DOCUMENT RESUME

ED 387 510 TM 023 713

AUTHOR Taube, Kurt T.
TITLE Critical Thinking Ability and Disposition as Factors of Performance on a Written Critical Thinking Test.
PUB DATE Apr 95
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability; *Critical Thinking; *Factor Structure; Grade Point Average; Higher Education; *Personality; *Thinking Skills; *Undergraduate Students
IDENTIFIERS AT 20 Ambiguity Tolerance Scale; Checklist of Educational Views; Confirmatory Factor Analysis; *Ennis Weir Critical Thinking Essay Test; Need for Cognition Scale; Scholastic Aptitude Test; Watson Glaser Critical Thinking Appraisal

ABSTRACT

Critical thinking has been conceptualized as a two-factor system in which critical thinking ability and critical thinking disposition combine to determine actual thinking performance. The present study used confirmatory factor analysis to investigate such a two-factor model empirically. One hundred ninety-eight Purdue University undergraduates completed the Watson-Glaser Critical Thinking Appraisal, the Ennis-Weir Critical Thinking Essay Test, the Need for Cognition Scale (NCS), the AT-20 ambiguity tolerance scale, and the Checklist of Educational Views (CLEV). The students' grade point averages (GPAs) and Scholastic Aptitude Test (SAT) Verbal and Mathematics scores were also collected. The NCS, AT-20, and CLEV served as measures of disposition; Watson-Glaser, SAT-Verbal, and SAT-Math served as measures of thinking ability; and Ennis-Weir and GPA were assumed to measure both ability and disposition. Confirmatory factor analysis indicated that the postulated two-factor model provided a more accurate fit with the data than did a model including only one latent factor and that Ennis-Weir, but not GPA, loads significantly on both factors. (Contains 6 tables and 80 references.) (Author)

CRITICAL THINKING ABILITY AND DISPOSITION AS FACTORS OF PERFORMANCE ON A WRITTEN CRITICAL THINKING TEST

Kurt T. Taube

Purdue University

ABSTRACT

Critical thinking has been conceptualized as a two-factor system in which critical thinking ability and critical thinking disposition combine to determine actual thinking performance. The present study used confirmatory factor analysis to investigate such a two-factor model empirically.

One hundred ninety-eight Purdue University undergraduates completed the Watson-Glaser Critical Thinking Appraisal, the Ennis-Weir Critical Thinking Essay Test, the Need For Cognition Scale (NCS), the AT-20 ambiguity tolerance scale, and the Checklist of Educational Views (CLEV). The students' grade-point averages (GPAs) and Scholastic Aptitude Test (SAT) Verbal and Math scores were also collected. The NCS, AT-20, and CLEV served as measures of disposition; Watson-Glaser, SAT-Verbal, and SAT-Math served as measures of thinking ability; and Ennis-Weir and GPA were assumed to measure both ability and disposition. Confirmatory factor analysis indicated that the postulated two-factor model provided a more accurate fit with the data than did a model including only one latent factor; and that Ennis-Weir, but not GPA, loads significantly on both factors.

INTRODUCTION

One of the major problems in the study of critical thinking has been the lack of a universally accepted definition of critical thinking. Galotti (1989) argued that critical thinking is no more than an instance of everyday, informal reasoning. The subject is generally faced with a poorly defined problem, must search for relevant information, and often has to determine what information is relevant. Halpern (1989, p. 5) described critical thinking as "thinking that is purposeful, reasoned, and goal directed. It is the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions." Lipman (1988, p. 39) stated that critical thinking is "skillful, responsible thinking that facilitates good judgment because it relies upon criteria, is self-correcting, and is sensitive to context." Lipman went on to compare ordinary thinking to good (critical) thinking by focusing on the differences in precision and rigor between the two. Thus, while guessing is characteristic of ordinary thinking, estimating characterizes critical thinking; supposing is contrasted with hypothesizing; and so on. Therefore, critical thinking is a sort of "cognitive accountability."

Ennis (1962, p. 83) defined critical thinking as "the correct assessing of statements," and went on to outline twelve key critical thinking skills. Browne, Haas, and Keeley (1978) identified a similar set of eight critical thinking skills, while Dick (1991) suggested a taxonomy of fifteen critical thinking skills organized into five categories. These researchers focused on critical thinking as a set of reasoning skills, with little or no attention paid to any motivational or dispositional component.

Others have addressed this issue. Dressel and Mayhew (1954) outlined a set of critical thinking abilities, but also defined critical thinking as "a point of view toward problems and their solutions and a way of thinking about basic problems faced by mankind (Dressel & Mayhew, 1954, p. 273)." McPeck (cited in Powell, 1987, p. 170) defined critical thinking as "the appropriate use of reflective scepticism [sic]." A critical thinking "attitude" is central to this conception.

Other theorists explicitly include dispositions in their definitions of critical thinking. Beyer (1985) stated that critical thinking has two important dimensions: a "frame of mind" and a number of specific mental operations. The frame of mind includes an alertness to the need to evaluate information, a willingness to test opinions, and a desire to consider all viewpoints. DeNitto and Strickland (1987) featured a critical "attitude" consisting of doubt, carefulness, objectivity, and determinism, as well as thinking strategies and thinking skills, in their model of critical thinking. Halpern (1989) drew a distinction between performance and competence, stating that critical thinking skills are of no value if they are not used. The development of a critical thinking attitude is as important as the development of the skills themselves.

Paul and Nosich (1991, p. 5) stated that "critical thinking entails the possession and active use of a set of traits of mind" including "independence of thought, fairmindedness, intellectual humility, intellectual courage, intellectual perseverance, intellectual integrity, curiosity, confidence in reason, the willingness to see objections, to enter sympathetically into another's point of view," and "to recognize one's own egocentricity or ethnocentricity." This is consistent with Paul's (cited in Siegel, 1988) contrast between "weak sense" and "strong sense" critical thinking. In the weak sense, one uses his/her battery of critical thinking skills to attack positions and assumptions that one has already rejected, and to fend off challenges to one's own deeply-held beliefs. In the strong sense, one is willing to challenge one's own beliefs and attitudes as well as those of others.

The role of dispositions is obvious in Facione's (1991) description of the ideal critical thinker. According to Facione:

> The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and circumstances of inquiry permit. (p. 14)

Ennis (1985, 1991) has incorporated lispositions into his theory of critical thinking.

Critical thinking is defined as "reflective and reasonable thinking that is focused on what to believe or do (Ennis, 1985, p. 45)." Ennis' latest conception (Ennis, 1991) included twelve critical thinking dispositions and sixteen critical thinking abilities. The abilities are closely allied to those outlined by Ennis (1962), Browne et al. (1978), and Dick (1991): to (1) identify the focus, issue, question, or conclusion; (2) analyze arguments; (3) ask and answer questions of clarification and/or challenge; (4) define terms, judge definitions, and deal with equivocation; (5) identify unstated assumptions; (6) judge the credibility of a source; (7) observe, and judge observation reports; (8) deduce, and judge deductions; (9) induce, and judge inductions; (10) make and judge value judgments; (11) consider and reason from premises, reasons, assumptions, positions, and other propositions with which one disagrees or about which one is in doubt, without letting the disagreement or doubt interfere with one's thinking; (12) integrate the other abilities and dispositions into making and defending a decision; (13) proceed in an orderly manner appropriate to the situation; (14) be sensitive to the feelings, level of knowledge, and degree of sophistication of others; (15) employ appropriate rhetorical strategies in discussion and presentation; and (16) employ and react to "fallacy" labels in an appropriate manner. The dispositions, however, are presented in considerably greater detail than had been the case in earlier papers. According to Ennis (1991, Table 1), the dispositions of the ideal critical thinker are as follows:

(1) To be clear about the intended meaning of what is said, written, or otherwise communicated.

(2) To determine and maintain focus on the conclusion or question.

(3) To take into account the total situation.

(4) To seek and offer reasons.

(5) To try to be well informed.

(6) To look for alternatives.

(7) To seek as much precision as the situation requires.

(8) To try to be reflectively aware of one's own basic beliefs.

(9) To be open-minded; consider seriously other points of view than one's own.

(10) To withhold judgment when the evidence and reasons are insufficient.

(11) To take a position (and change a position) when the evidence and reasons are sufficient to do so.

