A comparison of structured and non-structured methods for the evaluation of graphic design projects

Jen Yen
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Art Education Commons, Educational Assessment, Evaluation, and Research Commons, and the Other Education Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/10743

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
A comparison of structured and non-structured methods for the evaluation
of graphic design projects

by

Jen Yen

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Industrial Education and Technology
Major: Industrial Education and Technology

Approved:
Signature was redacted for privacy.

In Charge of Major Work
Signature was redacted for privacy.

For the Major Department
Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1995
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ v
LIST OF FIGURES ..................................................................................................... vii

CHAPTER I. INTRODUCTION ................................................................................. 1
Need for the Study ................................................................................................. 6
Statement of the Problem ...................................................................................... 9
Purpose of the Study ............................................................................................... 10
Variables of the Study ............................................................................................ 10
Questions of the Study ........................................................................................... 11
Hypotheses of the Study ......................................................................................... 11
Assumptions of the Study ....................................................................................... 12
Limitations of the Study ......................................................................................... 13
Definition of Terms ............................................................................................... 13

CHAPTER II. LITERATURE REVIEW .................................................................... 15
Introduction ........................................................................................................... 15
An Overview of Evaluation ..................................................................................... 15
The role and function of educational evaluation ................................................ 15
The development of educational evaluation ......................................................... 19
Evaluation framework .......................................................................................... 22
Bloom’s Taxonomy ............................................................................................... 25
Evaluation Methods .............................................................................................. 27
Quantitative and qualitative evaluation ............................................................... 28
Norm-referenced and criterion-referenced evaluation ......................................... 31
Summative and formative evaluation .................................................................. 33
Authentic evaluation .............................................................................................. 35
Grading .................................................................................................................. 37
Evaluation in Visual Arts/Design ......................................................................... 40
Measurement of creativity ..................................................................................... 40
Measurement of design ........................................................................................ 41
Selecting criteria..................................................................................................... 42
An Overview of Graphic Design ............................................................................ 44
Principles of design .............................................................................................. 45
Color ..................................................................................................................... 48
Typographic design ............................................................................................... 49
Summary ................................................................................................................. 57

CHAPTER III. METHODOLOGY .......................................................................... 59
Design of the Study ............................................................................................... 59
Population and Sample ......................................................................................... 60
LIST OF TABLES

Table 2.1  A taxonomy of major evaluation models ........................................ 24
Table 2.2  Attributes of the qualitative and quantitative paradigms .................. 29
Table 2.3  Examples of paradigms, design approach, and methods ...................... 30
Table 2.4  A comparison between summative and formative evaluation ............... 35
Table 3.1  Response of the instruments .......................................................... 60
Table 3.2  Coding format .............................................................................. 65
Table 4.1  Test for differences between structured and non-structured rating scores by student and professional evaluators combined ........................................... 73
Table 4.2  Test for differences between structured and non-structured rating scores in each project by student and professional evaluators combined ........................................ 75
Table 4.3  Test of differences in rating scores between student and professional evaluators using the structured method ......................................................... 76
Table 4.4  Test of the ten projects for differences in rating scores between student and professional evaluators using the structured method ...................... 77
Table 4.5  Test of differences in rating score between student and professional evaluators using the non-structured method .................................................... 79
Table 4.6  Test of ten projects for differences in rating scores between student and professional evaluators using the non-structured method ...................... 80
Table 4.7  Test for differences between structured and non-structured in rating scores by student evaluators ............................................................... 81
Table 4.8  Test for differences between structured and non-structured rating scores in each projects by student evaluators .............................................. 83
Table 4.9  Test for differences between structured and non-structured rating scores by professional evaluators ............................................................. 84
Table 4.10 Test for differences between structured and non-structured rating scores in each project by professional evaluators

Table 4.11 Correlation coefficients between subscores and total score

Table 4.12 The correlation coefficients between subscores and total score by student evaluators

Table 4.13 The correlation coefficients between subscores and total score by professional evaluators

