8-1985

Assessing Critical Thinking Processes in the Gifted: Predicting GRE Analytical Performance from Watson-Glaser Results

Susie Garrott
Western Kentucky University

Garrott,

Susie Newton

1985

ASSESSING CRITICAL THINKING PROCESSES IN THE GIFTED: PREDICTING GRE ANALYTICAL PERFORMANCE FROM WATSON-GLASER RESULTS

A Thesis

Presented to

the Faculty of the Department of Psychology

Western Kentucky University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Susie Newton Garrott

August 1985

AUTHORIZATION FOR USE OF THESIS

Permission is hereby

☒ granted to the Western Kentucky University Library to make, or allow to be made photocopies, microfilm or other copies of this thesis for appropriate research or scholarly purposes.

☐ reserved to the author for the making of any copies of this thesis except for brief sections for research or scholarly purposes.

Signed Susie N. Garrett

Date September 19, 1985

Please place an "X" in the appropriate box.

This form will be filed with the original of the thesis and will control future use of the thesis.

ASSESSING CRITICAL THINKING PROCESSES IN THE GIFTED:

PREDICTING GRE ANALYTICAL PERFORMANCE

FROM WATSON-GLASER RESULTS

Recommended August 23, 1985
Date

Richard Mills
Director of Thesis

Doris L Redfield

William R Johnl

Approved 9-19-85

Elmer Gray
Dean of the Graduate College

ASSESSING CRITICAL THINKING PROCESSES IN THE GIFTED: PREDICTING GRE ANALYTICAL PERFORMANCE FROM WATSON-GLASER RESULTS

Susie Newton Garrott August 1985 pages 45

Directed by: Richard L Miller, Doris L. Redfield, and William F. Pfohl, Jr.

Department of Psychology Western Kentucky University

The purpose of this study was to examine the relationship between the constructs measured by the Watson-Glaser Critical Thinking Appraisal and the Graduate Record Examination Analytical Scale in "gifted" adolescents 15 to 17 years of age. There were three hypotheses addressed in this study: 1) "gifted" adolescents would be able to think critically as measured by college level norms when measured by the Watson-Glaser and the GRE-Analytical; 2) significant differences would exist between different levels of gifted populations; and 3) a factor or group of factors of the Watson-Glaser subscales would significantly predict performance on the GRE-Analytical Scale.

The Watson-Glaser and the GRE-Analytical were administered to 104 high school students, most of whom were seniors in high school or in the summer preceeding their senior year. The subject pool labeled "national gifted" consisted of 50 students in summer programs for academically gifted sudents at the University of Indiana

and Western Kentucky University. The "local gifted" group consisted of 54 students in secondary schools in western Kentucky and northern middle Tennessee.

Analysis of Watson-Glaser total scores indicated that the national group mean was at the 60th percentile and the local group mean was at the 30th percentile when compared to college senior women. GRE-Analytical total raw scores converted to scaled scores were 580 for the national group and 440 for the local group. Stastical analysis confirmed the significantly superior performance of the national group over the local group on both instruments.

Pearson product-moment correlation coefficients were calculated to examine the relationship between the GRE-Analytical and the Watson-Glaser for both national and local samples. The relationship between the performance on the two instruments was highly significant for both groups.

The local group data were subjected to stepwise regression analysis to determine which individual subscale or group of subscales best predicted GRE-Analytical performane. In the local sample, Subscale 4 clearly emerged as the best single predictor.

Pearson product-moment correlation coefficients based on a median-split of data from each test indicated that lower half total (national plus local) and local group scores were slightly more consistent than were the upper

halves of these groups.

Implications of these results for expanding the cognitive processes and motivating the gifted student were discussed.

CHAPTER I

Introduction

At present there is interest among educators and researchers in the critical thinking ability of students, particularly those regarded as "gifted." Students in "gifted" programs are frequently described as reasoning at high levels; however, this has not been not been empirically investigated, especially in adolescents. Verification of the high level reasoning structure used by gifted adolescents could have important implications in designing curriculum for this population.

Additionally, the term critical thinkng is not clearly defined and does not have a single accepted definition. There are, however, some dimensions of critical thinking that have agreement among professionals. Ennis (1962), Watson and Glaser (1964), and Dressler and Mayhew (1954) agree that critical thinking includes the following characteristics: 1) identifying the problem, 2) selecting pertinent information for solving the problem, 3) recognizing stated and unstated assumptions, 4) formulating and selecting relevant hypotheses, and 5) recognizing the validity of inferences and conclusions.

Just as there are differences in defining critical thinking, there are diverse theories and methodologies

employed to examine this elusive concept. Among them are the theories and methodologies of Das, Meeker, Sternberg, and Piaget. Das (1973) approached thought processes as simultaneous and successive syntheses, which are governed by cortical functions. Meeker (1968) has developed Guilford's Structure of Intellect model (Guilford, 1967), a three dimensional classification of 120 distinct types of intellectual abilities, into a diagnostic and remediation program to help students learn. While valid in their approaches, these theorists do not address the hierarchal nature of cognitive structures which a person must use to critically analyze and solve increasingly more difficult problems. For this study the theoretical frameworks of Sternberg and Piaget have been selected as most relevant because of the hierarchal nature of their theories and their application to advanced thought processes.

In his theory of componential analysis of information-processing, Sternberg (1979) proposed a continuium extending from perception in very simple identification tasks to problem solving in very complex inferential tasks. He has shown that the strategies used in solving advanced analogies are those strategies which Piaget ascribed to the formal thought stage of development. Sternberg has concluded that strategies or processes continue to change within the stage of formal thought and possibly beyond

(Sternberg & Downing, 1982).

