1-2004

Psychological Versus Generic Critical Thinking as Predictors and Outcome Measures in a Large Undergraduate Human Development Course

Robert L. Williams
University of Tennessee

Renee Oliver
University of Tennessee - Chattanooga

Susan Stockdale
Kennesaw State University, sstockda@kennesaw.edu

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/facpubs>

Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Educational Psychology Commons](#)

Recommended Citation

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

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Faculty Publications by an authorized administrator of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

PSYCHOLOGICAL VERSUS GENERIC CRITICAL THINKING AS PREDICTORS AND OUTCOME MEASURES IN A LARGE UNDERGRADUATE HUMAN DEVELOPMENT COURSE

Author(s): Robert L. Williams, Renee Oliver and Susan Stockdale

Source: *The Journal of General Education*, Vol. 53, No. 1 (2004), pp. 37-58

Published by: Penn State University Press

Stable URL: <http://www.jstor.org/stable/27797975>

Accessed: 02-08-2016 13:40 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Penn State University Press is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of General Education*

PSYCHOLOGICAL VERSUS GENERIC CRITICAL THINKING AS PREDICTORS AND OUTCOME MEASURES IN A LARGE UNDERGRADUATE HUMAN DEVELOPMENT COURSE

Robert L. Williams, Renee Oliver, and Susan Stockdale

Introduction

Few concepts have attracted more attention in higher education than the notion of critical thinking. Although a variety of definitions has been advanced for critical thinking, most appear to emphasize the ability to construct and evaluate conclusions from available evidence and assumptions (Williams & Worth, 2001). This ability seems integral to success in college courses. In principle, critical thinking could serve both as a predictor of course performance and as an outcome of learning experiences in the course. However, the potential of various critical thinking measures to predict course performance has seldom been directly contrasted with changes in critical thinking ability as a result of course instruction.

Predictive Potential

The predictive capacity of critical thinking likely differs both by the type of critical thinking measure used and by the type of performance measure predicted. With respect to the first issue, critical thinking tests may be classified as either generic or subject-specific. One might expect a subject-specific measure of critical thinking to be more strongly linked to performance in a course than a generic measure. However, the literature reveals little research that has directly compared the predictive potential of the two types of critical thinking measures. One of the few studies to make this comparison found precourse statistical reasoning to be a better

JGE: THE JOURNAL OF GENERAL EDUCATION, Vol. 53, No. 1, 2004.
Copyright © 2004 The Pennsylvania State University, University Park, PA.

predictor of related postcourse measures than a precourse general critical thinking measure (Royalty, 1995).

Researchers have found critical thinking skills predictive of performance in a variety of college courses, such as physics and psychology (Gadzella, Ginther, & Bryant, 1997; McCammon, Golden, & Wuensch, 1988). Nonetheless, critical thinking may be more strongly related to some course measures than to others. Presumably, if a performance task requires a high level of critical thinking, then critical thinking measures should strongly predict performance for that task. For example, critical thinking might better predict performance on a test requiring inferential thinking than one requiring only recall or recognition of factual information. Consistent with the former possibility, Williams and Worth (2002) found generic critical thinking to be a stronger predictor of performance on multiple-choice tests requiring inferential reasoning than on other established predictors (student attendance and notetaking). In contrast, notetaking was the best predictor of performance on essay quizzes requiring direct recall of information.

Outcome Potential

A number of researchers and various commissions have proposed that critical thinking is among the most important outcomes of a college education (Halpern, 1988; Jones, 1995; Resnick & Peterson, 1991). However, the effects of individual courses on critical thinking remain somewhat equivocal. Some researchers (Allegretti & Frederick, 1995; Bensley & Haynes, 1995; Isaacs, 1991; Reed & Kromrey, 2001; Sandor, Clark, Campbell, Rains, & Cascio, 1998; Williams, Oliver, Allin, Winn, & Booher, 2003) have produced critical thinking gains in academic courses, but other researchers have failed to do so (Arburn, 1998; Forbes, 1997; Lierman, 1997; Lyle, 1958; Slaughter, Brown, Gardner, & Perritt, 1989).

Three factors may fundamentally affect the possibility of changing critical thinking in college courses: the nature of the critical thinking measure, the nature of the course experience, and the nature of the student. For example, one might expect a subject-specific measure of critical thinking to be more changeable than a generic measure. Subject-specific critical thinking could readily be targeted in tasks

required in a subject-matter course. Although subject-specific measures of critical thinking have been developed in such areas as psychology (Lawson, 1999), biology (McMurray, Beisenherz, & Thompson, 1991), and statistics (Royalty, 1995), these measures have mainly been used as predictors of performance rather than as outcome measures of course experiences.

