

Aligning Best Practices to Develop Targeted Critical Thinking Skills and Habits

Ilea E. Heft¹ and Lauren F. V. Scharff²

Abstract: This project evaluated the effectiveness of a course design within an upper-level biology course that incorporated what prior scholarship of teaching and learning (SoTL) research has suggested to be best practices for developing critical thinking skills while also managing the grading load on the instructor. These efforts centered on the development of a clearly articulated subset of skills identified by the Critical Thinking Assessment Test (CAT) and also incorporated learning experiences designed to instill what we refer to as a “habit of critical investigation.” In this study, we tested the hypothesis that a single semester of an aligned course utilizing active learning and multiple opportunities for practice and feedback would: (a) increase the extent to which students agreed with the importance of questioning the credibility of claims across the semester, (b) increase the frequency at which students reported personally questioning the credibility of claims across the semester, (c) increase the number of students reporting investigation techniques consistent with critical investigation across the semester and (d) result in significantly greater student performance on the CAT questions that assessed the sub-skills practiced in the course when compared to the performance of a representative group of senior students at our institution. We observed substantial and statistically significant gains in both the frequency at which students reported questioning claims and the degree to which their reported investigative actions were consistent with critical investigation. Furthermore, on the critical thinking sub-skills most aligned with what was practiced in the course, the experimental group significantly outperformed the comparison group.

Keywords: active learning, alignment, critical thinking dispositions, critical investigation, critical thinking assessment test

Improving students' critical thinking skills is widely referenced as *the most important* outcome of education (Bok, 2006). Equally important to skill development is student attainment of the inclination and sensitivities that support the application of critical thinking skills in day-to-day life (often termed critical thinking dispositions)(Perkins & Ritchhart, 2004). However, multiple studies have concluded that higher education at large has not been very effective in developing students' critical thinking skills (e.g. Arum & Roksa, 2011). We suspect that current general shortcomings in critical thinking development are not due to lack of motivation or caring on the part of most faculty. Rather, we posit that misconceptions and lack of awareness regarding effective techniques for developing critical thinking skills and dispositions lead to courses that are not well aligned for

¹Department of Biology, United States Air Force Academy, 2355 Faculty Dr, USAF Academy, CO, 80840, ilea.heft@gmail.com

²Center for Educational Innovation, United States Air Force Academy, 2354 Fairchild Dr, USAF Academy, CO, 80840, lauren.scharff@usafa.edu

such development.

For example, prior to this research effort, the lead author assigned a major research essay with the proud belief that it would develop students' critical thinking skills. Then, through her experience grading the Critical Thinking Assessment (CAT) at her institution and follow-on discussions with faculty development staff, she had the career-changing realization that a single, summative assignment did not do much to support the development of her students' critical thinking skills despite the fact it took a substantial amount of time for students to write and the author to grade. In response to this realization, we sought to incorporate within an upper-level biology course what prior scholarship of teaching and learning (SoTL) research has suggested to be best practices for developing critical thinking skills while also managing the grading load on the instructor. Importantly, these efforts centered on the development of a clearly articulated subset of skills identified by the CAT and incorporated learning experiences designed to instill what we refer to as a "habit of critical investigation." This article summarizes our explicit example of how best practices can be effectively implemented to achieve substantial and significant gains in critical thinking skills and dispositions. It also shares the lessons learned and recommendations derived from our observations and results.

Best practices for student learning and development

What do we know about best practices for student learning and development that apply to learning goals in general? We know that alignment between all elements of a course is *essential* to achieving desired outcomes. The effectiveness of any pedagogical technique designed to promote student learning will be limited if the course is not aligned. The concept of alignment has been referred to as constructive alignment (Biggs, 1996), backwards course design (Wiggins & McTighe, 2005), and learning-focused course design (Fink, 2003; Jones, Noyd, & Sagendorf, 2014) and has been shown to greatly improve the achievement of desired outcomes (Cohen, 1987). Elements of a course are aligned when the learning experiences involve the student performing the desired skill, and the assessment explicitly assesses that skill. For example, if a goal is for students to be able to provide alternative explanations for observations, then learning experiences should involve the students practicing this skill, and the assessment should ask the student to provide alternative explanations for observations. This degree of alignment is sometimes negatively interpreted as "teaching to the test." However, "teaching to the test" can in fact be a positive characteristic. If the test assesses your desired skills, your goal should be to develop those skills.

In addition to alignment between course elements, there is substantial evidence that active learning is more effective than lecture in promoting development of both critical thinking skills (Tayyeb, 2013; Yuan, Williams, & Fan, 2008) and content understanding (Freeman et al., 2014) across a range of disciplines. For optimal development, the practice that occurs during active learning should be incorporated into *multiple* cycles of practice and feedback ("*performance-feedback-revision-new performance*") that focus student learning on the desired skills (Bok, 2006; Fink, 2003; Wiggins, 1998). The type of feedback provided to students during these cycles is also important. Sadler (1989) stated that for improvement to occur the learner must "possess a concept of the standard/goal being aimed for, compare the actual level of performance with the standard, and engage in appropriate action which leads to some closure of the gap" (p. 121). The feedback that is provided must be sufficiently detailed and clearly understood by the student to support each of these steps that lead to development. Additional characteristics of effective feedback noted in the literature are that it is frequent, immediate, delivered supportively (Fink, 2003), and task-

focused (e.g. comments rather than grades; Butler, 1988).

What do we know about best practices for developing the inclination to think critically outside of the classroom? The body of evidence is less robust, however, studies suggest that activities consistent with a culture of thinking (e.g. challenging students to ask questions, seek justification, and probe assumptions) are more effective than traditional “teaching by transmission” activities (e.g. lecture). Multiple studies have reported that students randomly assigned to a course section employing problem-based learning significantly outperformed students assigned to a content-equivalent, lecture-based course on the California Critical Thinking Disposition Inventory (CCTDI) (e.g. Ozturk, Muslu, & Dicle, 2008; Yu, Zhang, Xu, Wu, & Wang, 2013). Additionally, students in an online course in which dispositions were fostered by highlighting exemplars, praising students when they demonstrated critical thinking dispositions, and requiring students to describe examples of critical thinking dispositions in action were reported to outperform, on the CCTDI, other students who had been in an equivalent course without these features (Yang & Chou, 2008). Overall, such practices support the development of a culture of thinking, which in turn provides motivation for challenging mental work such as critical thinking (Ritchhart, 2015).

