Stages of Intellectual Development and Associated Critical Thinking Skills in College Students

Robert A. Mines — Mines and Associates, P.C., Denver, Colorado

Patricia M. King — Department of College Student Personnel, Bowling Green State University

Albert B. Hood — Division of Counselor Education, University of Iowa

Phillip K. Wood — Department of Psychology, University of Missouri–Columbia

Students who reasoned at higher stages of reflective judgment also revealed better critical thinking skills, suggesting a developmental basis for the acquisition of critical thinking skills.

Since the early 1970s researchers have proposed various descriptions of intellectual development in adulthood (e.g., Arlin, 1984; Basseches, 1984; Fischer, 1980; Kitchener & King, 1981; Kramer, 1989; Labouvie-Vief, 1982; Perry, 1981; Riegel, 1973; Sinnott, 1981). These models have provided global descriptions of adult reasoning and many insights into the qualitative changes that characterize adult intellectual development. The current study attempted to identify specific reasoning skills that are associated with given stages of the Reflective Judgment model (Kitchener & King, 1981). Brabeck (1984) has noted that this model has the strongest data base of existing models of adult intellectual development.

The Reflective Judgment Model. The Reflective Judgment model (King, 1985; Kitchener, 1986; Kitchener & King, 1981; Kitchener, King, Wood, & Davison, 1989) describes a sequence of changing assumptions about knowledge and how those assumptions affect the ways a person reasons to a conclusion about problems that do not have verifiable right and wrong answers. These assumptions are summarized by stage in Table 1. They may be seen in individuals' responses to the following types of questions: What and how can we know? How certain can we be about what we know? How can we convincingly defend what we know or believe? Why do people hold different opinions about controversial issues? Answers to these questions offer useful information about students' reasoning styles, because students' assumptions about knowledge (e.g., what can be known and with what degree of certainty) are reflected in the strategies they use to gain knowledge; these, in turn, affect the adequacy with which students can solve complex and controversial problems.

Prior Research. A rather extensive research base exists for this model. Because most of the available studies have been reviewed elsewhere (King, Kitchener, Davison, Parker, & Wood, 1983; King, Kitchener, & Wood, 1985; Kitchener, 1986; Kitchener & King, 1990b; Kitchener et al., 1989), only one particularly noteworthy study is summarized here. Brabeck (1983) examined the relationship between reflective judg-

Robert A. Mines is president of Mines and Associates, P.C., and can be contacted at Mines and Associates, P.C., Counseling Psychology Services, Denver, CO 80203. Patricia M. King is an associate professor in the Department of College Student Personnel and can be contacted at Bowling Green State University, Bowling Green, OH 43402. Albert B. Hood is a professor in the Division of Counselor Education and can be contacted at the University of Iowa, Iowa City, IA 52242. Phillip K. Wood is an assistant professor in the Department of Psychology and can be contacted at the University of Missouri, Columbia, MO 65201.

This study was made possible through the financial support of the Division of Student Services, University of Iowa, and was based on Robert A. Mines's doctoral dissertation.

Requests for reprints should be sent to Dr. Robert A. Mines, Mines and Associates, P.C., 777 Grant, Suite 203, Denver, CO 80203.

TABLE 1
Reflective Judgment Stages

	What Can We Know?	How Certain Can We Know?	Through What Process Can We Know?	How May Beliefs Be Justified?	Differentiation/ Integration
1	Reality	Absolutely certain	By direct observation	Beliefs are a direct reflection of reality. No need to justify them.	Single category belief system: "What I believe is. . ."
2	True reality and false claims	Absolutely certain and certain but not immediately available.	By direct observation and via what authorities say is true.	Direct observation or via authorities.	Two category belief system; knowledge is true but some claims are false.
3	True reality, false claims, uncertainty	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 in the moment where knowledge is uncertain.	Three category belief system; knowledge is true, some claims are false, and others are uncertain.
4	While there is a reality, it can never be known. Knowledge is individually idiosyncratic.	No certainty because of situational variables (e.g., time).	Via our own and others' biases, data, and logic.	Via idiosyncratic evaluations of evidence and unevaluated beliefs.	Uncertain-knowledge becomes further differentiated into types of uncertainty and becomes overriding category, i.e., ultimately uncertain.
5	Personal interpretations of individual realities	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.	Greater differentiation within domains. Evidence integrated within specific domains.
6	Reality assumed. Evaluated personal interpretations.	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.	Evidence and opinion can be integrated across as well as within different domains. Greater differentiation.
7	Reality is never "given." Facts and assumptions may be constructed into evaluated knowledge claims about reality.	Certainty that some knowledge claims are better or more complete than others although they are open to reevaluation.	Via process of critical inquiry or synthesis.	As more or less reasonable conjectures about reality or the world based on an integration and evaluation of data, evidence, and/or opinion.	Viewpoint constructed by abstracting or synthesizing across as well as within different domains.

