

## RESEARCH ARTICLES

### **An Assessment of Critical Thinking: Can Pharmacy Students Evaluate Clinical Studies Like Experts?**

Donald R. Miller, PharmD

College of Pharmacy, North Dakota State University

Submitted July 10, 2003; accepted September 1, 2003; published March 2, 2004.

**Objective.** The objective of this study was to examine whether students in a drug literature evaluation class evaluated research studies similarly to the way in which experienced researchers (experts) evaluated research studies, and whether critical thinking skills predicted which students thought more like experts.

**Methods.** Sixty-six pharmacy students read 3 research summaries and indicated their belief about the study topic on a visual analog scale before and after each summary. Belief in the topic, after reading the study, was compared to their prior belief using likelihood ratios. Students also ranked the 3 study summaries from best to worst. Students' likelihood ratios were compared to those of experts. Ratios and study rankings were also matched to scores on the California Critical Thinking Test (CCTST).

**Results.** Students' change in belief after reading each study was modest and not significantly different between study types. Their ability to distinguish study quality was not as good as that of the experts. CCTST scores were not related to study evaluation ability.

**Conclusions.** Students in this class did not think like experts and the CCTST was not useful in predicting which students thought more expertly.

**Keywords:** critical thinking, CAPE, clinical studies, scientific research, assessment

---

## INTRODUCTION

The ability to think critically is one of the Center for Advancement of Pharmaceutical Education (CAPE) educational outcomes of the American Association of Colleges of Pharmacy.<sup>1</sup> Although there are numerous definitions of critical thinking (CT), a useful working definition by Ennis is "reasonable and reflective thinking that is focused on what to believe or what to do."<sup>2</sup> A construct of critical thinking based on a 1990 American Psychological Association Delphi Report defined it as purposeful, self-regulatory judgment that gives reasoned consideration to context, evidence, concepts, methods, and standards in deciding what to believe or what to do.<sup>3</sup> Critical thinking is often mentioned as an essential skill for pharmacy students and practitioners because of the common assumption that pharmacy practitioners who think critically will be better able to solve problems and think in an "expert" manner.

Several standardized instruments are available to assess CT skills. The California Critical Thinking Skills Test (CCTST) is a challenging, discipline-neutral meas-

ure of actual cognitive skills in critical thinking.<sup>3</sup> It is a standardized and validated 34-item, multiple choice test that targets those core critical thinking skills regarded to be essential elements in a College education. There is only one correct answer for each question and one point is given for each correct answer. The CCTST also reports 5 subscales: analysis, evaluation, inference, deductive reasoning, and inductive reasoning. The CCTST test manual describes normative scores in a large sample of college students, and provides evidence for content, construct, and criterion validity.<sup>3</sup>

The CCTST uses a construct of critical thinking based on the 1990 American Psychological Association Delphi Report, which describes the interactive skills of interpretation, analysis, inference, evaluation, and explanation. Other standardized tests use somewhat different working definitions of critical thinking and different test strategies. An older, frequently used test, is the Watson-Glaser Critical Thinking Appraisal.<sup>4</sup> Like the CCTST, it is a multiple-choice tool with text-based questions. However, it uses a more limited response format than the CCTST, as Watson-Glaser asks test takers to evaluate whether each purported conclusion must be true, is probably true, is probably not true, or must be false. No consensus exists concerning the best test. An individual's results on one test cannot be as-

---

**Corresponding Author:** Donald R. Miller, PharmD, Professor and Chair, Department of Pharmacy Practice, North Dakota State University, Fargo, ND 58105-5055. Tel: 701-231-7941. Fax: 701-231-7606. Email: Donald.Miller@ndsu.nodak.edu.

sumed to indicate what they would score on other tests because the tests use different constructs. An individual's results on the CCTST correlate moderately well ( $r = 0.405$  to  $0.544$ ) with those on the Watson-Glaser test, thus establishing concurrent validity between tests.<sup>5</sup> Reliability of the CCTST was established using Kuder-Richardson coefficients for internal consistency of about  $0.70$ .<sup>3</sup>