Challenges to implementing best practices

As indicated in our above review, prior research has identified best practices for student learning and development; however, they are often not implemented for several reasons. Some reasons include lack of knowledge of best practices by individual faculty, the time and effort required to change a course and to provide multiple iterations of feedback, and the (cognitively painful) realization that one’s current course may not be in line with best practices. Furthermore, pressure (perceived or real) to “cover” a certain quantity of content material within a semester may be a barrier to providing time and opportunities for effective skill development. This content-coverage-versus-skill-development challenge may be exacerbated by the assumption that students already have the desired skills and habits or that development of those skills and habits are the responsibility of a different course.

The implementation of best practices in support of critical thinking development is further hindered by expansive definitions of critical thinking. For example, the website for the *Foundation for Critical Thinking* (2015) defines critical thinking as:

that mode of thinking — about any subject, content, or problem — in which the thinker improves the quality of his or her thinking by skillfully analyzing, assessing, and reconstructing it. Critical thinking is self-directed, self-disciplined, self-monitored, and self-corrective thinking. It presupposes assent to rigorous standards of excellence and mindful command of their use. It entails effective communication and problem-solving abilities, as well as a commitment to overcome our native egocentrism and sociocentrism.

Similarly broad definitions exist for critical thinking dispositions (Facione, 1990). While a large set of discrete skills and dispositions can be considered to constitute critical thinking, broad definitions can be an impediment to the implementation of best practices because they are often overwhelming, do not paint a clear picture of what the student should be able to do, and mask the fact that individual sub-skills must be individually developed and assessed. In order to design an aligned course focused on critical thinking skills and to provide the targeted practice and feedback

that develops them, we must clearly define the specific skills we want to develop and the criteria for their assessment. We must also make these skills and criteria visible and explicit to students (Heijltjes, Van Gog, & Paas, 2014; Marin & Halpern, 2011).

Pedagogical misconceptions also hinder student development of skills. Two misconceptions we often encounter related to critical thinking development are the belief that a single summative assessment or stand-alone assignment (often an essay) develops critical thinking skills, and the assumption that students' skills will develop implicitly by completing assignments in the course that do not explicitly focus and provide feedback on the desired skills. A single summative assessment or stand-alone assessment is unlikely to be effective in developing students' skills because there is no requirement for students to use feedback on these assignments to improve their skills and try again. Rather than promoting development, a single assignment asks students only to demonstrate their skills at a particular point in time. Furthermore, even multiple cycles of practice and feedback will be of limited effectiveness unless the desired skills are made clear (Heijltjes et al., 2014; Marin & Halpern, 2011) and are made the primary focus of the assessment, e.g. they have a meaningful number of grade points dedicated to them. If the desired skills are not the primary focus of the assessment, it becomes possible for students to receive a high grade without demonstrating mastery of the desired skill (Wiggins, 1998) by demonstrating, for example, strong writing skills or content knowledge. The authors acknowledge that crafting and implementing multiple iterations of essay assignments that meet the above criteria can be a Herculean task that becomes difficult to implement due to the time required for students to write the essays and for the instructors to provide quality feedback. As a solution, below we describe an approach for developing and assessing critical thinking skills that incorporates best practices for skill development, is rigorous, and is manageable to implement.

Our approach for developing critical thinking skills and dispositions

As mentioned above, broad definitions of critical thinking make it challenging to design effective, aligned learning experiences. Further, there is a lack of guidance regarding how to assess these skills in a manner that is both rigorous and manageable. These challenges can be overcome by converting broad definitions into a set of discrete but rigorous skills with clearly defined assessment criteria. The Critical Thinking Assessment Test (CAT) (Center for Assessment and Improvement of Learning, 2016) provides an example for doing this. The CAT is a standardized critical thinking assessment created by researchers at the Tennessee Technological University through the support of a National Science Foundation grant.

One of the hallmarks of the CAT is that it is graded by faculty onsite at each institution, with very specific rubric instructions and guidance provided by trained facilitators. Institutional grading is submitted to the CAT organization, where a scoring check is performed. Data collected across several hundred institutions allow normative interpretations. While participating in one of our institution's CAT grading sessions, the lead author realized that the discrete skills assessed by the CAT could serve as a framework upon which to build targeted learning experiences and assessments that were manageable to implement and to assess. While the individual CAT sub-skills are concrete and explicit, they are not trivial or easy. When polled, most faculty agree that the questions on the CAT are valid measures of critical thinking (Stein, Haynes, & Redding, 2006). Additionally, the design of the CAT questions and scoring rubric model a way for these challenging skills to be probed and scored with relative ease, thus enabling implementation of multiple cycles of practice and feedback.

Seeing this approach, the lead author was inspired to design her Human Nutrition course to focus on the development of a subset of critical thinking skills from the CAT that naturally aligned with the goals she had for her students. Beyond these skills, the lead author also wanted her students to develop a complimentary set of critical thinking dispositions that we refer to as a “habit of critical investigation”. Our “habit of critical investigation” is conceptually similar to Perkin’s (1993) disposition “to seek and evaluate reasons” and includes: valuing the importance of questioning health/nutrition claims, frequently questioning health/nutrition claims, and when evaluating claims, taking actions that are most likely to result in a sound conclusion.

Thus, the key components of our approach were: (1) targeting discrete skills and habits associated with critical thinking, (2) aligning all elements of the course with development of those skills, (3) explicitly communicating intent and activity alignment with students, (4) giving students multiple opportunities throughout the semester to practice, receive prompt feedback, and be formally assessed on those skills, using examples that are highly relevant to them, (5) establishing clear scoring criteria so that the assessments could be easily and reliably evaluated. Through the use of multiple, salient activities and assessments, we hoped to create a classroom culture that clearly supported the value and development of critical thinking and the habit of critical investigation.

The purpose of our research was to investigate the effectiveness of this approach at developing the desired habits, attitudes, and skills. Specifically, we hypothesized that a single semester of this type of course would: (a) increase the extent to which students agreed with the importance of questioning the credibility of claims across the semester, (b) increase the frequency at which students reported personally questioning the credibility of claims across the semester, (c) increase the number of students reporting investigation techniques consistent with critical investigation across the semester and (d) result in significantly greater student performance on the CAT questions that assessed the sub-skills practiced in the course when compared to the performance of a representative group of senior students at our institution.

Methods

Participants

Three sections of the Fall 2013 Human Nutrition course, an upper-level course taught by the Biology Department, were used to evaluate the effectiveness of the learning experiences ($n=72$; 78% seniors; 22% juniors). This course was open to students from all majors who had completed the introductory Biology course. The sections for this study included large proportions of biology, management and civil engineering students. At the end of the semester, the CAT performance of the Human Nutrition (HN) cohort was compared to the CAT performance of a representative group of seniors who took the CAT in May of 2014, referred to as the senior (S) cohort ($n=96$; a representative sample of approximately 1000 students in the entire senior class). For all measures of student academic aptitude analyzed, independent sample *t*-tests revealed that the mean values for the Human Nutrition cohort were significantly lower ($p<.001$) than the senior cohort. These included Composite SAT (HN = 1247; S = 1338), Major’s GPA (HN = 2.79; S = 3.14), Cumulative GPA (HN = 2.8; S = 3.05), and an institution-designed Academic Composite score (HN = 3141; S = 3358) reflecting pre-college academic achievement.