Note. From Kitchener, K.S. (1986). The reflective judgment model: Characteristics, evidence and measurement (pp. 78–79). In R.A. Mines & K.S. Kitchener (Eds.), *Adult cognitive development: Methods and models* (pp. 76–91). New York: Praeger. Reprinted by permission.

ment stage and critical thinking. She matched female students at four educational levels (from high school seniors to master's level graduates) on high and low extremes of critical thinking scores using the Watson-Glaser Critical Thinking Appraisal (Forms A and B) and then administered the Reflective Judgment Interview (RJI). She found a moderate (r=.40, p<.001) correlation between the two measures. Despite the matching, she also found significant differences in reflective judgment scores between educational levels, with more educationally advanced students scoring higher. High-level critical thinkers also outscored low-level critical thinkers on the RJI. She concluded that "the development of reflective judgment is separate from, and involves something other than the acquisition of these skills" (p. 32) and suggested that critical thinking skills may be necessary but not sufficient for development of reflective judgment (see also Brabeck & Wood, 1990).

The Current Study. In this study, a more fine-grained approach was used to examine the relationships between reflective judgment and standardized critical thinking tests. We focused on the specific skills constituting these measures rather than use the more global overall score. This approach allowed us to determine whether component critical thinking skills are present at some reflective judgment stages and not others, and how important certain critical thinking skills are to the complex problem-solving abilities that are reflected in the more advanced stages of reflective judgment.

METHOD

Sample

The sample for this investigation was composed of 100 students at a large midwestern university: 20 freshmen, 40 seniors, and 40 graduate students (second year or beyond). (The groups were also balanced by gender and area of study [either mathematical or social sciences]; results for this companion study are reported in King, Wood, & Mines, 1990.)

The purpose of selecting students in this manner was to obtain a group of individuals who would represent a broad range of critical thinking skills and reflective judgment stages. Twice as many students were selected at the senior and graduate levels, as compared with the freshman level, to increase the likelihood of including students who would score at the middle and upper stages of the Reflective Judgment model. Students were contacted through courses, posters, advertisements, and departmental census lists until the cells were filled. Students were paid $10 each to participate in the study.

Instruments

Reflective Judgment Interview (RJI). The RJI consists of four intellectual problems and a series of standardized probe questions and uses a semistructured format; it takes about 1 hour to complete. Each problem consists of two contradictory points of view on an intellectual issue; respondents are asked to explain and defend their responses to these points of view. A certified interviewer presents the dilemmas (usually in random order) and asks the probe questions, following up for clarity when needed. Transcripts of the interviews are then independently analyzed by certified raters using the *Reflective Judgment Scoring Rules* (Kitchener & King, 1985). Acceptable interrater reliability and agreement rates, as well as internal consistency levels, have been found in previous studies (see Mines, 1982; Schmidt & Davison, 1981, for reviews).