Perhaps the most important issue in determining whether a course experience should promote critical thinking is the instructional format of the course. For example, courses involving tasks that require students to construct and evaluate conclusions from available evidence should promote critical thinking. Also, courses that allow students to interact with one another in evaluating arguments appear more conducive to critical thinking than those in which the teacher simply lectures about argument evaluation (Tsui, 1998). Additionally, Garside (1996) reported that group discussion produced higher performance on test questions requiring higher-order reasoning, whereas lecturing produced better performance on test questions requiring lower-order reasoning.

Another potentially important issue regarding course-based changes in critical thinking is the interaction between instruction and student characteristics. For example, a particular instructional model might be very effective in facilitating critical thinking for high-performing students but not for low-performing students (Lyle, 1958; Williams et al., 2003). Of particular interest in the current study is whether the same instructional strategy similarly affects the critical thinking skills of students who do well and who do poorly on the course exams. In a related study, Royalty (1995) reported that statistics students who made high scores on an end-of-the-course statistical knowledge test showed an increase in statistical reasoning during the course, whereas students who did poorly on the knowledge test did not improve their statistical reasoning.

The overall purpose of the current study is to compare the predictive and outcome status of a subject-specific versus a generic measure of critical thinking in a large undergraduate course. Specifically, the predictive potential of critical thinking was assessed with respect to test performance, both for tests requiring critical thinking and for those requiring only direct recall. The study also

examines the extent to which answering practice-exam questions with embedded psychological critical thinking issues produced improvement in critical thinking. Finally, the study assesses the impact of the treatment on the critical thinking of students who performed well or poorly on the course tests requiring considerable application of critical thinking.

Method

Participants

More than 200 students in five sections (ranging from 25 to 55 students per section) of an undergraduate Human Development course participated in various phases of the study. The gender ratio of participants favored women three to one. Although the course was taken by freshmen through seniors, close to 60% of the students were sophomores and juniors. Students earned a small amount of course credit for participating in the research, but equivalent credit was available for non-research activities. More than 100 students in three treatment sections of the course completed all pre- and postassessments, and approximately 70 students in two control sections also completed all assessments. To permit additional subgroup comparisons for critical thinking outcomes, we used the criterion-referenced grading standards in the course to identify high performers (students earning As) and low performers (students making Ds or Fs) on the course multiple-choice exams in both the treatment and control sections. The cell *ns* for these performance groups varied depending on the completion rate of the pretest and posttest measures of each critical thinking instrument; however, all performance-cell *ns* were low (ranging from 2 to 8).

Assessment Measures

The two types of assessment measures were critical thinking instruments and course tests. Critical thinking was assessed by both generic and subject-specific measures of critical thinking, with the same two critical thinking instruments administered at the beginning and end of

the course (approximately a four-month period between the critical thinking assessments). In addition, the two types of course tests used were multiple-choice exams requiring critical thinking and brief essay quizzes requiring only direct recall of information.

Critical thinking measures. The only critical thinking instrument designed specifically for the area of psychology was used to assess subject-specific critical thinking (Lawson, 1999). The *Psychological Critical Thinking* instrument uses an essay format consisting of 14 scenarios describing various psychological claims. Respondents judge whether each claim follows from the information given and, if not, what fallacies are embedded in the claim. All claims are counter to the principles of psychological science, relating to such issues as comparison groups, confounding variables, generalization of findings, and experimenter bias. Using a qualitative scoring procedure developed for this study, graduate teaching assistants rated each student's response to each scenario on a 0 to 3 scale: 0 = *no problem identified*, 1 = *a problem recognized but misidentified*, 2 = *some aspect(s) of the actual problem(s) specified*, and 3 = *actual problem(s) fully elaborated*. Overall inter-rater reliability for pairs of raters who rated approximately one-third of the inventories proved to be 0.88 for the pretest and 0.94 for the posttest.

The second measure of critical thinking used in this study (*Watson-Glaser Critical Thinking Appraisal—WGCTA*) is probably the most widely used generic measure of critical thinking at the college level (Watson & Glaser, 1980). The particular form used in the current study (Form S) is an abbreviated version of the original Form A (Watson & Glaser, 1994). Form S was designed primarily for adults, including college students. It uses a multiple-choice format, with the item options ranging from two to five. Respondents are instructed to judge the probable credibility of conclusions based on assumptions and information provided in the test. The test manual reports both the internal consistency and the test-retest reliability for Form S to be 0.81. The instrument also is reported to be moderately predictive of academic and professional indices of success. This instrument was selected as our measure of generic critical thinking because of its suitability for college-level students, its brevity, and its psychometric heritage within the Watson and Glaser tradition of critical thinking assessment.