Materials

Our assessment utilized two instruments, the CAT and an instructor-generated questionnaire created to assess students' habit of critical investigation. The CAT contains 15 multi-part questions (mostly open-ended) that assess different critical thinking skills. The CAT creators have grouped these 15 specific sub-skills into four categories: Evaluating Information, Creative Thinking, Learning and Problem Solving, and Communication. Examples of individual skills include "evaluate how strongly correlational-type data supports a hypothesis" and "provide alternative explanations for observations"; for the full list of sub-skills see Table 1.

Table 1

CAT Sub-skills, How They Aligned with the Course, and Student Performance for the Human Nutrition (HN) and Senior (S) Cohorts

Sub-skill	Skill Assessed	Alignment with HN	Max Possible Score	HN Cohort Mean	S Cohort Mean	ANCOVA F & p values*
1	Summarize the pattern of results in a graph without making inappropriate inferences.	High	1	0.94	0.75	F = 6.73 p=0.01
2	Evaluate how strongly correlational-type data supports a hypothesis.	High	3	2.1	1.83	F = 4.08 p=0.05
3	Provide alternative explanations for observations.	High/Partial	3	1.69	1.74	F = 0.03 p=0.86
4	Identify additional information needed to evaluate a hypothesis or particular explanation of an observation	High/Partial	4	1.15	1.58	F = 2.16 p=0.14
5	Evaluate whether spurious relationships strongly support a claim	Partial	1	0.94	0.92	F = 0.90 p=0.34
6	Provide alternative explanations for spurious relationships	High/Partial	3	1.85	2.15	F = 3.10 p=0.08
7	Identify additional information needed to evaluate a hypothesis/interpretation	Partial	2	0.41	0.58	F = 1.02 p=0.31
8	Determine whether an invited inference in an advertisement is supported by information	Partial	1	0.81	0.78	F=0.97 p=0.33
9	Provide relevant alternative interpretations of information.	High	2	1.26	1.08	F = 4.39 p=0.04
10	Separate relevant from irrelevant information when solving a real-world problem.	None	4	3.09	3.46	F = 6.77 p=0.01
11	Analyze and integrate information from separate sources to solve a real-world problem	None	2	0.87	1.03	F = 1.32 p=0.25
12	Use basic mathematical skills to help solve a real-world problem.	None	1	0.89	0.91	F = 0.03 p=0.87
13	Identify suitable solutions for a real-world problem using relevant information.	None	3	1.11	1.37	F = 0.16 p=0.69
14	Identify and explain the best solution for a real-world problem using relevant information	None	5	1.72	2.54	F = 5.16 p=0.03
15	Explain how changes in a real-world problem situation might affect the solution	None	3	0.94	0.98	F = 0.25 p=0.62
Total Score	--	--	38	19.79	21.69	F = 1.39 p=0.24

*df = 2 for all ANCOVA tests

Our Habit of Critical Investigation Questionnaire contained 5 items that asked for self-report responses on a combination of Likert-style and open-ended questions regarding specific behaviors associated with critical investigation. Importantly, one of the open-ended questions asked students to describe the behaviors they were likely to take when they questioned the credibility of a claim. The responses to this question provided a more direct assessment of students' behaviors than could be obtained from the Likert-style questions.

Responses to two questions indicated that they were not consistently interpreted across students, resulting in unusable data. The three remaining questions and their intent are described in Box 1.

Box 1. Questions utilized from the Habit of Critical Investigation Questionnaire

Question 1 (Assess frequency of questioning the credibility of information):

Question: “How frequently do you question the credibility of nutrition/health related claims that you hear/read?”

Response options: never, rarely, sometimes, most of the time, always.

Question 2 (Assess consistency of response with good critical investigation):

Question: Students were asked to provide a written response to the question “When you question the credibility of a nutrition related claim, what are you most likely to do?”

Question 3 (Assess agreement with the importance of questioning the credibility of information):

Question: Select the extent to which you agree or disagree with the following statement: “I think it is important to question the credibility of nutrition/health related information that I hear and read.”

Response options: Strongly Agree, Agree, Disagree, and Strongly Disagree

Course Design and Lesson Procedures

Incorporation of best practices. For the development of both the skills of critical thinking and the habit of critical investigation, the instructor explicitly shared the activity and lesson goals with students, and explicitly reinforced connections between the practice and assessment of those skills and habits throughout the semester. The course goals, learning experiences and assessments were all aligned with the development of critical thinking habits and skills (Box 2). The desired skills were explicitly communicated and substantial class time was devoted to their development: 2.5 lessons (out of 40) were entirely devoted to the development of critical thinking skills, plus small portions of at least 5 additional lessons included explicit reminders and/or activities. Multiple cycles of practice (active learning) and feedback occurred within the 2.5 devoted lessons as students worked through multiple problems in groups and as a class. Working through problems in class allowed students to receive frequent and immediate feedback. Application of the desired critical thinking skills was reinforced throughout the semester by beginning several lessons with a nutrition claim relevant to the day’s lesson and having students evaluate it as a class. The assessment of critical thinking skills on two formative, low-threat quizzes provided an additional means for each student to receive individual feedback. Students were allowed to retake these quizzes to improve their score, thus incentivizing the use of feedback for improvement. Critical

Box 2. Alignment between all course elements (Explicitly communicated with students)

Course Goals:

- Be able to determine the extent to which a nutrition related claim is supported by facts & data
- Be able to provide accurate dietary recommendations tailored to individuals and their goals
- Have an attitude of skepticism and a habit of critical investigation

Lesson Objectives (Lessons 4 & 5; additional half lesson of practice on Lesson 11)

- State the reasons that anecdotal evidence does not constitute strong support for a claim
- Identify if a study was observational or intervention and describe the limitations and benefits of each type of study
- Summarize the pattern of results, in a graph or in written form, without making inappropriate inferences.
- Evaluate how strongly a given claim is supported by the data provided
- Provide alternative explanations for a pattern of results that has many possible causes.
- Identify additional information needed to evaluate a claim.

