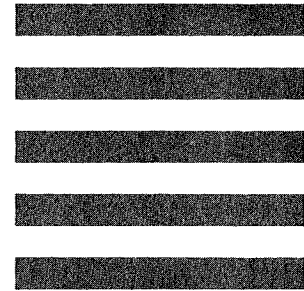


The Review of Higher Education
Winter 1990, Volume 13 No. 2
Pages 167-186
Copyright © 1990 Association for the
Study of Higher Education
All Rights Reserved



Critical Thinking Among College and Graduate Students

Patricia M. King, Phillip K. Wood,
and Robert A. Mines

John Dewey (1933) argued that reflective thinking, the careful collection and evaluation of evidence leading to a conclusion, should be a central aim of education. Recent national reports on the quality of post-secondary education in the United States have affirmed the centrality of teaching critical thinking skills to college students (National Institute of Education 1984; Association of American Colleges 1985; Gamson 1984). Yet the empirical evidence documenting progress toward this objective among undergraduate college students is minimal, and among graduate students, is virtu-

Patricia M. King is associate professor in the College Student Personnel Department, Bowling Green State University. Phillip K. Wood is assistant professor in the Psychology Department, University of Missouri-Columbia. Robert A. Mines is president of Mines and Associates, P.C., Counseling Psychology Services, Denver. This research was supported by the National Institute of Aging Traineeship Grant T32-AG00110 (02), by the Iowa Student Development Project (Office of Student Services), and by the College of Education, University of Iowa. Requests for reprints should be sent to Dr. Patricia King, Department of College Student Personnel, Bowling Green State University, Bowling Green, Ohio 43403.

ally nonexistent. Further, educators and measurement specialists alike are increasingly being asked to document the effectiveness of programs designed to improve critical thinking skills. This may involve assessing the effects of strengthened general education requirements, evaluating the effectiveness of new courses that focus on critical thinking skills, or conducting general institutional research monitoring student achievement.

However, research in this area has been hampered by the lack of agreement about what constitutes critical thinking. As Samuel Skinner has remarked, "After reading the various definitions of critical thinking, it becomes clear that agreement upon a single, concise definition of this concept is difficult, if not impossible" (1976, 293). This lack of consensus can be understood in light of recent conceptual insights into the structure of problems and corresponding decision-making strategies. Describing several developments in the conceptualization of critical thinking is the first purpose of this article. We then report the results of a study that investigated differences in critical thinking using multiple measures that reflect different definitions of the construct. Finally, we discuss the implications of these findings for researchers and educators in higher education.

Problem Structure and Tests of Critical Thinking

Robert Sternberg (1982) discusses several ways that problems differ from each other and urges his colleagues to develop better measures of ill-structured problem solving and ways of comparing how well people solve ill-structured problems compared to solving well-structured problems. Problem structure may be defined as the degree to which a problem can be described completely and the certainty with which a solution can be identified as true or correct (Wood 1983). For example, puzzles are problems that can be solved using deductive logic (e.g., All men are mortal; Plato was a man; therefore Plato was mortal). Such puzzles, which can be described with a high degree of certainty and correctness, are called "well-structured" problems (Churchman 1971).

In contrast, problems like overpopulation, hunger, pollution, and inflation are more complex. We cannot describe them completely nor solve them with a high degree of certainty; in fact, it is sometimes difficult to determine when a solution has been reached. West Churchman (1971) calls such problems "ill-structured" or "wicked." He further points out that they are more important than puzzles and that adults must daily make decisions about and solve ill-structured problems. We suggest that one reason there is such

confusion about what constitutes critical thinking is that well- and ill-structured problems have not been adequately distinguished.

The same is true for the instruments that measure the construct. Two of the most popular measures of critical thinking are the Cornell Critical Thinking Test (CCTT), which measures one's ability to solve well-structured problems (Ennis and Millman 1971), and the Watson-Glaser Critical Thinking Appraisal (WGCTA) which reflects a combination of well- and ill-structured problems (Watson and Glaser 1964). A more recent measure of critical thinking, the Reflective Judgment Interview (RJI) deals exclusively with ill-structured problems.

The Reflective Judgment model (Kitchener and King 1981) is a model of post-adolescent intellectual development and describes changes in the assumptions people hold about knowledge. In particular, the model focuses on people's assumptions about the certainty of a knowledge claim, how knowledge is acquired, and how beliefs or knowledge claims can be justified. Each of the seven stages of the model is associated with a different strategy for solving ill-structured problems. Stages 1-3 reflect the assumption that knowledge is gained either by direct, personal observation or transmitted from an authority figure. Such knowledge is assumed to be absolutely correct and certain. People who hold these assumptions cannot differentiate between well- and ill-structured problems, viewing all problems as though they were defined with a high degree of certainty and completeness.

By contrast, Stages 4-5 reflect the assumption that knowledge is gained through evaluating the available evidence and that judgments involve personal and often idiosyncratic evaluation of data. People holding these assumptions acknowledge differences in types of problems but are often at a loss when asked to solve ill-structured problems because of their inherent ambiguity.