Test measures. In addition to scores on the two critical thinking instruments, scores were determined for two types of test measures in the course: brief essay quizzes and unit multiple-choice exams. Near the end of each of five units in the course, students were presented two factual questions based strictly on the reading materials. Students chose one of the two questions to answer, with each question requiring an answer of no more than a paragraph. Students were given up to five minutes to formulate and submit their answers. Each question required recall of specific information from the reading materials. Graduate teaching assistants rated the answers on a 0 to 10 scale, with 0 = *no answer or totally inaccurate answer* and 10 = *complete and accurate answer*. Inter-rater reliability for past scoring of the quizzes has typically been at least 0.90 (Williams & Worth, 2002). Scores on the five unit quizzes were combined to provide a total quiz score, which constituted about 6% of the total course credit.

At the conclusion of each of the five course units within the semester-long course, students took a 50-item multiple-choice exam that addressed most major issues in the unit. Close to two-thirds of the items on the five exams emphasized logical reasoning regarding course information, with many of the remaining items requiring a combination of specific recall and logical reasoning (Wallace & Williams, 2003). Combined scores on the unit exams constituted about 50% of the total course credit.

Treatment Condition

Based on Bangert-Drowns and Bankert's (1990) recommendations regarding explicit instruction in critical thinking, we incorporated critical thinking practice into an existing course activity. Specifically, 25 practice questions per unit, similar in nature and difficulty to the items on the unit exam, were posted at the course website at the beginning of the course. Students printed the practice questions from the website and answered each set before attending the class session in which the questions were discussed. For the treatment sections, two to five of the practice questions per unit integrated notions from the *Psychological Critical Thinking* test with concepts in that unit. The 125 practice questions across units in the

treatment sections included 21 questions targeting critical thinking issues. In the control sections, the critical thinking practice questions were replaced with companion questions that addressed the same course concepts as the critical thinking questions but with no reference to critical thinking concepts. (See Appendix for samples of companion questions across the treatment and control groups.)

On the day before the official unit exam, students turned in a scan form with their answers to the practice items and kept their marked copy of the practice exam. The answer sheets were immediately scanned to identify the items missed by a substantial percentage of students (usually defined as 25% or more). The discussion leader then targeted the “most missed” items, including the critical thinking questions in the treatment sections, and invited students to share their answers to questions and explain how they arrived at these answers. The instructor underscored the reasoning involved in answering each question, especially the critical thinking questions for the treatment sections. In contrast to the practice exams in the treatment sections, the official unit exams for all sections had no items that specifically incorporated Lawson’s (1999) psychological critical thinking concepts.

Results

This section of the article presents the findings in three major areas: relationships between critical thinking measures and test performance, changes in critical thinking measures as a result of practice-exam questions, and patterns of change in critical thinking for students who did well or poorly on the multiple-choice exams in the course. Data analyses involved correlations, stepwise regression, and repeated measures designs.

Relationship between Critical Thinking and Test Performance

Table 1 indicates that all pretest and posttest measures of critical thinking significantly correlated with multiple-choice exam performance but correlated only minimally with quiz performance. Pretest measures of psychological critical thinking and generic critical thinking

(WGCTA) correlated comparably with unit exam performance. However, the posttest psychological measure correlated significantly higher ($t = 2.6$, $df = 176$, $p < .05$) with exam performance (0.62) than did the posttest generic measure (0.44). Neither critical thinking pretest measure correlated significantly with quiz performance, but both critical thinking posttest measures correlated significantly ($p < .01$) with quiz performance. In addition to the correlational information provided in Table 1, each of the critical thinking measures yielded essentially the same correlation between its pretest and posttest scores (0.61 for subject-specific and 0.60 for generic). The two measures of critical thinking also correlated similarly with each other at the pretest (0.41) and posttest level (0.49).

Table 1: Correlations Between Critical Thinking and Test Performance

Test performance	Critical thinking measures			
	Psych pretest ^a	Psych posttest ^b	Generic pretest ^c	Generic posttest ^d
Multiple-choice exams	.49*	.62*	.43*	.44*
Essay quizzes	.13	.33*	.07	.21*

^aPsych pretest = Psychological critical thinking pretest. ^bPsych posttest = Psychological critical thinking posttest. ^cGeneric pretest = Generic critical thinking pretest. ^dGeneric posttest = Generic critical thinking posttest.

* $p < .01$.

In addition to establishing correlational relationships between critical thinking and test measures, a stepwise regression analysis was done to determine the extent to which the two precourse measures of critical thinking predicted exam and quiz performance. Because a majority of the exam items required inferential reasoning and the quiz questions required only recall of specific content, we expected the precourse critical thinking measures to predict exam performance better than quiz performance. A stepwise regression analysis showed that neither of the precourse critical thinking instruments significantly predicted quiz performance, but both precourse measures accounted for a significant portion ($p < .001$) of

the variance in exam performance. Of the two precourse critical thinking instruments, psychological critical thinking better predicted exam performance (accounting for 23% of the variance in exam scores). The combination of precourse psychological critical thinking and precourse generic critical thinking accounted for 29% of the variance in exam performance.