Table 4.14 Means and standard deviations of rating scores using structured and non-structured methods by student and professional evaluators
LIST OF FIGURES

Figure 2.1 Representation of the role of evaluation in educational process .............. 18
Figure 2.2 A model of instruction ............................................................................. 22
Figure 2.3 Suggested components of evaluation ...................................................... 25
Figure 2.4 Overview of classroom evaluation and grade assignment ....................... 39
Figure 2.5 Assessment task map ............................................................................. 43
Figure 2.6 The construction of letterform .................................................................. 52
Figure 2.7 Five basic styles of typeface .................................................................... 54
Figure 2.8 Type family range ................................................................................... 55
Figure 2.9 A comparison of line spacing between two different typefaces and space ........................................................................................................ 56
Figure 4.1 A comparison of mean scores between structured and non-structured methods by student evaluators ................................................................. 91
Figure 4.2 A comparison of mean scores between structured and non-structured methods by professional evaluators ................................................................. 91
CHAPTER I. INTRODUCTION

"In every day life we tend to use numbers and scales of measurement very carelessly. In fact we commonly employ several quite distinct ways of using numbers, without really being aware of the differences" (Lawson, 1990, p. 49).

Evaluating student achievement is a complex task. What does one need to know about student learning? How does one use the gathered information toward meaningful work in education? Educators must be careful to assure that evaluation is driven by educational purposes. As goals are determined, indicators may be devised to estimate how many and how well accomplishments have been mastered (Archbald & Newmann, 1988).

Evaluation is a process of getting, assembling, and interpreting evidence to help evaluators make more valid judgments to meet objectives, goals, and values (Fults, Lutz, & Eddleman, 1972). Bloom, Hastings, and Madaus (1971) defined evaluation specifically as a method of acquiring and processing the evidence needed to improve the student learning and teaching. They saw evaluation as an aid to clarify the significant goals and objectives of education and as a process to determine the extent to which students were developing in these desired ways.

A good evaluation not only requires students to achieve an objective or complex activity, but it also reliably measures beyond the specific tasks that students are asked to complete. The results of a good evaluation identify how well students can perform in a broad knowledge or skill domain. In addition, the skills that students demonstrate in an evaluative
situation should transfer to other situations and other problems (Herman, Aschbacher, & Winters, 1992). The teacher plays an important role during the evaluation process, especially in the classroom. Evaluation in the classroom includes the full range of information teachers collect about their students, the instruction, and classroom climate to help them understand their students, plan and monitor instruction, and establish a viable classroom climate. Thus, in classroom evaluation, teachers or instructors are required to make a broad range of decisions. Wolansky (1985) offered a specific view of evaluation by stating it is anticipated that teachers develop teacher-made evaluation instruments just as they are required to plan and implement instruction, and stated that “... evaluation is an inescapable activity and responsibility of every teacher” (p. 3).

Touzel (1993) observed the problem of teacher-centered evaluation. There are many educators who know all too well that the evaluative data of student performance may provide convenient numbers with which to categorize students, but that numbers may not be the best indicators of student performance. In fact, the numbers often focus attention on goals that are not the most significant. Touzel believed that it is time to change how one evaluates students throughout the whole educational enterprise.

Archbald and Newmann (1988) indicated a similar problem when evaluating student performance relating to the question of educational purpose. The problem is that most traditional assessment indicators communicate very little about the quality or substance of a
student's specific accomplishments. If an assessment provides little information and lacks authenticity, it may depress student learning, teacher commitment, and public support.

Moreover, when evaluating student performance based on scientific theories, literary and artistic masterpieces, architectural and mechanical designs, or musical compositions, one must consider student performance as a whole, not just as collections of knowledge fragments. Archbald and Newmann (1988) emphasized that authentic academic achievement should integrate knowledge in two ways. First, in order to understand integrated forms of knowledge, the students must be involved in the production, not simply the reproduction of new knowledge. Second, as achievement has aesthetic value apart from determining the competence of the learner, the students should demonstrate their disciplined inquiries.