Stages 6-7 represent the most advanced sets of assumptions identified to date that are used in solving ill-structured problems. These stages reflect the assumptions that interpretations must be grounded in data and, more importantly, that the interpretations themselves must be evaluated to determine the truth-value of a knowledge claim using such criteria as conceptual soundness, degree of fit with the data, and parsimony. Certainty of judgment may be high but not absolute, and judgments are open to change if new data or new ways of interpreting the data become available.

People holding these assumptions point out that even well-structured problems are dependent upon grounding assumptions, for example, $2 + 2 = 4$ assumes that you are working in base 10 (King 1985). Because the Reflective Judgment model has been described

in detail elsewhere (Kitchener and King 1981; King et al. 1983; Kitchener 1985; Kitchener and King in press-a), only a brief summary is included here (see Table 1).

However, because the RJI is a relatively new measure, we will describe its psychometric properties and research base. Acceptable interrater reliability levels (calculated using Pearson product-moment correlations) and interrater agreement levels (the portion of time that two judges' scores are discrepant by less than one stage) have been attained, with most ranging from .70 - .90. Internal consistency reliabilities, measured by coefficient alpha, have ranged from .62 (Welfel 1982) to .96 (Kitchener and King 1981), reflecting the heterogeneity of the samples tested. (See Mines [1982] and Schmidt and Davison [1981] for more detailed reviews of the RJI's psychometric properties.)

In a recent review of cross-sectional and longitudinal studies conducted on the model, Karen Kitchener and Patricia King (in press-a) analyzed data from almost 1,000 individuals who had been tested on the RJI. On the cross-sectional studies, they found a consistent upward trend in Reflective Judgment scores according to age/educational levels (i.e., high school, freshman to senior years of college, and beginning to advanced levels of graduate school). Mean scores, corresponding to the stage numbers given in Table 1, for high school juniors, college freshmen, college seniors, and advanced doctoral students were 2.77, 3.60, 3.99, and 5.04, respectively.

Researchers have tested over 150 individuals longitudinally at one- to six-year intervals on the RJI, providing a stronger test of the developmental sequence predicted by the model. They, too, found consistent upward shifts in individual scores over time. The mean scores of nine of the ten samples tested increased significantly over one- to six-year intervals. (See Brabeck [1984] and Kitchener and King [in press-a].) Kitchener (1985) reviews studies that differentiate the Reflective Judgment model from other models of intellectual development.

Problem Structure and Disciplinary Emphasis

Teaching students to reason well about both well- and ill-structured problems is a common goal for post-secondary institutions. Different academic disciplines, however, seem particularly suited for teaching college students such thinking skills. For example, the ability to solve well-structured problems is essential in computer science and mathematics where the parameters of the problems can be specified with a high degree of certainty and where deduc-

Table 1

REFLECTIVE JUDGMENT: DEVELOPMENT OF
EPISTEMIC COGNITION

Stage	How certain is knowledge?	How is knowledge gained?	How are beliefs justified?
1	Absolutely certain.	By direct observation.	Beliefs are a direct reflection of reality. No need to justify them.
2	Absolutely certain but not immediately available.	By direct observation and via what authorities say is true.	Direct observation or via authorities.
3	Absolutely certain about some things; temporarily uncertain about others.	Via authorities in some areas; through our own biases when knowledge is uncertain.	Via authorities in some areas; via what feels right at the moment where knowledge is uncertain.
4	No certainty because of situational variables (e.g., data lost over time).	Via our own and others' biases, data, and logic.	Via idiosyncratic evaluations of evidence and unevaluated beliefs.
5	No certainty except via personal perspectives within a specific context.	Via evidence and rules of inquiry appropriate for the context.	By rules of inquiry for a particular context.
6	Some personal certainty about beliefs based on evaluations of evidence on different sides of the question.	Via personal assessment of arguments and data, via evaluated opinions of experts.	Via generalized rules of inquiry, personal evaluations that apply across contexts, evaluated views of experts.
7	Certainty that some knowledge claims are better or more complete than others, although they are open to re-evaluation.	Via process of critical inquiry or synthesis.	As more or less reasonable conjectures about reality of the world based on an integration and evaluation of data, evidence and/or opinion.

Source: King, Kitchener, and Wood (1985). Reprinted with permission of *Moral Education Forum*.

tive logic and complex, logical manipulations are central tools of the discipline. The social sciences, by contrast, address ill-structured problems which cannot be described so completely, where the certainty of solution is lower, and where different tools of inquiry are used. Thus, we would expect students in social science majors to score higher on tests of critical thinking that use ill-structured problems because they have been given more guidance and practice in solving ill-structured problems, while students in mathematics would score higher on tests of critical thinking that use well-structured problems. We might further expect graduate students to score higher than their undergraduate counterparts on measures of critical thinking because of their increased educational exposure to such thinking tasks or because graduate school selection favors those who have already demonstrated a high level of critical thinking.

