Teaching Innovation Projects

Volume 9 | Issue 1 | 2020

Quantitatively Assessing the Success of Your Critical Thinking Teaching Strategies

Sarah D. McCrackin

The University of Waterloo, sdmccrac@uwaterloo.ca

Follow this and additional works at: <https://ojs.lib.uwo.ca/index.php/tips>

Recommended Citation:

McCrackin, S. (2020). Quantitatively assessing the success of your critical thinking teaching strategies. *Teaching Innovation Projects*, 9(1), 1-15. <https://doi.org/10.5206/tips.v9i1.10342>

ABSTRACT

The teaching literature contains a wealth of articles discussing the success of interventions designed to foster critical thinking. However, most of these articles describe qualitative assessments of critical thinking interventions, with very few quantitative assessments (Abrami et al., 2008). While both qualitative and quantitative methods have unique value, there is a need for quantitative data to supplement qualitative claims that various strategies for teaching critical thinking are successful. The goal of this workshop is to provide a framework that members of all disciplines can use to quantitatively assess the success of critical thinking interventions in their own classroom. Participants will learn about common research approaches and assessments that have been used in published quantitative studies on critical thinking. They will also learn strategies for assessing the experimental rigor of previous critical thinking research and will work together to create a plan to apply this knowledge in their own classrooms.

KEYWORDS

quantitative assessment; critical thinking; higher education; domain-general; domain-specific

LEARNING OUTCOMES

By the end of this workshop, participants will be able to:

- Assess the quality of published studies on critical thinking interventions and their outcomes.
- Identify pre-experimental, quasi-experimental and experimental research designs, along with measurements used in published studies to evaluate critical thinking interventions.
- Design their own method of quantitatively measuring the outcome of their critical thinking interventions.

ANNOTATED BIBLIOGRAPHY

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102-1134.
<https://dx.doi.org/10.3102/0034654308326084>

The authors perform a meta-analysis investigating which critical thinking interventions are the most successful. The reader can refer to the paper for a discussion of their results, though in the present workshop the study is simply used to demonstrate that most of the critical thinking interventions in the literature are described qualitatively instead of quantitatively. The authors began with a pool of 3720 papers on critical thinking interventions. They eliminated papers which did not describe a critical thinking intervention lasting more than three hours and which did not have quantitative data sufficiency, leaving them with 117 studies.

Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching and Learning*, 8(2), 25-42.
<https://dx.doi.org/10.19030/tlc.v8i2.3554>

The authors perform a meta-analysis of critical thinking studies that quantified critical thinking improvements in response to various types of instruction. The description of the number of studies that did not have sufficient quantitative data is used to demonstrate that most of the critical thinking interventions in the literature are described qualitatively. The authors group their included studies by experimental design method and demonstrate that weaker study designs are more likely to report that the critical thinking intervention led to critical thinking improvements. The findings from this section of the paper are used to demonstrate that weaker experimental designs may lead authors to falsely conclude that their critical thinking interventions had a significant impact on critical thinking.

Brunt, B. A. (2005). Models, measurement, and strategies in developing critical-thinking skills. *The Journal of Continuing Education in Nursing*, 36(6), 255-262.
<https://dx.doi.org/10.3928/0022-0124-20051101-05>

The authors begin by stating that it is hard to measure critical thinking because the concept has been defined in many different ways. They describe in detail some standardized instruments that have been created to assess critical thinking. However, they emphasize that despite arguments that critical thinking is a skill that transfers across domains; these instruments may not be the best for measuring critical thinking in a given context. Domain-general tests do not seem to measure domain-specific improvements in critical thinking well. Their focus is on nursing, which may have very different critical thinking requirements than other disciplines. The authors call for the development of discipline-specific critical thinking metrics to help in evaluating which teaching strategies are the most effective at developing critical thinking skills. This paper is used in the workshop to contribute to the comparison of domain-general and domain-specific critical thinking measurements.