Changes in Critical Thinking in Treatment and Control Groups

Using the pretest and posttest measures of critical thinking as the repeated measure and the treatment versus control group as the between variable, a multivariate mixed-design analysis for psychological critical thinking yielded a significant interaction, where $F(1, 170) = 3.96, p < .05$, between the two independent variables. A Bonferroni extension of pairwise comparisons (using the “compare option” under the Statistical Package for the Social Sciences (SPSS) test of estimated marginal means) revealed that the only significant simple-effect difference ($p < .001$) was between the pretest and posttest critical thinking measures in the treatment condition (Table 2). The posttest mean for the treatment group was 2.58 points higher than the pretest mean, whereas the posttest mean for the control group was only 0.82 points higher than the pretest mean. A similar multivariate mixed-design analysis for generic critical thinking produced no significant interaction or main effects (Table 2). The means for generic critical thinking were similar for both treatment and control groups and for both pretest and posttest levels.

Critical Thinking Changes for High and Low Exam Performers

The patterns of change in critical thinking also were examined for students who scored high (made an A) on the combined unit tests versus those who scored low (made a D or F). A series of multivariate mixed-design analyses subsuming comparisons for each of two between variables (high versus low performance, treatment versus control) and one repeated measure (pretest versus posttest) were done separately for psychological and generic critical thinking. It should be noted that the treatment-control comparisons in this

Table 2: Critical Thinking Means by Cells for Repeated Measures Analysis of Treatment versus Control Group

Group designation	Psychological critical thinking ^a	
	Pretest ct	Posttest ct
Treatment ($n = 105$)	19.56 (6.13) ^b	22.14 (6.81)
Control ($n = 67$)	20.40 (6.46)	21.22 (6.57)
	Generic critical thinking ^c	
Treatment ($n = 115$)	26.37 (5.88)	26.06 (5.88)
Control ($n = 69$)	25.13 (5.66)	25.80 (5.23)

^aA significant interaction effect ($p < .05$) was obtained for psychological critical thinking, with the posttest score for the treatment group significantly ($p < .001$) higher than the pretest score. ^bNumbers in parentheses after pretest and posttest means are standard deviations. ^cNo significant interaction or main effects were obtained for generic critical thinking.

analysis included only the high and low performers for those groups, not for the total treatment and control groups (as described in the previous section). The current analyses began with the treatment versus control dimension included in a three-way mixed design, but subsequent comparisons of high versus low performers were done separately in the treatment and control groups (Table 3).

The analyses consistently yielded significant performance-group main effects but no treatment versus control group main effects or interaction effects. Across both treatment and control groups, both critical thinking measures, and both pretest and posttest levels, the high exam performers consistently scored higher than the low performers on critical thinking. Specifically, the multivariate mixed-design analyses yielded the following significant performance-group main effects for psychological critical thinking: performance-group difference for combined treatment and control groups, where $F(1, 15) = 36.87, p < .001$; performance-group difference for treatment group, where $F(1, 10) = 23.19, p < .001$; and performance-group difference for the control group, where $F(1, 5) = 26.57, p < .005$. The analyses for generic critical thinking yielded similar

Table 3: Critical Thinking Means by Cells for Repeated Measures Analysis of High and Low Exam Performers within Treatment versus Control Group

Performance group	Psychological critical thinking ^a			
	Treatment		Control	
	Pretest	Posttest	Pretest	Posttest
High performers	25.67 (4.80)	31.83 (7.33)	24.80 (4.02)	30.20 (2.59)
Low performers	14.17 (4.40)	16.00 (4.34)	10.50 (2.12)	17.50 (7.78)
Performance group	Generic critical thinking ^b			
	Pretest	Posttest	Pretest	Posttest
	Pretest	Posttest	Pretest	Posttest
High performers	36.14 (2.79)	36.00 (3.00)	34.00 (1.22)	33.00 (3.81)
Low performers	21.63 (3.89)	22.25 (4.06)	19.50 (2.12)	21.50 (2.12)

Note. Because the repeated measures analysis required that students take both the pretest and posttest for the target critical thinking instrument, the *ns* were relatively small for all cells in this table (ranging from a low of 2 to a high of 8 per cell). In most cases, the *n* for the high-performing group exceeded that of the low-performing group.

^aSignificant within groups (pre-to-post) and between groups (high versus low performers) main effects (ranging from $p < .05$ to $p < .001$) were obtained for psychological critical thinking in both treatment and control groups, but no significant interaction was obtained within either of these groups. ^bSignificant performance groups main effects ($p < .001$) were obtained for generic critical thinking in both treatment and control groups, but no significant interaction or within group main effects were obtained in either the treatment or control group. ^cNumbers in parentheses following means represent standard deviations.

significant performance-group main effects: performance-group difference for combined treatment and control groups, where $F(1, 16) = 90.07$, $p < .001$; performance-group difference for the treatment group, where $F(1, 13) = 83.99$, $p < .001$; and performance-group difference for the control group, where $F(1, 5) = 45.55$, $p < .001$.