Prior research on the effects of enrollment in selected academic disciplines on critical thinking has yielded mixed results. For both the WGCTA and the CCTT, some researchers have found differences between students with and without natural science backgrounds (e.g., Burns 1974; Bennett 1975/76) while others (e.g., Bietter 1970/71; Simon and Ward 1974) found no such effect. For the RJI, liberal arts or humanities fields typically have been compared to engineering or agriculture (King and Parker 1978; Schmidt 1983; Welfel 1982; Welfel and Davison 1986). These studies found no significant differences that could be attributed solely to major. Thus, to this point, academic major has not emerged as a strong predictor of students' abilities to think critically. However, none of these studies measured skills in solving both ill-structured and well-structured problems to see what differences existed between disciplines.

Mary Brabeck (1983) measured reasoning in both domains but did not investigate disciplinary differences. She selected students who scored in the upper and lower 15 percent of the WGCTA and compared their RJI scores. Students from the top 15 percent scored higher on the RJI and had more varied scores than those of low critical thinking subjects. (A two-year follow-up of this sample is reported in Brabeck and Wood [in press].)

The Effects of Gender and Academic Ability

Investigations of gender differences on critical thinking measures have also yielded conflicting results, with most reporting no differences (e.g., Burns 1974; Cooney 1976; Skinner 1971), and some reporting gender differences favoring either women (Schafer 1972) or men (Simon and Ward 1974). An inconsistent pattern of gender

differences has also been reported for the RJI, suggesting the need to take this variable into account in future research.

Because the WGCTA (Little 1973), the CCTT (Bennett 1975/76) and the RJI (Kitchener and King 1981; King et al. 1983; Kitchener et al. 1989) have each been shown to have a strong relationship to academic ability (as measured by such standardized tests as the ACT, SAT, or GRE), it is important to ask what role academic ability plays in developing critical thinking skills and whether it can account for observed differences in critical thinking. It may be, for example, that the role of academic aptitude is different depending on the problem structure of the critical thinking measure.

The purpose of our study was to investigate differences in critical thinking between undergraduate and graduate students in social science and mathematics, using multiple measures of critical thinking that reflect different definitions of the construct. The research questions were:

1. Do graduate students score higher on each test of critical thinking than undergraduate seniors?
2. Are there differences in critical thinking scores for each measure by discipline?
3. Are there differences in critical thinking scores for each measure between men and women?
4. Can observed differences in critical thinking for each measure be accounted for by academic ability?

METHOD

Subjects for this study included forty college seniors and forty graduate students in their second year or beyond. Each group was balanced by gender and academic discipline, with half of each group majoring in a social science (psychology or sociology) and half majoring in a mathematical science (computer science, applied mathematics, or pure mathematics). Subjects were paid \$10 each for participating in the study.

We obtained test scores for each subject on a measure of academic ability (ACT or SAT scores) and on three tests of critical thinking, the WGCTA, the CCTT, and the RJI. We counter-balanced the test order by subject for the critical thinking tests. Where SAT rather than ACT scores were available for the graduate sample, we used James Maxey and Oscar Lenning's (1974) concordance tables for conversion. Only GRE scores were available for four subjects; in these cases, we transposed GRE scores to ACT equivalents with this formula: $ACTCOMP = 9.85 + GRETOTAL \times .015$ (Mines 1980/81).

The Watson-Glaser Critical Thinking Appraisal (WGCTA), Form Z, is a power test of abilities thought to be important in critical thinking. It consists of 100 items and five subtests: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments (Watson and Glaser 1964).

The Cornell Critical Thinking Test (CCTT), Form Z, is a fifty-item power test of the ability to distinguish whether a statement follows from a premise, whether something is an assumption, whether an observation statement is reliable, whether an alleged authority is reliable, whether a simple generalization is warranted, whether a hypothesis is warranted, whether a theory is warranted, whether an argument depends on ambiguity, whether a statement is over-vague or overspecific, and whether a reason is relevant (Ennis and Millman 1971, 2).

The Reflective Judgment Interview (RJI) consists of four dilemmas from science, current events, history, and religion accompanied by a set of standardized probe questions. The interview is administered by a trained interviewer and takes forty-five to sixty minutes. Each dilemma is defined by two contradictory points of view, and subjects are asked to state and justify their points of view about the issues. Each dilemma is separately tape-recorded, transcribed, and independently rated by trained raters according to the Reflective Judgment Scoring Rules (Kitchener and King 1985). A mean score for each subject is derived by averaging scores from two raters over all dilemmas. The raters (the second and third authors of this article) were blind to class level and academic discipline.

RESULTS

Because the RJI uses an interview format and is subjectively scored by trained raters, we report its psychometric properties at this testing first. Interrater reliability and agreement were .97 and .90, respectively. The internal consistency of the RJI, using coefficient alpha, was .85.