Watson-Glaser Critical Thinking Appraisal (WGCTA). The WGCTA is a power test designed to assess abilities thought to be important in critical thinking. The test contains 100 items and is usually completed in about 50 minutes. Watson and Glaser (1964) list five subtests: (a) *inference,* the ability to discriminate among degrees of truth and falsity of inference from given data; (b) *recognition of assumptions,* the ability to recognize unstated assumptions or presuppositions that are taken for granted in given statements or assertions; (c) *deduction,* the ability to reason deductively from given statements or premises, to recognize the relation of implication between propositions, or to determine what may seem to be an implication or a necessary inference from given premises is indeed such; (d) *interpretation,* the ability to weigh evidence and distinguish between generalizations from given data that are or are not warranted beyond a reasonable doubt; and (e) *evaluation of arguments,* the ability to distinguish between arguments that are strong and relevant and those that are weak or irrelevant to a particular question or issue. Split-half reliabilities of the subtests

for Form Zm range from .40 to .55 (Watson & Glaser, 1964). These moderately low reliabilities likely reflect the brevity of some of the subtests. Internal consistency reliabilities are not reported in the manual.

Cornell Critical Thinking Test (CCTT). The CCTT, Form Z, was used in this study to assess the hypothetico-deductive critical thinking process. The CCTT is described as having a 50-minute time limit, although its authors (Ennis & Millman, 1971) suggest that it may be used as a power test, which was done in this study. Form Z consists of 52 items in seven sections, as follows: (a) *deduction*, determining whether a statement follows from premises in material that is emotionally loaded; (b) *detecting fallaciously ambiguous arguments* (circularity, nonsupporting emotive language, oversimplification of alternatives); (c) *judging the reliability of information* and authenticity of sources; (d) *judging whether or not a hypothesis or generalization is warranted*; (e) *choosing useful hypothesis-testing predictions* when planning experiments; (f) assumption-finding, *identifying a definition* that best expresses another person's usage of a term; and (g) assumption-finding, *identifying a statement that fills a gap in a deductive argument.* Ennis and Millman (1971) have reported KR-20 reliability indexes ranging from .61 to .67 for Form Z; no subtest reliabilities are reported. The CCTT also has short subscales, which may contribute to lower subtest reliabilities.

Procedure

The two written critical thinking tests (CCTT and WGCTA) were administered in a group setting. Each of these tests took 45 to 55 minutes to complete, and the test order was counterbalanced by a student during the testing session. The RJI was administered individually by one of two interviewers in a private office. References in the transcriptions to educational level and gender were deleted prior to rating.

Permission was secured from all participants to obtain their American College Test (ACT), Scholastic Aptitude Test (SAT), or Graduate Record Examination (GRE) scores from institutional records as measures of academic aptitude; this was of interest given the possibility that any obtained differences in critical thinking and intellectual development might be the result of differential levels of academic aptitude that are associated with different educational levels. Where necessary, students' SAT or GRE scores were converted to ACT composite score equivalents.

RESULTS

Psychometric Information

For the RJI, interrater reliability (the degree of consistency between the two raters) was .97, calculated using a Pearson's *r* correlation. Interrater agreement (the proportion of time the two raters assigned the same scores) was .90. Four internal consistency reliabilities (coefficient alphas) are reported next for each instrument, both overall and by educational level (these are listed in order: freshmen, seniors, and then graduate students). For the RJI, the alphas were .89 (overall), .69, .77, and .84. The corresponding alphas for the WGCTA were .82 (overall), .49, .74, and .72. For the CCTT, alpha levels of .70 (overall), .00, .49, and .62 were obtained.

Educational Level Differences

The means and standard deviations of the four measures are given by educational level in Table 2. (The RJI mean scores correspond to stages: a mean score of 4.0 indicates that the students' assumptions about knowledge and justification of beliefs reflect Stage 4 assumptions as described in Table 1.) Without exception, the means of each of the measures were ordered in increasing magnitude across educational levels.

The RJI scores for the seniors are comparable to those that have been reported for other samples of college seniors. The RJI scores for the freshmen and graduate students are somewhat lower than those that have been reported for other samples (Kitchener & King, 1990b). For the WGCTA, the freshmen in this study scored almost a standard deviation lower than did a norm group of liberal arts freshmen (M=70.2) reported by Watson and Glaser (1964). The seniors' scores were very close to the mean of 74.4 that was reported for a sample of 200 senior women. Watson and Glaser report no graduate norms. Ennis and Millman (1971) have reported CCTT norms from two postsecondary samples, one group of college students and one group of graduate students. The mean scores between these two groups did not differ and were comparable to the scores obtained here from the freshman sample.