Tiruneh, D. T., Weldelessie, A. G., Kassa, A., Tefera, Z., De Cock, M., & Elen, J. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Research and Development*, 64(3), 481-505. <https://dx.doi.org/10.1007/s11423-015-9417-2>

The authors define domain-specific and domain-general critical thinking skills, and list some domain-general and domain-specific critical thinking questionnaires. They also discuss two different stances on critical thinking. Generalists suggest that critical thinking skills extend across domains, while Specificists suggest that critical thinking skills are domain-specific, meaning that good critical thinking in one domain may not generalize to another. The authors then detail an experiment they conducted in introductory physics courses to test the effect of explicit versus non-explicit critical thinking instruction on learning domain-specific and domain-general critical thinking. At the end of the term, the explicit critical thinking group demonstrated larger improvements in domain-specific critical thinking than the non-explicit instruction group,

though there were no group differences in domain-general critical thinking. This paper is used in the workshop to provide evidence that teaching domain-specific critical thinking may not transfer to domain-general critical thinking.

Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity*, 3(2), 85-93.
<https://dx.doi.org/10.1016/j.tsc.2008.03.005>

The authors list some domain-specific and domain-general critical thinking tests that have been used previously. They state that earlier studies demonstrating improvements in domain-specific critical thinking did not necessarily find improvements in domain-general critical thinking. The authors then discuss an experiment they conducted in an Introductory Psychology class. Participants were randomly assigned to an experimental group (consisting of exposure to higher-order questions) or a control group. Each group filled out a domain-specific critical thinking test (textbook questions) and a domain-general test (Watson-Glaser Critical Thinking Appraisal test; WGCTA). After the experimental manipulation, participants took the two tests again. The participants in the experimental group showed greater improvement on the domain-specific test than those in the control group, while there was no difference in improvement on the domain-general test. The authors conclude that studies using domain-general estimates of critical thinking may be underestimating critical thinking improvements in students, and that higher order questioning may improve critical thinking in a domain-specific manner. This study is used in the workshop to provide further evidence that domain-specific critical thinking improvements may not be associated with domain-general critical thinking improvements.

Williams, R. L., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, 53(1), 37-58.
<https://dx.doi.org/10.1353/jge.2004.0022>

The authors designed an experiment investigating the effect of completing higher-order practice questions on critical thinking ability. Five Human Development classes completed a domain-general (Watson-Glaser Critical Thinking Appraisal, or WGCTA) and a domain-specific (Lawson, 1999) critical thinking test at the start of the term. Students were then randomly assigned to the experimental group, where they were exposed to higher-order questions, or to the control group. At the end of the term, students completed both critical thinking tests again. The experimental group showed greater improvement on domain-specific critical thinking than the control group, while neither group improved at domain-general critical thinking. Exam performance was found to be related to both domain-general and domain-specific critical thinking scores, with a stronger relationship to the domain-specific scores. The authors concluded that domain-specific measures of critical thinking seem to be more promising as a predictor of course performance than domain-general ones. This paper is used in the workshop as evidence that domain-specific critical thinking skills often do not generalize to domain-general skills.

ADDITIONAL REFERENCES

Defining Critical Thinking

Facione, P.A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: California Academic Press. (ERIC No. ED 315 423).

Suggested Papers for the Introduction to the Critical Thinking Literature Activity

Chau, J. P. C., Chang, A. M., Lee, I. F. K., Ip, W. Y., Lee, D. T. F., & Wootton, Y. (2001). Effects of using videotaped vignettes on enhancing students' critical thinking ability in a baccalaureate nursing programme. *Journal of Advanced Nursing*, 36(1), 112-119.
<https://dx.doi.org/10.1046/j.1365-2648.2001.01948.x>

Daley, B. J., Shaw, C. A., Balistreri, T., Glasenapp, K., & Piacentine, L. (1999). Concept maps: A strategy to teach and evaluate critical thinking. *Journal of Nursing Education*, 38(1), 42-47. <https://doi.org/10.3928/0148-4834-19990101-12>

Forneris, S. G., & Peden-McAlpine, C. (2007). Evaluation of a reflective learning intervention to improve critical thinking in novice nurses. *Journal of Advanced Nursing*, 57(4), 410-421.
<https://dx.doi.org/10.1111/j.1365-2648.2007.04120.x>

Information about Qualitative and Quantitative Investigations in Teaching

Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches*. Boston, MA: Allyn & Bacon.

Smith, J. K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher*, 12(3), 6-13.