(12) To use one's critical thinking abilities. (p. 12)

Siegel (1988) criticized Ennis' approach to critical thinking dispositions on the grounds that Ennis concentrated on "micro-dispositions" rather than a global "macro-disposition." Siegel identified two components essential to critical thinking: reason assessment, which parallels the thinking abilities outlined so far, and the "critical spirit," which includes attitudes, dispositions, habits of mind, and character traits, all of which combine to produce a "love of reason" characterized by intellectual honesty, objectivity, justice to evidence, and sympathetic and impartial consideration of interests. The critical thinker will base his/her beliefs and actions on the results of such inquiry.

The situation of dispositions and traits being an important and largely untested component of mental skills has been known for many years. Wechsler (1943, 1950) identified "non-intellective" factors as untested content in intelligence tests. These "non-intellective" factors include drive, energy, curiosity, anxiety, impulsiveness, persistence, and other temperamental factors. Anastasi (1982) stated that:

> In interpreting test scores, personality and aptitudes cannot be kept apart. An individual's performance...is influenced by his or her achievement drive, persistence, value system, freedom from handicapping emotional problems, and other characteristics traditionally classified under the heading of "personality." (p. 352)

Critical thinking disposition may thus be seen as a key non-intellective factor in the assessment of critical thinking ability.

Hudgins, Riesenmy, Ebel, and Edelman (1989) have presented such a two-factor (cognitive and motivational) model of "self-directed critical thinking" in children. The motivational component of the model stresses the spontaneity and independence of the child. The child spontaneously calls upon his/her intellectual skills without any prompting from an external

authority, such as a teacher. Indeed, Norris (1985, p. 40) stated that "one must have the disposition to think productively and critically about issues, or else no amount of skill in doing so will be helpful." It seems plausible that such personality traits as open-mindedness, cognitive complexity, need for cognition, tolerance of ambiguity, and reflectiveness represent the dispositional component of critical thinking, and are characteristic of effective critical thinkers.

Dispositional Factors

Critical thinking disposition implies a willingness to expend cognitive effort in solving everyday problems. Two personality variables which appear to be conceptually related to critical thinking disposition are need for cognition and tolerance of ambiguity.

Need for Cognition

Need for cognition can be summarized as the "tendency for an individual to engage in and enjoy thinking (Cacioppo & Petty, 1982, p. 116)." Cohen, Stotland, and Wolfe (cited in Cacioppo & Petty, 1982, p. 116-117) described the need for cognition as "a need to structure relevant situations in meaningful, integrated ways. It is a need to understand and make reasonable the experiential world." Cacioppo and Petty developed the Need For Cognition Scale (NCS) to measure the construct of need for cognition.

Cacioppo and Petty (1982) observed that the NCS correlated positively with ACT score and field dependence, and negatively with dogmatism. The authors also reported that subjects high in need for cognition tended to like a complex number-circling task more than a simple number-circling task, while those low in need for cognition showed the opposite pattern of results. Later research demonstrated that subjects high in need for cognition as measured by the NCS were more likely to discriminate between strong and weak arguments, recall more supporting information, and expend more cognitive effort than subjects low in need for cognition when asked to evaluate a persuasive communication (Cacioppo, Petty, & Morris, 1983).

Other researchers have demonstrated positive correlations between NCS and persistence,

cognitive complexity, confidence, masculinity, ACT score, and grade-point average (Waters & Zakrajsek, 1990); and between NCS and objectivism, defined as the tendency to base judgments and beliefs on empirical information and rational considerations (Leary et al., cited in Osberg, 1987). Olson, Camp, and Fuller (1984) reported significant positive correlations between NCS and ACT score, as well as with a number of measures of curiosity. Need for cognition was also found to correlate positively with "high cognition" choices of magazines, books, and leisure activities (Tolentino, Curry, & Leak, 1990).

Based on the previous research, need for cognition appears to be a variable related to the disposition to think critically. Indeed, Petty and Cacioppo designed the NCS in response to a perceived need for a scale assessing individual differences in the motivation to think (Petty, cited in Heesacker, 1985).

Tolerance of Ambiguity

> English and English (cited in MacDonald, 1970) defined ambiguity tolerance as:
>
> (The) willingness to accept a state of affairs capable of alternate interpretations, or of alternate outcomes, e.g., feeling comfortable (or at least not feeling uncomfortable) when faced by a complex social issue in which opposed principles are intermingled. Low ambiguity tolerance is shown by the desire to have everything reduced to black and white. (p. 791)

Budner (cited in Kirton, 1981, p. 408) defined intolerance of ambiguity as the "tendency to perceive ambiguous situations as sources of threat," whereas tolerance of ambiguity was defined as the "tendency to perceive ambiguous situations as desirable." Budner (cited in MacDonald, 1970, p. 791) went on to define an ambiguous situation as "one which cannot be adequately structured or categorized by the individual because of lack of sufficient cues."

MacDonald (1970) described the person high in tolerance of ambiguity as one who will seek out and enjoy ambiguity and excel in the performance of ambiguous tasks. Rydell and Rosen (cited in Kirton, 1981) extended the definition of tolerance of ambiguity by describing the

ambiguity-tolerant person as one who:

> Seeks (ambiguity) out in order to reduce it; he's tolerant of it only as long as he can do something about it, and his cognitive task is to get rid of it. Thus, the highly cognitive individual attracted to ambiguity should be tolerant of it only to the extent that the "ambiguous situation" and the cognitive field seem capable of analysis and restructuring to bring about closure...(permitting) lack of closure only long enough to bring about a resolution of the ambiguous situation. (p. 408)

Frenkel-Brunswik (cited in MacDonald, 1970) attempted to relate intolerance of ambiguity to authoritarianism. Rokeach (cited in MacDonald, 1970) later followed a similar line in developing the concept of dogmatism. Inversely, ambiguity tolerance has been equated with and measured by a scale of cognitive complexity (Peters & Amburgey, 1982).

MacDonald (1970) developed the AT-20 scale to measure the construct of ambiguity tolerance. MacDonald obtained significant negative correlations between AT-20 and dogmatism, rigidity, and church attendance; while Kirton (1981) observed significant negative correlations between AT-20 and dogmatism, inflexibility, conservatism, and Budner's ambiguity intolerance scale. Kroll (1988) found that the AT-20 correlated positively with cognitive complexity, open-mindedness, individualism, grade-point average, achievement motivation, and task-orientation, defined as the tendency to view learning as an end in itself; and negatively with ego-orientation, defined as the tendency to view learning as a means of obtaining wealth and status rather than as an end in itself. Tegano (1990) reported that the AT-20 correlated positively with creativity and playfulness.

Based on the previous research, ambiguity tolerance also appears to be a variable related to the disposition to think critically. Tegano (1990, p. 1049) stated that "individuals who view ambiguity as desirable and challenging might be likely to engage in problem finding, problem solving, and evaluation, avoiding premature decisions throughout the process."

Intellectual Development

Perry's (1970, 1981) model of intellectual development postulated nine positions, which could be grouped into four basic divisions: dualism, multiplicity, relativism, and commitment. Dualism, the lowest level of this continuum (Positions 1 and 2), is characterized by a view of the world in absolute polar terms of good-bad, right-wrong, or we-they. In Position 1, Truth is understood only by Authorities, and there is no tolerance for alternative points of view. Students exhibiting "pure" dualism (Position 1) are rarely found empirically (Moore, 1991). In Position 2, other perspectives and beliefs are acknowledged to exist, but are assumed by the student to be wrong. The Checklist of Educational Views excerpt (CLEV; Perry, Sprinthall, Wideman, & Jones, 1968) used in the present research is a measure of dualism.

Multiplicity (Position 3) introduces legitimate uncertainty, as some things are not yet known even to Authorities, but will be known in the future. In Position 4, the student realizes that there is much more that is unknown than is known, but also believes that "anything goes" since there is no nonarbitrary means of determining the truth. Contextual relativism (Positions 5 and 6) features generally context-bound reasoning and the emergence of the self as a legitimate source of knowledge. Finally, commitment within relativism (Positions 7-9) features choices of career, mate, and lifestyle made in the face of alternatives after experiencing genuine doubt (Perry, 1981). Few undergraduates reach this level of development (Perry, cited in Moore, 1991).