Next we ran two sets of analyses: (1) three three-way (educational level x discipline x gender) ANOVAs, one for each measure of critical thinking; and (2) an ANACOVA for each measure partialling out the effects of ACT/GRE scores to determine whether academic ability could statistically account for observed differences.

We found a significant main effect for educational level for the RJI, $F(1,72) = 26.87$, $p < .001$, for the WGCTA, $F(1,72) = 4.39$, $p < .05$, and for the CCTT, $F(1,72) = 10.25$, $p < .01$, with graduate students scoring higher on each measure than the undergraduate seniors.

When we removed the effects of academic aptitude, the differences by educational level remained significant for the RJI only ($p < .001$), suggesting that the differences observed using the other two measures might be accounted for by differences in academic ability, but not for the RJI. Table 2 reports mean scores and standard deviations by educational level, major, and gender for all three measures.

We obtained a significant main effect for academic discipline for the RJI, $F(1,72) = 6.47$, $p < .01$, but not for the WGCTA ($p = .09$) or the CCTT ($p = .08$). For the RJI, the graduate level social science (SS) majors earned higher scores than any other group ($p < .01$). There was no statistically significant difference between the scores of the seniors in the two disciplines. A test of the interaction between academic discipline and critical thinking test yielded no significant differences. For the WGCTA and the CCTT, the mathematical science (MS) majors earned higher scores. (See Table 2.) These trends

Table 2

TEST SCORES BY EDUCATIONAL LEVEL, DISCIPLINE, AND GENDER

Student Populations	RJI		WGCTA		CCTT		ACT/GRE	
	<i>M</i>	<i>s.d.</i>	<i>M</i>	<i>s.d.</i>	<i>M</i>	<i>s.d.</i>	<i>M</i>	<i>s.d.</i>
MS Seniors	4.14	.44	82.15	7.94	37.40	4.27	28.74	2.56
Male	4.24	.36	84.00	6.50	38.20	3.77	28.80	2.94
Female	4.05	.51	80.30	9.10	36.60	4.80	28.68	2.30
SS Seniors	4.01	.67	75.60	7.40	33.75	5.50	24.00	4.79
Male	4.21	.81	76.80	7.40	34.60	6.60	25.80	3.90
Female	3.81	.45	74.40	7.57	32.90	4.38	22.20	5.12
MS Graduate Students	4.36	.61	82.05	9.46	39.10	6.60	29.25	3.10
Male	4.71	.49	85.60	5.21	43.00	4.22	30.50	2.30
Female	4.01	.51	78.50	11.56	35.20	6.36	28.00	3.39
SS Graduate Students	5.17	.73	82.75	5.87	38.95	3.70	28.41	2.32
Male	5.35	.83	85.60	5.10	39.50	3.50	28.90	1.60
Female	4.99	.60	79.90	5.30	38.40	4.00	27.92	2.90

MS = Mathematical Sciences

RJI = Reflective Judgment Interview

SS = Social Sciences

WGCTA = Watson Glaser Critical Thinking Appraisal

CCTT = Cornell Critical Thinking Test

are consistent with our prediction that students in academic disciplines emphasizing a given problem structure would score higher on critical thinking tests that also emphasize this type of problem. The disciplinary differences on the RJI remained significant when the ACT/GRE scores were partialled out.

We obtained a significant educational level \times discipline interaction for the RJI, $F(1,72) = 12.71$, $p < .001$ and the WGCTA, $F(1,72) = 4.64$, $p < .05$. For the RJI, this finding was accounted for by the scores of the SS graduate students who scored higher than any other group. (See Table 2.) The SS seniors scored lowest on the WGCTA, accounting for the second interaction finding. As expected, the SS seniors also scored lowest on the CCTT, although not significantly so. We found no other significant interactions for any of the three measures.

We found a significant main effect for gender for each measure, with males scoring consistently higher ($p < .01$): RJI: $F(1,72) = 9.67$; WGCTA: $F(1,72) = 7.88$; CCTT: $F(1,72) = 8.01$. This finding suggests that gender differences exist in both well- and ill-structured problem solving ability. However, because all three instruments measure some degree of academic aptitude in addition to problem solving ability, these mean differences could be due to preexisting differences in academic aptitude.

Gender effects for the RJI remained statistically significant ($p < .05$) in analyses of covariance using ACT/GRE scores as a covariate, but not for the other critical thinking measures. Interpreting the covariance analyses is difficult, given that the design violates two assumptions of analysis of covariance. First, the covariate measure is not error free. Second, we found significant differences in ACT/GRE score across educational levels, disciplines, and gender, $F(7,72) = 6.27$, $p < .01$, suggesting that between-group differences in ACT/GRE are not due to chance and that the correlation between the covariate and the dependent variable may be different across experimental groups.

To further understand the gender differences we found on the RJI, we ran an ANACOVA test using RJI scores as the dependent variable, gender as the independent variable, and academic ability and educational level as the covariates. The gender effect remained: $F(1,76) = 4.77$, $p < .05$.