It is critical that teachers estimate the whole person and evaluate in all domains (Wolansky, 1985). Perry (1982) denoted that the instructor should observe student performance and literally pass judgment on whether or not the quality of the performance meets the requirements or standards. Perry also pointed out that the characteristics of an effective evaluation of student's performance should:

- Be fair and impartial
- Measure only what has been presented to the student
- Use sound principles of test construction
- Be valid
- Be reliable
- Be objective
- Be comprehensive
- Be discriminating
- Be systematic and continuous (p. 116).
However, it is very difficult for instructors to determine the appropriate evaluation method in order to include all the characteristics that will effectively evaluate student performance. Marshall (1968) argued the problem by saying:

Teachers are usually required to choose one of several symbols, such as ABCDEF, to express a judgment of each of their students. The system has long been under fire, for specific objections are numerous, but when it comes to vote this shorthand for expressing judgment is more often favored than altered. Increasing criticism in recent years has not thus far defeated the common system. No substitute which meets the functions these symbols serve has been strong enough to replace them. (p. 118)

According to Wolansky (1985), grades have been the subject of contention by educators for almost a century. The student is conditioned and influenced by the emphasis on grades and labels to the point that grades show learning progress or achievement. Wolansky added that “... grades provide a description index of an individual's achievement in a given area. They have interpretable meaning for the student, teacher, parent, registrar, and employer. Yet grades serve only as an approximate index to achievements” (p. 76).

Bloom et al. (1981) indicated “... there is no statistical or completely objective method that can be used to assign grades to a student's test score or a student's product; ultimately a judgment of the worth or value of that score or product must be made by the teacher” (p. 104). Thus, fairness in grading is a serious factor of the assignment of grades which requires a clear understanding of a student's performance or product and grading methods should be as explicit as possible.

Design education is in an unique position. According to Haider (1990), it involves the disciplines of art history, art criticism, and aesthetics, also the ideas derived from psychology.
sociology, and anthropology of the arts. Thus, evaluation in design is too complex to be cast as a single number or to be assessed on a standardized test.

Greer and Hoepfner (1986) suggested that measurement of studio skills should be efficient and cost-effective, and necessarily will have to focus on the simplest and most available methods. The authors emphasized that with some specification of the criteria, the scores can be determined in a fairly objective manner by assessing each production on the basis of whether it has met each criterion or how well it has met the overall objective.

To measure aesthetic content, the teacher should present students with a visual arts reproduction having the same or similar aesthetics characteristics as those used in instruction. Then the students can respond in such a way that shows their development of knowledge and skill in employing aesthetics (Greer & Hoepfner, 1986).

Thus, when evaluating student productions, careful consideration should be placed on the teacher's ability to determine whether students meet the objectives and to what extent students have reached the objective. In addition, the goal of vocational and technical education is "...helping the student to develop a broad range of knowledge, skills, attitudes, and values, each of which ultimately contributes in some manner to the graduate's employability" (Finch & Crunkilton, 1993, p. 13).

The content of vocational education should focus on students' ability to transfer knowledge, skills, and attitudes to the world of work. Byrne (1990) explained that the evaluation method can be broken up into two kinds—informal and formal—in design-related education. As Byrne defined them:
Informal evaluation is dependent on casual observation, implicit goals and unstated criteria, intuitive norms, and subjective judgments; it can sometimes be penetrating and insightful, other times superficial and distorted. Formal evaluation is dependent on checklists, structures, controlled comparisons, and use of standardized kinds of tests (p. 154).

Based on Byrne's definition, in the present study, this researcher employed both informal evaluation (non-structured) and formal evaluation (structured).

**Need for the Study**

Evaluation is a vital part of any educational activity. The teacher needs evaluation as a guide to measure not only the effectiveness of teaching, but also to collect evidence indicating where the instruction might be improved. On the other hand, the student needs to know how well he or she is achieving the objectives established by the teacher. Evaluation is important, especially as design curriculum content is “. . . typically based upon the actual worker's role with relevant tasks, knowledge, skills, attitudes, and values serving as a foundation for what is to be taught” (Finch & Crunkiton, 1993, p. 19).