Some researchers seem to have drawn a close parallel between intellectual development and ambiguity tolerance. Frenkel-Brunswik (cited in Rotter & O'Connell, 1982, p. 1210) defined intolerance of ambiguity as "the tendency to resort to black-white solutions, to arrive at premature closure as to valuative aspects, often at the neglect of reality, and to seek unqualified and unambiguous overall acceptance or rejection of other people," a characteristic which fits well with Perry's concept of dualism. Dualism, as measured by the CLEV, has also been observed to correlate negatively with open-mindedness and cognitive complexity (Taube, 1987).

Kitchener and King's (cited in Welfel & Davison, 1986) seven-stage model of the development of reflective judgment drew heavily from Perry's work. This model has been used by

career counselors in their work with college students (Welfel, 1982). Counselors reported that students at lower levels of Kitchener and King's model often believe in "myths" (Ellis, cited in Welfel, 1982). For example, a student could believe that changing one's mind regarding academic major or career goals is undesirable, or that one should have clear, concrete career plans and stick to them. These "myths" may contribute to client frustration, indecision, and disillusionment about the benefit of career counseling. Students at middle levels may believe absolute truth doesn't exist, so they are often reluctant to take risks and make decisions. The counselor can provide better service to the student if he/she addresses the beliefs and fears common at different developmental levels.

It has been suggested that the traditional medical school environment encourages dualistic thinking by requiring memorization of large amounts of information (Simpson, Dalgaard, & O'Brien, 1986). Later, medical residents may adopt a multiplistic outlook in order to avoid confrontations with instructors. True relativism often does not arise until late in residency, when the resident realizes that he/she must make reasoned decisions regarding medical treatment and diagnosis in the face of uncertainty. Similarly, nursing students tended to remain at lower levels on the Perry continuum as measured by Allen's essay test (Frisch, 1987) and the CLEV (Woodham, Taube, Elmore, Linehan, & Wagner, 1986) through their school years and not move to higher levels until after graduating and entering practice, most likely due to the different socialization experiences at work in the classroom and in the practice setting. Professional commitment should not be expected until after graduation (Collins, cited in Frisch, 1987). Registered Nurses (RNs) earned higher scores on the CLEV (higher scores indicate a lower level of dualism) than did Licensed Practical Nurses, perhaps due to the greater level of individual initiative and problem-solving expected of the RN role (Woodham et al., 1986). CLEV score correlated positively with internal locus of control, so it is possible that people choose roles for themselves based on the amount of responsibility and autonomy they are comfortable with.

Based on the previous research, it is plausible that students at the dualistic level of intellectual development would be less likely than students at higher levels of intellectual

development to think critically. If this is true, it is also plausible that this difference is a result of disposition rather than ability, because the Perry (1970, 1981) scheme addresses typical patterns of thought rather than thinking ability.

The Present Study

The subjects completed separate objective measures of critical thinking ability and attitudes believed to represent critical thinking disposition. The subjects also completed a written critical thinking task, the Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985). Performance on the Ennis-Weir was expected to reflect critical thinking disposition as well as critical thinking skills, because subjects could spend as much time and effort as they wished in reading the stimulus passage, formulating a response, and writing. According to Norris and Ennis (1989), essay tests provide a means of assessing disposition as well as ability because a subject's essay will reflect whether he/she sought reasons for his/her position, looked for alternatives, and was open-minded. If the subject did these things without being specifically instructed to do so, this indicates that the subject possesses the related thinking dispositions.

It was hypothesized that a model of critical thinking consisting of the two factors of ability and disposition (Ennis, 1985, 1991; Siegel, 1988) would provide a more accurate fit with the obtained data than would a model including only a single "critical thinking" factor. It was also hypothesized that performance on the Ennis-Weir is a product of both critical thinking ability and critical thinking disposition.

METHOD

Subjects

A total of 198 students participated in this study. All were enrolled in undergraduate educational psychology at Purdue University during 1989 and 1990. Nearly all were education majors, for whom this course is a requirement, and a majority were sophomores or juniors.

Thirty-nine of the students were male and 159 were female. Scores on all measures were available for 137 (69%) of the 198 students. The students fulfilled one course research requirement by participating in this study.

Instruments

The Need for Cognition Scale

Cacioppo and Petty (1982) developed the Need for Cognition Scale (NCS) by empirical keying. According to the authors:

> The...pool of 45 items was administered to groups who were known to differ along the dimension of need for cognition...Members of a large midwestern university faculty served as subjects in the high-need-for-cognition group, whereas members of assembly lines in factories in the surrounding community served as subjects in the low-need-for-cognition group. (p. 118)

The items were scored on a 9-point (strongly agree - strongly disagree) scale. Thirty-four of the 45 items discriminated significantly ($p < .10$) between the two groups. Factor analysis of these 34 items yielded one primary factor. The authors reported a corrected split-half reliability coefficient of .87. Cacioppo and Petty's approach may have been flawed in its assumption that the university faculty and assembly-line workers differed in need for cognition. While it is possible that the two groups differed in need for cognition, and even plausible that the two groups differed in intellectual ability, these hypotheses were not investigated empirically prior to the development of the NCS.

Heesacker (1985, pp. 472-473) evaluated the NCS quite positively, describing it as "a very useful instrument for both researchers and practitioners because of its accurate assessment of individual differences in the propensity to enjoy and engage in effortful thought." It should be noted that this evaluation was based on a short (18 items) version of the NCS (Cacioppo, Petty & Kao, cited in Heesacker, 1985), but the short form has been shown to correlate highly ($r = .95$)

with and produce the same factor structure as the original 34-item form.

In the present study, the items were scored on a 5-point (strongly agree - strongly disagree) Likert-type scale. It was believed that this scale would provide sufficient precision without requiring the fine distinctions necessary to complete the original 9-point scale. Total scores could range from 34 to 170. Higher scores indicate higher levels of need for cognition.

The AT-20

The AT-20 ambiguity tolerance scale (MacDonald, 1970) consists of 20 dichotomous (true-false) items; thus, total scores could range from 0 to 20. Higher scores indicate greater tolerance of ambiguity. The AT-20 includes 16 items from the Rydell-Rosen Ambiguity Tolerance Scale, two items from the California Personality Inventory, and two items from the Barron Conformity Scale. MacDonald reported KR-20 reliability coefficients of .73 and .63 on different samples, and Kroll (1988) observed a Cronbach alpha reliability coefficient of .68. Kirton (1981) has described the AT-20 as an "operationally useful" measure of ambiguity tolerance.

The Checklist of Educational Views

Perry and his associates developed the Checklist of Educational Views (CLEV) to measure the degree to which students preferred "black-white, right-wrong, thinking in an authority-oriented outlook as against their preference for contingent, relativistic thinking in an outlook of greater individual judgment (Perry et al., 1968, p. 101)." Perry drew heavily from Adorno's work on the authoritarian personality in constructing the CLEV and developing the concept of dualism, so it is not surprising that the items centered on fact-oriented, dogmatic ways of thinking in which hard work, distrust of others, and loyalty to one's school and family are paramount.

Factor analysis of data obtained from a sample of male undergraduates revealed that 33 of the 46 items loaded on one major factor. This factor was named Adherence, and was defined as

the desire for "explicit and externally sanctioned structures of rightness (Perry et al., 1968, p. 102)." Adherence correlated negatively with SAT score (Perry et al., 1968), so students who score well on the SAT tend to be less dualistic. The 15 items which loaded highest on Adherence formed the CLEV excerpt used in the present study (Perry et al., 1968, Table 16, pp. 104-105). The items were originally scored on a 6-point Likert-type scale; however, it was believed that sufficient precision would be obtained if a 5-point (strongly agree - strongly disagree) scale was used. Total scores could range from 15 to 75. Higher scores indicate a lower amount of dualism, thus a higher level of intellectual development. Earlier research with the 15-item CLEV excerpt has yielded Cronbach alpha reliability coefficients of .87 for undergraduates (Taube, 1987) and .77 for practicing nurses (Woodham et al., 1986). Perry (Perry et al., 1968) reported split-half reliability coefficients of .66 and .73 for the full 46-item scale.

The Watson-Glaser Critical Thinking Appraisal

The Watson-Glaser (Watson & Glaser, 1980, Form A) provides an objective measure of critical thinking ability. It consists of five 16-item subscales: Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. Total scores can range from 0 to 80. The authors reported split-half reliability coefficients ranging from .69 to .85. Split-half coefficients were on the high side of this range for undergraduate samples, ranging from .80 to .83.