Because the three critical thinking measures reflect different definitions of the construct and different degrees of problem structure, the relationships between the measures are of interest. Table 3 shows the Pearson product-moment correlations between each pair of the three measures and between each measure and the ACT/GRE scores, as well as the partial correlations obtained when the effects

Table 3

**INTERCORRELATION MATRIX OF RJI, WGCTA, CCTT, AND
ACT/GRE SCORES, INCLUDING AND REMOVING THE
EFFECTS OF ACADEMIC ABILITY**

	RJI	WGCTA	CCTT
RJI	—	.27*	.27*
WGCTA	.46**	—	.54**
CCTT	.46**	.71**	—
ACT/GRE	.44**	.59**	.62**

* $p < .05$ ** $p < .01$

Note: Correlations to the right of the diagonal are partial correlations, corrected for the effects of academic ability.

of academic ability were removed.

The WGCTA and the CCTT are highly correlated with each other as well as with the ACT/GRE; they remain highly correlated when academic ability is partialled out. The relationship between the RJI and the other measures is significant but moderate (when ACT/GRE scores are included) to low (when they are removed). The three measures are clearly related and share a low to moderate degree of variance. However, the WGCTA and the CCTT appear to be more similar to each other and to the ACT/GRE tests than they are to the RJI.

Another way to examine these relationships is to compute the reliabilities for each instrument and correct the correlations for attenuation due to unreliability. We performed this function using the following formula: corrected reliability = obtained reliability / (square root of reliability of test 1) x (square root of reliability of test 2) (Lord and Novick 1968, 69). This formula yielded the following corrected correlations: RJI/WGCTA, $r = .51$; RJI/CCTT, $r = .59$; WGCTA/CCTT, $r = .94$. These correlations represent the highest correlation coefficient which could be attained between the measures, given these estimates of reliability and correlation for this sample. The WGCTA and CCTT correlation is very high using this

procedure—much higher than those involving the RJI—indicating again that while the three measures are clearly related, the RJI appears to measure a different aspect of critical thinking than the other two measures.

DISCUSSION

The purpose of this study was to examine whether the critical thinking scores of college and graduate students would differ by educational level, academic discipline, and gender, using tests of critical thinking that reflect different degrees of problem structure. We found significant main effects for educational level on each of the three critical thinking tests (graduate students scored higher than the undergraduate seniors), for discipline on the RJI only (graduate social science majors scored higher than any other group), and for gender on all three measures (males scored higher than females).

We found educational level by discipline interactions for the RJI and the WGCTA. Academic aptitude statistically accounted for differences by educational level and by gender for the WGCTA and the CCTT only. The pattern of strong correlations between the CCTT and WGCTA may be due to the lower reliability of WGCTA subtests that deal with ill-structured problem solving (e.g., evaluation of arguments), yielding a composite measure of critical thinking on the WGCTA which predominantly reflects well-structured problem solving ability. KR-18 reliabilities for the evaluation of arguments subtests of the WGCTA were .47 and .64 for the senior and graduate samples. By contrast, KR-18 reliabilities for the deduction subtests were .75 and .71, respectively (Wood 1980).

Differences between these measures may result from two characteristics which, at this point, are confounded. First, the two measures address different types of problems. Second, the two methods of assessment have different task demands. For example, the WGCTA and the CCTT are recognition tasks, while the RJI, with its semi-structured interview format, is a production task. Stuart Keeley, Neil Browne, and Jeffrey Kreutzer (1982) have suggested that production measures of critical thinking may provide more generalizable indicators of rational competencies. James Rest (1979), however, has argued that both abilities (recognizing as well as producing well-reasoned arguments) are important in the kinds of daily decisions that adults must make. He further points out that being able to recognize more adequate judgments precedes the ability to spontaneously produce them. Thus, the RJI is the more demanding task, and these levels of difficulty should be taken into account in

interpreting the results. A measure of Reflective Judgment that uses a recognition task format is being developed (King 1983) but is not yet available. When it is, we can control for this factor.

Critical Thinking as an Aim of Education

The college seniors in this sample do not appear to have mastered the skills of reflective thinking that John Dewey (1933) identified as a central aim of education. They did not consistently base their arguments on evidence and did not demonstrate an understanding of the role of evidence in making interpretations and judgments.

Their scores on the RJI fall predominantly around Stage 4, which has as a major characteristic the assumption that because there are many possible answers to every question and no absolutely certain way to adjudicate between competing answers, knowledge claims are simply idiosyncratic to the individual. In other words, an answer to an ill-structured problem is seen as merely an opinion. Another opinion may be quite different and even contradictory; but from a Stage 4 perspective, both approaches can be simultaneously judged as correct because of the assumption that any answer (or knowledge claim) is strictly an individual opinion. Such reasoning is common among college seniors (Kitchener and King [in press-a]), but it is a far cry from high-level critical thinking.