Examples of Domain-Specific Critical Thinking Measurements:

Lawson, T. J., Jordan-Fleming, M. K., & Bodle, J. H. (2015). Measuring psychological critical thinking: An update. *Teaching of Psychology*, 42(3), 248-253.
<https://dx.doi.org/10.1177/0098628315587624>

Tiruneh, D. T., De Cock, M., Weldeclassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663-682. <https://dx.doi.org/10.1007/s10763-016-9723-0>

Examples of Domain-General Critical Thinking Measurements:

Ennis, R., & Millman, J. (1971). *Manual for Cornell critical thinking test, level X, and Cornell critical thinking test, Level Z*. Urbana, IL: University of Illinois, Illinois Critical Thinking Project.

Facione, P. A. (1991). *Using the California critical thinking skills test in research, evaluation, and assessment*. Millbrae, CA: California Academic Press. (ERIC No. ED337498).

Gadzella, B. M., Hogan, L., Masten, W., Stacks, J., Stephens, R., & Zascavage, V. (2006). Reliability and validity of the Watson-Glaser critical thinking appraisal-forms for different academic groups. *Journal of Instructional Psychology*, 33(2), 141-143.
<https://dx.doi.org/10.1007/s10734-006-9002-z>

Information about Different Experimental Designs

Dawson, T. E. (1997, January). *A primer on experimental and quasi-experimental design*. Austin, TX: Annual Meeting of the Southwest Educational Research Association, January 23-25. (ERIC Document Reproduction Service No. ED406440).

WORKSHOP CONTENT AND ORGANIZATION

DURATION (MIN)	SUBJECT	ACTIVITY	PURPOSE
5	Introduction	Welcome the participants and facilitate an introduction. Ask participants to give their names, their background, and explain what they hope to take away from the workshop.	To familiarize the facilitator and participants with each other. To give the facilitator an idea of the knowledge that participants are hoping to gain.
5	Rationale and Intended Learning Outcomes	Review the intended learning outcomes and overview for the workshop. Then share a list of some key critical thinking behaviours and abilities (Facione, 1990). Ask participants to respond to the following questions:	To give participants an overview of what key points they should take away from the workshop. To develop a working definition of critical

		<ul style="list-style-type: none"> • Why are these behaviours and abilities important for students to learn? • Why is it important to be able to measure critical thinking improvement in students? 	<p>thinking for the workshop.</p> <p>To get participants thinking about why critical thinking is a valuable skill, and why it is important to assess whether or not students are improving this skill.</p>
30	Introduction to the Critical Thinking Literature	<p>Have the groups rotate through four example papers from the critical thinking literature (see Presentation Strategies).</p> <p>Provide groups with the handout (see Appendix A, Activity 1) to discuss and take guided notes on each paper.</p> <p>Rotate groups every 5 minutes. Use the remaining 10 minutes at the end to ask for observations about the papers.</p>	<p>To better understand how critical thinking interventions and outcomes have been documented, and to give participants some insight into the current critical thinking literature.</p> <p>To introduce participants to different methods used to measure critical thinking.</p>
40	Measuring Critical Thinking	<p>Briefly outline what qualitative and quantitative data look like (Smith, 1983).</p> <p>Ask participants to identify the benefits and constraints of the qualitative and quantitative approaches used in the papers in the previous activity.</p> <p>Present data demonstrating that most of the critical thinking literature is qualitative and not quantitative (Abrami et al., 2008; Behar-Horenstien & Niu, 2011).</p> <p>Define the two categories of quantitative critical thinking</p>	<p>To ensure that participants understand the difference between qualitative and quantitative data, and that each approach has benefits and constraints.</p> <p>To appreciate that critical thinking is measured and reported very differently across disciplines and in the literature, which could constrain</p>