There are limited data on how critical thinking skill relates to pharmacy student performance. It is unclear whether strong critical thinkers, based on the above definitions, actually solve pharmacy problems more expertly than weak critical thinkers. Adamcik et al found the Watson-Glaser Critical Thinking Test correlated with pharmacy student grade point average (GPA), but not with clinical problem-solving skills.<sup>6</sup> Others have reported a positive correlation between CCTST score and performance in a Pharmacy Health Care and Behavior class,<sup>7</sup> and between CCTST score and clerkship success.<sup>8</sup> Duncan-Hewitt found that a student's performance on an essay test of problem-solving ability in pharmacy correlated well with the student's disposition toward critical thinking, but there was no correlation between performance and the student's CCTST score.<sup>9</sup>

Evaluation of clinical drug studies is an essential skill for pharmacists. "Evidence-based medicine" is the standard for making decisions about patient care. The author teaches a 2-credit semester course titled "Drug Literature Evaluation" at North Dakota State University, and was interested in how well graduates of the class could evaluate information compared to experts. Assessment of educational outcomes (ie, abilities and skills) is expected by the American Council on Pharmaceutical Education.

Most teachers probably hope that students will develop excellent skills by the completion of a course but it is difficult to find a gold standard against which to compare student performance. The author of this paper came across a study that allowed comparison of study evaluation skills to those of an expert group.<sup>10</sup> This study examined the impact of the quality of evidence on clinician's beliefs. Since the course was designed to teach critical thinking and study evaluation skills, it seemed reasonable to use experts as a reference group with which to compare students. Many students would not be expected to have expert evaluation skills; however, it was hypothesized that students with superior critical thinking skills, as assessed by the CCTST and its evaluation subscale, would be more likely to have expert evaluation ability.

Therefore, the author tested the proposition that pharmacy students who completed the drug literature evaluation course are able to evaluate medical evidence

in a manner similar to experienced researchers. Specific research objectives of this study were (1) to determine whether students completing a course on drug literature evaluation could rank the quality of 3 studies similarly to a panel of experienced researchers, and (2) to determine whether the California Critical Thinking Skills Test, and in particular, the evaluation subscale, could predict which students thought more like experienced researchers.

## METHODS

The students in this study were enrolled in a 2-credit semester class titled, "Pharmacy 480: Drug Literature Evaluation." The course is a required class offered in the spring semester of the second professional year. Goals and objectives for the class are listed in Appendix 1, with grading criteria given in Appendix 2, and topics covered in Appendix 3. A major goal of the class is to have students critically evaluate scientific literature. They are taught to use a systematic approach when reading studies and to recognize the strengths and weaknesses of various study designs. Class time is devoted to small group discussion of study summaries. Part of the grade comes from the students' ability to evaluate 2 separate research papers. Higher-level thinking skills, especially the ability to draw appropriate conclusions from data, are emphasized throughout the course. In addition, part of the final examination was given in open-book format so that higher-level thinking skills would be a priority in assessment.

In order to compare a student's ability to rank the quality of research studies to that of an expert group, this study used the research of de Vet et al.<sup>10</sup> Those investigators surveyed 275 first authors of recent clinical research papers by sending them 1 of 3 different study summaries on the same subject to determine how research results altered their belief in that subject. The three study summaries consisted of a strong experimental study, a well-done case-control study, and a very poorly designed, uninformative, experimental study, all on the subject of the ability of beta-carotene supplements to inhibit the progression of cervical dysplasia to cancer. In each study summary, the results were negative (did not support a beneficial effect of beta-carotene). Their hypothesis was that changes in beliefs on the subject among researchers would depend on the strength of the methods used in the studies. Prior beliefs about the efficacy of beta-carotene should have been changed by the strong experimental study, changed less by the case-control study, and not changed at all by the badly flawed study (which was considered by de Vet et al to be a "placebo control"). Since preexisting beliefs on the subject differed among the practitioners surveyed, change in belief was expressed as a likelihood

ratio; that is, the ratio of posterior belief to prior belief on the subject, with each belief measured on a visual analog scale.<sup>11</sup> Because each of the study summaries provided negative findings, a lower likelihood ratio meant a stronger change in beliefs.