The graduate students, whose reasoning approached Stage 5, also fell short of John Dewey's standard. The logic at this stage is that different perspectives (e.g., different academic disciplines) have different rules of inquiry and thus yield different but equally legitimate interpretations. Beliefs are strictly relative to a particular perspective. Students holding these assumptions often cannot identify criteria by which to judge one interpretation as being more adequate or useful than another. Making informed decisions about controversial problems obviously requires such evaluative judgments. Our data suggest that even advanced doctoral students often lack the advanced levels of reasoning required in making such judgments. Mary Brabeck and Elizabeth Welfel (1985) lament that graduate students may learn this "unexamined eclecticism" from their textbooks, which thus contribute to their unwillingness or inability to evaluate differing interpretations or to rationally defend their own judgments.

How can students be taught to think reflectively? We assume that instructors need to be familiar with students' abilities to solve both well- and ill-structured problems and to understand that not all students differentiate between these types of problems. For ex-

ample, a student proficient in solving well-structured problems may need extra encouragement to tackle interpretive courses that address ill-structured issues. Further, if this student holds Stage 3 assumptions, the professor could actively explain how evidence differs from opinions or conclusions and how evidence can be used to support a conclusion or interpretation. She or he could then select assignments that require the student to evaluate or construct arguments using these skills. (For examples, see Davison, King, and Kitchener [in press]; Kitchener and King [in press-b]; and Kroll [in press].)

Many faculty members take such skills for granted in college students. Jonathan Dings (1989), for example, found that faculty members assumed that college seniors held more sophisticated epistemic beliefs (i.e., consistent with Stage 5 and 6 reasoning) than prior researchers had found (Kitchener and King [in press-a]). This inconsistency could lead to potentially frustrating and ineffective teaching practices, such as encouraging students to formulate their own synthesis of the literature on a given topic when they don't yet understand how evidence leads to conclusions or how different people evaluating the same body of evidence could arrive at different conclusions.

Disciplinary Differences

While the small to moderate correlations between the RJI and both the WGCTA and the CCTT may be explained in part by differing task demands, this explanation is unlikely for the observed disciplinary differences. Rather, it is more likely that academic disciplines differ in the emphasis they give to solving well- and ill-structured problems. However, our study did not investigate actual teaching practices across disciplines. We cannot say whether these differences result from specific educational experiences, but we recommend this topic for further research. For example, Robert Mines, Patricia King, and Phillip Wood (in press) report that different critical thinking skills are associated with different Reflective Judgment stages, a factor that may also explain these disciplinary differences. The RJI results reported here suggest that the social science disciplines may place greater emphasis on solving ill-structured problems at the graduate level. Furthermore, the fact that the RJI is administered verbally may also have given the social science students an advantage, reflected in their higher scores.

Nor can we overlook the possibility of unknown selection effects. The selection criteria for graduate departments and undergraduate majors vary widely. Admissions committees may have

weighted students' demonstrated ability to think critically about well- and/or ill-structured problems. Also, because academic departments requested but did not require students' participation in this study, the sample may have been unrepresentative.

Gender Differences

The significant and consistent effects for gender obtained in this study suggest that the rate of development of critical thinking may differ for men and women or that differential educational experiences may encourage men to become good critical thinkers, discourage women from acquiring such skills, or both. In light of the finding that the gender main effects on the WGCTA and the CCTT were eliminated for this sample when the ACT/GRE scores were partialled out, we urge that future studies investigate these relationships more fully using populations matched on academic aptitude or ability.

Although the group differences in critical thinking measures can be linearly accounted for by differences in academic aptitude, this should not be interpreted to mean that critical thinking is equivalent to academic aptitude. Rather, we propose either that critical thinking ability and academic aptitude develop synchronously or that relative standing in academic aptitude provides information on relative standing in critical thinking ability. However, the failure for Reflective Judgment differences to account linearly for group differences in ACT/GRE and, conversely, for ACT/GRE to account linearly for group differences in Reflective Judgment imply that the two measures assess different abilities.

CONCLUSION

Researchers or evaluators attempting to document the success of educational efforts designed to promote critical thinking should evaluate the assessment instruments they use or recommend in terms of the completeness and certainty of the knowledge claims they allow. As Karen Kitchener (1985) has noted, some instruments that purport to measure high levels of abstract thinking actually consist of well-structured problems. This inconsistency between what professors are attempting to teach (abstract thinking) and what the test may measure (skill at deductive reasoning) severely limits the usefulness of the research findings. To increase the usefulness and quality of research on critical thinking, we must have a clearly defined construct, specifications of the desired degree of problem structure, and instruments that reflect these elements.