The current version of the Watson-Glaser (Forms A and B) has earned generally favorable reviews. Berger (1985, p. 1693) stated that "this reviewer knows of no similar test that is on a par" with the Watson-Glaser, but expressed concern about the narrow range of content included. Helmstadter (1985, p. 1693) qualified his approval of the Watson-Glaser, describing it as a "good, solid measure of adequate - but not outstanding - reliability." While Helmstadter criticized the lack of evidence of construct validity, he also praised the careful definition of critical thinking, the psychometric methodology used to decide which items were to be included, the face validity of the items, and the availability of parallel forms, as well as the directions and norms. In contrast,

Woehlke (1985, p. 685) criticized the available norms, but recommended the Watson-Glaser as "the best available instrument for measuring critical thinking ability."

The Ennis-Weir Critical Thinking Essay Test

The Ennis-Weir (Ennis & Weir, 1985) is a written test designed to measure argument evaluation. The examinee is presented with a fictional letter to the editor of a newspaper. This letter includes a series of eight arguments in favor of a ban on overnight parking on city streets. Six of the eight arguments are flawed in some way, and the examinee is to respond to each argument in turn, identify each as strong or weak, and to describe any flaws in detail. A scoring rationale is provided, but subjects may also gain or lose points for unanticipated strengths or weaknesses. Scores can range from -1 to +3 for responses to each of the eight numbered arguments, and from -1 to +5 for an overall summary evaluation of the letter. Total scores can thus range from -9 to 29. The authors reported interrater reliability coefficients of .86 and .82.

Some researchers (Keeley & Browne, 1986) have suggested that multiple-choice tests are not valid indicators of critical thinking ability because subjects are not free to apply their own criteria and ask their own questions. Norris (1988) pointed out that objective tests cannot take the knowledge and beliefs possessed by the subjects into account. It is possible for subjects to arrive at a logical conclusion but not receive credit for an item if the author of the test followed a different line of reasoning in developing that item.

Norris (1985, 1986) advocated the use of essay tests of critical thinking, because they provide greater insight into subjects' thought processes and dispositions than do objective tests. In fact, Norris and Ennis (1989, p. 94) described the Ennis-Weir as "the only commercially available test that tests for critical thinking dispositions." While it was a prediction of the present study that the Ennis-Weir would be more influenced by dispositions than would the Watson-Glaser, nothing in the Ennis-Weir test or manual indicates that it was designed to explicitly test for dispositions.

Examples of research using the Ennis-Weir are scarce. Taube (cited in McDaniel, 1991)

reported a positive correlation between the Ennis-Weir and the "Bomb Factories" instrument (McDaniel and Lawrence, cited in McDaniel, 1991), a relatively unstructured written task based on Schroder, Driver, and Streufert's (1967) model of cognitive complexity. Unrau (1992) measured significant gains on the Ennis-Weir among students who were taught to use a thinking strategy in which they thought through the elements of a persuasive essay. Gibbs, Browne, and Keeley (1989) used the Ennis-Weir as a measure of the effectiveness of a critical thinking training program offered to faculty at a midwestern state university.

Poteet (1989, p. 290) described the Ennis-Weir as a "cleverly constructed" task that was "not meant to be considered as a norm-referenced test, but more as an informal assessment technique that lends itself well both to informal assessment and to instruction in the area of critical thinking." Tompkins (1989) criticized the paucity of validity and reliability data in the manual, expressed concern that writing ability and writing anxiety may influence examinees' scores, and suggested that the Ennis-Weir would be better used as an instructional tool than as a test. However, Tompkins stated that:

> The Ennis-Weir has the potential to be a valuable test of critical thinking ability for high school and college students. The authors are to be commended for developing an open-ended and content-specific test that allows students to respond to the arguments presented in the test in a variety of ways. The content of the test minimizes the artificiality of the testing situation, and an interesting and fairly realistic situation is presented in the test. (p. 291)

Additional Data

The subjects were also asked to release their Scholastic Aptitude Test (SAT) scores and grade-point averages. Glaser (1985) reported that it was possible for an examinee to earn a high score on a general mental ability test and a low score on a test of critical thinking, but it was quite rare for the opposite to occur. Thus, while critical thinking and intelligence are not identical, intelligence may be a prerequisite for the development of critical thinking. Grade-point average

was expected to reflect motivation, intelligence, difficulty of course of study (which was expected to be reasonably constant among the students in this sample), and other factors (Cacioppo & Petty, 1982). Subjects could refuse to grant access to these data without penalty. Twenty-six students refused to grant access. An additional 17 students had not taken the SAT, so SAT scores were unavailable for 43 students.

Data Analysis

The primary questions of interest were:

(1) Does a two-factor model provide a plausible description of critical thinking?

(2) Is performance on the Ennis-Weir a product of both critical thinking ability and critical thinking disposition?

Confirmatory factor analysis was used to investigate the proposed two-factor model. Factor analysis may be either exploratory or confirmatory in nature. In exploratory factor analysis, a detailed model relating latent factors to observed variables is not specified in advance (Bollen, 1989), thus, the obtained pattern may result from chance correlations as well as theoretically significant correlations. Exploratory factor analysis identifies a pattern that fits the data, which the researcher must interpret after the fact. No particular hypotheses can be tested (Biber, 1992). All latent factors are assumed to affect all observed variables, all latent factors are assumed to be correlated, and no assumptions of structure are made except for specification of the number of latent factors (Long, 1983).

In contrast, confirmatory factor analysis requires a detailed initial model. The researcher must specify the number of latent factors, which latent factors directly affect which observed variables, and which pairs of latent factors and/or measurement errors, if any, are correlated (Biber, 1992; Bollen, 1989). Statistical tests are performed to verify that the data "confirm" the model. Confirmatory factor analysis can be used in the exploratory sense if competing models are tested to see which provides the best fit with the data (Long, 1983). Thus, confirmatory factor analysis provides directly interpretable results, because the tested model or models are based on

theory and prior research (Biber, 1992). The current research is an example of confirmatory factor analysis, because explicit predictions were made as to which observed variables were directly affected by which latent factors.

LISREL (LInear Structural RELationships), a software package designed to solve for structural models (Joreskog & Sorbom, 1984), was used in this research. The complete LISREL model consists of the structural equation model and two measurement models. The structural equation model defines the relationship between latent independent and latent dependent variables, and the measurement models define the relationship between the latent variables and the observed variables by which they are inferred (Bollen, 1989; Joreskog & Sorbom, 1984). Latent variables are assumed to be unidimensional and free of error. The observed variables are imperfect measures of the latent variables (Bollen, 1989).

Confirmatory factor analysis requires only one measurement model, so that the number of latent and observed variables and the predicted relationships between them can be specified. A structural equation model is unnecessary, because no causal relationship is proposed among the latent variables (Joreskog & Sorbom, 1984). This was the case in the present research, because no causal relationship between the factors of ability and disposition was predicted.

LISREL calculates an estimate of the population covariance matrix Σ (Σ^*) based on estimates of factor loadings, factor intercorrelations, and residual variances; compares this estimated matrix with the obtained covariance matrix S; adjusts the estimated loadings, intercorrelations, and residuals; and recalculates Σ^*. This process is carried out iteratively until the closest approximation of S is obtained. The degree of "fit" of a given model is determined by a comparison of S and Σ^*. If the matrices are sufficiently similar, a given model is said to fit a given data set.

LISREL provides a number of statistics intended to summarize the overall fit of the tested model. Mulaik, James, van Alstine, Bennett, Lind, and Stilwell (1989) define a goodness-of-fit index as one whose value can range from 0 to 1, with 0 indicating a complete lack of fit and 1 indicating perfect fit. Indices in which 0 indicates perfect fit and larger values indicate

progressively poorer fit, such as chi-square, are more properly called lack-of-fit indices. The more similar the observed and implied matrices are, the better the fit. Note that satisfactory fit indices mean only that the model is plausible and has not been disconfirmed. Other models may fit the data better than the tested model (Loehlin, 1987).