In addition to the between-groups main effects, significant pre-to-post main effects were found for psychological critical thinking. Both high and low exam performers scored significantly higher on psychological critical thinking at the posttest than the

pretest level for the combined treatment and control groups, where $F(1, 15) = 20.08, p < .001$, as well as in the separate treatment, where $F(1, 10) = 10.45, p < .01$, and the control groups, where $F(1, 5) = 9.37, p < .05$. In contrast, the repeated measures analyses for generic critical thinking yielded no significant pre-to-post effects.

Discussion

This section explores the implications of the current study with respect to the predictive and outcome potential of subject-specific and generic critical thinking measures. In general, psychological critical thinking appeared to have more promise as a predictor of exam performance than did generic critical thinking. In addition, psychological critical thinking proved more amenable to change than did generic critical thinking. A treatment condition that infused psychological critical thinking concepts in a regular course activity produced significant gains in subject-specific critical thinking but no significant gains in generic critical thinking.

Predictive Potential of Critical Thinking

Our findings are consistent with prior research (Williams, Oliver, Allin, Winn, & Booher, in press, 2003; Williams & Worth, 2002) showing that precourse tests of critical thinking can significantly predict performance on academic tasks involving the use of critical thinking strategies. Precourse critical thinking better predicted performance on multiple-choice exams requiring considerable critical reasoning than on essay quizzes requiring only recall of information. Also, subject-specific critical thinking better predicted exam performance than did generic critical thinking. Furthermore, the correlation between subject-specific critical thinking and exam performance increased from the beginning to the end of the target course, whereas the correlation between generic critical thinking and exam performance remained virtually unchanged from the beginning to the end of the course.

Precourse assessment of critical thinking could help instructors identify both students who might need special assistance in course

activities requiring critical thinking and other students who could excel in the same activities. For example, in attempting to assist students with low critical thinking skills, we currently are exploring a peer coaching program in which students high in both exam performance and critical thinking assist students low in both exam performance and critical thinking with their multiple-choice exam performance. The high-performing students meet with the low-performing students on a one-on-one basis to review previously taken exams item by item, with the latter students explaining their reasoning for choosing answers for missed items. The high-performing coaches then pinpoint the informational/reasoning fallacies that adversely affected the low-performing students' choice for each item.

Despite some reason for optimism, helping low-performing students with both their critical thinking and exam performances represents a formidable challenge. Low-performing students in the current study, on the average, scored at the 3rd percentile according to college-level norms for generic critical thinking at both the beginning and the end of the course. Students with such low levels of critical thinking could have extreme difficulty on tasks requiring advanced critical thinking. Although all students in the low-performance group made Ds or Fs on the combined unit exams, which accounted for about 50% of their course grade, only 20% of them made as high as a C in the course (all others made Ds or Fs). In working with students who score low on critical thinking and on initial exams in our courses, we have been more successful in helping them improve their knowledge of course content than their reasoning regarding that content (Williams et al., in press).

Outcome Potential of Critical Thinking

Although critical thinking may be a relatively stable cognitive ability, some research has shown that the cumulative effect of a college education may be to upgrade critical thinking (McMillan, 1987; Terenzini, 1993). Two principal patterns have emerged regarding the improvement of critical thinking in college courses: gains tend to be low to moderate within specific courses; and are achieved primarily

when specific critical thinking skills are directly taught. Thus, the prospect of improving subject-specific critical thinking appears much better than the prospect of improving generic critical thinking, unless a course is designed specifically to teach generic critical thinking strategies (Dansereau et al., 1979; Facione, 1990; Halpern, 1993).

Gains in psychological critical thinking in the current study were achieved in a cost-effective way by infusing the study of critical thinking into an existing course activity. We estimate that students spent a maximum of one hour outside of class selecting their responses to the critical thinking practice questions and no more than one hour in class discussing their responses to these questions. The critical thinking questions fit comfortably into the fabric of an existing course activity, making it less likely that a specialized treatment condition would dominate course time. This arrangement contrasts with past studies that have achieved critical thinking gains by building the entire course around a particular critical thinking model (Allegretti & Frederick, 1995; Isaacs, 1991; Reed & Kromrey, 2001).

A discouraging finding of several studies that have attempted to develop critical thinking skills is that the students who enter with low critical thinking skills are the least likely to improve those skills (Williams et al., in press, 2003). In fact, those who enter with low critical thinking sometimes get worse in their critical thinking by the end of the course (Williams et al., in press). We have observed that students low in critical thinking appear to find critical thinking activities somewhat disconcerting, often characterizing critical thinking demands as tricky and even unfair. Other researchers (Halpern, 1998; Keeley, Shemberg, Cowell, & Zinnbauer, 1995) also have commented on some students' resistance to critical thinking activities. Because critical thinking can be hard work, students with minimal critical thinking skills probably have to expend great effort and overcome considerable frustration in course activities involving critical thinking. Halpern (1998) states that "learners need to understand and be prepared for the effortful nature of critical thinking so that they do not abandon the process too soon, believing that the thinking should have been easier and accomplished more quickly" (p. 452).

