This study first determined how well two types of critical thinking measures, generic and subject-specific, predicted performance on course tests. Secondly, the study examined the extent to which critical thinking changed from the beginning to the end of the course. Two generic and one subject-specific measure of critical thinking were used in the study. A total of 428 college students participated in various phases of the study across 3 semesters. All the critical thinking measures better predicted performance on multiple-choice examinations requiring critical thinking than on essay quizzes requiring only recall of course information. The Psychological Critical Thinking test (subject specific; T. Lawson, 1999) and the Watson-Glaser Critical Thinking Appraisal–Form S (generic; G. Watson and E. Glaser, 1980) were the best predictors of examination scores. Students also improved their scores on these two critical thinking measures significantly from the beginning to the end of the course. The pattern of change on critical thinking was somewhat different for high and low performers on examinations. (Contains 4 tables and 28 references.) (Author/SLD)
Critical Thinking as a Predictor and Outcome Measure in a Large Undergraduate Educational Psychology Course

Robert L. Williams
The University of Tennessee
Abstract

The study first determined how well two types of critical thinking measures, generic and subject-specific, predicted performance on course tests. Secondly, the study examined the extent to which critical thinking changed from the beginning to the end of the course. Two generic and one subject-specific measure of critical thinking were employed in the study. All the critical thinking measures better predicted performance on multiple-choice exams requiring critical thinking than on essay quizzes requiring only recall of course information. The Psychological Critical Thinking test (subject-specific) and the Watson-Glaser Critical Thinking Appraisal-Form S (generic) were the best predictors of exam scores. Students also significantly improved their scores on these two critical thinking measures from the beginning to the end of the course. The pattern of change on critical thinking was somewhat different for high and low performers on the exams.
Critical Thinking as a Predictor and Outcome Measure in a Large Undergraduate Educational Psychology Course

Objectives of the Research

The purposes of the research were twofold: (a) determine how well generic and psychological critical thinking instruments, given both at the beginning and end of a large undergraduate educational psychology course, predicted performance on course tests (some of which required only direct recall of information and others that involved extensive use of critical thinking) and (b) assess the extent to which the critical thinking scores change from the beginning to the end of the course, especially for students who obtained either high or low grades on course examinations designed to require critical thinking.

Perspectives

Few concepts have attracted more attention in higher education than the notion of critical thinking. Although a variety of definitions have 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 would appear to have a figural role in college courses. On the one hand, critical thinking may serve as a predictor of performance on a variety of performance measures and, on the other hand, it may represent an important outcome measure in college courses. However, its potential as a performance predictor has seldom been directly contrasted with its potential as an outcome measure in specific college courses.

Predictive Potential

The predictive capacity of critical thinking likely differs both by the type of critical thinking measure used and the type of performance measure predicted. With respect to the former issue, critical thinking tests may be classified as either generic or subject-specific.
Popular generic measures of critical thinking include the *Watson-Glaser Critical Thinking Appraisal* (Watson & Glaser, 1980, 1994) and the *California Critical Thinking Skills Test* (Facione & Facione, 1994). Subject-specific measures of critical thinking have been developed for psychology (Lawson, 1999), statistics (Royalty, 1995), and biology (McMurray, Beisenherz, & Thompson, 1991). One might expect a subject-specific measure of critical thinking to be a better predictor of course performance than would a generic measure. However, the critical thinking literature reveals few studies that have directly compared the predictive potential of generic versus subject-specific critical thinking measures. In one of few studies to address this comparison, Royalty (1995) reported that his test of statistical reasoning better predicted end-of-the-course statistical critical thinking than did a generic test of critical thinking.

Several studies have indicated that critical thinking is a moderate predictor of course success (Gadzella, Ginther, & Bryant, 1997; McCammon, Golden, & Wuensch, 1988; Wilson & Wagner, 1981). However, critical thinking may be more strongly related to some performance measures than to others. Presumably, if a performance task requires a high level of critical thinking, then critical thinking measures should strongly predict performance on 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 critical thinking to be a stronger predictor of performance on multiple-choice tests requiring inferential reasoning than was either attendance or student notetaking.

**Outcome Potential**

A variety of researchers and 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, in press) have produced critical thinking gains in academic courses, but others (Arburn, 1998; Forbes, 1997; Lierman, 1997; Lyle, 1958; Slaughter, Brown, Gardner, & Perritt, 1989) have failed to do so.