		<p>measurements: domain-specific (e.g., Lawson, 2015; Tiruneh et al., 2017) and domain-general (e.g., Ennis & Millman, 1971; Facione, 1991; Gadzella et al., 2006) measurements.</p> <p>Go through examples of questions asked in domain-specific and domain-general assessments (see Presentation Strategies). Ask participants which category they believe they belong in and why.</p> <p>Discuss literature showing that domain-specific critical thinking improvements do not seem to generalize to domain-general critical thinking improvements (Brunt, 2005; Renaud & Murray, 2008; Tiruneh et al., 2016; Williams, Oliver, & Stockdale, 2004).</p>	<p>interpretations of certain results.</p> <p>To understand the distinction between domain-general and domain-specific measurements.</p> <p>To reflect on the type of critical thinking skills we develop among our students, and how students use these skills inside and outside of the classroom.</p>
35	Research Designs Used to Measure Critical Thinking Improvements	<p>Present the three most common research designs used in papers quantitatively assessing critical thinking: pre-experimental, quasi-experimental and experimental (Behar-Horenstein & Niu, 2011). Sample discussion points about the strengths and weakness (Dawson, 1997) are provided in Presentation Strategies.</p> <p>Present the results from a meta-analysis demonstrating that the better the study design, the less likely the experimenters were to find critical thinking improvements (Behar-Horenstein & Niu, 2011).</p> <p>In partners, ask participants to discuss how they could use the different study designs to investigate the effectiveness of their own approaches</p>	<p>To emphasize the importance of research methodology, and to emphasize that some of the current literature may be reporting findings that are incorrect due to weak study design.</p> <p>To have participants consider how they could design a method to quantitatively assess critical thinking among their own students. This is also a critical thinking exercise in itself.</p> <p>To give participants a chance to receive peer feedback on their</p>

		to teaching critical thinking. Ask for volunteer responses from the group. Working with the same partner, ask participants to outline a research design process that could quantitatively evaluate attempts to foster critical thinking among their students (see Appendix A, Activity 2).	ideas to quantitatively measure critical thinking improvements among their students.
5	Conclusion/ Questions	Summarize the take-home message and provide a chance to discuss any final questions or thoughts from participants.	To reinforce the lesson and to clarify anything that was not made clear.
Total Time: 120 minutes			

PRESENTATION STRATEGIES

This workshop is designed for 20-30 participants per session. The room should be set up with four tables, and participants should be seated at in groups of 4 or 6 (even numbers for pairing up.)

The facilitator is advised to bring/have access to:

- Copies of the handout for each participant (Appendix A).
- Copies of each study chosen for analysis in the first activity. Each table should have enough copies of the same study for each person at that table to read, and no two tables should have the same study.
- An overhead projector and laptop to display the slide deck. Slide suggestions are provided below in italics.

Rationale and Intended Learning Outcomes

PowerPoint Slide Suggestion: 1) Present critical thinking behaviours and abilities from Facione (1990).

Introduction to the Critical Thinking Literature

The papers used in this activity should include both interventions that are described quantitatively and qualitatively. While the workshop facilitator is welcome to pick different papers, suggestions are provided in Additional References. Participants should be assigned these papers ahead of time as required reading, as they will only have time to review them in the workshop.

At each table there should be copies of one of the studies. Participant groups begin by reviewing the study at their table, and after five minutes they rotate to another table to look at a different study. This continues until all four studies have been examined by each group.

Measuring Critical Thinking

PowerPoint Slide Suggestion: 1) Show examples of qualitative (e.g., types of comments made on a discussion forum, interview questions, self-reported impressions of a teaching method) and quantitative (e.g., test scores, drop-out rates, IQ) data. 2) Use the inclusion criteria from Abrami et al. (2008) and Behar-Horenstien and Niu (2011) to depict the number of studies that were excluded from each meta-analysis for having insufficient quantitative data. 3) Display the definitions of Domain-specific and Domain-general measurements, followed by sample questions for participants to categorize.

Sample questions from common critical thinking assessments can be found in the published articles for each assessment. Tiruneh et al. (2017) and Lawson (2015) have domain-specific questions and Gadzella et al. (2006), Ennis and Millman (1971) and Facione (1991) have domain-general questions.

Research Designs used to Measure Critical Thinking Improvements

PowerPoint Slide Suggestion: 1) Present a slide showing how each research design works. 2) Behar-Horenstien and Nui (1990) have a graph showing that studies with weaker designs are more likely to report significant critical thinking improvements, which is a great visual to show that many of the papers which use quantitative measures might be reporting false improvements due to poor study designs.

Sample design strengths and weaknesses (Dawson, 1997) adapted to fit the critical thinking literature:

Pre-experimental design: a group is observed both before and after the intervention. Any critical thinking improvement that occurs between pretest to posttest is presumed to be due to the critical thinking intervention. In these designs, no control or comparison group is employed. This makes this design a weaker one, as it is unclear if improvements may have been due to something else other than the intervention. However, this design is easy to implement in the classroom.