In the existing educational system, the primary purpose of evaluation is used to grade and classify students. Most evaluation procedures are designed to find those who have failed (D or F), those who have succeeded (A or B), and those who have gotten by (C). However, the grading method of this system contributes little to the improvement of teaching and learning, and it hardly serves to ensure that all (or almost all) learn what the school system regards as the important tasks and goals of the education process (Bloom et al., 1971).

Wilson (1986) noted that “. . . testing, assessment, and evaluation in arts education are at a primitive level of development and application” (p. 6). There is no systematic effort to
determine what individual students, the students in a particular school, or the students in a school system learn as the result of arts instruction (Wilson, 1986, 1992). Greer and Hoepfner (1986) indicated that achievement tests in visual arts have disappeared gradually from the market over the last twenty years. The reason may be that “... no sequence of learner expectations has been apparent as in other subjects, or that experts in art education have been unable to agree on what they are” (p. 44). Another reason, according to Zerull (1990), may be that it “... lacks of any widespread, comprehensive, and systematic assessment of student achievement in the arts” (p. 19).

A comprehensive and systematic assessment of student products also lacks in design evaluation. There is a long tradition of design evaluation employing non-structured evaluation methods in the classroom. In this method most instructors assign a letter grade to each student’s project, which may not adequately describe the complexity of student achievement. This method may be used well by experienced teachers who are confident in their understanding of the design problems in hand, and in the performance and attributes of the proposed solutions. However, less experienced teachers possibly acquire this intuition through a combination of observing experts at work and by trial-and-error (Blandford, 1993).

In fact, use of the structured evaluation method (listing all the required criteria of the product) is becoming more necessary in the design field. Archbald and Newmann (1988) pointed out that evaluation should measure not only any kind of achievement, but also valuable or meaningful forms of mastery. A valid assessment system should also provide information about the particular tasks in which students succeed or fail. It should present
tasks that are worthwhile, significant, authentic, and meaningful (Grace, 1992). The greatest need of student evaluation is that it should be clearly formulated "... based on expectations that were clarified in a given assignment or period of instruction, and points the way to next stages of learning" (Waanders, 1986, p. 17)

On the other hand, Mayall (1979) indicated that design is full of paradoxes, and most designs are related to the principle of totality. Totality is fundamental to any design task, yet it is often overlooked by the designers themselves. Mayall emphasized that totality should be at the heart of and definition of design. However, no matter what sort of product or system be used, all design characteristics are interrelated. Therefore, when designing products, the designers should know the interrelationship among all characteristics and all the requirements necessary to obtain these characteristics. In other words, a design product should be evaluated as a whole, with all characteristics instead of individual parts.

Thus, when evaluating a design project, does the use of the structured form of evaluation employing a clear list of each criterion of the objective derive the same results as the evaluation that uses a non-structured form based on an evaluation of the totality of the product? Although the investigation of a meaningful evaluation method in design instruction is needed, little research has been done to study the relationship between these two evaluation methods. In addition, to what extent do both evaluation methods affect the evaluation of design projects toward the improvement of graphic design instruction? Thus, it is desirable to investigate whether differences exist between the use of these two evaluation methods—
structured and non-structured—and whether differences of perception exist between students, teachers, and professionals when employing either method.

**Statement of the Problem**

Most evaluations conducted in design education are based on systems developed by the instructor. Student performance and rating scales which are developed by instructors affect the results, and thus, directly influence the quality of evaluation (Perry, 1982). In fact, measurement in design involves both quantities and qualities (Lawson, 1990). The student has a right to be told that the grade reflects effort, growth, or a determination of the degree to which the student is successful in attaining the course objectives.

When evaluating graphic design projects, the method selected for evaluation is important and will affect both student performance and teaching. Furthermore, an understanding of whether or not differences exist between professionals and students may enhance the quality of the instructional objective toward a realistic approach to graphic design education and evaluation. The problems addressed in this study were:

1. Do any differences exist between structured and non-structured evaluation methods when evaluating graphic design projects?