Formal thought is the most advanced level, or stage, of Piaget's (1954) developmental theory of intelligence. In formal thought, the subject is able to 1) consider all possible sollutions to a problem, 2) manipulate propositions about data, and 3) generate all possible combinations of variables inherent in a problem. Piaget does not distinguish between levels within formal thought, whereas Sternberg does (Sternberg & Downing, 1982). The thought processes required for critical thinking seem to be those of formal thought and higher level information processing.

To evaluate gifted students' levels of critical thinking, an appropriate test instrument must be selected which validly relates to the critical thinking process as previously defined. Of those instruments currently availabale, the Cornell Critical Thinking Text, Level X, the Cornell Critical Thinking Test, Level Z, and the Watson-Glaser Critical Thinking Appraisal are comprehensive tests of critical thinking ability. Of these three, the Watson-Glaser, a widely used instrument with five subfactor scales, was selected for this study because of its appropriate focus on the critical thinking process and the well documented reliability and validity of its subscales. Although a new instrument with largely undetermined inferences, the Graduate Record Examination Analytical

Scale was chosen as an additional instrument. The research saff at ETS describe the characteristics of this scale as requiring critical thinking processes. Additionally, it will be encountered by the majority of gifted students as they apply to graduate programs to continue their education. These tests were used to explore the relationship between the Watson-Glaser and the GRE-Analytical for adolescent populations labeled "gifted." These populations were expected to reflect those higher order mental processes required in critical thinking

Two potentially distinct "gifted" adolescent populations were available for the study, one group labeled "national gifted" and the other "local gifted." "Gifted" education has been mandated by many states. Determining the level at which students in these programs think and the processes they use in reaching conclusions could have implications for curriculum development. "Gifted" adolescsents are in the formal thought stage of Piaget's theory and can be expected to perform at an advanced level of information processing according to Sternberg's theory of componential analysis.

The purpose of this study was to examine the relationship between the constructs measured by the Watson-Glaser Critical Thinking Appraisal and the Graduate Record Examination Analytical Scale in "gifted" adolescents. It was expected that "gifted" adolescents

will be able to reason critically as measured by college level norms as measured by the Watson-Glaser and the GRE-Analtical and that there would be significant differences between different levels of gifted students. Additionally, it was expected that a factor or group of factors of the Watson-Glaser subscales would significantly predict performance on the GRE-Analytical Scale.

CHAPTER II

Review of the Literature

Definitions of Critical Thinking

The most comprehensive concept of critical thinking is that of Robert Ennis (1962), who defines critical thinking as the "correct assessment of statements" (p. 15). Ennis's (1962) "Concept of Critical Thinking" includes twelve aspects of critical thinking which are not mutually exclusive. Critical thinking includes 1) grasping the meaning of a statement; 2) judging whether there is ambiguity in a line of reasoning; 3) judging whether certain statements contradict each other; 4) judging whether a statement follows necessarily from given data; 5) judging whether a statement is specific enough; 6) judging whether a principle establishes a statement that is alleged to be an application of it; 7) judging whether a statement made by an observer is reliable; 8) judging whether a conclusion is justified; 9) judging whether the problem has been identified; 10) judging whether something is an assumption; 11) judging whether a definition is adequate; and 12) judging whether a statement made by an alleged authority is acceptable.

Another approach to critical thinking is that of Watson and Glaser (1964), who define it as a composite of

attitudes, knowldge, and skills. This composite included (1) the ability to recognize the existence of problems and an acceptance of the need for evidence in support of what is stated to be true; (2) knowledge of the nature of valid inferences, abstractions, and generalizations in which the accuracy of different kinds of evidence is logically determined; and (3) skill in applying these attitudes and knowledge.

Yet another definition of critical thinking is that of the Cooperative Study of Evaluation in General Education (Dressel & Mayhew, 1954) which lists five abilities that are related to the concept of critical thinking: 1) the ability to define a problem; 2) the ability to select pertinent information for the solution of a problem; 3) the ability to recognize stated and unstated assumptions; 4) the ability to formulate and select relevant hypotheses; 5) the ability to draw conclusions validly and to judge the validity of inferences.

Although there are differences in these definitions, there is also agreement. They each agree that critical thinking includes the following characteristics: 1) identifying the problem, 2) selecting pertinent information for solving the problem, 3) recognizing stated and unstated assumptions, 4) formulating and selecting relevant hypotheses, and 5) recognizing the validity of inferences

in conclusions.

Component Analysis of Critical Thinking

While Ennis, Watson and Glaser, and Dressler and Mayhew have defined critical thinking, Sternberg has developed a theory to explain the processes that underlie thinking. His theory is based on information-processing in which data are taken in, analyzed, and organized in to an appropriate output. In his theory of componential analysis of information-processing, Sternberg (1979) proposed a continuum of levels of information-processing, extending from perception in very simple identification tasks to problem solving in very complex inferential tasks. He has shown that the strategies used in solving analogies are the strategies of formal thought as conceptualized by Piaget (Sternberg & Downing, 1982). He also concluded that strategies or processes continued to develop in the stage of formal thought and possibly beyond. He studied these strategies by requiring adolescents to evaluate higher order analogies or analogies between analogies. Thought processes involved in these tasks seemed to involve the characteristics of critical thinking such as formulating and selecting relevant hypotheses and recognizing the validity of inferences and conclusions.

Sternberg (1979) organized mental abilities into a hierarchy of four levels 1) composite tasks, 2) subtasks, 3) information-processing components, and 4) information-

processing metacomponents. Tasks used to measure mental abilities can be arranged hierarchically with successively higher vertical levels of the hierarchy handling higher task complexity. Composite tasks can be decomposed into subtasks and subtasks into components. Metacomponents control the use of components in composite tasks and subtasks.