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 students. For example, one might expect a subject-specific measure of critical thinking to be more changeable than a generic measure, given the direct linkage between issues embedded in the subject-specific measure and the content of the particular course. Thus, subject-specific critical thinking could readily be targeted in the context of tasks embedded in a subject-matter course.

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

Another important issue regarding changes in critical thinking in selected courses is the
interaction between instructional methodology and student characteristics. A particular instructional model might be very effective in facilitating critical thinking for one kind of student but not for another. For example, Lyle (1958) reported that students with high academic aptitude improved their critical thinking more under a problem-based format, whereas low-aptitude students improved more under a lecture format. Of particular interest in the current study is the student’s academic performance on outcome measures assumed to require critical thinking. For example, if the multiple-choice tests in a course emphasize critical thinking, will the pattern of change in critical thinking differ for students who do well as opposed to those who do poorly on the exams?

**Methods**

Over a period of three semesters, students in large sections (minimum of 50 students per section) of an undergraduate educational psychology course were given one of three critical thinking instruments at the beginning and end of the semester. Students received a small amount of course credit for taking the instruments, but equivalent credit was available through non-research activities. In the first semester of data collection, students took *The California Critical Thinking Skills Test-Forms A and B* (generic measures of critical thinking); in the second semester, they took the *Psychological Critical Thinking* instrument (a domain-specific measure of critical thinking); and in the third semester, they took the *Watson-Glaser Critical Thinking Appraisal-Form S* (a generic measure of critical thinking). All three semesters, students also took a variety of course performance measures, including brief essay quizzes and comprehensive multiple-choice exams. A total of 428 students participated in various phases of the study across the three semesters.

The primary instructional component targeting critical thinking was the inclusion of
higher-order questions in class discussion. All sections of the course used the same set of instructor notes (prepared by the supervising instructor) that included higher order questions strategically placed for class discussion. These questions required synthesis and application of course information in reaching conclusions regarding viewpoints on various course issues. (See Table 1 for a sample of these questions.)

Data Sources

Critical Thinking Instruments

The California Critical Thinking Skills Test (Facione & Facione, 1994) has 34 items in a multiple-choice format, with related assumptions/information provided on which to base answers to the questions. Respondents were instructed to choose options most consistent with the information provided in the test. Scores could range from 0 to 34. The mean and standard deviation for the local sample approximated companion metrics reported for the normative sample. Internal consistency was reported in the test manual to be .70, and test scores were reported to be moderately correlated with a variety of cognitive instruments (e.g., SAT-Verbal, SAT-Math, and Nelson-Denny Reading Test).

The second critical thinking instrument, Psychological Critical Thinking (Lawson, 1999), uses an essay format that consists of 14 scenarios describing various psychological claims. 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. In responding to each scenario, students indicate whether the claim is consistent with the information presented and then identifies any fallacies in the claim. Using a qualitative scoring procedure, graduate teaching assistants (GTAs) 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. Scores could range from 0 to 42. Inter-rater reliability for pairs of raters proved to be .91 for the pretest and .92 for the posttest (Williams et al., in press).

The third measure of critical thinking (Watson-Glaser Critical Thinking Appraisal—WGCTA) is probably the most widely used 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 contains 40 multiple-choice items, with the item options ranging from 2 to 5. All the information needed to respond to the items is provided in the test itself. Respondents are asked to judge the credibility of potential conclusions linked to the presented information. Scores on the WGCTA-Form S can range from 0 to 40. The test manual reports both the internal consistency and the test-retest reliability for Form S to be .81. The instrument also is reported to be moderately predictive of academic and professional indices of success.

Performance Measures

In addition to scores on the three critical thinking instruments, scores were determined for two performance measures in the course: brief essay quizzes and comprehensive 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 no more than a paragraph to respond. Students were given up to five minutes to formulate and submit their answers. Each question required recall of specific information in the reading materials. Pairs of GTAs rated the answers on a 0 to 5 scale, with 0 = no answer or totally inaccurate answer and 5 = complete and accurate answer. Inter-
rater reliability for past scoring of the quizzes has typically been at least .90 (Williams & Worth, 2002). Scores on the five unit quizzes were combined to constitute a total quiz score, which constituted about 6% of the total course credit.

At the conclusion of each of five units in the course, students took a 40-item multiple-choice exam that addressed most major issues in the unit. Then at the end of the course, students took a 75-item multiple-choice exam that sampled issues throughout the course. Close to two-thirds of the items emphasized logical reasoning regarding course information, with many of the remaining items requiring a combination of specific recall and logical reasoning (Wallace & Williams, 2003). Scores on the unit exams and the final exam were combined to constitute a total exam score, which constituted about 70% of the total course credit.