2. Were the results of evaluating a graphic design project consistent when rated by students and professional evaluators using structured and non-structured methods?
Purpose of the Study

The purpose of the study was to compare the judgments of student and professional evaluators when evaluating ten graphic design projects by using structured and non-structured methods. Moreover, a comparison was made between the judgment of student and professional evaluators employing the same method. The evaluations were carried out by distributing two instruments (simple form and evaluation matrix) to 32 senior graphic design students at Iowa State University as student evaluators, and 20 graphic design professionals in central Iowa as professional evaluators who included teachers, graphic designers and graduate students who were majoring in graphic design. Selected criteria related to effective typographic communication were included in the evaluation instrument which was used to evaluate the projects.

Variables of the Study

The variables of this study were as follows:

1. Non-structured method ratings by student evaluators
2. Structured method ratings by student evaluators
3. Non-structured method ratings by professional evaluators
4. Structured method ratings by professional evaluators
5. Individual subscore ratings in the structured method by student evaluators
6. Individual subscore ratings in the structured method by professional evaluators
Questions of the Study

The following research questions were developed for this study:

1. Do the evaluations differ between the structured and non-structured methods completed by both student and professional evaluators?

2. Do the evaluations differ between professional and students evaluators using the structured or non-structured method?

3. Do the evaluations completed by student or professional evaluators differ between structured and non-structured methods?

4. What is the degree of relationship of each subscore to the total score in the evaluation matrix of the structured evaluation method when evaluations are completed by student evaluators or professional evaluators?

Hypotheses of the Study

Based on the preceding questions of this study, the following null hypotheses were formulated:

*Hypothesis 1:* There are no differences between non-structured and structured in the mean rating scores when evaluating graphic design projects.

*Hypothesis 2:* There are no differences in the mean rating scores when evaluated by professional and student evaluators using the structured method.

*Hypothesis 3:* There are no differences in the mean rating scores when evaluated by professional and student evaluators using the non-structured method.
Hypothesis 4: There are no differences between the structured and non-structured method in the mean rating scores when evaluated by professional evaluators.

Hypothesis 5: There are no differences between the structured and non-structured method in the mean rating scores when evaluated by student evaluators.

Hypothesis 6: There is no correlation between each subscore and total score when evaluated by student and professional evaluators using the structured method.

Hypothesis 7: There is no correlation between each subscore and total score when evaluated by student evaluators using the structured method.

Hypothesis 8: There is no correlation between each subscore and the total score when evaluated by professional evaluators using the structured method.

Assumptions of the Study

The assumptions of this study were:

1. The evaluators attempted to make reasonable, professional judgments.

2. The collected data reflected the actual knowledge and judgments of evaluators in the area of graphic design.

3. The sample size was sufficient to provide a reasonable degree of control over type II errors.

4. The non-random sample of this study yielded an acceptable estimate of the population studied.
Limitations of the Study

Graphic design projects vary with emphases such as symbology, 3-D organization, and typography. Gottschall (1989) indicated that typography in graphic design is a vital element in making communication more effective and more efficient. Thus, 10 graphic design projects were evaluated emphasizing typographic problems of sophomore students in a graphic design course.

Definition of Terms

Non-Structured Evaluation Evaluation dependent on casual observation, implicit goals, and unseated criteria, intuitive norms, and subjective judgment.

Structured Evaluation Evaluation dependent on checklists, structures, controlled comparisons, and the use of standardized tests.

Evaluation "The systematic collection of evidence to determine whether in fact certain changes are taking place in the learners as well as to determine the amount or degree of change in individual students" (Bloom et al., 1981, p. 5). "A method through which the instructor can determine how effective the instruction system has been helping the individual student gain knowledge and skills" (Perry, 1982, p. 113).

Measurement The process of using numbers to describe quantity, quality, or frequency according to a set of rules.