TABLE 2

RJI, WGCTA, CCTT, and ACT Means and Standard Deviations by Educational Level

Educational Level	RJI		WGCTA		CCTT		ACT[1]	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Freshman	3.31	.37	61.55	6.71	27.35	3.33	19.40	3.70
Senior	4.08	.56	78.87	8.27	35.57	5.20	26.37	4.50
Graduate	4.76	.78	82.40	7.79	39.02	5.28	28.83	2.70

Note. RJI = Reflective Judgment Interview; WGCTA = Watson-Glaser Critical Thinking Appraisal; CCTT = Cornell Critical Thinking Test; ACT = American College Test.
[1]ACT composite score.

A series of ANCOVAs, one for each measure, was run to see whether or not scores differed by educational level and to examine concurrently the role of academic aptitude on these scores. Using the ACT composite score as the covariate, significant differences in scores were found for all three measures between the groups after linearly adjusting for the effects of academic aptitude: $F(2, 96)=52.11$, $p<.001$ for the RJI; $F(2, 96)=8.86$, $p<.01$ for the WGCTA; $F(2, 96)=6.76$, $p=.02$ for the CCTT. In other words, the educational level differences reported in Table 2 cannot be attributed to academic aptitude.

Table 3 reports the means and standard deviations of the WGCTA and CCTT subtest scores, listed by reflective judgment stage. For both the WGCTA and CCTT subtests, the highly consistent pattern of increasing subtest scores across reflective judgment stages was reported. On the

TABLE 3

Means and Standard Deviations of the WGCTA and CCTT Subtests by Reflective Judgment Stage

Subtests	Reflective Judgment Stage									
	3 ($n=19$)		4 ($n=40$)		5 ($n=30$)		6 ($n=9$)		7 ($n=1$)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*[1]
WGCTA										
Inference	9.05	2.34	12.71	2.88	13.13	1.94	14.33	1.58	15.00	
Recognition of assumptions	11.26	2.70	13.10	1.96	13.37	1.56	14.00	1.12	15.00	
Deduction	16.84	2.12	21.12	3.23	22.27	2.27	22.67	1.94	23.00	
Interpretation	15.00	3.04	19.71	3.40	22.03	1.65	21.33	2.34	23.00	
Evaluation of arguments	10.37	1.74	10.95	2.01	11.47	2.05	11.56	2.18	14.00	
CCTT										
Does statement follow from premise?	6.47	1.43	7.80	2.03	8.51	1.16	8.56	1.59	10.00	
Detecting ambiguous arguments	4.37	1.50	6.39	2.40	7.77	1.77	8.56	1.59	8.00	
Judging reliability of information	1.84	1.12	2.54	1.19	2.90	0.71	3.00	0.87	4.00	
Is hypothesis or generalization warranted?	8.32	1.11	9.78	1.51	9.87	1.61	9.78	1.92	12.00	
Making predictions	1.84	1.26	2.19	1.12	2.47	1.04	2.67	0.87	4.00	
Determining definitions	1.84	0.69	2.73	1.07	3.10	0.66	2.67	0.87	4.00	
Identifying assumptions	2.89	1.29	3.73	1.55	4.23	1.17	3.89	1.69	4.00	

Note. See Table 2 *Note.*
[1]Insufficient data to compute standard deviations.

WGCTA, an examination of subtest scores between adjacent reflective judgment stages reveals that 19 out of 20 comparisons formed a perfect Guttman scale. For the CCTT, 28 of 32 comparisons followed this pattern. Of the five inconsistencies, four involved slight mean score reversals between Stages 5 and 6. This pattern is consistent with the theoretical assumptions of the Reflective Judgment model. That is, if the critical thinking skills measured by these two tests are associated with progressively higher stages of reflective judgment, an improvement in the complexity of critical thinking skills should be exhibited. If the critical thinking subtest scores had exhibited nonlinear patterns, further analysis would have been impossible to interpret, and the use of discriminant analysis (discussed below) would have been inappropriate.