Quizzes and Exams

- Formative, low-threat learning quizzes on Lessons 6 & 12
- Higher stakes exams on Lessons 14 & 28

Example practice/assessment question set for a nutrition claim related to Attention-Deficit/Hyperactivity Disorder (ADHD):

1. A researcher has noticed that more children seem to be diagnosed with ADHD and is wondering if it might have something to do with changes in the food supply. *Summarize the results of the graph below.*
2. *How strongly do the results support the claim that a change in the food supply is causing increased rates of ADHD?*
3. *Provide an alternative explanation for the results in the graph.*
4. *What additional information would you need to evaluate the claim that a change in the food supply is causing increased cases of ADHD?*

thinking skills were also evaluated on two higher stakes exams. The evaluation of critical thinking skills on these high stakes assessments reinforced the primacy of these skills to the course goals. The feedback provided on both high and low-stakes exams was delivered through detailed grading rubrics designed to help the student understand exactly where they needed to improve. Additionally, all instruction and practice of the critical thinking skills was done actively, with students applying critical thinking skills to evaluate authentic and relevant nutrition claims.

Developing Critical Thinking Skills. The critical thinking skills that we focused on in the course came from the CAT and included: summarize the pattern of results in a graph without making inappropriate inferences (Sub-skill 1), evaluate how strongly correlational-type data supports a hypothesis (Sub-skill 2), provide alternative explanations/interpretations (Sub-skill 3,

Sub-skill 6, Sub-skill 9), and identify additional information needed to evaluate a hypothesis or particular explanation of an observation (Sub-skill 4). Note however that in Table 1 we identify Sub-skill 3, Sub-skill 4, and Sub-skill 6 as having “high/partial” rather than “high” alignment. This reflects our post-hoc assessment of what actually occurred in the course. While we intended for each of these Sub-skills to be highly aligned, we later realized that key components were not well aligned or not sufficiently practiced. Specifically, Sub-skill 3 of the CAT included a particular component (articulation of change over time) that was only practiced one time in the course. For Sub-skill 4, correct answers were accepted during the course that focused on a description of sound experimental design rather than a clear articulation of information required to evaluate a hypothesis. Lastly, Sub-skill 6 required students to clearly articulate their responses at a level beyond what was focused on in the course. In Table 1 we also identify four sub-skills as having partial alignment because they were lightly utilized and closely related to the primary skills, but not explicitly discussed, practiced or given feedback for improvement.

Those sub-skills with high alignment had lessons 4 and 5 at the beginning of the course dedicated to their instruction, practice, and feedback. Students spent the majority of time during these lessons engaged in active learning of the concepts. For example, students were given a worksheet containing popular claims and worked individually and in groups to develop alternative explanations. The instructor (first author of this paper) then led the class through a discussion of possible alternative explanations with students providing the answers. The desired skills were practiced and assessed with questions similar in style to those on the CAT, which provided strong and clear alignment between what was practiced and what was assessed. In addition to these two lessons, these skills were integrated into multiple lessons and assessments throughout the semester. When the first formative assessment quiz revealed that many students still struggled with the skills, an additional half lesson was devoted to additional practice on these skills. As mentioned above, the integration of these skills into lessons throughout the semester reinforced the development of these skills as well as the habit of critical investigation.

Developing the Habit of Critical Investigation. In order to develop a habit of critical investigation, learning experiences were designed with the intent that, through the repeated application of critical thinking skills and critical investigation to relevant claims, students would realize that personal experience, anecdotes, and unsupported claims that had previously guided their decision-making were not valid. We posited that through these experiences, students would come to see critical investigation as personally valuable to them, useful for realistically achieving their health and/or performance goals, and for avoiding scams. The hope was that if students viewed critical investigation as personally valuable to them, that they would adopt such behaviors in their own life outside of the classroom. Specifically, lessons four and five of the course included explicit instruction on the limitations of anecdotal evidence, personal experience, and observational studies and how to identify flaws in experimental design that affect the reliability of a conclusion. Students were also taught a specific method for critically investigating claims, were provided with specific credible websites that they could use to verify claims, and were given time to practice using these websites in class to research claims in which they were interested.

Assessment Procedure

Prior to data collection, the researchers received ethics approval for the study from the Institutional Review Board. The Critical Thinking Assessment Test (CAT) was administered on the second-to-last lesson of the semester, and students were given the full period to complete it. The majority of

students finished within 45 minutes, while a few finished sooner than that and a few used the full class period.

The Habit of Critical Investigation Questionnaire was administered on lessons 1, 20, and 38 (of 40 total lessons), and took approximately 10 minutes for the students to complete. To remove any perception on the students' part that their answers on the questionnaires could affect their grade, a person not affiliated with the course collected the forms and assigned a random number to each set of questionnaires. This individual separated the students' names from the questionnaires so that the responses could be reviewed and general trends discussed during class.

Results and Discussion

Critical Thinking Skill Development

In order to determine if the systematic implementation of critical thinking development practices led to a significant improvement in critical thinking as measured by the CAT, we compared end-of-semester results from the Human Nutrition (HN) cohort and the senior (S) cohort. Specifically, we used two-group comparisons to examine total CAT scores as well as each sub-score. Of particular interest was whether or not those sub-skills that were targeted in the HN course showed greater group differences than those sub-skills that were not targeted. Following the guidelines and rubric provided by the CAT creators, student performance was first scored by a group of our faculty. Scores were then verified by the CAT office.

Academic Composite was used as a covariate because pre-analyses indicated that the two groups were initially different on this measure and for both groups combined it significantly correlated with overall CAT scores $r(145) = 0.34, p < .05$. Therefore, CAT tests for which the Academic Composite score of the test taker could not be identified (by linking student number on the CAT test to an academic database) were omitted (9 HN CAT scores and 3 S CAT scores) leaving 54 HN CAT scores and 93 S CAT scores.

The right three columns of Table 1 show mean scores for the HN and S cohorts and the F and p values, respectively. For the HN cohort, the average total score, low and high scores were 19.79, 11, and 30, respectively, and for the S cohort they were 21.69, 12, and 31, respectively. On three of the sub-skills of critical thinking most aligned with the skills practiced in Human Nutrition (Sub-skills 1: summarize graph, 2: correlational data, and 9: alternative interpretations), students who took the Human Nutrition course significantly outperformed the representative sample of seniors. Specifically, the HN cohort outperformed the senior cohort by 19% on Sub-skill 1 and by 9% on both Sub-skills 2 and 9. However, the HN students did not perform better than the senior cohort on those Sub-skills with high/partial alignment (Sub-skills 3, 4, 6) or any of those with only partial or no alignment.