Although high exam performers in the current study scored higher on critical thinking than the low exam performers, the latter group generally made raw-score gains on psychological critical thinking comparable to those of the high exam performers. Our past research has confirmed that some low critical thinkers can improve even their generic critical thinking skills and perform at a high level in the course targeted in this study (Williams & Stockdale, 2003). For several semesters, we have monitored cognitive and study-habits differences between low critical thinkers who do well in the course and those who do poorly. These groups differ more in their study habits than in their initial cognitive skills. The differences mainly relate to completeness and accuracy of notetaking, performance on practice exams similar to those used in the current study, and improvement of generic critical thinking skills. Students who enter the course with low critical thinking skills seldom conclude the course as outstanding critical thinkers, but some low critical thinkers significantly improve their critical thinking skills, apply superior study habits, and perform at a high level in the course.

Limitations of the Study

Even though psychological critical thinking showed modest potential as a predictor and outcome variable in the target course, further avenues for strengthening both psychometric possibilities should be explored. Although significantly predicting exam performance scores, psychological critical thinking explained only 23% of the variance in exam scores. The treatment condition was effective in significantly increasing psychological critical thinking, but the gain in psychological critical thinking in the treatment group compared to gain in the control group yielded an effect size of only 0.28, which is near the low end of the practically useful range. Because only 17% of the practice-exam questions included psychological critical thinking issues, a more liberal inclusion of these issues in practice-exam items might have produced a greater treatment effect. Nonetheless, if most courses could produce the level of subject-specific gain in critical thinking achieved with the current treatment condition, a sizeable cumulative gain in subject-specific critical thinking would accrue across courses.

Gains in psychological critical thinking under the treatment condition were not paralleled by gains in generic critical thinking. The treatment group did no better than the control group and neither group showed improvement in generic critical thinking from the beginning to the end of the course. The treatment approach used in the current study likely would have to be geared specifically to the dimensions of generic critical thinking for researchers to expect gains on this variable. Dimensions subsumed in the generic critical thinking test used in the current study (i.e., inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments) probably would need to be meticulously highlighted in course activities to achieve a pre-to-post difference in generic critical thinking. Though achievable, the infusion of these specific dimensions into the regular content of a course would be far more labor intensive than the inclusion of the relatively broad psychological critical thinking notions in the practice-exam questions.

Attempting to identify the exact source of the treatment effect for improvement in psychological critical thinking raises questions regarding specific cause-effect relationships. Was the treatment effect produced simply by the availability of practice-exam questions requiring psychological critical thinking strategies or by the discussion of these items in class or by a combination of the two? Our speculation is that discussion of these items was fundamental to their impact on psychological critical thinking. That hunch is based on the observation that many students had difficulty reasoning their way through these items before they were analyzed and explained in class. Because the psychological critical thinking issues often were encountered for the first time in the practice-exam questions, many students would have had difficulty appropriating those issues apart from their analysis in class. Therefore, the treatment approach used in this study might be considered a blend of an embedded and a direct instruction approach. The treatment began with critical thinking notions embedded in selected practice-exam questions but culminated with direct explanation of those strategies by the instructor or students advanced in critical thinking.

Although this study mainly compared the treatment and control groups with respect to changes in psychological and generic critical thinking, additional analyses examined the linkage between exam

performance and critical thinking skills in both the treatment and control groups. In general, high exam performers did better than low exam performers on both critical thinking measures at both the pre- and postcourse levels. However, because of the exceedingly small *ns* in the treatment by performance by time cells, significant critical thinking differences between the performance groups should be interpreted with caution. Even though low exam performers gained in psychological critical thinking to a comparable degree to that of high exam performers, several of our previous studies (Williams et al., in press, 2003) have shown that high exam performers consistently make greater gains in critical thinking than do low exam performers.

Some might question whether additional analyses could have extended the findings of the study. For example, no critical thinking comparisons were done by gender and academic classification. However, past research reviewed by Williams and Worth (2001) has yielded mixed results for the linkage between such demographic variables and critical thinking. Because dividing the current sample by gender and academic classification would have appreciably reduced the cell *ns* and possibly obscured the major comparisons of the study, we focused on the treatment- versus control-group comparison without regard to gender and academic classification.

Also missing from the current study is an evaluation of inter-rater reliability for quiz scoring. The decision not to include this analysis in the current study was based on three considerations: (a) we had found inter-rater reliability for quiz scoring to be generally high in our past research (Williams & Worth, 2002); (b) the process of establishing inter-rater reliability for the rating of written products typically is very labor intensive; and (c) quiz performance was secondary to exam performance as a criterion for assessing the predictive potential of the critical thinking instruments. Critical thinking was expected to predict exam performance but to be minimally related to quiz scores.