For the current research, pharmacy students were given the same study summaries used by de Vet and asked to rate their belief in the topic before and after reading each of the 3 summaries. A difference with de Vet et al was that those authors only sent 1 of the 3 study summaries to each of the experts in their study. This research exercise was included and graded as part of the final examination in the course. Therefore, all 66 students in the class were present and all had an incentive to do their best. Informed consent for this particular exercise was not requested from the students because (1) the exercise was a graded portion of the examination, and (2) they each had previously given consent for the author to collect and use their CCTST scores for research purposes.<sup>12</sup> The research protocol had approval from the NDSU Institutional Review Board.

Before reading a summary, students indicated their belief on the topic by making a mark on a 15 cm visual analog scale (VAS) that was anchored at both ends and in the middle. They were asked to describe their own belief in the ability of beta-carotene to prevent cervical cancer by placing an X on the VAS anywhere between 0% and 100%. After reading each study summary, the students were again asked to indicate the degree of their belief in the subject on the VAS. An identical but new VAS was used each time. Study summaries were provided in random order, with each summary appearing first on one third of the tests. After each study summary, the posterior belief was measured on the VAS and the likelihood ratio (the posterior belief on a scale of 0 to 100, divided by the prior belief) was calculated. For example, if a student's belief changed from 45% prior to the exercise to 20% after reading a particular study, the corresponding likelihood ratio would be 0.20/0.45 divided by 0.45/0.55, which would equal 0.31. The likelihood ratio is analogous to analyzing posttest versus pretest probabilities in medical diagnosis.<sup>11</sup> A likelihood ratio of 1 indicated no change of belief, a ratio less than 1 indicated a lower belief in beta-carotene's ability to prevent cervical cancer, and a ratio above 1 indicated a greater belief in the subject.

Data on critical thinking ability, including each subscale, were collected at a separate time during the same semester as part of an ongoing project.<sup>12</sup> Form A of the CCTST was used for all students.

Change in belief as measured by likelihood ratios was compared between study types by analysis of variance. Likelihood ratios given to each of the studies were correlated with the CCTST score of each student, based

on the hypothesis that better critical thinkers would give lower likelihood ratios to the experimental study and higher likelihood ratios to the placebo study (since this was how the experienced researchers in the de Vet study approached the studies). Pearson product-moment correlations were used to examine correlations between variables.

Students were also asked to rank the studies from most useful to least useful. Those who ranked the studies in correct order (experimental > case-control > flawed study) were compared to those who ranked them incorrectly on overall CCTST scores, and to scores on the CCTST evaluation subscale, by independent t tests. The relationship of critical thinking to the student's overall grade in the course and to the student's final examination score in the course was tested by Pearson product-moment correlations.

The a priori level of statistical significance was 0.05. All statistical analysis was done with Statgraphics Plus for Windows software (Manugistics, Rockville, Md).

## RESULTS

Sixty-six students completed the final examination (Table 1). Students' change in belief about the topic was modest for each study and not statistically different between studies (Table 2). This is in contrast to the changes in belief that occurred among the experts surveyed by de Vet et al, who clearly placed more weight on the findings in the experimental study, and less

Table 1. Demographic Characteristics of Students (n = 66)

Variable	Mean (SD)	Frequency
Gender		
Male		25
Female		41
Age (years)	23.7 (4.3)	
Race		
White		58
Asian		3
Native American		5
Admission GPA	3.61 (0.26)	
Pharmacy GPA	3.52 (0.35)	
CCTST score	19.9 (4.6)	
Grade in DLE course (%)	88.5 (4.4)	
Grade on final exam (%)	81.4 (7.6)	
Previous CT course		
Yes		6
No		60
Prior Baccalaureate Degree		
Yes		6
No		60

SD = standard deviation, GPA = grade point average, CT = critical thinking, CCTST = California Critical Thinking Skills Test, DLE = drug literature evaluation.

