|
| 1. | In this course, we have played multiple historical role-playing games. Have you ever played similar role-playing games in other courses? Yes/no. If yes, please describe the course and the game. |
| 2. | What skills do you believe that you have developed over the course of playing these games? Please list and explain.   |
| 3. | What was your emotional experience of this course like? Please explain.   |
| 4. | If you have any other comments you wish to make at this time, please do so here.  |

Figure 7: Survey 1 Questions

To the first question, four of the nine students responded that they had played similar games, some of them in my courses. Two of them, in fact, had previously played an older and less sophisticated version of the Valladolid Debate game.

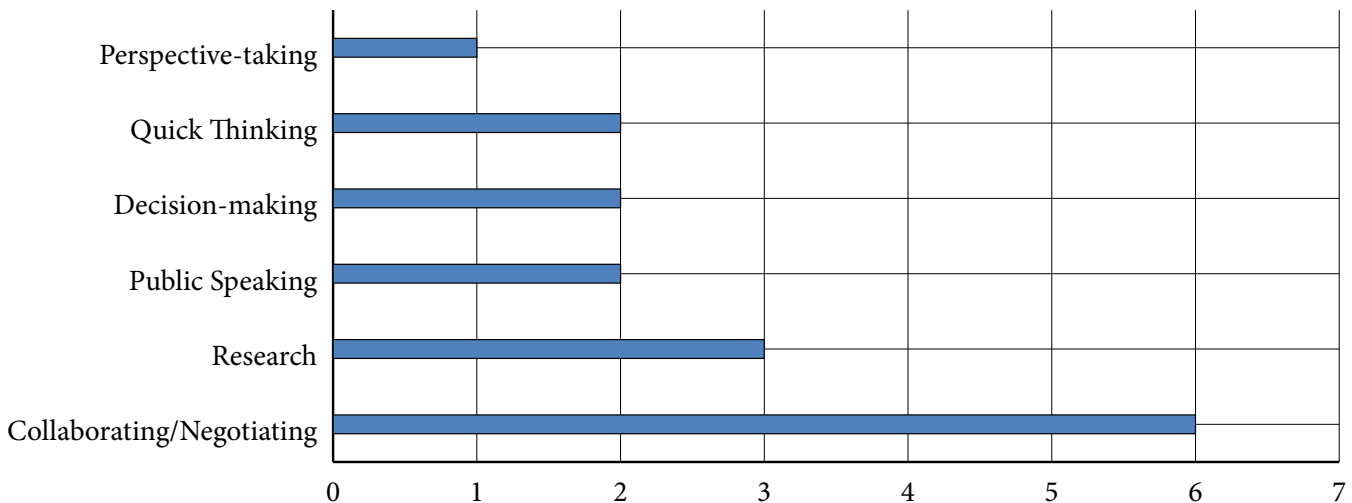


Figure 8: Survey 1 Skills (Q2)

To the second question, student answers here were generally focused on the in-person aspects of the game. They focused on communication, particularly the more social or speaking aspects. Six noted something to do with

collaborating and/or negotiating with their peers. Only three mentioned research skills, though all students had been doing some research for their roles. Two discussed having to think quickly in the moment. Two discussed working on their decision-making. Strangely enough, only two explicitly addressed working on their public speaking. Finally, only one discussed learning how to take on someone else's perspective.

For the third question, I had anticipated for many students to say that they found the games stressful, but the answers were generally positive. Students seem to have found the experiences energizing. Five students indicated having at least some negative emotions, but were still enjoying themselves. Two indicated some distress at the "heavy" topics of the games or taking on roles of unsavory historical figures, but both of those students indicated their enjoyment of the course in the following question. The other three noted some anxiety or frustration in speaking in front of or dealing with their fellow students in the games, but all indicated that they were enjoying the course. One student noted that it "can get to you a little bit when other people are roll [sic] playing more aggressively, but its [sic] just part of the game." Another noted that the "emotional experience was good but made me nervous becaus [sic] I'm not used to speaking in front of others." The third student also noted that the games were enjoyable, but that "I would get frustrated with other players both during and outside of game sessions, but I feel that those were still productive emotions, and were beneficial to critical thinking and acting under pressure." This was reminiscent of Weldenfeld and Fernandez's finding that negative emotions could support student learning.<sup>30</sup>