Quasi-experimental design: there is a control group and an experimental group. The experimental group receives the critical thinking intervention, while the control group receives a control intervention. If the critical thinking intervention group shows greater improvements in critical thinking than the control group, then the critical thinking intervention is considered to be a success. The addition of a control group makes this design a stronger one than the pre-experimental design. However, note that the two groups are not randomly assigned.

Experimental design: there is a control group and an experimental group, which students are randomly assigned to. This ensures that there is nothing about which students are in which groups that could be responsible for the critical thinking improvements seen in one group but not the other. For example, there might be something fundamentally different about students who signed up for the night class versus students who picked the morning class. The experimental group receives the critical thinking intervention, while the control group receives

a control intervention. If the critical thinking intervention group shows greater improvements in critical thinking than the control group, then the critical thinking intervention is considered to be a success. This is the strongest design, though it might be the hardest to implement because it requires strict experimental methods to control for extraneous factors which may affect the outcome of your study. For example, it would be hard to randomly assign students to each condition within your class because you wouldn't want to disadvantage students should one treatment be more successful than the other. However, there are certain ways around this. For example, give group one treatment A first, while group two gets treatment B. Then, measure their improvement, and give group one treatment B and group two treatment A. Measure their improvements again.

APPENDIX A: Handout

Activity 1: Reviewing Examples from the Literature on Critical Thinking

For each of the four articles, answer the following questions below. 1) How is the critical thinking intervention described? 2) How were critical thinking skills assessed among students? 3) How were improvements measured?

Article 1 Title:

Article 2 Title:

Article 3 Title:

Article 4 Title:

Activity 2: Quantitatively Measuring Critical Thinking in Your Classroom

What would critical thinking among my students look like?

How will I quantify critical thinking? What types of assessments will I use to measure it?

What design will I use to assess the effectiveness of my critical thinking instruction and why?

References for More Information (Handout)

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102-1134.
<https://dx.doi.org/10.3102/0034654308326084>
- Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching and Learning*, 8(2), 25-42.
<https://dx.doi.org/10.19030/tlc.v8i2.3554>
- Brunt, B. A. (2005). Models, measurement, and strategies in developing critical-thinking skills. *The Journal of Continuing Education in Nursing*, 36(6), 255-262.
<https://dx.doi.org/10.3928/0022-0124-20051101-05>
- Dawson, T. E. (1997). *A primer on experimental and quasi-experimental design*. Austin, TX: Annual Meeting of the Southwest Educational Research Association, January 23-25. (ERIC No. ED406440).
- Ennis, R., & Millman, J. (1971). *Manual for Cornell critical thinking test, level X, and Cornell critical thinking test, level Z*. Urbana, IL: University of Illinois, Illinois Critical Thinking Project.
- Facione, P. A. (1991). *Using the California critical thinking skills test in research, evaluation, and assessment*. Millbrae, CA: California Academic Press. (ERIC No. ED337498).
- Facione, P.A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: California Academic Press. (ERIC No. ED 315 423).
- Gadzella, B. M., Hogan, L., Masten, W., Stacks, J., Stephens, R., & Zascavage, V. (2006). Reliability and validity of the Watson-Glaser critical thinking appraisal-forms for different academic groups. *Journal of Instructional Psychology*, 33(2), 141-143.
<https://dx.doi.org/10.1007/s10734-006-9002-z>
- Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches*. Boston, MA: Allyn & Bacon.
- Lawson, T. J., Jordan-Fleming, M. K., & Bodle, J. H. (2015). Measuring psychological critical thinking: An update. *Teaching of Psychology*, 42(3), 248-253.
<https://dx.doi.org/10.1177/0098628315587624>

- Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity*, 3(2), 85-93.
<https://dx.doi.org/10.1016/j.tsc.2008.03.005>
- Smith, J. K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher*, 12(3), 6-13. <https://dx.doi.org/10.2307/1175144>
- Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663-682. <https://dx.doi.org/10.1007/s10763-016-9723-0>
- Tiruneh, D. T., Weldeslassie, A. G., Kassa, A., Tefera, Z., Cock, M., & Elen, J. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Research and Development*, 64(3), 481-505.
<https://dx.doi.org/10.1007/s11423-015-9417-2>
- Williams, R. L., Oliver, R., & Stockdale, S. L. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, 53(1), 37-58.
<https://dx.doi.org/10.1353/jge.2004.0022>