Questions also could be raised about our using the same pre- and postmeasures of critical thinking. No alternate forms of either critical thinking instrument were available. Although the original WGCTA has alternate forms, Form S of the WGCTA does not. In addition, the approximate four-month spacing between pretest and posttest was

judged adequate to minimize any learning or practice effect from the first testing. A kindred concern relates to subtest scoring for the WGCTA, which would have added greater specificity to our results. However, the developers of the WGCTA Form S (Watson & Glaser, 1994) do not recommend scoring this instrument by subtest because the individual subtests lack adequate reliability.

The limitations of the current study point to areas of needed research regarding the critical thinking patterns in the current study. As is the case with most research studies, no final answers were generated regarding the predictive and outcome potential of subject-specific versus generic critical thinking measures. At this point, subject-specific measures appear to have greater promise as both predictor and outcome variables. Nonetheless, finding ways to strengthen the predictive and outcome potential of generic critical thinking in subject-matter courses could increase the generalization of critical thinking skills across subject-areas.

References

- Allegretti, C.L., & Frederick, J.N. (1995). A model for thinking critically about ethical issues. *Teaching of Psychology*, 22, 46–48.
- Arburn, T.M. (1998). Assisting at-risk community college students' acquisition of critical thinking strategies in human anatomy and physiology. *Dissertation Abstracts International*, 59(6-A), 1969.
- Bangert-Drowns, R.L., & Bankert, E. (1990). *Meta-analysis of effects of explicit instruction for critical thinking*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA, April.
- Bensley, D.A., & Haynes, C. (1995). The acquisition of general purpose strategic knowledge for argumentation. *Teaching of Psychology*, 22, 41–45.
- Dansereau, D.F., Collins, K.W., McDonald, B.A., Holley, C.D., Garland, J., Diekoff, G., & Evans, S.H. (1979). Development and evaluation of a learning strategy training program. *Journal of Educational Psychology*, 71, 64–73.
- Facione, P.A. (1990). *The California Critical Thinking Skills Test—college level. Technical report #1. Experimental validation and content validity*. Millbrae, CA: California Academic Press. (ERIC Document ED327549).
- Forbes, C.A. (1997). Analyzing the growth of the critical thinking skills of college calculus students. *Dissertation Abstracts International*, 58(6-A), 2101.
- Gadzella, B.M., Ginther, D.W., & Bryant, G.W. (1997). Prediction of performance in an academic course by scores on measures of learning style and critical thinking. *Psychological Reports*, 81, 595–602.
- Garside, C. (1996). Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills. *Communication Education*, 45, 212–227.

- Halpern, D.F. (1988). Assessing student outcomes for psychological majors. *Teaching of Psychology*, 15, 181–186.
- Halpern, D.F. (1993). Assessing the effectiveness of critical-thinking instruction. *Journal of General Education*, 42, 239–254.
- Halpern, D.F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53, 449–455.
- Isaacs, L.G. (1991). A study of the effect of teaching critical thinking in an introductory nursing course. *Dissertation Abstracts International*, 51(10-A), 3317.
- Jones, E.A. (Ed.). (1995). *The national assessment of college student learning: Identifying college graduates, essential skills in writing, speech and listening, and critical thinking*. Washington, DC: National Center for Education Statistics.
- Keeley, S.M., Shemberg, K.M., Cowell, B.S., & Zinnbauer, B.J. (1995). Coping with student resistance to critical thinking: What the psychotherapy literature can tell us. *College Teaching*, 43, 140–145.
- Lawson, T.J. (1999). Assessing psychological critical thinking as a learning outcome for psychology majors. *Teaching of Psychology*, 26, 207–209.
- Lierman, J.A. (1997). *Effects of instructional methods upon the development of critical thinking skills in baccalaureate nursing students*. Unpublished doctoral dissertation, University of Missouri-Kansas City.
- Lyle, E. (1958). An exploration in the teaching of critical thinking in general psychology. *Journal of Educational Research*, 52, 129–133.
- McCammon, S., Golden, J., & Wuensch, K. (1988). Predicting course performance in freshman and sophomore physics courses: Women are more predictable than men. *Journal of Research in Science Teaching*, 25, 501–510.
- McMillan, J.H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, 26, 3–27.
- McMurray, M.A., Beisenherz, P., & Thompson, B. (1991). Reliability and concurrent validity of a measure of critical thinking skills in biology. *Journal of Research in Science Teaching*, 28, 183–192.
- Reed, D.H., & Kromrey, J.D. (2001). Teaching critical thinking in a community college history course: Empirical evidence from infusing Paul's model. *College Student Journal*, 35, 201–215.
- Resnick, D.P., & Peterson, N.L. (1991). *Evaluating progress toward goal five: A report to the National Center for Education Statistics*. Washington, DC: National Center for Education Statistics. (ERIC Document ED340764).
- Royalty, J. (1995). Evaluating knowledge-based statistical reasoning. *Psychological Reports*, 77, 1323–1327.
- Sandor, M.K., Clark, M., Campbell, D., Rains, A.P., & Cascio, R. (1998). Evaluating critical thinking skills in a scenario-based community health course. *Journal of Community Health Nursing*, 15(1), 21–29.
- Slaughter, D.S., Brown, D.S., Gardner, D.L., & Perritt, L.J. (1989). Improving physical therapy students' clinical problem-solving skills: An analytical questioning model. *Physical Therapy*, 69, 441–447.
- Terenzini, P.T. (1993). *Influences affecting the development of students' critical thinking skills*. Paper presented at the Association for Institutional Research, New Orleans, LA, May.
- Tsui, L. (1998). *A review of research on critical thinking*. Paper presented at the annual meeting of the Association for the Study of Higher Education, Miami, FL, November.
- Wallace, M., & Williams, R.L. (2003). Multiple-choice exams: Explanations for student choices. *Teaching of Psychology*, 30, 136–138.