Chi-square (X^2) is the most common index of model fit (Mulaik et al., 1989). In contrast to the typical statistical significance test, a well-fitting model will yield a nonsignificant X^2 value. Accepting the null hypothesis amounts to accepting the tested theory, because the tested model is not significantly different from the "perfect" model which would reproduce the obtained covariance matrix perfectly, and differences between the actual and calculated covariance matrices are small enough to be sampling fluctuations. Chi-square is directly proportional to n. Thus, larger samples yield larger X^2 values and are more likely to reach significance, because smaller differences are considered large enough to be meaningful even if the model is "minimally false," that is, the matrix of residuals ($S-\Sigma^*$) contains trivial discrepancies (Bentler & Bonett, 1980; Hayduk, 1987). The model may not be as poor as the significant X^2 would suggest (Mulaik et al., 1989). Therefore, the more conservative test will include a larger sample size.

The ratio of X^2 to its degrees of freedom (X^2/df; Joreskog & Sorbom, cited in Loehlin, 1987) has been suggested as a possible index of fit. If this ratio exceeds 2, improvement is probably possible. If this ratio is decreased to 1 or less by altering the model solely to take advantage of model-specified modifications, the fit may be "too good" because the researcher has capitalized on chance at the expense of theory. Small X^2/df ratios are acceptable if they do not result from data-driven modifications.

The Goodness-of-Fit Index (GFI) represents the proportion of variance and covariance explained by the tested model. Values will generally range between 0 and 1, but negative values are possible. Values closer to 1 indicate better fit (Joreskog & Sorbom, cited in Anderson & Gerbing, 1984). The Root Mean Square Residual (RMR) is calculated by squaring the residuals, finding the mean, and taking the square root of the result (Loehlin, 1987; Marsh, Balla & McDonald, 1988). If model fits the data exactly, the residual matrix will contain all zeroes, and

RMR will equal 0. Greater deviation between the estimated and observed matrices leads to an increased RMR (Rupp & Segal, 1989). RMR must be evaluated relative to the size of elements in the sample covariance matrix (Anderson & Gerbing, 1984).

Chi-square, GFI, and RMR are omnibus tests of model fit. They do not evaluate individual components of a given model. Thus, it is impossible to determine if one or more individual relationships are poorly determined based on a global measure of goodness-of-fit (Joreskog & Sorbom, cited in Bentler & Bonett, 1980). Bollen (1989, p. 281) stated that "nonsense results for individual parameters can occur in conjunction with good overall fit measures, and these would be missed if a researcher only examined the overall fit." Hayduk (1987) suggested that each coefficient in a model be tested for significance.

RESULTS

Descriptive Statistics and Reliability

Numbers of cases, means, standard deviations, and ranges of observed scores for all eight observed variables are reported in Table 1. Cronbach alpha reliability coefficients and standard errors of measurement obtained for the Watson-Glaser, AT-20, NCS, and CLEV are also included in Table 1.

Insert Table 1 about here

Internal consistency is an inappropriate index of reliability for the Ennis-Weir due to the small number of "items" (paragraphs) and the fact that Paragraph 9, at least in part, depends on the other paragraphs (Norris & Ennis, 1989). An interrater reliability coefficient of .83 and a standard error of measurement of 2.52 were obtained on a sample of 36 (20% of total) subjects.

Correlations Among Observed Variables

Intercorrelations of the eight observed variables are reported in Table 2. All correlations except for Watson-Glaser/AT-20, AT-20/SAT-Math, and AT-20/GPA were significant ($p < .05$). Each correlation coefficient was calculated on all cases in which values existed for both variables, thus the number of cases included in each calculation ranged from 147 to 194.

Insert Table 2 about here

It was postulated that Ennis-Weir would load on both the ability and disposition factors, while Watson-Glaser would load on the ability factor only. Therefore, it was expected that the correlations between the three dispositional measures (NCS, AT-20, CLEV) and Ennis-Weir would exceed the corresponding correlations between the dispositional measures and Watson-Glaser. The correlations between the dispositional measures and Ennis-Weir were larger than those involving Watson-Glaser; however, this difference was significant only in the case of AT-20 (Fisher's $Z = 2.58$).

Evaluation of Models

The covariance matrix of the eight observed variables was computed by LISREL, using the observed standard deviations of and correlations between each variable. The standard deviations of SAT-Verbal and SAT-Math were divided by 10 so that the scale of these variables would more closely approximate the scale of the other variables.

It was hypothesized that critical thinking would be best described by a two-factor (ability and disposition) model. It was expected that a model including only a single latent "critical thinking" factor would not reproduce the obtained covariance matrix as accurately as a two-factor model would. Thus, it was necessary to generate and test a single-factor model to provide a

comparison with the two-factor model. The coefficients and fit indices obtained for the single-factor model are presented in Table 3. Note that TD (Theta Delta) corresponds to the residual variance for a given variable, R^2 indicates the percentage of that variable's variance accounted for by the model, GFI represents the Goodness-of-Fit Index, and RMR represents the Root Mean Square Residual. Loadings and residuals are expressed in raw score units.

Insert Table 3 about here

The single-factor model accounts for 37% (GFI = .869) of the variance and covariance of the observed variables. While this model is itself plausible, the high X^2/df ratio indicates that considerable improvement is possible.

The two latent factors included in the two-factor model have been conceptualized as ability and disposition. Watson-Glaser, SAT-Verbal, and SAT-Math were loaded on Factor 1 (ability); AT-20, NCS, and CLEV were loaded on Factor 2 (disposition); and Ennis-Weir and GPA were cross-loaded on both Factors 1 and 2. All other loadings were set equal to zero. The coefficients and fit indices for the two-factor model are presented in Table 4.

Insert Table 4 about here

The two-factor model provided a much better fit with the data than did the single-factor model. Therefore, a conception of critical thinking as a two-factor system appears more plausible than a conception which includes only a single "critical thinking" factor. GFI increased from .869 to .955; RMR decreased by 63%; and X^2/df decreased by 55%. X^2 was significant ($p = .003$ at n

= 194, p = .042 at n = 147), but this is understandable due to the relatively large sample size. The correlation between the two factors (Phi) was equal to .585.

Bentler & Bonett (1980) stated that the difference in fit between two models can be tested by dividing the difference in X^2 by the difference in df. The difference between two X^2s is a X^2, and is evaluated against df equal to the difference in df between the two models. This test was highly significant ($p < .001$), illustrating that the two-factor model provided a significant increase in fit compared to the single-factor model.

Table 5 presents the factor loadings and residuals obtained for the two-factor model expressed in standard deviation units rather than raw-score units. The relative magnitude of the loadings is more apparent, as the variability of each observed variable is taken into account. SAT-Verbal had the highest standardized loading on Factor 1, and CLEV had the highest standardized loading on Factor 2.

Insert Table 5 about here

Hayduk (1987) stated that each coefficient in a model should be tested for significance. Investigation of the coefficients obtained on the two-factor model revealed that the loading of GPA on Factor 2 (disposition) was nonsignificant. This was surprising, because grades were expected to reflect motivation as well as ability (Cacioppo & Petty, 1982). All other loadings were significant ($p < .05$). The two-factor model was modified by setting the loading of GPA on Factor 2 equal to zero. The differences in overall fit indices between the modified model and the original two-factor model were trivial. X^2/df decreased by 5%; GFI was unchanged; and RMR increased by 2%. X^2 remained significant (p = .005 at n = 194, p = .058 at n = 147). The difference in X^2 between the two models was not significant ($p > .80$). In addition, the factor loadings obtained from the modified and original two-factor models were nearly identical. Thus,

there is little to be gained by model respecification. The coefficients and fit indices for the modified two-factor model are presented in Table 6.

Insert Table 6 about here

The matrix of residual covariances ($S-\Sigma^*$) is an important source of information about the overall quality of a model (Hayduk, 1987). Large residuals may pinpoint a model's deficiencies. Dividing a residual by the square root of its asymptotic variance yields a standardized residual, which defines the size of the residual in standard deviation units. If residuals were truly uncorrelated, only 4-5% of standardized residual covariances (one or two of 28) would exceed +/-2.0. Twelve standardized residuals exceeded +/- 2.0 in the single-factor model, compared to five in the original two-factor model and four in the modified two-factor model. All four of these residuals involved the AT-20 scale. It is likely that the relatively poor reliability of the AT-20 contributed to the large residuals associated with this scale. Taken together, the residuals indicate a fairly well-fitting model.