In Sternberg's (1979) theory, the first level was that of the composite task, or the full task as the subject sees it. The composite tasks Sternberg chose to investigate mental abilities with include 1) analogies; 2) classifications; 3) series completions; 4) metaphorical completions and ratings; 5) linear syllogisms; 6) categorical syllogisms; and 7) conditional syllogisms. In the first four tasks, the subject was presented with three terms and a blank term and asked to select the best solution from given terms. In the fifth task, the subject was required to discern the answer in the premise; and in the last two tasks, the subject was asked to decide if the conclusion followed logically from the premises (Sternberg, 1979). Composite tasks can be related to the definition of critical thinking used in this study as identification of the problem to be solved.

The second level of the theory was that of the subtask. When related to critical thinking as defined earlier, subtasks were involved in selecting pertinent

information for solving various aspects of the problem. Composite tasks could be decomposed into subtasks in many different ways and by different methods. Sternberg chose to consider composite tasks in two groups, induction tasks (analogies, classifications, series completions, metaphorical completions) and deduction tasks (linear syllogisms, categorical syllogisms, conditional syllogisms). These composite tasks were decomposed into subtasks by different methods.

1. Induction tasks. Sternberg generally decomposed induction tasks by pre-cuing. In the pre-cuing method, presentation trials were divided into two parts. The first consisted of precuing that enabled problem solution and the second consisted of the full problem, which allowed solution of the problem. First, the subject processed the precuing information as completely as possible. Second, the subject solved the problem. Pre-cuing consisted of presenting as many terms of the problem as necessary for the solution of the problem.

2. Deduction tasks. The method of partial tasks were used to decompose subtasks in deductive reasoning. In this method either full or partial tasks were presented. To solve partial tasks required that a subset of the information-processing components be inolved in the full task.

The third level of the theory was that of the

information-processing component. Subtasks were decomposed into information-processing components. The component was of primary importance in Sternberg's theory. A component was an elementary information process which could translate sensory imput into a conceptual representation, transform one conceptional representation into another, or translate that representation into a motor output (Newell and Simon, 1972). Sternberg (1979) used the term "componential metatheory" to indicate the schematization of the nature of mental abilities and the term "componential analysis" to refer to the methodology employed to describe that schematization (Sterberg, 1979).

There were three kinds of components. General (G) components were necessary for all tasks within a given universe of tasks. Class, or group, components (C components) were necessary for performance of classes of tasks within a task universe. Specific (S) components were necessary for performance of specific tasks within a task universe and were involved in solving analogical reasoning problems and the strategies of inference, mapping, application, justification, encoding, and response--all of which are inherent in critical thinking.

The fourth level of Sternberg's (1979) theory was that of metacomponents. Although Sternberg called this the fourth level of his theory, essentially metacomponents affected three other levels (i.e. composite tasks,

subtasks, and components) by controlling what happened at the componential level. Metacomponents were processes which controled components, representations, and strategies used in each problem solution. Metacomponents also controled the rates of component executions and the probability that certain components would be used in specific situations.

In his research, Sternberg (1979) found that there was an increase in the correlation between response times on analogical reasoning subtasks and on composite reasoning tasks as the amount of information-processing required by the subtasks decreased. The increase in response times as the amount of information-processing decreased seemed to be a function of metacomponents, i.e., in deciding how a problem would be solved, not in the actual solution of the problem. Sternberg also found that better reasoners tended to spend longer time in encoding the terms of an analogy than did poorer reasoners. A third finding was that older children tended to perform the component processes of analogical reasoning more comprehensively than did younger children (Sternberg, 1979). Increased use of comprehensive information-processing appeared to be a general characteristic of cognitive development (Brown and DeLoache, 1978).

In his work with analogies, Sternberg (1979) found that most errors made in analogy solution could be traced

to self-terminating component processes--that is, processes that ended before all relevant aspects have been identified or compared. This process appeared to be a function of the metacompential level. Sternberg concluded that the metacompential processes were responsible for problem solutions in an intelligent way.

Level of Cognitive Processing in Critical Thinking

The abilities necessary for critical thinking can be related to the formal thought stage of Piaget's developmental theory of intelligence as "if-then" or hypothetico-deductive reasoning. The stage of formal thought begins at about twelve years of age, is consolidated in adolescence, and is the most advanced stage of Piaget's theory (Ginsburg & Opper, 1979).

The most important characteristic of formal thought involves the real versus the possible. The formal thinker begins the solution of a problem by considering all the possible relationships inherent in the problem and tries, by experimentation and analysis, to find which of the possible relations is true. This type of thinking is fundamentally hypothetico-deductive in character. In other words, in trying to discover the real among the possible, what is possible is considered as a set of hypotheses to be confirmed or denied (Flavell, 1963), recognizing the validity of inferences.

Formal thinking is also propositional thinking. The

formal thinker no longer manipulates only the raw data, but assertions or statements - propositions - which contain these data. In the stage preceding formal thought, concrete operations, the 7 to 11 year old organizes objects and events by putting them into classes, seriating them, setting them into correspondence, etc. The formal thinker also performs these first order operations, then takes the results of these concrete processes, puts them in the form of propositions, and continues to operate on them by making various logical connections between them such as implication, conjunction, identity, disjunction, etc. (Flavell, 1963). This property of formal thought seems to be involved in critical thinking in formulating and selecting relevant hypotheses and in recognizing the validity of inferences and conclusions. These strategies also seem to be used in analogical reasoning as defined by Sternberg.

Another property of formal thought is its combinatorial ability. In solving a problem, the formal thinker is able to consider all possible combinations of variables in the problem. Each possible combination has to be tested before the analysis is complete (Flavell, 1963). By considering all possible combinations, the formal thinker is able to select pertinent information for solving problems and formulate hypotheses. Combinatorial ability seems to be at the subtask level of Sternberg's theory.