The four students who only noted positive emotional experiences pointed to their emotional investment in the course. One noted how "consuming" the games were, "like it's always in the back of your mind," particularly when other students were also taking the game seriously. Another student noted how "exciting" the games were and called them "the most impactful form of Active learning I have experience." The third student noted how "productive" the games were for developing "real-life skills" like "ways to cooperate and collaborate in preparation of real-world problems." The fourth expressed his/her enjoyment and that "I felt more emotionally connected to the content and more invested."

Seven students responded to the fourth question. All expressed their appreciation for the games and generally wanted more. They wished to see them in other classes, to try new ones, to have more participants, or to have me run the course again, for the sake of future students.

Overall, these surveys expressed what *Reacting* instructors have generally reported. Students emotionally engage with the games and develop their social and communication skills. To a lesser extent, they also reported developing their research, decision-making, and perspective-taking. There are two important caveats to the above data. First, at the end of the semester, multiple students said that they found the last game, *The Needs of Others*, to be far more stressful than the other two because of the subject matter and their frustration at being unable to intervene in Rwanda—or because they were one of the characters who actively prevented intervention. This could indicate that the material and gameplay of individual games should be taken into account when considering the emotional impact of role-playing games. Second, these students self-selected into a role-playing course. The emotional experience would likely be different for students who had not.

The second survey used a 1-5 point scale, with one 1 being "much less," 2 "somewhat less," 3 "about the same," 4 "somewhat more," and 5 "a lot more." Students were asked to rate their experience in the *Reacting* course relative to non-role-playing courses. The first six questions, listed below, were intended to test student engagement in the course: stress, skill development, learning, and workload.

---

30 Weldenfeld and Fernandez, "Does *Reacting to the Past* Increase Student Engagement," 57-58.

1. How stressful was this course, relative to other courses?
2. How much of the material have you learned in this course, relative to other courses?
3. To what extent have your writing skills improved as a results of this course, relative to other courses?
4. To what extent have your speaking skills improved as a result of this course, relative to other courses?
5. To what extent have your reading skills improved as a result of this course, relative to other courses?
6. How much work have you done for this course, relative to other courses?

Figure 9: First six questions of Survey 2

Questions 7-21 of the survey asked them to rate their skill development using the same 1-5 rating scale. However, students were given CAIL's descriptions of the skills tested by each of the 15 questions on the CAT. The survey instructions noted that these skill "descriptions were not designed for this course, so they may or may not apply. Just consider how much you have developed that skill in this course and compare it to how much you have developed that skill in other courses." Discussion with students after the survey indicated that they did not recognize them as the basis of the CAT.

7. Summarize the pattern of results in a graph without making inappropriate inferences.
8. Evaluate how strongly correlational-type data supports a hypothesis.
9. Provide alternative explanations for a pattern of results that has many possible causes.
10. Identify additional information needed to evaluate a hypothesis.
11. Evaluate whether spurious information strongly supports a hypothesis.
12. Provide alternative explanations for spurious associations.
13. Identify additional information needed to evaluate a hypothesis.
14. Determine whether an invited inference is supported by specific information.
15. Provide relevant alternative interpretations for a specific set of results.
16. Separate relevant from irrelevant information when solving a real-world problem.
17. Use and apply relevant information to evaluate a problem.
18. Use basic mathematical skills to help solve a real-world problem.
19. Identify suitable solutions for a real-world problem using relevant information.
20. Identify and explain the best solution for a real-world problem using relevant information.
21. Explain how changes in a real-world problem situation might affect the solution.

Figure 10: Last 15 questions of Survey 2

Student responses to the first six questions generally reflected their responses in the first survey. Based on the first survey, students should report the most development of their communication skills, particularly oral communication. Their emotional investment in the course would suggest that they are doing more work, but their positive and negative emotional experiences would suggest that their stress level should be average, or perhaps slightly elevated. These predictions bear out in the second survey, with speaking skills claiming the top ranking and writing skills, and workload tying for second. Student stress was the lowest score, but still slightly elevated from a non-role-playing course.