BIBLIOGRAPHY

- Association of American Colleges. *Integrity in the College Curriculum: A Report to the Academic Community: The Findings and Recommendations of the Project on Redefining the Meaning and Purpose of Baccalaureate Degrees*. Washington, D.C.: Association of American Colleges, 1985.
- Bennett, Marguerite H. "A Study of the Relationship Between Curricula Taken and the Critical Thinking Abilities of High School Seniors and University of Illinois Freshmen, Sophomores, and Seniors, Majoring in Elementary Education." Ph.D. diss., University of Illinois at Urbana-Champaign, 1975. In *Dissertation Abstracts International* 36 (1975/76): 5799A.
- Bietter, John P. "A Study to Determine the Effect of Selected Variables on the Critical Thinking Ability of Social Studies Methodology Students at the University of Idaho." Ph.D. diss., University of Idaho, 1970. In *Dissertation Abstracts International* 31 (1970/71): 3383A.
- Brabeck, Mary. "Critical Thinking Skills and Reflective Judgment Development: Redefining the Aims of Higher Education." *Journal of Applied Developmental Psychology* 4 (1983): 23-34.
- _____. "Longitudinal Studies of Intellectual Development During Adulthood: Theoretical and Research Models." *Journal of Research and Development in Education* 17, no. 3 (1984): 12-27.
- Brabeck, Mary, and Elizabeth R. Welfel. "Counseling Theory: Understanding the Trend Toward Eclecticism from a Developmental Perspective." *Journal of Counseling and Development* 63 (1985): 343-48.
- Brabeck, Mary, and Phillip K. Wood. "Cross-sectional and Longitudinal Evidence for Differences Between Well-structured and Ill-structured Problem Solving Abilities." In *Adult Development: Models and Methods in the Study of Adolescent and Adult Thought*, Vol. 2, edited by Michael L. Commons, Jan D. Sinnott, Francis A. Richards, and Cheryl Armon. New York: Praeger Press, in press.
- Burns, Robert L. "The Testing of a Model of Critical Thinking Ontogeny Among Central State College Undergraduates." Ph.D. diss., University of Connecticut, 1974. *Dissertation Abstracts International* 54 (1974): 5467A.
- Churchman, C. West. *The Design of Inquiring Systems: Basic Concepts of Systems and Organizations*. New York: Basic Books, 1971.
- Cooney, Timothy M. "The Effects of Value Clarification Techniques in College Physical Science on Critical Thinking Ability, Open-

- mindfulness, and Achievement." Ph.D. diss., University of Northern Colorado, 1976. *Dissertation Abstracts International* 54 (1976): 5467A.
- Davison, Mark L., Patricia M. King, and Karen S. Kitchener. "Developing Reflective Thinking Through Writing." In *Becoming Readers and Writers During Adolescence and Adulthood*, edited by Richard Beach and Susan Hynds. Norwood, N.J.: Ablex, in press.
- Dewey, John. *How We Think: A Restatement of the Relation of Reflective Thinking to the Education Process*. Boston: Heath, 1933.
- Dings, Jonathan G. "Faculty Members' Assumptions About College Students' Reasoning Using the Reflective Judgment Model." M. A. thesis, Bowling Green State University, 1989.
- Ennis, Robert J., and Jason Millman. *Cornell Critical Thinking Test Manual*. Critical Thinking Project, The University of Illinois, Urbana, 1971.
- Gamson, Zelda. *Liberating Education*. San Francisco: Jossey-Bass, 1984.
- Keeley, Stuart M., M. Neil Browne, and Jeffrey S. Kreutzer. "A Comparison of Freshmen and Seniors on General and Specific Essay Tests of Critical Thinking." *Research in Higher Education* 17, no. 2 (1982): 139-54.
- King, Patricia M. *Reflective Judgment Questionnaire: Technical Report No. 1*. Unpublished manuscript. Department of College Student Personnel, Bowling Green State University, 1983.
- _____. *From Solving Puzzles to Resolving Problems: The Many Faces of Critical Thinking*. Paper presented at the National Invitational Conference on Pedagogy and Practice for Student Intellectual Development: High School/College Partnership, Davidson College, Davidson, N.C., June 1985.
- King, Patricia M., and Clyde A. Parker. *Assessing Intellectual Development in the College Years*. Unpublished manuscript. Instructional Improvement Project, 1976/77, University of Minnesota, Minneapolis, 1978.
- King, Patricia M., Karen S. Kitchener, Mark L. Davison, Clyde A. Parker, and Phillip K. Wood. "The Justification of Beliefs in Young Adults: A Longitudinal Study." *Human Development* 26 (1983): 106-16.
- King, Patricia M., Karen S. Kitchener, and Phillip K. Wood. "The Development of Intellect and Character: A Longitudinal-sequential Study of Intellectual and Moral Development in Young Adults." *Moral Education Forum* 10, no. 1 (1985): 1-13.
- Kitchener, Karen S. "The Reflective Judgment Model: Characteristics, Evidence and Measurement." In *Cognitive Development in*