Results

Predictive Potential of Critical Thinking

Analyses indicated that the critical thinking measures were minimally related to performance on essay quizzes but significantly ($p < .01$) and moderately related to performance on exams (see Table 2). The strongest overall predictor of test performance proved to be the WGCTA. Also, the post-course critical thinking scores tended to be more strongly related to quiz and exam performance than were the pre-course scores. Thus, critical-thinking measures became more strongly linked to the performance measures from the beginning to the end of the course. This pattern was generally consistent across all three critical thinking measures. A series of stepwise regression analyses assessing how well various critical thinking measures predicted exam and quiz performance showed that both psychological and WGCTA posttests were moderately good predictors of exam performance, but none of the critical thinking measures contributed substantially to the prediction of the quiz scores (see Table 3).
Outcome Status of Critical Thinking

Paired samples t-tests indicated that students significantly increased their critical thinking scores both for the Psychological Critical Thinking instrument and for the WGCTA-Form S instrument. However, raw score pre-to-post differences were more apparent for the former than for the latter instrument (see Table 4). For the California Critical Thinking Skills Test, student scores actually declined slightly (but significantly) from the beginning to the end of the course. When subgroups of high-grade (those making As on the multiple-choice exams) and low-grade students (those making Ds and Fs) were contrasted for changes in critical thinking, the high-grade students showed more favorable changes across all three critical thinking instruments (see Table 4). This was especially the case for the Psychological Critical Thinking measure, with the high-grade group gaining significantly \((p < .001)\) from pretest to posttest and the low-grade group staying the same. All three critical thinking measures yielded a significant \((p < .001)\) between-groups main effect, with the high group scoring significantly higher than the low group for both pretest and posttest assessment.

Educational and Scientific Importance

The findings of this study suggest that measures of critical thinking may have more potential as predictors (correlates) of academic performance than as outcome measures in college courses. The study indicated that performance measures can be designed to require or minimize critical thinking and that critical thinking instruments differentially predicted scores on such performance measures. With respect to the outcome potential of critical thinking measures, courses involving frequent use of higher-order thinking questions appear to have the best potential for promoting critical thinking, especially critical thinking specific to a subject area. However, the questioning/interactive approach highlighted in this report is likely to be more
effective in improving the critical thinking skills of students who do well on major performance measures in a course than of those who do poorly on those performance measures.

Addendum

Because the findings of this study appear to link low exam performance to low pre-course critical thinking and minimal improvement in critical thinking during the course, a followup analysis of grades was done for students who scored either in the bottom quartile on the CCTST-A or at the 5th percentile and lower on the WGCTA-S at the beginning of the course. Low critical thinkers who eventually made As or Bs in the course were contrasted with those who made Ds and Fs on a number of credit (exam and non-exam dimensions) and support variables (e.g., attendance, notetaking, and critical thinking improvement).

Effect sizes show that the high-grade low critical thinkers (HG-LCT) consistently outperformed the low-grade low critical thinkers (LG-LCT) on practically all credit and support variables in the course. The former group did significantly better than the latter on both exam and non-exam dimensions. With respect to support variables, the HG-LCT group had better attendance, did much better notetaking, and made greater critical thinking gains than the LG-LCT group. In fact, the HG-LCT group did better on most support variables (including all aspects of notetaking and critical thinking gains) than a comparison group of high-scoring critical thinkers who also made high grades in the course. Clearly, low critical thinking did not doom one to low grades in the target course. By working hard and smartly, some low critical thinkers performed almost as well as high critical thinkers. What is most hopeful about the HG-LCT group is that across subsamples in our database the HG-LCT students consistently improved their critical thinking more than both the low-grade low critical thinkers and the high-grade high critical thinkers.
References


Table 1

* Selected Critical Thinking Questions Interspersed in Class Discussion

* What would be the long-range economic effects of heavily taxing cigarettes?
* Are IQ tests vital to best serving the academic needs of children?
* Is self-concept best interpreted as a cause or an effect of one's performance?
* Why would emotion-focused coping generally be less adaptive than problem-focused?
* Which of the following parenting styles is likely to have the most adverse effects on children: authoritarian, indulgent, or laissez-faire?
* Why has the crime rate for adolescents increased so much more than that for adults?
* What level of gun control would be in the best interest of this society?
* Are the findings on character education more consistent with a behavioristic or humanistic view of humankind?
### Table 2