Question	Mean Response
1. Stress	3.11
2. Learning	3.44
3. Writing Skills	3.89
4. Speaking Skills	4.11
5. Reading Skills	3.33
6. Workload	3.89

Figure 11: Mean responses to first six questions of Survey 2

Responses to the second half of the survey, on the other hand, indicate that students do not know which critical thinking scores they have been developing. Student CAT scores and their ratings of how much they have developed those skills do not show correlation. To make comparison easier, I have rendered both CAT scores and the student ratings for their skill development by percentage change. For the CAT, this meant comparing the average pretest and posttest scores and finding the percent increase and decrease for each question. For the 1-5 rating skill, I marked 1 as -100%, 2 as -50%, 3 as 0%, 4 as 50%, and 5 as 100%. X-axis values represent the number of the survey question and the relevant CAT “Q” question. To see predictability in student report and real outcomes in the below charts, we would need to see a similar shape, though perhaps not values.<sup>31</sup>

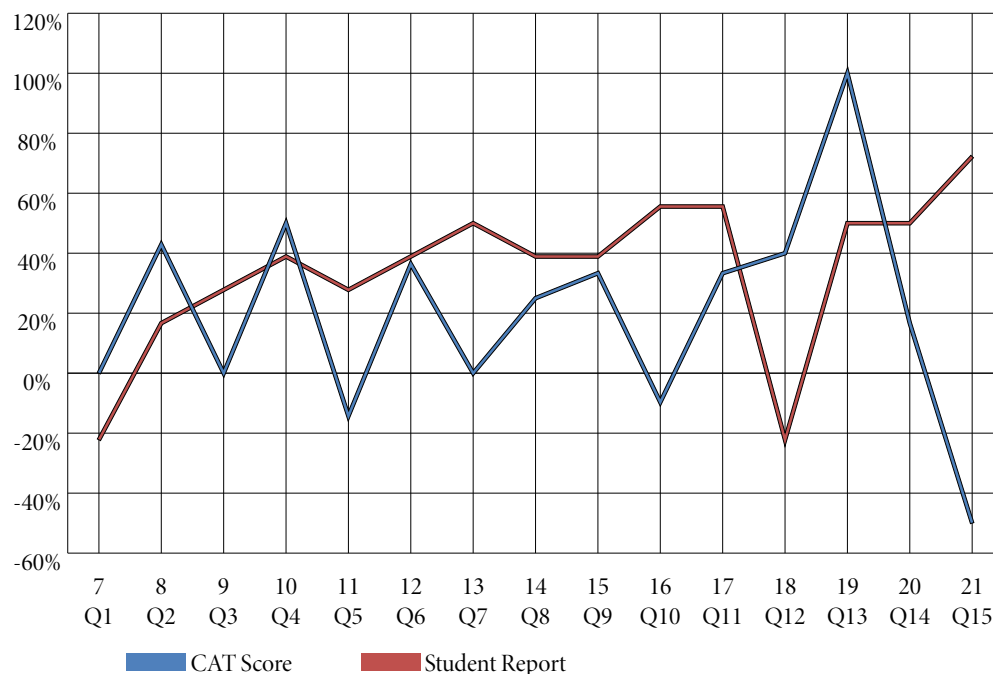


Figure 12: Mean percentage changes in CAT scores and student-reported skill development

The above figure does not suggest that students can self-report their own skill development. In fact, the highest and lowest data points, questions 13 and 15 from the CAT, demonstrate the problems with relying on faculty observations and student self-report. In question 13, students have to “identify suitable solutions for a real-world problem using relevant information.” It makes sense that they should develop this skill while playing *Reacting*, and the CAT scores reflect that, but students do not perceive that this was their most developed skill. One would

<sup>31</sup> Data on Q1 was included, but students all scored the maximum in both the pre- and posttest. Therefore it was impossible to see improvement on it. Similarly, all students also scored zero points on Q7 in both the pre- and posttest, and so it was impossible to measure a decline.

also expect them to improve for question 15, in which students are expected to “explain how changes in a real-world problem situation might affect the solution.” In fact, students marked it as their most developed skill, but their CAT scores dropped by 50%. What faculty and students experience or expect to see does not necessarily align with objective assessment.