The large range of sample sizes (147 to 194) resulting from pairwise deletion of cases in calculation of the correlation and covariance matrices raised concerns that those cases in which certain data were missing may have differed systematically from those cases in which all data were available. The three models outlined previously were recalculated on samples for which values existed for all variables ($\underline{n}$ = 137), and for which access to SAT scores and grade-point averages was granted, even though 17 of these students had not taken the SAT ($\underline{n}$ = 172). The loadings, residuals, and fit indices obtained in both cases were nearly identical to those described previously, thus this issue did not appear to pose a problem.

DISCUSSION

A two-factor model of critical thinking consisting of the separate factors of ability and disposition was investigated. Need for cognition, tolerance of ambiguity, and intellectual development represented the disposition factor, and SAT scores, grade-point average, and performance on the Watson-Glaser Critical Thinking Appraisal represented the ability factor. Performance on the Ennis-Weir Critical Thinking Essay Test was expected to result from both factors. Confirmatory factor analysis confirmed the plausibility of this two-factor model, and the Ennis-Weir loaded significantly on both the ability and disposition factors. The two-factor model provided a more accurate fit with the data than did a model including only one latent factor, thus a conception of critical thinking as a two-factor system appears more plausible than a single-factor conception of critical thinking.

The Ennis-Weir was used in the present study because it provides a "realistic" assessment of everyday critical thinking behavior (Norris & Ennis, 1989; Tompkins, 1989). The Ennis-Weir is less constrained than objective measures such as the Watson-Glaser, thus disposition was expected to affect performance to a greater degree. However, the format of the Ennis-Weir imposes a fairly rigid structure on the examinees' responses. It is possible that disposition may affect performance on a less-structured written measure of critical thinking to an even greater extent. Two such written measures are the Bomb Factories instrument (McDaniel & Lawrence, cited in McDaniel, 1991) described previously, and the Illinois Critical Thinking Essay Test (Ennis & Finken, 1993).

It is noteworthy that only two of the eight measures included in the present study (Watson-Glaser and Ennis-Weir) measured critical thinking explicitly. The ability and disposition factors identified here may actually be general intellectual ability and disposition applied in a critical thinking situation, rather than ability and disposition specific to critical thinking. Indeed, critical thinking has been characterized as an instance of everyday, informal reasoning (Galotti, 1989). The standardized factor loadings (loadings expressed in standard deviation units) of SAT-

Verbal and SAT-Math exceeded that of Watson-Glaser on the ability factor of the two-factor model. The Ennis-Weir's loading on the ability factor was more than twice its loading on the disposition factor, thus the Ennis-Weir may also be a measure of general intellectual ability.

The SAT was designed to measure generalized reasoning and thinking abilities. Cameron (1989, p. 2) states that "society demands citizens who can read with depth and comprehension, who can reason and discern complex relationships, and who can think critically. The SAT measures these skills." Wilson and Wagner (1981) noted that performance in an accelerated physics course designed to stress critical thinking correlated higher with combined SAT score than with total score on the Watson-Glaser. Although the difference between these correlations was not significant (a fact not mentioned in Wilson & Wagner's article), it is significant that the SAT performed at least as well as a test designed to measure critical thinking in predicting performance in a critical thinking course.

Similarly, the Watson-Glaser correlates positively with measures of achievement and general mental ability such as the SAT, the Otis-Lennon Mental Abilities Test, and the Stanford Achievement Test (Watson & Glaser, 1980). The authors caution that the overlap between intelligence and critical thinking is not complete. Factor analysis (Landis, cited in Watson & Glaser, 1980) indicates that the Watson-Glaser reflects a dimension of intellectual functioning separate from other traditional measures of intelligence.

Facione (1990) reported significant correlations between the California Critical Thinking Skills Test (CCTST) and SAT Verbal ($r = .55$) and Math ($r = .44$) scores, and with college GPA ($r = .20$). These correlations are similar to those obtained between SAT-Verbal, SAT-Math, GPA, and Watson-Glaser in the present study. A combination of SAT-Verbal, SAT-Math, college GPA, and high-school GPA accounted for 41% of the variance in CCTST scores. Unlike Watson and Glaser (1980), Facione treats these findings as evidence of the concurrent validity of the CCTST.

Facione and Facione (1992) developed the California Critical Thinking Dispositions Inventory (CCTDI) as a measure of critical thinking dispositions. The CCTDI consists of seven

subscales: Truth-Seeking, or the disposition to seek the truth, ask questions, and be honest and objective about pursuing inquiry even if the findings contradict one's interests or opinions; Open-Mindedness, or the tendency to be tolerant of divergent views and sensitive to one's own bias; Analyticity, defined as being alert to potentially problematic situations, anticipating possible results and consequences, and prizing the use of reason or evidence; Systematicity, or the tendency to be organized, orderly, focused, and diligent in inquiry; Self-Confidence, or trust in one's own reasoning processes; Inquisitiveness, or intellectual curiosity; and Maturity, or the disposition to make reflective judgments. Preliminary research has revealed a positive correlation ($r = .67$) between total score on the CCTDI and score on the CCTST (Facione, Facione, & Sanchez, 1994). This correlation is comparable to the intercorrelation between the ability and disposition factors obtained in the present study.

Given the strong correlation observed between ability and disposition in the present study and by Facione, Facione, and Sanchez (1994), it is plausible that one is a cause of the other. According to motivational theory (Lewin, cited in Facione, Facione, & Sanchez, 1994, p. 346), "the disposition to value and use critical thinking would impel an individual to achieve mastery over critical thinking skills, being motivated to close the gap between what is valued and what is attained." The individual will practice these skills, eventually allowing him/her to perform them. Anastasi (1982) postulated a reciprocal relationship between personality and ability factors. Improved performance leads to a more positive self-concept, which leads to greater effort, which leads to a greater improvement in performance, and so on.

In the present study, performance on the Ennis-Weir was a function of both ability and disposition. Therefore, it appears necessary that attempts to foster critical thinking address dispositions and attitudes as well as abilities. Indeed, a number of researchers treat development of critical thinking dispositions as the key issue in fostering critical thinking. Hudgins, Riesenmy, Ebel, and Edelman (1989, p. 329) maintained that "the end goal for teaching children to become critical thinkers is the development of a disposition to do so." Facione, Facione, and Sanchez (1994) stated that:

> Educating good critical thinkers is more than developing critical thinking skills. A complete approach to developing good critical thinkers includes nurturing the disposition toward critical thinking, an effort...integral to insuring the use of critical thinking skills outside the narrow instructional setting. (p. 346)

Nicholls' (1983, 1984) concept of task-orientation is analogous to critical thinking disposition as defined in the present study. The goal of a task-oriented student is to increase understanding or mastery, thus, learning or knowledge is an end in itself. In contrast, the goal of an ego-oriented student is to outperform others or avoid appearing incompetent (Nicholls, cited in Kroll, 1988). Exploratory research (Taube, 1989) revealed that a scale designed to measure task-orientation (Nicholls, Patashnick, & Nolen, 1985) correlated positively with the NCS. Nonsignificant positive correlations were also noted between this scale and Watson-Glaser, AT-20, and CLEV. Kroll (1988) reported positive correlations between this scale and measures of thoughtfulness and open-mindedness, while a similar scale measuring ego-orientation (Nicholls, Patashnick, & Nolen; cited in Kroll, 1988) correlated negatively with the AT-20, as well as measures of complexity, thoughtfulness, and individualism.

By adolescence, the conception of ability as capacity rather than effort has been attained by most students (Nicholls, 1984). Thus, ego-oriented students may refuse assistance or refrain from seeking assistance, because this may be seen as lack of ability. Therefore, classrooms which are ego-oriented maximize the inequality of motivation, resulting in unequal development of intellectual potential. Task-orientation and ego-orientation are also evident in the tasks chosen by students when given the opportunity to choose a task (Nicholls, Patashnick, & Nolen, 1985). An ego-oriented student defines task difficulty in terms of peer performance. An ego-oriented student will choose either a very easy task, on which success is nearly certain, or a very difficult task, because failure, while likely, is not seen as damaging. These tasks are not beneficial to learning. In contrast, a task-oriented student will choose challenging tasks where neither success or failure is certain.