Relationships Between Measures

Pearson product moment correlations were run between all pairs of the three measures. The WGCTA and CCTT correlated highly with each other (r=.71, p<.01) and with the ACT score (r=.59 and .62, p<.01). The correlation between each of these measures and the RJI was moderate (r=.46, p<.01). When the effects of academic ability were removed, the resulting partial correlation between the WGCTA and CCTT was .54 (p<.01); the correlation between these measures and the RJI decreased to .27 (p<.05).

The structural relationship between critical thinking subtest scores and reflective judgment stage was investigated by means of a discriminant analysis. This procedure is designed to distinguish between two or more groups (in this case, between students who score at different reflective judgment stages) by assigning weights to variables (here, the critical thinking skills) and then linearly combining the discriminating variables to make the groups as distinct as possible.

First, a global discriminant analysis (based on residualized scores) was conducted across all educational levels, with statistically significant results, $\chi^2(12, 100)=76.9$, p<.001. This indicates that critical thinking scores can be used to predict reflective judgment stage at a rate greater than chance. Subsequent analyses were also run by educational level; none of these analyses achieved statistical significance.

The discriminant analysis yielded four variables that significantly distinguished between reflective judgment stages. These variables, along with their respective standardized discriminant function coefficients, were as follows: WGCTA, interpretation (–.52); CCTT, detecting ambiguous arguments (–.30); WGCTA, deduction (–.28); and WGCTA, inference (–.21). Of these, the interpretation subtest was by far the most potent contributor to the function.

The utility of the discriminant function was tested by classifying each of the students into a given reflective judgment stage on the basis of their discriminant function scores. The percentage of correct classifications was compared with the probability of being assigned to a given reflective judgment stage based on the distribution of scores on which the student's overall reflective judgment score was based. The results of the classification procedure are reported in Table 4. The discriminant function correctly classified 50% of the students (e.g., Stage 3 with Stage 3). Percentages of correct classification by stage were 74 (Stage 3), 46 (Stage 4), 57 (Stage 5) and 0 (Stage 6). Scores assigned to adjacent stages (e.g., Stage 3 with Stage 4) account for

TABLE 4
Classification Effectiveness of the Discriminant Function for Reflective Judgment Stages 3–6

Actual Group Membership by Stage	n	Predicted Group Membership (by Stage)			
		3 (%)	4 (%)	5 (%)	6 (%)
3	19	14 (74)	5 (26)	0 (0)	0
4	41	9 (22)	19 (46)	12 (29)	1 (2)
5	30	0	13 (43)	17 (57)	0
6	10	0	4 (40)	6 (60)	0
Total	100	23	41	35	1

Percent of cases correctly classified: 50%

100% of the Stage 3 assignments, 98% for Stage 4, 100% for Stage 5, and 60% for Stage 6. Although these are not all correct classifications, they show the clustering of classifications around given stages. Based on the square of the canonical correlation, this model accounted for 50% of the variance of reflective judgment stage scores.

A comparison of the classification of each stage with that which would have occurred by chance yielded another indication of the effectiveness of the discriminant function. As shown in Table 4, the discriminant function correctly classified 74% of the students who scored at Stage 3, as compared with 19% if they were assigned by chance. None of the other stages were classified as effectively. Students who scored at Stage 4 were correctly classified 46% of the time, above the chance level of 41%. Stage 5 was the only other stage in which students were classified at an accuracy level that was well above its corresponding chance probability (57% versus 30%). The discrimination function classified students who scored at Stages 3 and 5 at only two stages (the actual one and one adjacent stage). Four students who scored at Stage 6 and one who scored at Stage 4 were classified at stages that were up to two stages discrepant from the correct stage; in these cases, the model was not accurate in classifying stage membership.