Table 2. Mean Likelihood Ratios for Change in Belief After Reading Each Study

Study Type	Experts	Students (SD)
Experimental	0.33	0.67 (0.38)
Case-control	0.45	0.78 (0.34)
Flawed study	0.75	0.77 (0.39)

SD = standard deviation

weight on the findings in the flawed study.<sup>10</sup> Likelihood ratios did not differ by the order in which the studies were presented to the students. Therefore, it was concluded that these students did not evaluate the studies like the experts did when assessing the quality of the evidence presented to them.

When ranking studies in order of their quality, only 18 of the 66 students put them in the correct order. There was no significant difference in overall CCTST scores between students who did and did not rank the studies in the correct order of quality. Students who ranked the studies correctly had an average CCTST score of 20.0, versus a score of 19.9 for students who ranked them incorrectly. On the evaluation subscale of the CCTST, average score of those ranking studies correctly was 8.06, versus a score of 7.81 for those who ranked them incorrectly ( $P > 0.05$ ).

Overall CCTST scores correlated with the students' likelihood ratios for the case-control study, but not with the other study types (Table 3). Scores on the CCTST evaluation subscale did not correlate significantly with any study type likelihood ratios. Therefore, the students' critical thinking ability was not related to their ability to assess study quality.

On the other hand, there was a good correlation between the overall CCTST score and the student's final grade in the class ( $r = 0.50, P < 0.001$ ), and between the CCTST score and the student's final examination grade ( $r = 0.36, P = 0.004$ ). Student scores on the CCTST evaluation subscale also correlated with the student's final class grade ( $r = 0.45, P < 0.001$ ) and final examination grade ( $r = 0.31, P = 0.02$ ).

## DISCUSSION

The CCTST requires no discipline-specific knowledge, but is set in contexts familiar to college-aged students or older. It does not specifically assess a student's ability to apply their skills in pharmacy contexts. There is some research that suggests general critical thinking skills can be transferred to new situations.<sup>13</sup> However, some pharmacy investigators were unable to find a correlation between students' CT scores and pharmacy-specific problem-solving exercises.<sup>6,9</sup> Nursing investigators have similarly found no or weak correlations between scores on standardized tests of critical thinking and a student's skill in problem-solving in nursing contexts.<sup>14</sup> A drawback of

all these investigations, including the current one, is that results may depend on the specifics of a particular course and the exercises used within it.

In the 1970s, medical educators thought that clinical reasoning was distinguished by the presence of generic problem-solving skills, and some authors believed critical thinking was quite independent of content.<sup>15</sup> If this were true, then development of critical thinking for its own sake would transfer to improved problem-solving ability in a pharmacy context, and the students with high CT ability would be expected to perform better than those with low ability. However, research has revealed that expertise lay predominately in the knowledge that experts bring to a problem, and in particular, in how experts structure that knowledge.<sup>15</sup> When people become better thinkers about a topic, their internal representations of the topic-related knowledge will become similar to the representations, called schemata, of experts in the field.<sup>15</sup> General problem-solving processes are only used for problems in unfamiliar domains. The results of this study are consistent with a failure to transfer generic, critical-thinking ability to the task of rating research quality. The CT ability of our students is strongest in the evaluation and induction subscales.<sup>12</sup> This should provide them with the ability to evaluate the relative strength of information and to draw probabilistic inferences. However, according to the above theories, although the class had above-average CT skills, even the highest scoring students apparently had not yet developed expert-level knowledge schemata for this task, and therefore performed poorly on it.

That the CCTST score correlated with course grade and final examination grade is relevant and suggests that critical thinking ability is a useful predictor of course performance in general. It also suggests that the course instructor did indeed teach and assess students at a high cognitive level, and implies that comparison of students to an expert group was reasonable. Grades in the course were high, (Table 1) which suggests that students did have an adequate knowledge base to use in assessing the studies.