Assessment A process of collecting, interpreting, and synthesizing information to and in decision making.
Grading  The process of assigning a symbolic label to accumulated evaluations of student performance at specific times.

Matrix  A kind of structure which places information on coordinates, usually—though not necessarily—horizontal and vertical. Connections among elements of information can be read across the whole field in any direction. A matrix is semi- or non-hierarchical. It is a pattern of connections (Hiebert, 1989).

Test  An instrument, device, or hands-on tasks relative to the prestated goals.

Evaluator  Anyone who accepts and executes responsibility for planning, conducting, and reporting evaluations.

Graphic Design  Graphic design is problem-solving on a flat two-dimensional surface. The designer conceives, plans, and executes designs that communicate a specific message to a specific audience within given limitations—financial, physical, or psychological.
CHAPTER II. REVIEW OF LITERATURE

Introduction

The main purpose of this study was to investigate the relationship between structured and non-structured methods in evaluating graphic design projects. An understanding of theory, concepts, functions, learner performance, instructional objectives, evaluation methods, data collection, etc., are related to a meaningful evaluation. Furthermore, the content of the evaluation instrument regarding the objective—to understand typographic design—is also important. Thus, this chapter presents a review of the literature related to the purpose of this study. Numerous categories of information are examined. Four sections are discussed as follows: (1) An Overview of Evaluation; (2) Evaluation Methods; (3) Evaluation in Visual Arts/Design; and (4) An Overview of Graphic Design.

An Overview of Evaluation

The role and function of educational evaluation

Evaluation is defined in a broad view by Quinn and Hennelly (1981) as "... an inevitable part of any human undertaking and by the belief that sound evaluation can promote the understanding and improvement of education, while faulty evaluation can impair it" (p. 5). Biggs and Collis (1982) stated that the word evaluation "... contains the root word value" (p. 6). They believed that the value is seen in the decisions made by the educator. The educator not only makes decisions of whether to pass or fail student, but also makes decisions about how well students learn.
In his book, *Evaluating Student Performance in Vocational Education*, Wolansky (1984) clarified the differences between measurement and evaluation. Educational measurement is "... the process employed to obtain a quantified representation of the degree to which a student reflects a trait or behavior" (p. 3). When measurement is related to evaluation, it can provide valuable information to the evaluator. Wolansky indicated that evaluation may be viewed as a systematic process for the collection of information and use this information to interpret the results and to make value judgments and decisions. Wolf (1990) pointed out that the distinction between evaluation and measurement is in the different objectives. This means that evaluation is used to describe the effect of treatments. On the other hand, measurement is used for description and comparison of individuals.

Although some differences exist between the researchers views, Popham (1972) offered a more specific definition of educational evaluation and stated that evaluation means appraising the worth of an educational undertaking with a view of making decisions such as in curriculum or a particular instructional procedure. These decisions are the act of assessing merit by judgmentally comparing the performance data that come from a desired standard or criterion of acceptability. Nitko (1983) defined the measurement as "... a procedure for assigning numbers (usually called scores) to a specified attribute or characteristic of persons in such a manner ..." (p. 5). In other words, measurement indicates the quantitative aspects of describing the characteristics or attributes of persons.

The primary task of the educational process is to change the learners in desirable ways, as well as the teachers and curriculum designers/writers. There are a series of decisions that
teachers must make if they are to be effective in helping learners change in desired ways. The role of evaluation is to provide appropriate evidence to help both teachers and learners attain the goals of instruction (Bloom et al., 1971).

Popham (1972) stated that there are two roles of educational evaluation:

The first role of educational evaluation occurs with respect to what is increasingly referred to these days as needs assessment. In the needs assessment operation an educator attempts to identify the goals toward which an educational system ought to be directed. Another way of putting it is that the educator is deciding on the objectives for the educational system. A second main role for educational evaluation is in treatment adequacy assessment, that is, determining the quality of educational means which were designed to accomplish the ends originally decided on through needs assessment. (pp. 3-4)

The representation of the role of evaluation in the educational process is shown in Figure 2.1, which presents the relationships between the objective, learning experiences, and learner appraisal (Wolf, 1990).