Overall these group comparison results have some implications regarding the development of critical thinking. First, despite the limitation that the CAT was given only at the end of the semester (and not pre-post) for the HN cohort, we believe the HN cohort achieved meaningful development in the critical thinking sub-skills that were well targeted by the course activities. Because the HN cohort was composed of 22% juniors and performed worse (in two cases significantly so) than the S cohort on most CAT items not practiced in class, it seems reasonable to conclude that they were not any better than the S cohort on Sub-skills 1, 2, and 9 prior to taking the class. It can therefore be cautiously concluded that the level of development that occurred on those targeted items over the semester is at least as great as the final difference in performance between the HN cohort and the S cohort, and that the development was due to the targeted practice

throughout the semester.

A second conclusion is that, even in a class with an overt focus on critical thinking, students develop little in the areas for which they do not receive multiple cycles of aligned practice and feedback. On those sub-skills of critical thinking for which the activities in class were less well-aligned with what was assessed on the CAT, HN student performance was not significantly different from the S cohort. Not surprisingly, for the sub-skills not at all practiced in the class, student performance either was not significantly different across the two cohorts, or the senior cohort performed significantly better. A pattern of significant improvement on only a few sub-skills is consistent with previous studies that used the CAT to assess critical thinking skills (e.g. Frisch, Jackson, & Murray, 2013; Gasper & Gardner, 2013). These findings reinforce the idea that critical thinking is made up of a large set of diverse skills and that multiple rounds of explicit practice and feedback is required with each individual skill.

Habit of Critical Investigation

In order to examine the development of habits of critical investigation, responses from the pre, mid, and end-of-semester questionnaire were scored and analyzed as follows. The data from all three Human Nutrition sections were treated as a single group. If data from all three time-points was not available for a participant on a particular question then all responses from that participant, for that question, were dropped. Responses to each question were scored as described in Box 3.

Box 3. Analysis methods for Habit of Critical Investigation questions

Question 1 (Assess frequency of questioning the credibility of information):

Analysis Method: Responses were given a score of 1-5 with “1” representing an answer of “never” and 5 representing an answer of “always.”

Question 2 (Assess consistency of response with critical investigation):

Analysis Method: As detailed below, student responses were categorized as being: consistent with critical investigation, unclear with respect to being consistent with critical investigation, or inconsistent with critical investigation. The types of responses in each category are shown below.

Consistent with Critical Investigation

Consult a credible source
Evaluate the credibility of the source
Evaluate the study design
Evaluate it using biological knowledge
Evaluate the nutrition facts label

Consistency with Critical Investigation Unclear

Consult another source (credibility not mentioned)
Consult a teacher/professional/nurse/M.D.
Consult a Coach/Trainer

Inconsistent with Critical Investigation

Try it for myself
Consult someone who has tried it
Do not use the product/Do No Research
Consult a biased/non-credible source
Ask a friend or family member (rationale not stated)
Don't worry about/"Eat it anyways"
Gut Instinct/Guess
Consult someone who is "very health-conscious"
Consult the most convenient source (e.g. internet)

Responses in which all statements were consistent with critical investigation were awarded a value of “2”. If the consistency of the response with critical investigation was unclear and did not include any inconsistent statements, the response was awarded a value of “1.” If the response included a combination of “consistent” and “unclear” statements, the response was also awarded a value of “1”. Any response that included statements not consistent with critical investigation was awarded a value of “0.”

Question 3 (Assess level of agreement with the importance of questioning the credibility of claims):

Analysis Method: Responses were awarded a score of 1-4 with “1” representing an answer of “Strongly Disagree” and 4 representing an answer of “Strongly Agree.”

The frequency with which respondents stated they “questioned the credibility of claims” increased steadily over the course with the percentage of respondents selecting “Always” increasing from 9% to 25% by the middle of the course and to 41% by the end of the course. See Figure 1 for a summary of all responses for the three time periods. Mean scores with a score of 5 being the highest possible were 3.64 (pre), 3.98 (mid), and 4.33 (post). A single-factor (time) within-subjects ANOVA suggests that the shifts in the student responses were highly reliable; $F(2,126) = 27.86$ ($p < 0.001$). Post-hoc Tukey’s tests revealed that all paired differences (pre-post; pre-mid; mid-end) were significant ($p < .01$).

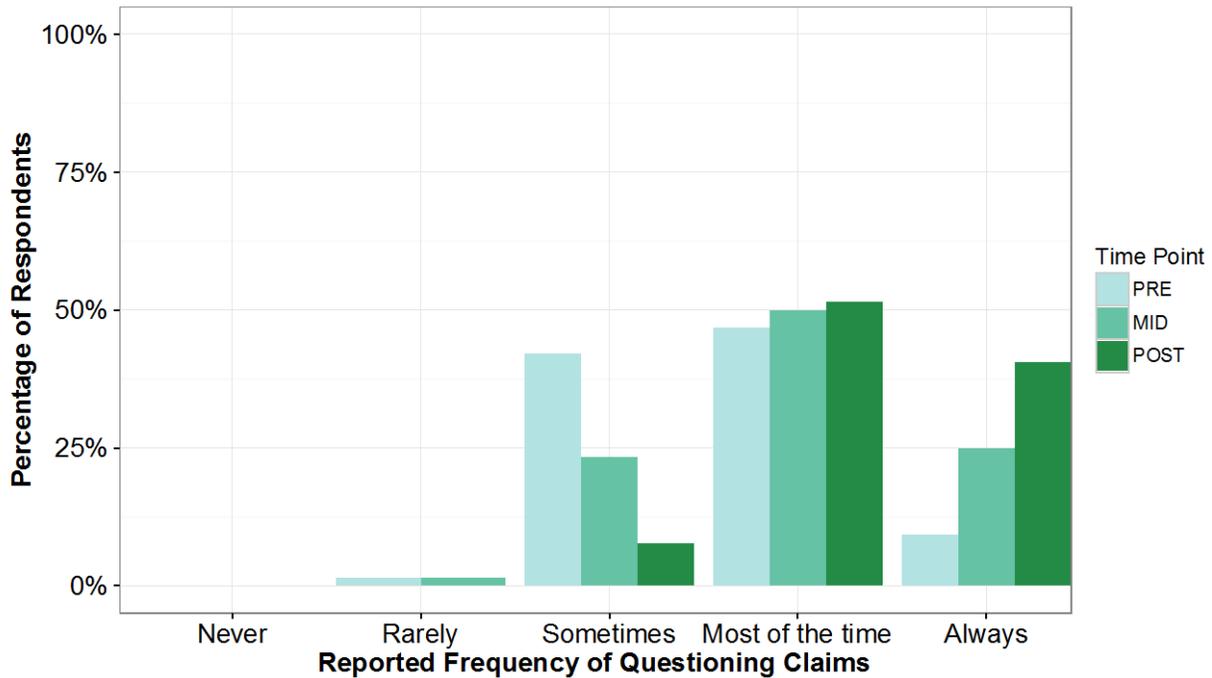


Figure 1. At pre, mid, and post semester, the percentage of respondents (N = 64) reporting each level of frequency in response to the question “How frequently do you question the credibility of nutrition/health related claims that you hear/read?”