Four follow-up discriminant analyses were then run to determine whether or not a different pattern of critical thinking skills might distinguish Stage 3 reasoning from that characterizing the more advanced stages. The first comparisons were between Stage 3 and Stages 5 and 6. The results of these analyses are reported in Table 5. Three of the four variables that were significant on the global analysis were also significant here; the exception was the WGCTA inference variable, which was not significant for either comparison. With only two exceptions, the same subtests were significant for both sets of comparisons. Furthermore, the coefficients were either very similar or stronger when Stage 3 scores were compared with Stage 6 scores (with one exception: the WGCTA inference variable). Variations in the patterns are also apparent. For example, the interpretation subtest, which had a coefficient of −.52 in the overall analysis was comparably high (−.55) for the Stage 3 versus 5 comparison, but it was tied for the lowest weighted variable (−.30) for the Stage 3 versus 6 comparison. A classification effectiveness of 100% was achieved for these two discriminant analyses; the probabilities of arriving at these classifications by chance were 19%, 30%, and 10% for Stages 3, 5, and 6 (see Table 6). Different critical thinking skills seem to be associated with the assumptions of the three reflective judgment stages compared here.

DISCUSSION

Educational Level Differences

Without exception, the overall scores for each measure increased across the three educational levels. In each case, the more educationally advanced students scored higher than did their counterparts at earlier educational levels. Despite their higher RJI scores, however, the col-

TABLE 5

Discriminant Function for Reflective Judgment Stages 3 and 5, and 3 and 6

Variable	Standardized Discriminant Function Coefficient	
	Stages 3 and 5	Stages 3 and 6
Detecting ambiguous arguments (CCTT)	−.22	−.49
Does statement follow from premise? (CCTT)	−.24	
Recognition of assumptions (WGCTA)	−.25	−.51
Determining definitions (CCTT)	−.32	−.30
Judging reliability of information (CCTT)	−.34	−.74
Identifying assumptions (CCTT)	−.35	−.33
Deduction (WGCTA)	−.36	−.68
Is hypothesis or generalization warranted? (CCTT)		−.44
Interpretation (WGCTA)	−.55	−.30

Note. See Table 2 *Note.*

TABLE 6

Classification Effectiveness of the Discriminant Function for Reflective Judgment Stages

Stages 3 and 5 Actual Group	*n*	Prior Probability	Predicted Group (Stage) Membership 3 (%)	5 (%)
3	19	.39	19 (100)	0
5	30	.61	0	30 (100)
Ungrouped Cases	51		13 (26)	38 (74)
Percentage of Cases Correctly Classified: 100%				

Stages 3 and 6 Actual Group	*n*	Prior Probability	Predicted Group (Stage) Membership 3 (%)	6 (%)
3	19	.66	19 (100)	0
6	10	.34	0	10 (100)
Ungrouped Cases	71		18 (25)	53 (75)
Percentage of Cases Correctly Classified: 100%				

lege and graduate students did not reflect the skills of critical thinking that are commonly associated with the intended outcomes of higher education discussed in several recent national reports (Association of American Colleges, 1985; Gamson, 1984; National Institute of Education, 1984). (See King et al., 1990, for a more detailed discussion of this point.) The acquisition of higher order cognitive skills has been shown (Fischer & Kenny, 1986; Fischer & Pipp, 1984) to be related to environmental opportunities to learn and practice one's reasoning skills. In such environments, skills are modeled and taught, and students are given opportunities to practice and receive feedback about their success in applying new skills. The college students in this sample may not have had sufficient opportunities to refine their thinking skills (a difficult undertaking in a large university with many large classes), or they may not have had the self-confidence to take advantage of such opportunities.

Academic Ability

It is noteworthy that ACT scores also increased by educational level. Nevertheless, academic ability did not statistically account for educational level differences in the three measures of reasoning. In other words, it seems that the development of these skills is more strongly related to students' educational experiences than to their academic aptitude at the time they entered college. Educators attempting to teach reasoning skills to college and graduate students may find this reassuring. Cross-sectional research designs such as this one offer a preliminary basis for this conclusion; evidence from longitudinal studies, however, would provide a stronger data base from which to examine questions regarding the development of these sets of skills over time. Furthermore, it should be noted that in the analyses regarding the effect of academic ability (the ANCOVAs), the ACT scores of all participants were covaried out, and the relationships remained significant. It may be that the effect of academic ability is different for each educational level (or for each educational level by gender combination). Examining this possibility, however, would require a larger sample than was feasible in this study.