Eisner (1985) emphasized the major importance of evaluation. First, one can make some judgments about the educational significance of the content to which students will be exposed. Second is the quality of teaching that is provided. The third, and the characteristic most attended to, is the evaluation of student outcomes.

The characteristics of evaluation, according to Marshall (1972), are:

1. Evaluation is continuous, cooperative and comprehensive.
2. Evaluation is involved with both means and ends. Unless appropriate means are used, ends will not be effectively achieved.
3. Evaluation is concerned with valuing.
4. All of those involved in the process of education contribute to the evaluation.
5. Evaluation leads to next steps. It helps us to answer the questions: From where have we come? Where are we now? Where do we need to go? (p. viii)
Evaluating involves several methods such as comparing, analyzing, and the utilization of criteria. The judgments of evaluation may be qualitative, or quantitative which is based on the nature of the related evidence and objective. Whatever the selection of a proper judgment, the most important point of evaluation for the evaluator should be, as Dave (1971) stated:

... clarifying what is to be evaluated and the purpose of the evaluation; clarifying the uses, functions, or objectives, or roles of the item to be evaluated; deciding whether a qualitative or quantitative judgment is needed; defining or selecting standards; gathering evidence related to each standard; and stating why a given judgment has been made. (p. 117)

The function of evaluation in education, according to Eisner (1985), is to perform a wide variety of assessments, such as to diagnose, revise curricula, compare, anticipate educational needs, and determine if objectives have been achieved. In other words, the main
function of evaluation is to take all the information about student progress and school program effectiveness and then to establish valid judgments using a set of objectives. Bloom et al. (1971) indicated similar views by saying that evaluation serves several functions:

1. Evaluation as a method of acquiring and processing the evidence needed to determine the student's level of learning and the effectiveness of the teaching.
2. Evaluation as including a great variety of evidence beyond the usual final paper-pencil examination.
3. Evaluation as an aid in clarifying the significant goals and objectives of education and as a process for determining the extent to which students are developing in these desired ways.
4. Evaluation as a system of feedback-corrective to determine at each step in the teaching-learning process whether the process is effective or not, and if not, what changes must be made to ensure its effectiveness before it is too late.
5. Evaluation as a tool in educational research and practice for ascertaining whether or not alternative procedures are equally effective in achieving a set of educational ends. (p. 7)

The development of educational evaluation

The primary reforms in educational evaluation have focused on the methods of evaluation. From the 1930s through 1950s, improvements in evaluation were made primarily in evaluating student performance. Especially in the 1930s, the progression of the education movement was focused on new content, new methods, and new materials. Evaluating student performance was developed by new evaluation designs, approaches, and instruments (Quinn & Hennelly, 1981). In fact, according to Wolf (1990), the development of educational evaluation was an integral part of educational process since the late 1920s and early 1930s.

The late 1940s and the 1950s were a time to forget the war throughout American society. There was no particular interest on the part of society in holding educators accountable. There was little requirement for educators to present efficiency and
effectiveness. Educators did collect considerable amounts of data, but there is little evidence to show that these data were used to judge and improve the quality of course or program (Madaus et al., 1983). During the 1940s and 1950s, Ralph W. Tyler developed a rationale which profoundly influenced the practice of educational evaluation. Tyler proposed that 

"... educators should carefully define their objectives and gather the data needed to determine whether they had been achieved" (cited in Quinn & Hennelly, 1981, p. 2).

Tyler believed evaluation was central to the educational process. The rationale included three major elements in the educational process: objectives, learning experiences, and appraisal procedures (cited in Wolf, 1990). Objectives, according to Wolf (1990), were defined as "... one's intentions for an educational endeavor" (p. 11). The objectives represented desired, or valued, performances or behaviors that individuals in a program can develop.