Piaget considered formal thought to be of a propositions-about-propositions nature which he termed interpropositional thought. He also considered formal thought to be second degree operations because operations are performed on the results of the first degree, or concrete, operations. Piaget did not distinguish among levels of thought within formal thought (Flavell, 1963). By using higher order analogies, Sternberg (1979) found that adolescent strategies can continue to develop within the formal thought stage and possibly beyond.

Gifted Thinking

Adolescents displaying substantial amounts of formal operational thinking are generally regarded as intellectually "gifted" by educators and researchers. The ability to reason is an integral part of intelligence. Terman (1921) defined intelligence as the ability "to carry on abstract thinking" (p.128.) Resnick and Glasser (1980) interpret intelligence as "the ability to solve problems" (p.205). Sternberg (1982) stated that "reasoning, problem solving, and intelligence are so closely related that it is often difficult to tell them apart" (p.115). Therefore, those individuals labeled "intellectually gifted" may, in fact, be critical thinkers.

According to Newland (1976) "high intellectual potential--a superior capability to make and work on the basis of abstractions, to grasp and use complex

relationships, and to generalize meaningfully--is taken to be the primary ingredient in genius" (p.60). Hildreth (1966) believed that "the outstanding trait of children who rate high in intelligence is their ability to think, reason, and generalize beyond their years" (p.76). Although this exceptional thinking and reasoning ability is recognized by educators and researchers as an essestial attribute of giftedness, little empirical data on this topic is found in the professional literature.

Measurement of Critical Thinking

One of the emphases of this study was to examine the critical thinking ability of "gifted" adolescents. To evaluate "gifted" students' levels of critical thinking selection of appropriate instruments was necessary. Further, these instruments had to reliably and validly measure to critical thinking as defined by Ennis, Watson and Glaser, and Dressler and Mayhew.

Several instruments were available to assess critical thinking ability. Among those currently available which measure critical thinking from a comprehensive perspective (containing several narrow dimensions) were the Cornell Critical Thinking Test (Level X and Level Z) and the Watson-Glaser Critical Thinking Appraisal. The Graduate Record Examination Analytical Scale (ETS, 1981) was a new instrument developed to assess the analytical reasoning ability of students entering graduate education. The

inferences of this test were largely undetermined; however, the content seemed conceptually similar to tests of critical thinking.

Cornell Critical Thinking Test, Level X. This test was developed by Ennis and Millman (1961). It was designed to assess mastery of four aspects of critical thinking ability: (1) inductive reasoning--evaluation of evidence and its direction of support for a given hypothesis; (2) judging reliability of observation statements and inferences made from them; (3) deductive reasoning ability--judging what follows necessarily from given premises; and 4) assumption finding (Stewart, 1979). This test was written in story form about a group of explorers who land on an unfamiliar planet and must deal with the situation there. Ennis, Millman, and Tomko (1979) suggest it is appropriate for students from junior high (age 13) and up.

Cornell Critical Thinking Test, Level Z. Level Z of the Cornell Critical Thinking Test (1961) was substantially different from Level X and somewhat more difficult. This test was divided into seven sections and tested for the following critical thinking abilities: (1) deductive reasoning, determining whether a statement follows from given premises; (2) identifying faulty reasoning, recognition of informal fallacies such as circularity, over simplification of alternatives, and non-supporting emotive

language; (3) judging reliability of statements; (4) evaluation of evidence and its direction of support for a given hypothesis; (5) choosing useful hypothesis-testing predictions; (6) assumption finding, determining what a speaker probably had in mind by particular use of a term; (7) assumption-finding (gap-filling), identification of the unspoken premise in a speaker's arguments (Stewart, 1979). Ennis, Millman, and Tomko (1979) indicated that this instrument may be too difficult for the average secondary school student and is best suited for college level and advanced secondary school students.

Although the Cornell Critical Thinking Tests, Level X and Level Z, were tests which meassure comprehensive critical thinking ability, they were not chosen to be used in this study. Level X was directed toward junior high students, age 13 and up, which was not considered challenging for gifted students. Level Z was for advanced high school students and college students and would have been appropriate; however, the authors caution that individual subtests were likely to be unreliable (Stewart, 1979), perhaps because several of them contained only 3 or 4 items.

Watson-Glaser Critical Thinking Appraisal. This test was a widely used instrument for measuring critical thinking ability. It was one of the few currently available tests which assess comprehensive critical

thinking ability rather than a single aspect (Stewart, 1979). It was designed to measure skills and yield subscale values in five areas of critical thinking: (1) inference--the ability to discriminate among degrees of truth or falsity in inferences; (2) assumption recognition--the ability to recognize unstated assumptions and presuppositions taken for granted in statements; (3) deduction--the ability to reason according to the principles of class and conditional logic; (4) interpretation--the ability to judge whether given generalizations are reasonably supported by the data; and (5) evaluation of arguments--the ability to discriminate between strong, relevant arguments and weak, irrelevant arguments (Watson and Glaser, 1964). These five subscales collectively yielded a total general measure of critical thinking imbedded throughout the test.

The Watson-Glaser was composed of two different kinds of item content. Some questions dealt with "neutral" topics such as the weather, scientific facts or experiments, and other non-prejudical subjects. Other questions concerned political, economic, and social issues about which people have strong emotional feelings and prejudices. Controversial issues are included to provide a sample of an individual's ability to think critically about issues in an emotional context.

The Graduate Record Examination Analytical Scale. Although the Watson-Glaser is a widely used test of general critical thnking, little is known about the inferences of the results of the Analytical Scale of the Graduate Record Examination. This instument was developed to test the abstract reasoning ability of students applying to graduate school. It was designed to measure advanced levels of analytical reasoning. Exactly which dimensions of thinking are involved and their relationship to different levels of critical thinking had not been specified. No particular construct or constructs were formulated in the development of this scale except to select questions that measured analytical resasoning (C. Wild, Personal Communication, August, 1984).