Relationships Between Measures

A major finding of this study was that students who reason using the assumptions of the higher stages of reflective judgment demonstrate better critical thinking skills than do those who use lower stage assumptions. The near-perfect orderings of subtest scores across reflective judgment stages (Table 3) offer preliminary support for the argument that there is a developmental basis for the acquisition of critical thinking skills.

Other results clarify some of the cognitive skills necessary to reason at the various stages of reflective judgment. The critical thinking skills that distinguished between reflective judgment stages were the following: (a) interpretation, weighing evidence and identifying gener-

alizations that are warranted beyond a reasonable doubt; (b) detecting fallaciously ambiguous arguments (e.g., evaluating arguments to determine if they violate laws of a valid argument); (c) deduction, reasoning deductively from premises to conclusions; and (d) inference, analyzing the degree of accuracy of inferences drawn from given statements. Specifically, the students holding Stage 3 assumptions showed less mastery of these skills than did their counterparts who held more advanced assumptions. Mastering these skills may be a necessary prerequisite for continued intellectual development through the reflective judgment stages; evidence of these central critical skills is certainly consistent with the reasoning that is characteristic of the more advanced stages of reflective judgment.

The four skills identified in the discriminant analyses are theoretically consistent with the characteristics of higher stage reflective judgment reasoning. Two of these skills (inference and interpretation) seem particularly well-matched to the assumptions of Stages 5, 6, and 7. For example, the interpretation subscale involves weighing evidence and identifying generalizations that are warranted beyond a reasonable doubt. Using a reasoning style that does this explicitly is a major hallmark of the upper stages. A major characteristic of Stage 4 reasoning, by contrast, is that evidence is used inconsistently to support a point of view, and evidence is not assumed to entail a conclusion.

Implications

Student affairs practitioners as well as faculty members have many opportunities to help students examine their assumptions about what and how they claim to know. Furthermore, they have many opportunities to create environments expressly designed to teach critical thinking skills, environments that include many opportunities to practice and receive feedback about these skills. If in fact there is a developmental basis for the acquisition of critical thinking skills, as this study suggests, then those who create and work in such learning environments would be well-advised to attend to the developmental characteristics of the students they attempt to serve and teach. For example, students who hold different reflective judgment assumptions translate these into different expectations for the learning environment, and as a consequence, perceive different challenges and supports in their educational tasks (e.g., see Kitchener & King, 1990a). Trying to teach critical thinking without taking these factors into account would not only contradict our knowledge of developmental processes but would probably also result in less effective practice (Strange & King, 1990) and less success in achieving the central educational goal of teaching students to reason critically.

REFERENCES

Arlin, P. K. (1984). Adolescent and adult thought: A structural interpretation. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations: Late adolescent and adult cognitive development* (pp. 258–271). New York: Praeger.

Association of American Colleges. (1985). *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, DC: Association of American Colleges.

Basseches, M. (1984). *Dialectical thinking and adult development.* Norwood, CA: Ablex.

Brabeck, M. (1983). Critical thinking skills and reflective judgment development: Redefining the aims of higher education. *Journal of Applied Developmental Psychology, 4,* 23–34.

Brabeck, M. (1984). Longitudinal studies of intellectual development during adulthood: Theoretical and research models. *Journal of Research and Development in Education, 17,* 12–27.

Brabeck, M., & Wood, P. (1990). Cross-sectional and longitudinal evidence for differences between well-structured and ill-structured problem solving abilities. In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, T. A. Grotzer, & J. D. Sinnott (Eds.), *Adult development: Models and methods in the study of adolescent and adult thought* (Vol. 2, pp. 133–146). New York: Praeger.

Ennis, R. H., & Millman, J. (1971). *Cornell critical thinking test manual.* Urbana, IL: The University of Illinois, Critical Thinking Project.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87,* 477–531.

Fischer, K. W., & Kenny, S. L. (1986). Environmental conditions for discontinuities in the development of abstractions. In R. A. Mines & K. S. Kitchener (Eds.), *Adult cognitive development: Methods and models* (pp. 57–75). New York: Praeger.