Task-orientation is ideal from the educational perspective, because it should lead to

greater effort, which, in turn, should lead to more learning and greater ability (Nicholls, Patashnick, & Nolen, 1985). One's feelings of competence are directly related to learning. Nicholls (1984, p. 64) stated that "classrooms that sustain task-involvement are more likely to maintain equal and optimum motivation for intellectual development, particularly the development of logical thought." Nicholls (1983) also noted that task-oriented students tend to be more creative, critical, and inquisitive, and less concerned with learning content quickly. Teachers should try to sustain task-orientation in their classrooms by deemphasizing rewards, coercion, and comparison and competition among students, and encouraging student input and participation (Nicholls, 1983). Regrettably, the general trend is in the opposite direction, as students are typically compared with others and praised for effort or improvement (Nicholls, 1984). Students who do not receive praise when they see others receiving praise may view themselves as unable (Nicholls, 1983).

REFERENCES

Anastasi, A. (1982). Psychological testing, fifth edition. New York, NY: MacMillan.

Anderson, J.C. & Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49, 155-173.

Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Berger, A. (1985). Review of the Watson-Glaser Critical Thinking Appraisal, Forms A and B. In Ninth mental measurement yearbook (pp. 1692-1693), Lincoln, NE: Buros Institute of Mental Measurement.

Beyer, B.K. (1985). Critical thinking: what is it? Social Education, 270-276.

Biber, D. (1992). On the complexity of discourse complexity: a multidimensional analysis. Discourse Processes, 15, 133-163.

Bollen, K.A. (1989). Structural equations with latent variables. New York, NY: John Wiley and Sons.

Browne, M.N., Haas, P.F., & Keeley, S. (1978). Measuring critical thinking skills in college. Educational Forum, 219-226.

Cacioppo, J.T. & Petty, R.E. (1982). The need for cognition. Journal of Personality and Social Psychology, 42, 116-131.

Cacioppo, J.T., Petty, R.E., & Morris, K.J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. Journal of Personality and Social Psychology, 45, 805-818.

Cameron, R.G. (1989). The common yardstick: a case for the SAT. New York, NY: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 315 438)

DeNitto, J. & Strickland, J. (1987). Critical thinking: a skill for all seasons. College Student Journal, 21, 201-204.

Dick, R.D. (1991). An empirical taxonomy of critical thinking. Journal of Instructional Psychology, 18, 79-92.

Dressel, P.L. & Mayhew, L.B. (1954). General education: explorations in evaluation. Washington, DC: American Council on Education.

Ennis, R.H. (1962). A concept of critical thinking. Harvard Educational Review, 32, 81-111.

Ennis, R.H. (1985). A logical basis for measuring critical thinking skills. Educational Leadership, 43, 44-48.

Ennis, R.H. (1991). Critical thinking: a streamlined conception. Paper presented at the American Educational Research Association national convention, Chicago, IL.

Ennis, R.H. & Finken, M.M. (1993). A pilot study of essential school assessment. Champaign, IL: Illinois Alliance of Essential Schools and Illinois Critical Thinking Project.

Ennis, R.H. & Weir, E. (1985). The Ennis-Weir Critical Thinking Essay Test. Pacific Grove, CA: Midwest Publications.

Facione, N.C., Facione, P.A., & Sanchez, C.A. (1994). Using critical thinking dispositions to predict competent clinical judgment: the development of the California Critical Thinking Dispositions Inventory. Journal of Nursing Education, 33, 345-350.

Facione, P.A. (1991). Critical thinking: what it is and why it counts. Millbrae, CA: California Academic Press.

Facione, P.A. (1990). The California Critical Thinking Skills Test: College Level. Technical report #2 - factors predictive of CT skills. Millbrae, CA: California Academic Press. (ERIC Document Reproduction Service No. ED 327 550)

Facione, P.A. & Facione, N.C. (1992). The California Critical Thinking Dispositions Inventory. Millbrae, CA: California Academic Press.

Frisch, N.A. (1987). Cognitive maturity of nursing students. Image: The Journal of Nursing Scholarship, 19, 25-27.

Galotti, K.M. (1989). Approaches to studying formal and everyday reasoning. Psychological Bulletin, 105, 331-351.

Gibbs, L.E., Browne, M.N., & Keeley, S.M. (1989). Critical thinking: a study's outcome. Journal of Professional Studies, 13, 44-59.

Glaser, E.M. (1985). Critical thinking: educating for responsible citizenship in a democracy. National Forum: The Phi Kappa Phi Journal, 65, 24-27.

Halpern, D.F. (1989). Thought and knowledge: an introduction to critical thinking. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hayduk, L.A. (1987). Structural equation modeling with LISREL: essentials and advances. Baltimore, MD: Johns Hopkins University Press.

Heesacker, M. (1985). Need For Cognition Scale. In Keyser, D.J. & Sweetland, R.C., Test critiques, volume III (pp. 466-474), Kansas City, MO: Test Corporation of America.

Helmstadter, G.C. (1985). Review of the Watson-Glaser Critical Thinking Appraisal, Forms A and B. In Ninth mental measurement yearbook (pp. 1693-1694), Lincoln, NE: Buros Institute of Mental Measurement.

Hudgins, B.B., Riesenmy, M., Ebel, D., & Edelman, S. (1989). Children's critical thinking: a model for its analysis and two examples. Journal of Educational Research, 82, 327-338.

Joreskog, K.G. & Sorbom, D. (1984). LISREL VI: analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods. Mooresville, IN: Scientific Software.

Keeley, S.M. & Browne, M.N. (1986). How college seniors operationalize critical thinking behavior. College Student Journal, 20, 389-395.

Kirton, M.J. (1981). A reanalysis of two scales of tolerance of ambiguity. Journal of Personality Assessment, 45, 407-414.

Kroll, M.D. (1988). Motivational orientations, views about the purpose of education, and intellectual styles. Psychology in the Schools, 25, 338-343.

Lipman, M. (1988). Critical thinking - what can it be? Educational Leadership, 46, 38-43.

Loehlin, J.C. (1987). Latent variable models: an introduction to factor, path, and structural analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.

Long, J.S. (1983). Confirmatory factor analysis: a preface to LISREL. Beverly Hills, CA: Sage.

MacDonald, A.P., Jr. (1970). Revised scale for ambiguity tolerance: reliability and validity. Psychological Reports, 26, 791-798.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indices in confirmatory factor analysis: the effect of sample size. Psychological Bulletin, 103, 391-410.

McDaniel, E. (1991). Levels of cognitive complexity: a framework for the measurement of thinking. Paper presented at the American Educational Research Association national convention, Chicago, IL. (ERIC Document Reproduction Service No. ED 331 868)

Moore, W.S. (1991). The Perry scheme of intellectual and ethical development: an introduction to the model and major assessment approaches. Paper presented at the American Educational Research Association national convention, Chicago, IL.

Mulaik, S.A., James, L.R., van Alstine, J., Bennett, N., Lind, S., & Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Nicholls, J.G. (1984). Conceptions of ability and achievement motivation. In R. Ames & C. Ames (eds.), Research on Motivation in Education: Student Motivation (pp. 39-73), San Diego, CA: Academic Press.

Nicholls, J.G. (1983). Conceptions of ability and achievement motivation: a theory and its implications for education. In S.G. Paris, G.M. Olson, & H.W. Stevenson (eds.), Learning and Motivation in the Classroom (pp. 211-237), Hillsdale, NJ: Lawrence Erlbaum Associates.

Nicholls, J.G., Patashnick, M., & Nolen, S.B. (1985). Adolescents' theories of education. Journal of Educational Psychology, 77, 683-692.

Norris, S.P. (1988). Controlling for background beliefs when developing multiple-choice critical thinking tests. Educational Measurement: Issues and Practice, 7, 5-11.

Norris, S.P. (1986). Evaluating critical thinking ability. History and Social Science Teacher, 21, 135-146.

Norris, S.P. (1985). Synthesis of research on critical thinking. Educational Leadership, 42, 40-45.

Norris, S.P. & Ennis, R.H. (1989). Evaluating critical thinking. Pacific Grove, CA: Midwest Publications.

Olson, K., Camp, C., & Fuller, D. (1984). Curiosity and need for cognition. Psychological Reports, 54, 71-74.

Osberg, T.M. (1987). The convergent and discriminant validity of the Need For Cognition Scale. Journal of Personality Assessment, 51, 441-450.