When asked to “describe the actions they would take when investigating a nutrition related claim,” the percentage of respondents providing responses consistent with critical investigation increased from 20% to 44% by the middle of the semester and to 77% by the end of the semester. Further, by the end of the semester, the percentage of respondents whose actions were clearly inconsistent with critical investigation had dropped from 52% to 16.39% (Figure 2). Mean scores, with a score of 2 being the highest possible (answers consistent with critical investigation), were 0.67 (pre), 1.16 (mid), and 1.61 (post). A single-factor (time) within-subjects ANOVA suggests that the shifts in the student responses were highly reliable; $F(2,120) = 21.23$ ($p < 0.001$). Post-hoc Tukey’s tests revealed that all paired differences (pre-post; pre-mid; mid-end) were significant ($p < .01$).

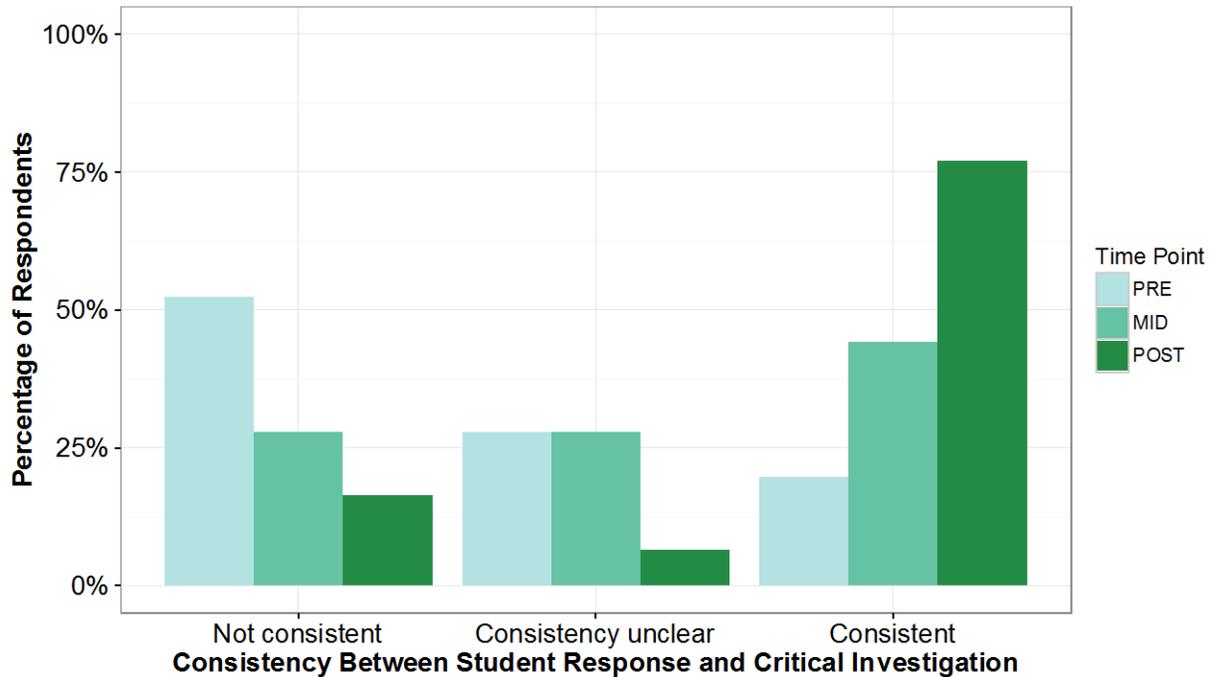


Figure 2. At pre, mid, and post semester, the percentage of respondents (N = 61) reporting habits consistent with critical investigation, habits for which the consistency with critical investigation is unclear, and habits which are clearly not consistent with critical investigation.

When asked to indicate their level of agreement with the statement “I think it is important to question the credibility of nutrition/health related information that I hear and read,” the majority of participants at all time periods indicated strong agreement. Mean scores, with a score of 4 being the highest possible (strongly agree), were 3.67 (pre), 3.67 (mid), and 3.63 (post). A single-factor (time) within-subjects ANOVA indicated no reliable changes in perceived value across the semester (Figure 3).

Taken together, these results have several implications regarding the development and assessment of habits of critical investigation. One of the more striking findings was the disconnect between students’ reported agreement with the importance of questioning claims, their self-reports of actually practicing it, and their self-reported actions when investigating a claim. Even at the beginning of the semester the vast majority of students reported strongly agreeing with the importance of questioning claims, but only 10% reported engaging in such practice all of the time and 47% most of the time. Further, when their actual self-reported actions were evaluated for consistency with good critical investigation, only 20% at the beginning of the semester described actions that were completely consistent with critical investigation. This pattern indicates that reports of agreement with the importance of an action are not reliable indicators of behaviors, and even more crucial, self-reports of behavioral frequencies are not necessarily a reliable indicator of whether their actions would be consistent with critical investigation. Therefore, if you want to assess the effectiveness of a course in developing students’ habits of critical investigation, it is