The GRE-Analytical scale used two kinds of questions, i.e., analytical reasoning and logical reasoning, to test the ability to think analytically. Analytical reasoning questions emphasized the ability to analyze a given set of arbitrary relationships among fictional entities and to derive new information from them. Logical reasoning questions emphasize the ability to understand and to analyze relationships among arguments or parts of an argument.

The GRE-Analytical was of particular interest in this study for several reasons. Most gifted students will be required to take the GRE for graduate education; however

the GRE-Analytical was not well understood. By exploring the relationship of the Watson-Glaser to the GRE-Analytical, the conceptual nature of the GRE-Analytical might be clarified. As more is understood about this instrument, practical applications for its results are more likely to occur.

Purpose of the Present Study

The purpose of this study was to determine the relationship between the construct(s) measured by the Watson-Glaser Critical Thinking Appraisal and the construct(s) measured by the Graduate Record Examination Analytical Scale among intellectually gifted adolescents with possible implications for "gifted" education. The hypotheses of this study were 1)gifted adolescents can think critically compared to college level norm groups as measured by the Watson-Glaser and the GRE-Analytical scale; 2) significant differences would exist between differeent levels of gifted populations; and 3) a factor or group of factors of the Watson-Glaser subscales would signifcantly predict performance on the GRE-Analytical Scale.

CHAPTER III

Method

Subjects

The total subject pool was composed of one hundred and four high school students, most of whom were in their senior year or in the summer preceeding their senior year. One student was in the eleventh grade and three students were entering the tenth grade. The subject pool labeled "national gifted" consisted of 50 students in summer programs for academically gifted students held at two universities. These students were from fourteen different states. There were 15 students from a student science training program in psychobiology at Western Kentucky University and thirty-five students from the University of Indiana High School Student Science Training Program at Bloomington, Indiana. The "local gifted" group consisted of 54 students in secondary schools in western Kentucky and northern middle Tennessee. Twenty four students were in an honors humanities class. There were 13 students in a seminar for honor students in physics. Although these two classes were different in scope, they were both for advanced students. Seventeen students from three different high schools, all in the upper ten percent of their classes, also agreed to participate after school. All

subjects were volunteers in this study.

Instruments

Watson-Glaser Critical Thinking Appraisal. One of the instruments administered to subjects was the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1964). The Watson-Glaser tests the application of abilities involved in critical thinking. Although there are two forms, Ym and Zm, form Ym was used for all subjects. The test consists of five subtests designed to measure different, but interdependent, aspects of critical thinking (Watson and Glaser, 1964). Although subtests are timed to help pace students, as much time as is necessary to complete the test may be used for this instrument. Most subjects finish within 50 minutes. The five subtests of the Watson-Glaser and a description of each follows:

Test 1. Inference (29 items) samples the ability to discriminate among degrees of truth or falsity of inferences drawn from given data.

Test 2. Recognition of Assumptions (16 items) samples the ability to recognize unstated assumptions or presumptions which are taken for granted in given statements or assertions.

Text 3. Deduction (25 items) samples the ability to reason deductively from given statements or

premises; to recognize the relation of implication between propositions; to determine whether what may seem to be an implication or a necessary inference from given premises is indeed such.

Test 4. Interpretation (24 items) samples the ability to weigh evidence and to distinguish between generalizations from given data that are not warranted beyond a reasonable doubt.

Test 5. Evaluation of Arguments (15 items) samples the ability to distinguish between arguments which are strong and relevant and those which are weak or irrelevant to a particular question.

Reliability data consist of split-half reliability coefficients from test scores of the various normative groups. The test was normed on test scores from 20,312 high school students, grades 9 through 12; 5,197 liberal arts college freshmen; and 554 college senior women. Odd-even split-half reliability estimates, corrected by the Spearman-Brown formula, range from .77 (Form Zm, grade 9) to .87 (Form Ym, grade 12). Odd-even split-half reliability estimates for individual subtests, corrected by the Spearman-Brown formula, range from .41 (deduction subtest, Form Zm) to .74 (assumption recognition subtest, Form Ym). These estimates are from the grade 10 normative

sample (N=2947, Form Ym; N=2995, Form ZM) (Watson and Glaser, 1964).

Watson and Glaser (1964) stated that critical thinking is not clearly defined and that the content validity of the Watson-Glaser should be evaluated in the context in which it is being used. Construct validity is inferred from the low subtest inter-correlations (grade 10 normative sample). The low inter-correlations (from r=.21 to r=.50) "support the contention that relatively distinct abilities are being measured with sufficient overlap to warrant their inclusion in one total score" (p. 14). Evidence of construct validity is supported by subtest to total test score correlation which vary from .56 (argument evaluation, Form Zm) to .79 (interpretation subtest, Form Ym) based on the grade 10 normative sample. The authors suggest that the predictive validity of the instrument should be established empirically in each situation in which the test is used.

Graduate Record Examination, Analytical Scale

The Graduate Record Examinations were developed by the Carnegie Foundation for the Advancement of Teaching and the graduate school deans of four eastern United States universities. They were first administered in 1937. The GRE Aptitude Test measures the general verbal, quantitative, and analytical abilities students need to be successful in graduate education. The analytical measure is the instrument used in this study to measure critical

thinking ability because it is believed to exist at high levels in "gifted" adolescents. The Analytical Scale was developed and added to the Graduate Record Examinations in 1977 in order to expand the scope of the Aptitude Test and to enable those taking the exam to demonstrate a wider range of academic abilities than those measured by the verbal and quantitative scales of the tests (ETS, 1981). The Analytical Scale is composed of two sections of twenty-five questions each. Thirty minutes is allowed for completion of each section.