*Correlations between Pretest/Posttest Critical Thinking Measures and Test Performance*

<table>
<thead>
<tr>
<th>Critical Thinking Instrument</th>
<th>Test Performance</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quizzes</td>
<td>Exams</td>
<td></td>
</tr>
<tr>
<td>California Critical Thinking Skills Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest (n = 147)</td>
<td>.08</td>
<td>.31**</td>
<td></td>
</tr>
<tr>
<td>Posttest (n = 141)</td>
<td>.19*</td>
<td>.38**</td>
<td></td>
</tr>
<tr>
<td>Psychological Critical Thinking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest (n = 129)</td>
<td>.11</td>
<td>.41**</td>
<td></td>
</tr>
<tr>
<td>Posttest (n = 121)</td>
<td>.08</td>
<td>.49**</td>
<td></td>
</tr>
<tr>
<td>Watson-Glaser Critical Thinking Appraisal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest (n = 164)</td>
<td>.20**</td>
<td>.42**</td>
<td></td>
</tr>
<tr>
<td>Posttest (n = 158)</td>
<td>.28**</td>
<td>.57**</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
### Table 3

*Stepwise Regression Analyses for Significant Critical Thinking Predictors of Test Performance*

<table>
<thead>
<tr>
<th>Critical thinking measure</th>
<th>Test variance&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Exam total</th>
<th>Quiz total</th>
</tr>
</thead>
<tbody>
<tr>
<td>California Critical Thinking Skills Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td>.13</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Posttest and pretest</td>
<td>.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological Critical Thinking Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td>.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson-Glaser Critical Thinking Appraisal</td>
<td></td>
<td>.31</td>
<td>.07</td>
</tr>
</tbody>
</table>

<sup>a</sup> Amount of test performance variance explained by each critical thinking measure.
Table 4

Pre- to Post-Changes in Critical Thinking Measures

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>Pre-mean</th>
<th>Post-mean</th>
<th>Significant difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>California Critical Thinking Skills Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>16.62</td>
<td>15.38</td>
<td>.001</td>
</tr>
<tr>
<td>High grade</td>
<td>15</td>
<td>18.87</td>
<td>19.07</td>
<td>NS</td>
</tr>
<tr>
<td>Low grade</td>
<td>24</td>
<td>14.54</td>
<td>12.66</td>
<td>.01</td>
</tr>
<tr>
<td>Psychological Critical Thinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>16.45</td>
<td>18.81</td>
<td>.001</td>
</tr>
<tr>
<td>High grade</td>
<td>14</td>
<td>19.29</td>
<td>25.93</td>
<td>.001</td>
</tr>
<tr>
<td>Low grade</td>
<td>11</td>
<td>15.27</td>
<td>15.45</td>
<td>NS</td>
</tr>
<tr>
<td>Watson-Glaser Critical Thinking Appraisal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>149</td>
<td>26.21</td>
<td>27.61</td>
<td>.005</td>
</tr>
<tr>
<td>High grade</td>
<td>20</td>
<td>31.85</td>
<td>32.20</td>
<td>NS</td>
</tr>
<tr>
<td>Low grade</td>
<td>20</td>
<td>21.20</td>
<td>21.70</td>
<td>NS</td>
</tr>
</tbody>
</table>

Note. A repeated measures analysis for each critical thinking dimension yielded a significant main effect difference between the performance groups. In addition, Psychological Critical Thinking produced a significant interaction effect, resulting in a significant pretest to posttest increment in critical thinking for the high-grade group but not for the low-grade group.
Title: Critical Thinking as a Predictor and Outcome Measure in a Large Undergraduate Educational Psychology Course

Author(s): Robert L. Williams

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Robert L. Williams
Printed Name/Position/Title: Robert L. Williams Professor

Organization/Address:

Telephone: 865-974-4625
Fax: 865-974-0135
E-Mail Address: robwilliams@utk.edu

Date: June 3, 2003
III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

<table>
<thead>
<tr>
<th>Publisher/Distributor:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address:</td>
</tr>
<tr>
<td>Price:</td>
</tr>
</tbody>
</table>

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

<table>
<thead>
<tr>
<th>Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address:</td>
</tr>
</tbody>
</table>

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

**ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION**
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**ERIC Processing and Reference Facility**
4483-A Forbes Boulevard
Lanham, Maryland 20706

- Telephone: 301-552-4200
- Toll Free: 800-799-3742
- FAX: 301-552-4700
- e-mail: ericfac@inet.ed.gov
- WWW: http://ericfacility.org