Instead of comparing the pharmacy students to experienced researchers, a better comparison group might have been pharmacy or medical practitioners. Previous studies have shown that many practitioners have difficulty interpreting quantitative data.<sup>16</sup> The experienced researchers used in de Vet's study probably had much greater ex-

Table 3. Correlation of Likelihood Ratios With Critical Thinking Scores

Likelihood Ratio	Overall CCTST	Evaluation Subscale
Experimental Study	-0.09	-0.02
Case-Control Study	-0.33	-0.15
Flawed Study	-0.04	-0.03

CCTST = California Critical Thinking Skills Test.

expertise in analyzing study methods than typical practitioners, though even some of them inappropriately changed their belief after reading the badly flawed, uninformative study. Still, the students in this study clearly did not think like experts. More importantly, their ability to classify the studies correctly was not related to their performance on the CCTST or on its evaluation subscale.

A study on the ability of medical students to appraise evidence has recently been published.<sup>17</sup> Taking an approach similar to that of this study, they used students' change in belief about an appropriate action to determine whether students reacted correctly to published evidence. Like the current study, medical students' change in belief was not correctly related to the quality of the evidence presented.

A key assumption of this study is that a student's ability to think critically about a study is measured by the change in belief it induces in those students. The assumption seems to have face validity because one's evaluation of evidence ought to be reflected by the change in belief about that topic. The study by de Vet et al indicated that the likelihood ratios were a valid way of assessing a person's ability to assess the strength of medical evidence.

Although including the exercise within the final examination was a way to get full and honest effort from all students, it may have actually had some perverse incentives. They were unaware of the research objectives and in informal conversations with some of the students after the examination, the students suggested they were trying to please the teacher by not placing much confidence on any of the individual studies because they had been taught that all clinical studies have weaknesses and therefore they should be somewhat skeptical of all studies. Also, some students appeared to have been confused about how to use the visual analog scales for rating their belief, or the instructions about ranking studies.

Another problem is that the CCTST is not necessarily a perfect representation of critical thinking skill. The scores on the CCTST might have been inaccurate due to low student motivation to do well. However, CCTST data were collected as part of a longitudinal study in which careful attention was paid to motivation and appropriate test conditions.

## CONCLUSIONS

After this particular drug literature evaluation class, students did not think like experts and the CCTST did not predict which students thought more like experts, even though the students did well in the class. Critical-thinking skill was related to student grades in the class, but not to expert level use of evidence. However, this study was an initial investigation into the subject, involved students at a single school, and had several limitations as discussed above.

More importantly, this experiment provides a model

for assessment of pharmaceutical education outcomes by reference to expert standards. Future studies should compare student performance in authentic assessments to performance by pharmacy practice residents, and pharmacy and medical practitioners, as well as to expert reference groups. In the field of evidence-based medicine, some investigators are developing validated tests of skills that can be used to compare performance by each of the above groups.<sup>18,19</sup> Therefore, additional investigation might determine what level of performance can be realistically expected by undergraduate students at the end of a class on drug literature evaluation. Other models of critical thinking might also be used to determine how expert performance is related to high-level cognitive skills.