Research has shown that two item types in the original 1977 analytical measure, analysis of explanations and logical diagrams, are affected by short term practice and instruction. The scale was revised in 1981, deleting these items and increasing the number of analytical reasoning and logical reasoning items. According to the test guide,

> Analytical reasoning questions test the ability to understand a given structure of arbitrary relationships among ficticious persons, places, things or events; to deduce new information from the relationships given; and to assess the conditions used to establish the structure of relationships. Logical reasoning questions test the ability to understand, analyze, and evaluate arguments; recognizing the assumptions on which an argument is based, drawing conclusions from given premises,

inferring material missing from given passages, applying to one argument principles governing another, identifying methods of argument, evaluating arguments and counter arguments, and analyzing evidence" (ETS, 1981).

The reliability coefficient for the GRE-Analytical measure administered primarily to college seniors in October 1981 was .87. The correlation between the analytical score and the verbal score for college seniors and unenrolled college graduates range from .63 to .68 and the correlations between the analytical score and the quantitative score range from .63 to .71. Educational Testing Service (1981) states that content validity is based on extensive experience in developing aptitude tests. The predictive ability of the analytical measure of the GRE can be substantiated only when those taking the September 1981 test demonstrate success in graduate school.

Procedures

The national gifted groups were administered both instruments in single evening sessions at each university. Half of each group took the Watson-Glaser while the other half took the GRE-Analytical. After a ten minute break the groups each took the remaining test. Three of the local gifted groups were administered the tests in one session after school at their high schools. The remaining two local gifted groups were administered the tests during

class time on two consecutive days with the Watson-Glaser given the first day and the GRE-Analytical the second day.

The tests were administered by the researcher with the assistance of a monitor when two groups were tested simultaneously. Each subject was assigned a number to protect anonymity. All test results were confidential; only group data were reported. Subjects were told they would be given their individual scores if they wanted them.

Analyses

Mean values and standard deviations were obtained for national and local groups for the GRE-Analytical and Watson-Glaser total scores. GRE-Analytical raw scores were converted to standard scores as provided by ETS for its 1981 data. Watson-Glaser raw scores were converted to percentile performance derived from norms on college senior women. A t-test for independent means (Ary and Jacobs, 1976) was used to determine the significance of the difference between the means of the two groups. The degree of relationship between the GRE-Analytical and the Watson-Glaser, total and subtest scores, was determined by calculating Pearson's product-moment correlation. Then, stepwise regression analysis procedures of the Statistical Analysis System, (1983) determined the predictive relationships of Watson-Glaser subtests to GRE-Analytical total score. Pearson's product-moment median-split correlations on total, national, and local groups were used

to establish the degree of consistency in the upper and lower halves of each group. Stepwise regression analysis was performed on each half to examine possible relationship differences between the GRE-Analytical and the Watson-Glaser for the upper and lower halves of the local and national samples.

CHAPTER IV

Results

Analysis of the data on overall Watson-Glaser total scores for the national and local samples indicated raw score mean/standard deviation performances of 77.42/8.28 and 69.02/11.89, respectively. Compared to college senior women, the national sample mean was at the 60th percentile and the local sample mean at the 30 percentile (Watson & Glaser, 1964). Raw score means/standard deviations on the GRE-Analytical scales for national and local students were 31.3/9.24 and 20.83/8.6, respectively. The scaled mean score was 580 for the national sample and 440 for the local sample. Comparison of performances on the two instruments by the two groups indicated that the national sample's performance was superior to that of the local sample on both instruments ($t_{wg} = 3.4$, $p < .01$; $t_{gre} = 5.51$, $p < .01$). These results are summarized in Table 1.

Pearson product-moment correlation coefficients were calculated to examine the relationship between the GRE-Analytical and Watson-Glaser total scores for local and national samples. As can be seen in Table 2, the relationships between the performances on the two instruments were highly significant for both the local and

Table 1

Statistical Comparisons of Local and National Gifted Students on the Watson-Glaser and GRE Analytical Scale

	Watson-Glaser	GRE Analytical
Local	$\bar{X}$ = 69.02	$\bar{X}$ = 20.83
	SD = 11.89	SD = 8.6
National	$\bar{\bar{X}}$ = 77.42	X = 31.3
	SD = 8.28	SD = 9.24
	t = 3.4 (p<.01)	t = 5.51 (P<.01)

Table 2

Correlations Between Watson-Glaser and GRE Analytical Performance for Local and National Gifted Students

	1982 Data	
Local	$r_{gre/wg} = .74$	($p < .0001$)
(n=55)		
National	$r_{gre/wg} = .55$	($p < .0001$)
(n=50)		
	1984 Data	
National	$r_{gre/wg} = .42$	($p < .01$)
(n=34)		

national groups (rgre/wg.loc=.74, $p < .0001$; $r_{gre/wg.nat} = .55$, $p < .0001$).

As a preliminary step to the planned regression analyses on data from each test, zero order correlation coefficients were calculated between each of the Watson-Glaser subscales and the GRE-Analytical scores for students in the local sample. These data are found in Table 3. As can be seen, all subscales were positively correlated with GRE performance. Subscale 4 (Interpretation) displayed the highest relationship (r=.57) followed by Subscale 2 (Recognition of Assumptions) at r=.53. After analysis of the data on total scores had been completed, the national gifted group data were accidently destroyed, prohibiting further analysis.