- Young Adults*, edited by Robert A. Mines and Karen S. Kitchener, 76-91. New York: Praeger, 1985.
- Kitchener, Karen S., and Patricia M. King. "Reflective Judgment: Concepts of Justification and Their Relationship to Age and Education." *Journal of Applied Developmental Psychology* 2 (1981): 89-116.
- _____. *Reflective Judgment Scoring Rules*. Unpublished manuscript. Bowling Green State University and University of Denver, 1985.
- _____. "The Reflective Judgment Model: Ten Years of Research." In *Adult Development: Models and Methods in the Study of Adolescent and Adult Thought*, Vol. 2, edited by Michael L. Commons, Cheryl Armon, Lawrence Kohlberg, Francis A. Richards, Tina A. Grotzer, and Jan D. Sinnott. New York: Praeger Press, in press-a.
- _____. "The Development of Reflective Judgment: Instructional Implications." In *Fostering Critical Self-reflection: Tools for Transformative Learning*, edited by Jack Mezirow. San Francisco: Jossey-Bass, in press-b.
- Kitchener, Karen S., Patricia M. King, Phillip K. Wood, and Mark L. Davison. "Sequentiality and Consistency in the Development of Reflective Judgment: A Six-year Longitudinal Study." *Journal of Applied Developmental Psychology* 10 (1989): 73-95.
- Kroll, Barry M. "Teaching English for Reflective Thinking." In *Becoming Readers and Writers During Adolescence and Adulthood*, edited by Richard Beach and Susan Hynds. Norwood, N.J.: Ablex, in press.
- Little, Thomas L. "The Relationship of Critical Thinking Ability to Intelligence, Personality Factors, and Academic Achievement." Ed.D. diss., Memphis State University, 1973. In *Dissertation Abstracts International* 33 (1973): 5554-5555A.
- Lord, Frederic M., and Melvin R. Novick. *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley, 1968.
- Maxey, E. James, and Oscar T. Lenning. "Another Look at Concordance Tables Between ACT and SAT Scores." *Journal of College Student Personnel* 15 (1974): 300-304.
- Mines, Robert A. "Levels of Intellectual Development and Associated Critical Thinking Skills in Young Adults." Ph.D. diss., University of Iowa, 1980. *Dissertation Abstracts International* 41 (1980/1981): 1495A.
- Mines, Robert A. "Student Development Assessment Techniques." In *New Directions for Student Services*. Measuring Student Development, No. 20, edited by Ursula Delworth and Gary R. Hanson, 65-92. San Francisco: Jossey-Bass, 1985.

- Mines, Robert A., Patricia M. King, and Phillip K. Wood. "Stages of Intellectual Development and Associated Critical Thinking Skills in College Students." *Journal of College Student Development*, in press.
- National Institute of Education. *Involvement in Learning: Realizing the Potential of American Higher Education*. Stock no. 065-000-00213-2. Washington, D.C.: U.S. Government Printing Office, 1984.
- Rest, James R. *Development in Judging Moral Issues*. Minneapolis: University of Minnesota Press, 1979.
- Schafer, Paul J. "An Inquiry into the Relationship Between Critical Thinking Ability of Teachers and Selected Variables." Ed.D. diss., University of Pittsburgh, 1972. In *Dissertation Abstracts International* 33 (1972): 1066A.
- Schmidt, Janet A. "The Intellectual Development of Traditionally and Nontraditionally Aged College Students: A Cross Sectional Study with Longitudinal Follow-up." Ph.D. diss., University of Minnesota, 1983. In *Dissertation Abstracts International* 44 (1983): 2681A.
- Schmidt, Janet A., and Mark L. Davison. "Does College Matter? Reflective Judgment: How Students Tackle the Tough Questions." *Moral Education Forum* 6, no. 3 (1981): 2-14.
- Simon, A., and L. D. Ward. "The Performance on the Watson-Glaser Thinking Appraisal of University Students Classified According to Sex, Type of Course Pursued, and Personality Score Category." *Educational and Psychological Measurement* 34 (1974): 957-60.
- Skinner, Samuel B. "The Myth of Teaching for Critical Thinking." *The Clearing House* 45 (1971): 373-76.
- _____. "Cognitive Development: A Prerequisite for Critical Thinking." *The Clearing House* 49 (1976): 242-99.
- Sternberg, Robert. "Reasoning, Problem Solving and Intelligence." In *Handbook of Human Intelligence*, edited by Robert Sternberg, 225-307. New York: Cambridge University Press, 1982.
- Watson, Goodwin, and Edward M. Glaser. *Watson-Glaser Critical Thinking Appraisal Manual*. New York: Harcourt, Brace and World, 1964.
- Welfel, Elizabeth R. "How Students Make Judgments: Do Educational Level and Academic Major Make a Difference?" *Journal of College Student Personnel* 23 (1982): 490-97.
- Welfel, Elizabeth R., and Mark L. Davison. "The Development of Reflective Judgment During the College Years: A Four-year Longitudinal Study." *Journal of College Student Personnel* 27 (1986): 209-16.

Wood, Phillip K. *"An Analysis of the Structural Relationships Between Two Tests of Critical Thinking and Reflective Judgment."* M.A. thesis, University of Iowa, 1980.

_____. "Inquiring Systems and Problem Structure: Implications for Cognitive Development." *Human Development* 26 (1983): 249-65.