## REFERENCES

1. American Association of Colleges of Pharmacy Center for Advancement of Pharmaceutical Education. Educational Outcomes, 1998.
2. Ennis RH. *Critical Thinking*. Upper Saddle River, NJ: Prentice-Hall; 1996.
3. Facione PA, and Facione NC. *The California Critical Thinking Skills Test Manual*. California Academic Press, Millbrae, CA, 1998.
4. Watson G and Glaser EM. *Watson-Glaser Critical Thinking Appraisal Manual*. San Antonio, TX: The Psychological Corporation; 1980.
5. Facione NC. *Critical Thinking Assessment in Nursing Education Programs: An Aggregate Data Analysis*. Millbrae, CA: California Academic Press; 1997.
6. Adamcik B., Hurley S, Erramouspe J. Assessment of pharmacy students' critical thinking and problem-solving abilities. *Am J Pharm Educ*. 1996;60:256-65.
7. Odedina F, Dukes N, Clemmons C, Relationship between critical thinking and pharmacy students' academic performance. *Am J Pharm Educ*. 2001;65:107S.
8. Allen DD, Bond CA. Prepharmacy predictors of success in pharmacy school: Grade point averages, pharmacy college admission test, communication abilities, and critical thinking skills. *Pharmacother*. 2001;21:842-9.
9. Duncan-Hewitt WC. Designing admissions criteria: A framework. *Am J Pharm Educ*. 1996;60:109-21.
10. de Vet HCW, Kessels AGH, Leffers P, Knipschild PG. A randomized trial about the perceived informativeness of new empirical evidence. *J Clin Epidemiol*. 1993;46:509-17.
11. Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: Getting diagnostic testing into perspective. *Q J Med*. 1998;91:247-58.
12. Miller DR. Longitudinal assessment of critical thinking skills in pharmacy students. *Am J Pharm Educ*. 2003;67: Article 120.
13. Halpern DF. Assessing the effectiveness of critical-thinking instruction. *J Gen Educ*. 1993; 42:239-54.
14. Morrison S, Free KW. Writing multiple-choice test items that promote and measure critical thinking. *J Nurs Educ*. 2001;40:17-24.
15. Winslade N, A system to assess the achievement of doctor of pharmacy students. *Am J Pharm Educ*. 2001;65:363-392.
16. Hux JE, Naylor CD. In the eye of the beholder. *Arch Intern Med*. 1995;155:2277-80.
17. Schwartz A, Hupert J. Medical students' application of published evidence: randomized trial. *BMJ*. 2003;326:536-8.
18. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence-based medicine. *BMJ*. 2003;326:319-21.
19. Fritsche L, Greenhalgh T, Falk-Ytter Y, Neumayer HH, Kunz R. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. *BMJ*. 2002;325:1338-41.

## Appendix 1. Course Goals and Objectives

### Course Goals

The goals of this course are to achieve a thorough understanding of the structure of the literature and its inherent strengths and weaknesses, such that the student may retrieve and critically evaluate scientific studies and utilize the literature to support a point of view.

### Objectives

At the completion of the course the student shall be able to:

1. Define the terms primary, secondary and tertiary literature, be able to distinguish one type from another and describe the relative advantages of each type.
2. Utilize a systematic approach to the literature to retrieve information on a given topic.
3. Describe the "scientific journal" and its publication process.
4. Define and discuss the scientific method; explain why scientific evidence is more reliable than the alternatives.
5. Summarize the salient points of a published article.
6. Apply the principles of drug literature evaluation to specific articles representing a variety of study designs by describing positive and negative aspects of an article, and assessing appropriateness of study design.
7. Assess the appropriateness of statistical analysis in a study.
8. Interpret and use data from scientific or lay publications to draw appropriate conclusions.
9. Correctly use and apply all of the following terms/concepts: placebo, placebo effect, blinding, control group, sample, random sample, randomization, inclusion and exclusion criteria, bias, hypothesis, crossover design, parallel design, within-patient design, cohort study, case-control study, longitudinal study, prospective study, retrospective study, cross-sectional study, validity, reliability, scientific method, sensitivity, specificity, surrogate criterion, correlation, regression, relative risk, odds ratio, absolute risk, attributable risk, cost-minimization, cost-benefit analysis, cost-effectiveness analysis.
10. Describe ethical concerns surrounding scientific clinical studies.
11. Analyze and synthesize researched information in order to explain a concept or argue a point in writing.

## Appendix 2. Class Assignments and Grading

1. Review of assigned journals and books (5%)
2. Evaluation of internet information (10%)
3. Evaluation of two research reports (first one in group, 5%; second, individual 10%)
4. Preparation of drug review paper (25%)
5. Critique of partner drug review paper (5%)
6. Two quizzes (5% each)
7. Final Examination (30%)

A = 90%, B = 82%, C = 75%, D = 68%

## Appendix 3. Drug Literature Evaluation Topics

### General approach

- how do we know what is true? Evidence based medicine
- the scientific method and the concept of research
- evaluation outlines, with discussion of key aspects of studies

### Research design

- study types
- placebos
- blinding
- control groups
- sampling and randomization
- statistical analysis
- measurements
- interpretation of results
- ethical concerns

### Specific study types

- clinical trials
- epidemiological studies
- pharmaco-economic studies