The local gifted data were then subjected to stepwise regression analyses (SAS, 1983) to determine which individual subscale or group of subscales best predicted the GRE-Analytical performance. In the local sample Subscale 4 (Interpretation) clearly emerged as the best predictor ($p < .0001$), accounting for 33% of the GRE-Analytical variance. The addition of Subscale 2 (Recognition of Assumptions) accounted for an additional 9% of the variance. The inability of the second factor to account for a more significant amount of the GRE-Analytical variance can be explained by much of Subscale 2's being a

Table 3

Correlations Between Watson-Glaser Subscales and GRE Analytical Performance for Local and Gifted Students

$r_{gre/wg-t} = .720$ ($p < .0001$)

$r_{gre/wg-1} = .458$ ($p < .0020$)

$r_{gre/wg-2} = .530$ ($p < .0003$)

$r_{gre/wg-3} = .390$ ($p < .0100$)

$r_{gre/wg-4} = .577$ ($p < .0001$)

$r_{gre/wg-5} = .401$ ($p < .0085$)

$n = 42$

subset of Subscale 4. Interpretation largely accounts for Recognition of Assumptions. Subscale 4 has subsequently been substantiated to be the best predictor of GRE-Analytical in national gifted students in 1984 on 34 participants in Indiana's and Western Kentucky University's summer programs for nationally gifted students ($p < .005$) (R. L. Miller, personal communication, August, 1984).

Pearson product-moment correlations based on a median-split were calculated on Watson-Glaser total to GRE-Analytical scores of national and local groups to assess the degree of consistency within groups. Total group ($N = 104$) median-split correlation coefficients for upper and lower groups were $r = .31$ and .56, respectively. National group ($N=50$) median-split correlations were $r_{upper}=.42$, $p < .038$; $r_{lower}= .43$, $p < .031$. Local group ($N = 42$) median split correlations were r upper= .31, $p < .170$; $r_{lower}= .42$, $p < .057$.

Local group median-split data ($N = 21$ each) were subjected to stepwise regression analyses to determine if the same or different factors were operating to predict the GRE-Analytical from the Watson-Glaser in each half. The best single subscale predictor was Subscale 1 for local upper half ($p < .06$), Subscales 1 and 3 for local lower half ($p < .025, .03$), Subscale 2 for national upper ($p < .07$); and Subscales 2 and 4 for national lowerl ($p < .02, .01$).

CHAPTER V

Discussion

The purpose of this thesis was to examine the relationship between the constructs measured by the Watson-Glaser Critical Thinking Appraisal and the constructs measured by the Graduate Record Examination Analytical Scale among intellectually gifted adolescents. Three hypotheses were addressed in this study: 1) gifted adolescents would be able to think critically as determined by college level norms, 2) significant differences would exist between different levels of gifted populations, and 3) a factor or group of factors of the Watson-Glaser subscales would significantly predict performance on the GRE-Analytical Scale.

College Level Performance

Analysis of data on overall Watson-Glaser performance for national and local groups yielded mean performances which placed the groups in the 60th and 30th percentile respectively when compared to college senior women. Also, mean raw scores for the GRE-Analytical, when converted to scaled scores, were 580 (national) and 440 (local). These results confirmed that these students were able to think critically compared to college level norms.

Superior Performance of the National Sample

Comparison of performance on the two instruments by national and local groups indicated significantly superior performance of the national group over the local group on both tests. This significant difference in performance by the national group might have been attributable to a number of possible factors. The difference may have been the result of 1) differences in quantity or quality of educational background of the participants, 2) the selection criteria for inclusion in each group, or 3) conditions within the study. Students in the national programs routinely attend schools with broad advanced placement offerings in science, math, history, English, etc. These schools also stress academic excellence and press students for performance. A more restricted offering and less demand for superior performance at the local levels are common. Additionally, national programs for the gifted student select students in the top 1% or 2% academically as indicated by grades, class standing, and national standardized tests. Local gifted programs have less rigorous criteria, selecting from the top 10% of an existing school class, or an IQ of 125 or higher. Lastly, testing conditions which may have affected performance included a high level of noise during one testing session, necessitating a move to a quieter room for the last half of

the test. In another testing situation, the total time allowed for the GRE-Analytical was shortened because of scheduling. This time reduction could have lowered performance for that particular group.

Relationship Between Watson-Glaser and GRE-Analytical Tests

Correlation coefficients calculated to determine the relationship of the Watson-Glaser to GRE-Analytical performance of national and local samples showed the instruments to be significantly correlated. Because each test was designed to assess advanced thinking ability, this result was to be expected. The Watson-Glaser includes questions that require subjects to make inferences, recognize assumptions, use deductive logic, interpret data, and evaluate arguments (Watson & Glaser, 1964). The GRE-Analytical asks subjects to analyze data, relationships, and arguments, to make inferences, recognize assumptions, and draw conclusions (ETS, 1981). The conceptual similarity of the questions on each instrument could account for the high correlation.

Zero order correlation coefficients calculated between Watson-Glaser subscales and GRE-Analytical total scores for the local group were all strongly, positively correlated. Subscale 4 (Interpretation) exhibited the highest correlation, followed by Subscale 2 (Recognition of Assumptions). Stepwise regression analyses of local data to determine the best Watson-Glaser subscale factor or

group of subscales to predict GRE-Analytical performance indicated Subscale 4 (Interpretation) as the best single predictor, supporting that zero order correlation. As reported in the results section, Subscale 2 (Recognition of Assumptions) contributed only a small amount of additional variance and weakened the prediction of GRE-Analytical when combined with Subscale 4 (Interpretation) for a two factor solution. A single factor predictor was the best solution.

Pearson product-moment correlations based on a median-split were calculated to determine consistency in total, national, and local groups. In the total sample (national plus local) and in the local sample, the lower halves of each group displayed slightly higher correlations than the upper halves. These disparate results possibly suggest that factors other than Watson-Glaser performance were differentially operating for the two groups. However, it is more likely that the median-split produced the lower correlations as a result of restriction of range of scores in a sample which was already somewhat small prior to halving it.