Fischer, K. W., & Pipp, S. L. (1984). Processes of cognitive development: Optimal level and skill acquisition. In R.J. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 45–80). New York: (W. H.) Freeman.

Gamson, Z. (1984). *Liberating education.* San Francisco: Jossey-Bass.

King, P. M. (1985). Choice-making in young adulthood: A developmental double-bind. *Counseling and Human Development, 18*(Whole no. 3).

King, P. M., Kitchener, K. S., Davison, M. L., Parker, C. A., & Wood, P. K. (1983). The justification of beliefs in

young adults: A longitudinal study. *Human Development, 25,* 106–116.

King, P. M., Kitchener, K. S., & Wood, P. K. (1985). The development of intellect and character: A longitudinal-sequential study of intellectual and moral development in young adults. *Moral Education Forum, 10,* 1–13.

King, P. M., Wood, P. K., & Mines, R. A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education, 13,* 167–186.

Kitchener, K. S. (1986). The reflective judgment model: Characteristics, evidence and measurement. In R. A. Mines & K. S. Kitchener, (Eds.), *Adult cognitive development: Methods and models* (pp. 76–91). New York: Praeger.

Kitchener, K. S., & King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology, 2,* 89–116.

Kitchener, K. S., & King, P. M. (1985). *Reflective judgment scoring rules.* Unpublished manuscript, Bowling Green State University and University of Denver.

Kitchener, K. S., & King, P. M. (1990a). The reflective judgment model: Transforming assumptions about knowing. In J. Mezirow (Ed.), *Fostering critical reflection in adulthood: A guide to transformative and emanicaptory learning* (pp. 159–176). San Francisco: Jossey-Bass.

Kitchener, K. S., & King, P. M. (1990b). The reflective judgment model: Ten years of research. In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, R. A. Grotzer, & J. Sinnott (Eds.), *Adult development: Models and methods in the study of adolescent and adult thought* (Vol. 2, pp. 63–78). New York: Praeger.

Kitchener, K. S., King, P. M., Wood, P. K., & Davison, M. L. (1989). Sequentiality and consistency in the development of reflective judgment: A six-year longitudinal study. *Journal of Applied Developmental Psychology, 10,* 73–95.

Kramer, D. (1989). Development of an awareness of contradiction across the life span and the question of postformal operations. In M. L. Commons, J. D. Sinnott, F. A. Richards, & C. Armon (Eds.), *Adult development, Vol. 1: Comparisons and applications of developmental models* (pp. 133–159). New York: Praeger.

Labouvie-Vief, G. (1982). Dynamic development and mature autonomy: A theoretical prologue. *Human Development, 25,* 161–191.

Mines, R. (1982). Student development assessment techniques. In G. R. Hanson (Ed.), *New directions for student services: Measuring student development* (No. 20, 65–92). San Francisco: Jossey-Bass.

National Institute of Education. (1984). *Involvement in learning: Realizing the potential of American higher education* (Stock No. 065-000-00213-2). Washington, DC: U.S. Government Printing Office.

Perry, W. G., Jr. (1981). Cognitive and ethical growth: The making of meaning. In A. Chickering & Associates (Eds.), *The modern American college* (pp. 76–116). San Francisco: Jossey-Bass.

Riegel, K. F. (1973). Dialectic operations: The final period of cognitive development. *Human Development, 16,* 346–370.

Schmidt, J. A., & Davison, M. L. (1981). Does college matter? Reflective judgment: How students tackle the tough questions. *Moral Education Forum, 6*(3), 2–14.

Sinnott, J. (1981). The theory of relativity: A metatheory for development? *Human Development, 24,* 293–311.

Strange, C. C., & King, P. M. (1990). The professional practice of student development. In D. Creamer (Ed.), *College student development: Theory and practice in the 1990s* (Vol. II, pp. 9–24). Alexandria, VA: ACPA Media Publications #49.

Watson, G., & Glaser, E. M. (1964). *Watson-Glaser critical thinking appraisal manual.* New York: Harcourt, Brace, and World.