- Watson, G.B., & Glaser, E.M. (1980). *Watson-Glaser Critical Thinking Appraisal manual*. Cleveland, OH: The Psychological Corporation.
- Watson, G.B., & Glaser, E.M. (1994). *Watson-Glaser Critical Thinking Appraisal manual—Form S*. San Antonio, TX: The Psychological Corporation.
- Williams, R.L., Oliver, R., Allin, J.L., Winn, B., & Booher, C.S. (2003). Psychological critical thinking as a course predictor and outcome variable. *Teaching of Psychology, 30*, 220–223.
- Williams, R.L., Oliver, R., Allin, J.L., Winn, B., & Booher, C.S. (In press). "Knowledge and critical thinking as course predictors and course outcomes." *Inquiry: Critical Thinking Across the Disciplines*.
- Williams, R.L., & Stockdale, S.L. (2003). High-performing students with low critical thinking skills. *Journal of General Education, 52*(3), 199–225.
- Williams, R.L., & Worth, S.L. (2001). The relationship of critical thinking to success in college. *Inquiry: Critical Thinking Across the Disciplines, 21*, 5–16.
- Williams, R.L., & Worth, S.L. (2002). Thinking skills and work habits: Contributors to course performance. *Journal of General Education, 51*, 200–227.

Appendix: Companion Practice-Exam Items for Treatment and Control Groups

Treatment item (Unit B: Cognitive Development)

In attempting to evaluate the efficacy of a child-centered preschool experience, researchers followed the development of a group of students who had attended a Piagetian kindergarten. These students were assessed every four years until age 20 on a variety of life-adjustment measures. In the main, the students obtained favorable scores on all the adjustment measures through their 20th birthday. The researchers could reasonably conclude from their findings that

- a. the child-centered approach is superior to most other approaches in promoting life adjustment.
- b. children just naturally obtain higher adjustment scores as they get older.
- c. participation in the Piagetian kindergarten was associated with favorable scores on life adjustment measures.
- d. they had established important causal relationships between an early Piagetian experience and later adjustment.

Control item (Unit B: Cognitive Development)

Research on child-centered preschool programs most strongly points to which of the following conclusions?

- a. Children in these programs progress through Piaget's developmental stages at an accelerated rate.
- b. The child-centered approach is more efficacious in promoting mastery of phonetical skills than are more teacher-centered approaches.
- c. Children in these programs may benefit with respect to long-term emotional and social adjustment.
- d. When given a choice, most children prefer a teacher-centered over a child-centered program.

Treatment item (Unit B: Cognitive Development)

A medical researcher who believes quite strongly in the efficacy of Ritalin for ADHD children conducted a survey of parental reactions to the use of this drug. A team of research assistants called the identified parents and asked, "Aren't you encouraged by how much calmer your children are when taking Ritalin?" Most parents answered this question affirmatively. One could conclude from this finding that

- a. the effects of Ritalin on ADHD children significantly boosts the morale of their parents.
- b. the question clearly invited parents to express their true feelings about the effectiveness of Ritalin.
- c. the question elicited the kind of direct evidence needed in evaluating the efficacy of medication in treating ADHD.
- d. none of the above would necessarily follow from the finding.

Control item (Unit C: Psychological Development)

A medical specialist on the treatment of ADHD would most likely make which of the following claims regarding the use of Ritalin?

- a. Ritalin is being used less frequently with children diagnosed with ADHD.
- b. A small dosage of Ritalin is more likely to calm the child than to help the child focus.
- c. Practically all children with ADHD respond well to Ritalin.
- d. Most children with ADHD are helped by Ritalin.