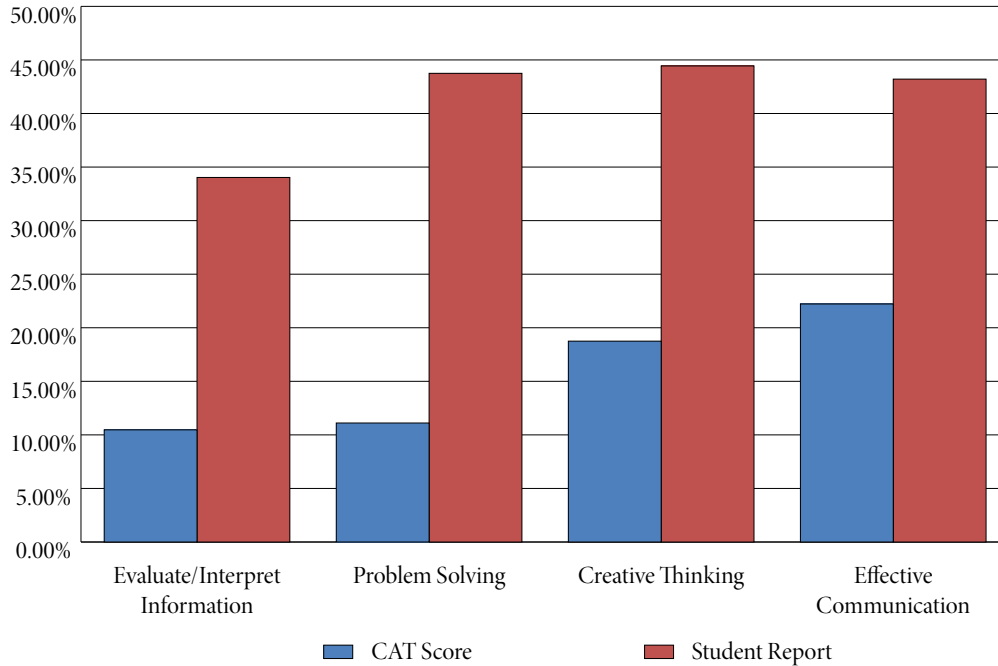


Figure 13: Mean percentage changes in CAT scores and student-reported skill development by skill set

Student expectations are slightly better when considered by skill set, but still do not align with CAT scores. Students believed that they improved most on problem solving, creative thinking, and effective communication. Student ratings for these three categories were extremely close, but we know that the changes in creative thinking and effective communication were substantially higher than for problem solving.

## Limitations

The limitations on this study indicate paths forward. First, the most significant limitation on the study was its size. While the student pretest averages did not indicate that these students were extraordinary as a group, even small changes in CAT scores could provide misleading results. A larger, sustained study, using multiple faculty and courses and with its own control group would be able to provide more reliable data.

Second, student selection could also lead to unusual outcomes. Students in this study were told what they would be doing and signed up for the course. Around half had already played similar games. Some of the studies discussed previously, particularly the one by Stroessner, Beckerman, and Whitaker, have found that individual students may react differently based on their personality traits. It is possible that a larger pool of students would react more negatively or experience greater levels of stress than this self-selected group of students. A broader study of *Reacting* across multiple courses that are not specifically designated as *Reacting* courses could overcome this limitation.

Finally, there are limitations on the instrument used to assess critical thinking. While the CAT is a validated instrument, it may not be best suited to evaluate the effects of *Reacting*. Student self-report emphasized the work they had done to improve their ability to speak, collaborate, and negotiate. Much of the structure of the games, both the sessions and the writing assignments, focus on quality of oral and written communication and the construction of arguments. The CAT does not assess communication very deeply and it may also not be sensitive

to other kinds of thinking done in history courses, such as perspective-taking. If indeed critical thinking has a dispositional element, such as Delphi's self-regulation, then that would be missed entirely. On the other hand, some skills assessed by the CAT may not be relevant to many *Reacting* games. For instance, students may not need to know how to read a graph or even how to use basic mathematical skills to solve an in-game problem. Certainly, there are games where these skills would be useful, but they are not central to role-playing games as a whole. With all of that said, the CAT has still shown that it can detect changes in critical thinking as a result of students playing two role-playing games.