Stepwise regression analyses based on a median-split for local data indicated Subscale 1 (Inference) the best Watson-Glaser predictor of GRE-Analytical performance for both groups. Although the two factor solutions were not impressive when compared to the one factor solution, there was a difference in upper and lower group results. The

best two factor solution for the upper group was Subscale 1 (Inference) plus Subscale 5 (Evaluation of Arguments) and for the lower half Subscale 1 (Inference) plus Subscale 3 (Deduction). These median-split results were derived from very marginal group sizes which may well be insufficient to produce reliable results. Any conclusions from these findings should be withheld pending replication with larger samples.

Implications for Future Research

Future research should include stabilizing the study by construction of homogeneous groups balanced on grade point average, national test scores, and/or other more definitive data which were not available for this study. Because of the small number of subjects in this study, additional research using the same design is needed to verify Subscale 4 (Interpretation) as the best Watson-Glaser subscale predictor of GRE-Analytical performance. If Subscale 4 continues to be the best predictor of GRE-Analytical performance in subsequent research, it may be concluded to be the best predictor model. Other populations could be assessed to determine if Subscale 4 continues to predict GRE-Analytical performance. These include non-gifted adolescents and younger gifted students, such as middle school students. Additional unanswered questions deal with male-female differences and age-related performance on the two instruments.

Implications for Education

Critical thinking as measured by the Watson-Glaser and the GRE-Analytical seem to be at the upper level of Sternberg's continuum in information-processing. The tasks involved in solving the problems in these instruments are complex inferential tasks, which involve processing the composite task, subtasks, and information-processing components under the control of metacomponents. Sternberg used analogies as higher order tasks through which one can investigate thought processes and found the strategies necessary for their solution to be the same as those in Piaget's formal thought stage. The strategies of formal thought include the ability to consider all possible hypotheses when solving a problem, to combine these hypotheses in all possible combinations, and to manipulate propositions about data, not the data alone. These strategies also are necessary for critical thinking tasks which include 1) identifying the problem, 2) selecting pertinent information for solving the problem, 3) recognizing stated and unstated assumptions, 4) formulating and selecting relevant hypotheses, and 5) recognizing the validity of inferences and conclusions.

The results of this study have provided evidence that gifted adolescents are able to think critically at college level. Students with these abilities can be easily discouraged by a low-level curriculum which does not

stretch their cognitive capacities. They can also be challenged and motivated by curricula which stretches their abilities.

REFERENCES

Ary, D., & Jacobs, L. C. (1976). Introduction to statistics: purposes and procedures. New York: Holt, Rinehart, and Winston.

Brown, A. L., & DeLoach, J. S. (1978). Skills, plans, and self-regulation. In R. Siegler (Ed.), Children's thinking: What develops? Hillsdale, NJ: Erlbaum.

Das, J. P. (1973). Structure of cognitive abilities: evidence for simultaneous and successive processing. Journal of Educational Psychology, 65, no. 1, 103-108.

Dressler, P. L. & Mayhew, L. B. (1954). Critical Thinking in social science. Dubuque, IA: William C. Brown.

Ennis, R. H. (1962). A concept of critical thinking. Harvard Educational Review, 32, 157-187.

Ennis, R. H., Millman, J., & Tomko, T. N. (1979). Manual for Cornell critical thinking test, level X and Cornell critical thinking test, level Z. Illinois Rational Thinking Project. Urbana, IL.: University of Illinois.

Flavell, J. H. (1963). The developmental psychology of Jean Piaget. Princeton, NJ: D. Van Nostrand Co.

Ginsburg, H. & Opper S. (1979). Piaget's theory of intellectual development. (2nd ed.). Englewood

Cliffs, NJ: Prentice-Hall, Inc.

Guide to the use of the graduate record exainations. (1981). Princeton, NJ: Educational Testing Service.

Guilford, J. P. (1967). The nature of intelligence. New York: MacGraw-Hill.

Hildreth, G. H. (1966). Introduction to the gifted. New York: McGraw-Hill.

Inhelder, B. & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books.

Meeker, M. M. (1969). Structure of intellect. Columbus, OH: Charles E. Merrill Publishing Company.

Newell, A. & Simon, H. A. (1971). Human problem solving Englewood Cliffs, NJ: Prentice-Hall, Inc.

Newland, T. E. (1976). The gifted in socioeducational perspective. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Resnick, L. B. & Glasser, R. (1976). Problem solving and intelligence. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Erlbaum.

SAS introductory guide, revised edition. (1983). Cary, NC: SAS Institute, Inc.

Sternberg, R. H. (1979). The nature of mental abilities. American Psychologist, 34 (3), 214-230.

Sternberg, R. H. (1981). A componential theory of intellectual giftedness. Gifted Child Quarterly, 25

(2), 86-93.

Sternberg, R. H. and Downing, C. J. (1982). The development of higher-order reasoning in adolescence. Child Development, 53, 209-221.

Sternberg, R. H. (1982). Reasoning, problem solving, and intelligence. In R. H. Sternberg (Ed.) Handbook of Human Intelligence. Cambridge: Cambridge University Press.

Sternberg, R. J. & Salter, W. (1982). Conceptions of intelligence. In R. H. Sternberg (Ed.) Handbook of Human Intelligence. Cambridge: Cambridge University Press.

Stewart, B. L. (1979). Testing for critical thinking: a review of the resources. Rational thinking reports number 2. The Illinois Rational Thinking Project. Urbana, IL: The University of Illinois

Terman, L. M. (1925). Genetic studies in genius: Mental and physical traits of a thousand gifted children (Vol 1). Stanford, CA.: Stanford University Press.

Watson, G. & Glaser, E. M. (1964) Watson-Glaser critical thinking appraisal: Manual, forms YM and ZM. New York: Harcourt Brace Jovanovich, Inc.