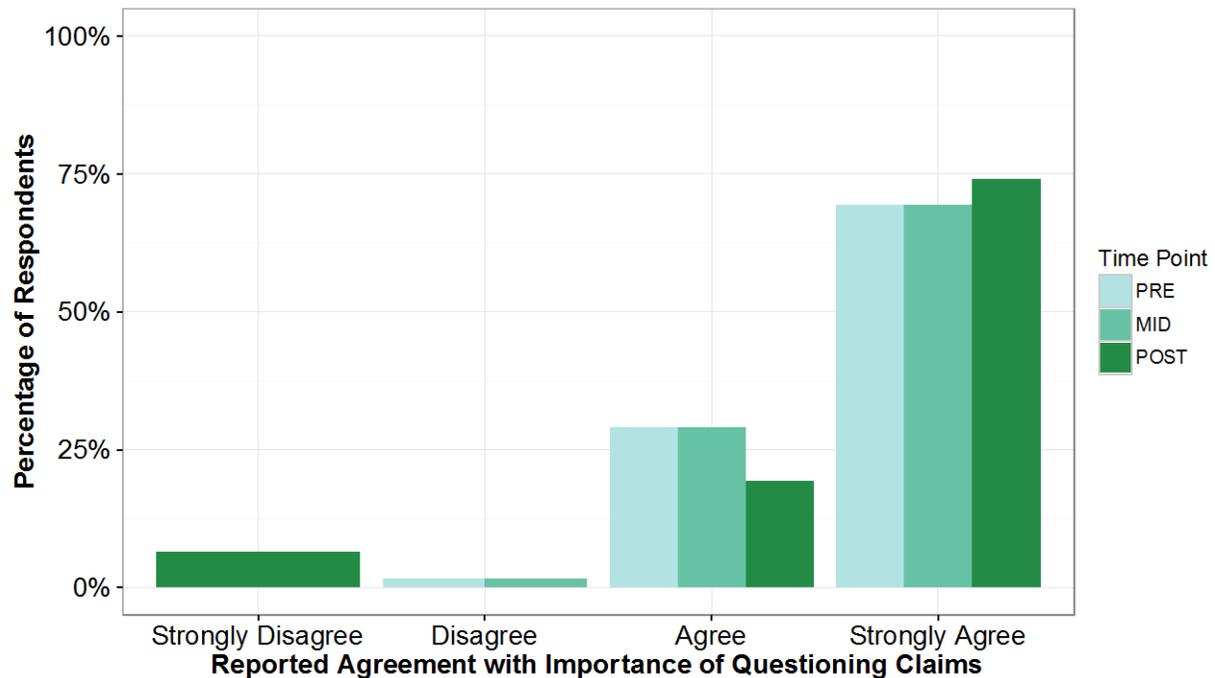


Figure 3. At pre, mid, and post semester, the percentage of respondents (N = 62) reporting each level of agreement with the statement “I think it is important to question the credibility of nutrition/health related information that I hear and read.”

important to examine students’ actions rather than just their intent. This can be facilitated by the use of open-ended rather than Likert-style questions.

A second implication of our results is that when provided with intentional focus on the value of critical investigation and opportunities to practice with feedback, students report significant changes in the frequency with which they engage in critical investigation and show significant increases in self-reported actions that are consistent with good critical investigation.

A third implication of our results is that the gain in self-reported frequency of actions and the number of students reporting actions consistent with a habit of critical investigation doesn’t happen in just a few weeks. Our students showed significant gains both from pre- to mid-semester and from mid-semester to the end of the semester. Thus, if you desire to develop this habit in your students, continue to focus on developing it throughout the entire semester (or until an assessment shows you that your students have fully achieved your desired outcome).

The lack of any significant increase in the reported agreement with the importance of questioning the credibility of information is potentially due to a ceiling effect for the majority of the respondents (68% of respondents selected the highest possible response on the pre-class questionnaire and 98% selected one of the top two responses). Surprisingly, 4 individuals selected the lowest possible response on the post-course questionnaire. However, these same students reported high frequencies of questioning claims and actions consistent with critical investigation. Thus, we believe they might have misread the response options and thought they were indicating strong agreement rather than strong disagreement. It seems unlikely that someone who “Strongly Disagrees” with the statement that it is important to question the credibility of claims they encounter would report that they “Always” question the credibility of claims. In the future, we

recommend including an open-ended question that asks students to describe the degree to which their attitude has changed and how.

General Conclusions

Based on the above findings, we conclude that an aligned course utilizing active learning and multiple opportunities for practice and feedback is an effective means of developing students' targeted critical thinking skills as well as their habit of critical investigation. Specifically, for the targeted sub-skills of the CAT that had high alignment, the HN cohort significantly outperformed the senior-level control group that had not received the targeted practice and feedback. Further, the percentage of students who reported taking actions consistent with critical investigation when they question a claim increased from 20% to 77% over the course of the semester.

Our results reinforce the need to break expansive definitions into discrete sub-skills and provide opportunities for explicit practice and feedback with each individual sub-skill. Because of the large number of sub-skills within the expansive definitions, it would be extremely difficult to effectively target them all within a single semester course. We believe instructors should choose the specific components that best fit their course goals. If such targeted efforts were combined with other courses targeting other components, then across a curriculum, a student should develop broader critical thinking skills and habits.

We also believe that the targeting of discrete sub-skills through the cycles of activities, discussion, and feedback led to a more manageable load for the instructor and the students when compared to previous semesters when the lead author assigned a long research paper that was intended to require critical thinking. The total effort put into critical thinking development across the semester of this study was at least as great as the writing and grading of the longer essay papers in previous semesters, however, because it was distributed over time and occurred in smaller chunks, it was more manageable for both the instructor and the students and followed best practices for skill development. Further, the targeted nature of the skill development made the learning goals clearer, which led to clearer evaluation of performance and allowed feedback to be more focused and rigorous. An approach consisting of small, manageable chunks is also more likely to be adopted by other instructors than one which involves a large overwhelming assignment. One additional benefit of our approach was that students did not display the negative responses often associated with major papers, but rather, they were engaged and seemed to enjoy the critical thinking activities and assignments.

Despite these encouraging results, we must also acknowledge some limitations of our study. First, our measures of the habit of critical investigation are self-reports, which means some caution is warranted when making claims about those results due to possible demand characteristics. However, we derive some comfort from the fact that fewer of our participants reported engaging in critical investigation than the number who reported that it was important to do so, i.e. they didn't chose to max out their responses on both questions in order to portray themselves in the most positive light. Further, our open-ended question required students to describe how they would approach a claim requiring critical investigation, thus minimizing their ability to inaccurately claim proficiency.

Second, we also acknowledge that we did not have pre-class measures of CAT performance for the Human Nutrition Cohort. Therefore, the exact level of development that occurred over the course of the semester cannot be quantified. However, given that the Human Nutrition Cohort had lower initial measures of academic ability than the Senior Cohort, it is reasonable to conclude that

their initial CAT sub-skill scores would not have been significantly higher than the Senior Cohort at the beginning of the course, and that they may have been lower. Given this logic, we are comfortable concluding that the difference in performance between the two groups indicates the minimum level of development in the HN cohort that occurred over the semester.

As we encourage others to implement similar approaches in their courses and curriculum we have several recommendations. In addition to incorporating multiple cycles of well-aligned practice and feedback focused on well-defined sub-skills and creating a culture of thinking, we recommend that instructors do the following. They should focus on developing the desired skills and dispositions throughout the entire semester, or until an assessment confirms the desired level of mastery has been attained. This recommendation is supported by our results showing that, over the second half of the semester, students made substantial and significant improvement in the habit of critical investigation above and beyond the gains made in the first half of the semester. Furthermore, some students may not achieve proficiency by the end of the semester (e.g. 17% of respondents at the end of the course still reported taking actions inconsistent with critical investigation). Thus, faculty, program directors, and general curriculum designers should ensure that multiple courses target desired skills, both to reinforce skills for those students who have achieved competency and to provide additional opportunities for development for those students who do not achieve competency after a single course.