Paul, R.W. & Nosich, G.M. (1991). A proposal for the national assessment of higher-order thinking at the community college, college, and university levels. Washington, DC: National Center for Educational Statistics. (ERIC Document Reproduction Service No. ED 340 762)

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

Perry, W.G., Jr. (1970). Forms of intellectual and ethical development in the college years: a scheme. New York, NY: Holt, Rinehart, & Winston.

Perry, W.G., Jr., Sprinthall, N.A., Wideman, J.W., & Jones, F.J. (1968). Patterns of development in thought and values of students in a liberal arts college...a validation of a scheme. Washington, DC: U.S. Department of Health, Education, & Welfare, Office of Education, Bureau of Research. (ERIC Document Reproduction Service No. ED 024 315)

Peters, W.H. & Amburgey, B.S. (1982). Teacher intellectual disposition and cognitive classroom verbal reactions. Journal of Educational Research, 76, 94-99.

Poteet, J.A. (1989). Review of the Ennis-Weir Critical Thinking Essay Test. In Tenth mental measurement yearbook (pp. 289-290), Lincoln, NE: Buros Institute of Mental Measurement.

Powell, S. (1987). Improving critical thinking: a review. Educational Psychology, 7, 169-185.

Rotter, N.G. & O'Connell, A.N. (1982). The relationships among sex-role orientation, cognitive complexity, and tolerance for ambiguity. Sex Roles, 8, 1209-1220.

Rupp, M.T. & Segal, R. (1989). Confirmatory factor analysis of a professionalism scale in pharmacy. Journal of Social and Administrative Pharmacy, 6, 31-38.

Schroder, H.M., Driver, M.J., & Streufert, S. (1967). Human information processing: individuals and groups functioning in complex social situations. New York, NY: Holt, Rinehart, & Winston.

Siegel, H. (1988). Educating reason: rationality, critical thinking, and education. New York, NY: Routledge.

Simpson, D.E., Dalgaard, K.A., & O'Brien, D.K. (1986). Student and faculty assumptions about the nature of uncertainty in medicine and medical education. Journal of Family Practice, 23, 468-472.

Taube, K.T. (1987). [Checklist of Educational Views: Reliability estimates and correlations with AT-20 and measures of open-mindedness and cognitive complexity]. Unpublished raw data.

Taube, K.T. (1989). [Task-Orientation Scale: Correlations with Watson-Glaser Critical Thinking Appraisal, Need for Cognition Scale, AT-20, and Checklist of Educational Views]. Unpublished raw data.

Tegano, D.W. (1990). Relationship of tolerance of ambiguity and playfulness to creativity. Psychological Reports, 66, 1047-1056.

Tolentino, E., Curry, L., & Leak, G. (1990). Further validation of the short form of the Need For Cognition Scale. Psychological Reports, 66, 321-322.

Tompkins, G.E. (1989). Review of the Ennis-Weir Critical Thinking Essay Test. In Tenth mental measurement yearbook (pp. 290-291), Lincoln, NE: Buros Institute of Mental Measurement.

Unrau, N.J. (1992). The TASK of reading (and writing) arguments: a guide to building critical literacy. Journal of Reading, 35, 436-442.

Waters, L.K. & Zakrajsek, T. (1990). Correlates of need for cognition total and subscale scores. Educational and Psychological Measurement, 50, 213-217.

Watson, G. & Glaser, E.M. (1980). Watson-Glaser Critical Thinking Appraisal, Form A. Cleveland, OH: Psychological Corporation.

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. American Psychologist, 5, 78-83.

Wechsler, D. (1943). Non-intellective factors in general intelligence. Journal of Abnormal and Social Psychology, 38, 101-103.

Welfel, E.R. (1982). The development of reflective judgment: implications for career counseling of college students. Personnel and Guidance Journal, 60, 17-21.

Welfel, E.R. & Davison, M.L. (1986). The development of reflective judgment during the college years: a 4-year longitudinal study. Journal of College Student Personnel, 27, 209-216.

Wilson, D.G. & Wagner, E.E. (1981). The Watson-Glaser Critical Thinking Appraisal as a predictor of performance in a critical thinking course. Educational and Psychological Measurement, 41, 1319-1322.

Woehlke, P.L. (1985). Watson-Glaser Critical Thinking Appraisal. In Keyser, D.J. & Sweetland, R.C., Test critiques, volume III (pp. 682-685), Kansas City, MO: Test Corporation of America.

Woodham, R., Taube, K.T., Elmore, J., Linehan, M., & Wagner, D. (1986). Earn/learn mobility track: LPN-RN socialization. Washington, DC: U.S. Department of Health & Human Services, Division of Nursing, Special Project Grant 1D10 NU 25164.

Table 1

Descriptive Statistics and Reliability Estimates

Variable	N	Mean	S.D.	Range	Alpha	SE_{meas}
Watson-Glaser	189	54.89	8.08	37-76	.78	3.79
Ennis-Weir	187	14.63	6.14	-3-27		
AT-20	190	8.69	3.20	2-16	.62	1.98
NCS	194	115.08	16.52	71-159	.90	5.21
CLEV	194	50.08	7.55	30-70	.79	3.48
SAT-Verbal	155	431.48	75.29	240-620		
SAT-Math	155	495.48	91.46	250-740		
GPA	171	2.81	.51	1.68-4.00		

Table 2

Intercorrelations of Observed Variables

	W-G	E-W	AT20	NCS	CLEV	SATV	SATM	GPA
Watson-Glaser	---	.37	.07	.15	.33	.43	.39	.30
Ennis-Weir		---	.33	.24	.35	.40	.28	.28
AT-20			---	.45	.47	.22	.09	.00
NCS				---	.51	.26	.22	.16
CLEV					---	.43	.32	.34
SAT-Verbal						---	.50	.40
SAT-Math							---	.34
GPA								---

Table 3

Summary of Single-Factor Model

Variable	Loadings F1	TD	R^2
Watson-Glaser	4.331	46.515	.287
Ennis-Weir	3.431	25.605	.315
AT-20	1.358	8.371	.180
NCS	8.265	204.435	.250
CLEV	5.243	29.523	.482
SAT-Verbal	5.228	29.355	.482
SAT-Math	5.154	57.089	.318
GPA	.255	.204	.241

Note. GFI = .869, RMR = 5.895.

X^2 (n = 194) = 98.69, 20 df, p = .000. X^2/df = 4.93.
X^2 (n = 147) = 74.66, 20 df, p = .000. X^2/df = 3.73.

Table 4

Summary of Two-Factor Model

Variable	Loadings F1	F2	TD	R^2
Watson-Glaser	4.730		42.895	.343
Ennis-Weir	2.492	1.191	26.275	.297
AT-20		1.856	6.769	.337
NCS		10.342	165.801	.392
CLEV		6.332	16.924	.703
SAT-Verbal	5.772		23.367	.588
SAT-Math	5.782		50.219	.400
GPA	.264	.013	.196	.274

Note. GFI = .955, RMR = 2.202, Phi = .585.

X^2 (n = 194) = 37.38, 17 df, p = .003. X^2/df = 2.20.
X^2 (n = 147) = 28.28, 17 df, p = .042. X^2/df = 1.66.

Table 5

Two-Factor Model Expressed In Standardized Form

Variable	Loadings F1	F2	TD
Watson-Glaser	.586		.811
Ennis-Weir	.408	.195	.838
AT-20		.581	.814
NCS		.626	.780
CLEV		.839	.545
SAT-Verbal	.767		.642
SAT-Math	.632		.775
GPA	.508	.025	.852

Table 6

Summary of Modified Two-Factor Model

Variable	Loadings F1	F2	TD	R^2
Watson-Glaser	4.726		42.939	.342
Ennis-Weir	2.499	1.184	26.251	.298
AT-20		1.865	6.737	.340
NCS		10.373	165.156	.394
CLEV		6.308	17.224	.698
SAT-Verbal	5.764		23.467	.586
SAT-Math	5.775		50.299	.399
GPA	.273		.195	.277

Note. GFI = .955, RMR = 2.240, Phi = .589.

X^2 (n = 194) = 37.43, 18 df, p = .005. X^2/df = 2.08.
X^2 (n = 147) = 28.32, 18 df, p = .057. X^2/df = 1.57.