## Conclusions

History needs objective, validated assessment to stay competitive. Lendol Calder and Tracy Steffes called for a standard history assessment tool, making the case that it would benefit history as a discipline. In part, it would be defensive: "if historians do not come to the table for conversations about assessment, decisions will be made without us." However, it would also be one means whereby "historians could help us gather important evidence about student learning according to the criteria that we as historians deem most important."<sup>32</sup> Calder and Steffes were suggesting an assessment based on core elements of historical thinking that would explain and justify history's place in education to students, administration, and future employers, but this study suggests that assessing critical thinking could be just as valuable. If critical thinking is commonly desired by students, faculty, and employers, and history classes can show that it provides effective training, assessment would prove how valuable history courses are. While this study was limited in scale, it gives cause to feel optimistic about the effects of *Reacting* on students' critical thinking development and justifies *Reacting's* use in the history classroom and beyond. Students' scores, which started above GGC average, but not exceptionally so, improved by more than the practice effect. This study confirms something that was assumed: *Reacting* and *Reacting*-style games lead to a measurable improvement in critical thinking.

This study also found the limitations of previous studies on *Reacting*. While student critical thinking improved, it did not do so evenly. Creative thinking and effective communication scores increased at a faster rate than evaluating/interpreting information and problem solving. This was not necessarily surprising, but it also could not have been stated with any certainty based on previous studies. Surveys found that while students were positive about their experience and skill building, they were not very accurate in predicting which of their skills had developed. This casts doubt on self-report as a means of analyzing the effects of *Reacting*. Lastly, student responses to the open-ended survey focused mostly on in-game communication more than the writing and perspective-taking elements often discussed in the *Reacting* literature. This may be a sign of faculty and students understanding the experience of the games differently, casting doubt on the validity of faculty observations. All of this underscores the need for more studies of *Reacting* using validated, standardized measures.

The gains in creative thinking and effective communication were heartening for a few reasons. Whereas GGC students tend to score close to the national average in evaluating and interpreting information, they lag far behind the national average in the former two areas, averaging only 51.2% and 61% of the national norm, respectively. Interventions that target those areas specifically could be extremely beneficial, as they would be at any school in a similar situation. In addition, some areas of critical thinking may be more beneficial than others for student success. As discussed previously, CAII's creative thinking skill set bears resemblance to the Delphi Report's definition of inference. Liam O'Hare and Carol McGuinness found that "the critical thinking skill of inference has particularly strong validity in terms of higher education outcomes."<sup>33</sup> They examined students' marks and found that the inference component of the California Critical Thinking Skills Test, which is based on Delphi, had clear,

---

32 Lendol Calder and Tracy Steffes, "Measuring College Learning in History," in *Improving Quality in American Higher Education: Learning Outcomes and Assessments for the 21st Century*, ed. Richard Arum, Josipa Roksa, and Amanda Cook (San Francisco: Jossey-Bass, 2016), 40, 41.

33 Liam O'Hare and Carol McGuinness, "The validity of critical thinking tests for predicting degree performance: A longitudinal study," *International Journal of Educational Research* 72 (2015): 170.

if modest, predictive power. Thus, it is possible that targeting this area of critical thinking could improve student outcomes overall. Schools in a similar situation, not just GGC, should take note.

Finally, *Reacting* pushes students to build a wide variety of skills, many of which may not be properly termed critical thinking. For instance, students may learn how to cope with distress created by openly disagreeing with their classmates. Even if critical thinking gains were smaller than those already measured, that would not necessarily discount *Reacting's* value as a pedagogy because of the variety of skills it hones. The value of *Reacting* is that it can improve student skills in many areas, including critical thinking.

This preliminary study was intended to test the viability of assessing the development of critical thinking as a result of *Reacting* pedagogy. While the small size of the study makes it difficult to draw any definite conclusions, the evidence suggests that *Reacting* is an effective pedagogy for improving student critical thinking.