Based on our findings and on the literature, we plan to modify our pedagogy and incorporate additional best practices. In an effort to promote deeper thinking (De L'Etoile, 2008) and minimize the likelihood that students memorize a formulaic type of answer, which appears to have occurred on Sub-skill 4, we plan to add open-ended justification components to skills assessment questions. We also plan to incorporate activities designed to increase students' sensitivity to opportunities for critical thinking (Halpern, 1998), as low sensitivity has been reported to be a large contributor to low performance on measures of ability (Perkins, Tishman, Ritchhart, Donis, & Andrade, 2000).

The scope of our intervention may seem daunting, but the evidence suggests that there is broad need for such efforts. Even at our highly selective institution, roughly half of our upper-level students described taking actions clearly inconsistent with critical investigation at the beginning of the semester, and national data (e.g. Arum & Roksa, 2011) suggests this is not isolated to our institution. We believe that the critical thinking skills and approach outlined here are applicable across the disciplines and levels of expertise. The media provides a continuous stream of claims to which critical thinking should be applied, and even among experts in a field there is often disagreement on the proper interpretation of data. The current climate of information overload and readily available claims in the media underscore the need to implement best practices for developing critical thinking skills and creating a culture of thinking. We hope that instructors in all disciplines are inspired by the clarity and viability of this approach to incorporate these practices within their own classrooms.

Acknowledgements

We thank Dr. Steve Jones for leading the CAT grading session, for providing invaluable advice as to the implementation of CAT analogs and best practices in education, and for providing valuable feedback on this manuscript. We also thank our colleagues for assistance in grading the CAT assessment and their general support of the project.

References

- Arum, R., & Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. Chicago: The University of Chicago Press.
- Biggs, J. (1996). Enhancing Teaching through Constructive Alignment. *Higher Education*, 32(3), 347–364. Retrieved from <http://www.jstor.org/stable/3448076>
- Bok, D. (2006). *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*. Princeton: Princeton University Press.
- Butler, R. (1988). Enhancing and Undermining Intrinsic Motivation: the Effects of Task-Involving and Ego-Involving Evaluation on Interest and Performance. *British Journal of Educational Psychology*, 58(1), 1–14. <http://doi.org/10.1111/j.2044-8279.1988.tb00874.x>
- Center for Assessment and Improvement of Learning. (2016). Critical Thinking Assessment Test. Retrieved June 7, 2016, from <https://www.tntech.edu/cat/>
- Cohen, A. S. (1987). Instructional Alignment: Searching for a Magic Bullet. *Educational R*, 16(8). <http://doi.org/10.3102/0013189X016008016>
- De L'Etoile, S. K. (2008). Applying Perry's Scheme of Intellectual and Ethical Development in the College Years to Undergraduate Music Therapy Education. *Music Therapy Perspectives*, 26(2), 110–116. <http://doi.org/10.1093/mtp/26.2.110>
- Facione, P. A. (1990). *Critical Thinking : A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. The California Academic Press. Retrieved from http://www.insightassessment.com/pdf_files/DEXadobe.PDF
- Fink, L. D. (2003). *Creating Significant Learning Experiences*. San Francisco: Jossey-Bass.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 111(23), 8410–8415. <http://doi.org/10.1073/pnas.1319030111>
- Frisch, B. J. K., Jackson, P. C., & Murray, M. C. (2013). WikiED: Using Web 2.0 Tools to Teach Content and Critical Thinking. *Journal of College Science Teaching*, 43(1).
- Gasper, B. J., & Gardner, S. M. (2013). Engaging Students in Authentic Microbiology Research in an Introductory Biology Laboratory Course is Correlated with Gains in Student Understanding of the Nature of Authentic Research and Critical Thinking. *Journal of Microbiology & Biology Education*, 14(1), 25–34. <http://doi.org/10.1128/jmbe.v14i1.460>
- Halpern, D. (1998). Teaching Critical Thinking for Transfer Across Domains. *American Psychologist*, 53(4), 449–455.
- Heijltjes, A., Van Gog, T., & Paas, F. (2014). Improving students' critical thinking: Empirical support for explicit instructions combined with practice. *Applied Cognitive Psychology*, 28(4), 518–530. <http://doi.org/10.1002/acp.3025>

Jones, S., Noyd, R., & Sagendorf, K. (2014). *Building a Pathway to Student Learning*. Sterling: Stylus Publishing.

Marin, L. M., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, 6(1), 1–13. <http://doi.org/10.1016/j.tsc.2010.08.002>

Ozturk, C., Muslu, G. K., & Dicle, A. (2008). A comparison of problem-based and traditional education on nursing students' critical thinking dispositions. *Nurse Education Today*, 28(5), 627–632. <http://doi.org/10.1016/j.nedt.2007.10.001>

Perkins, D., & Ritchhart, R. (2004). When is good thinking? In D. Dai & R. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*. Mahwah, NJ: Routledge.

Perkins, D., Tishman, S., Ritchhart, R., Donis, K., & Andrade, A. (2000). Intelligence in the Wild. *Educational Psychology Review*.

Ritchhart, R. (2015). *Creating cultures of thinking*. San Francisco: Jossey-Bass.

Stein, B., Haynes, A., & Redding, M. (2006). Project CAT: Assessing Critical Thinking Skills. *Proceedings of the National STEM Assessment Conference*, 290–299.

Tayyeb, R. (2013). Effectiveness of problem based learning as an instructional tool for acquisition of content knowledge and promotion of critical thinking among medical students. *Journal of the College of Physicians and Surgeons Pakistan*, 23(1), 42–46. <http://doi.org/01.2013/jcpsp.4246>

Wiggins, G. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass.

Wiggins, G., & McTighe, J. (2005). *Understanding By Design*. Alexandria: Association for Supervision and Curriculum Development.

Yang, Y. T. C., & Chou, H. A. (2008). Beyond critical thinking skills: Investigating the relationship between critical thinking skills and dispositions through different online instructional strategies. *British Journal of Educational Technology*, 39(4), 666–684. <http://doi.org/10.1111/j.1467-8535.2007.00767.x>

Yu, D., Zhang, Y., Xu, Y., Wu, J., & Wang, C. (2013). Improvement in Critical Thinking Dispositions of Undergraduate Nursing Students Through Problem-Based Learning: A Crossover-Experimental Study. *Journal of Nursing Education*, 52(10). <http://doi.org/10.3928/01484834>

Yuan, H., Williams, B. A., & Fan, L. (2008). A systematic review of selected evidence on developing nursing students' critical thinking through problem-based learning. *Nurse Education Today*, 28(6), 657–663. <http://doi.org/10.1016/j.nedt.2007.12.006>