In the late 1950s and early 1960s, evaluation of large-scale curriculum development projects were supported by the federal government. When the United States Congress passed the Elementary and Secondary Education Act of 1965 (ESEA), many developments related to the assessment of educational programs, projects, and materials were increased during the 1960s (Quinn & Hennelly, 1981). This profound change would see evaluation expand as an industry and into a profession dependent on taxpayer support.

There was considerable technical development related to the Tylerian view of evaluation during the 1950s and early 1960s. Tyler's approach in an evaluation required a clear statement of the objectives that could help educators and other professionals to do a
better job. The Tyler rationale also was used extensively to train teachers in test development. Madaus et al. (1983) reported that there were four approaches to evaluation during this period. First, the Tyler approach was employed to help define objectives for the new curricula. Second, new nationally standardized tests were developed to reflect better the objectives and content of the new curricula. Third, the professional-judgment approach was utilized to rate proposals. Finally, field experiments were used to evaluate curriculum development by many evaluators.

In the late 1960s and early 1970s, there were a great number of discussion and concern about how evaluation should be conceived. A number of important evaluations concluded in negative findings and raised serious questions about evaluation in general and certain methodologies in particular.

During the 1970s, to the 1980s, the field of evaluation began to solidify and emerge as a distinct profession related to its offspring of research and testing. This field has advanced notably as a profession. However, evaluators faced an identity crisis. They could not identify themselves as researchers, teachers, testers, administrators, or philosophers. There was no professional organization dedicated to evaluation as a field and no specialized journals to provide or exchange information about evaluation. In addition, during this period, evaluators increasingly realized that the techniques of evaluation must serve "... the information needs of clients; address the central value issues; deal with situational realities; meet the requirements of probity; and satisfy needs for veracity" (Madaus et al., 1983, p. 16).
Evaluation framework

Since educational evaluation is an integral part of instruction, it is necessary to show the role of evaluation in an instruction model to gain a better understanding of educational evaluation. The model in Figure 2.2 demonstrates the determinants of the outcome of learning from two perspectives: the teacher (external), and the learner (internal). Both domains affect each other (Biggs & Collis, 1982).

There are several models for evaluation based on different assumptions of liberal ideology (Madaus et al., 1983). A brief discussion of each evaluation model is stated.

Systems analysis: This approach assumes a few quantitative output measures, such as test scores. The data are usually survey data, and the objective measures are related to the program through corralational analyses.

![Figure 2.2 A model of instruction (Biggs & Collis, 1982, p. 9)](image-url)
Behavioral objectives: The objectives of a program are illustrated in terms of specific student performances. These performances are evaluated by tests using norm-referencing or criterion-referencing. Ralph Tyler initiated this approach.

Decision making: In this approach the evaluators provide information for decision making.

Goal free: This model is concerned primarily with reducing the bias of evaluation.

Art criticism: This model is applied from the traditions of art and literary criticism. The evaluators should be accommodated by experience and training to judge an educational program.

Accreditation: This model is ordinarily employed by a team of outside professionals visiting on-site.

Adversary: This model employing quasi-legal procedures has been used by several people. The evaluation often takes the form of trial-by-jury.

Transaction: This model focuses on the educational processes themselves, such as the classroom, the school, and the program. It uses numerous informal methods of examination and utilizes the case study as the major methodology.

The major evaluation models discussed in the preceding paragraphs are shown in Table 2.1. In addition, Wolansky (1985) proposed a framework in the form of a relatively simple diagram to help the teacher or evaluator identify the components involved in evaluation (Figure 2.3).
<table>
<thead>
<tr>
<th>Model</th>
<th>Proponents</th>
<th>Major Audiences</th>
<th>Assumes Consensus</th>
<th>Methodology</th>
<th>Outcome</th>
<th>Typical Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems Analysis</td>
<td>Rivlin</td>
<td>Economists</td>
<td>Goals; known cause &amp; effect; quantified variables.</td>
<td>PPBS, linear programming, planned variation; cost benefit analysis.</td>
<td>Efficiency</td>
<td>Are the expected effects achieved? Can the effects be achieved more economically? What are the most efficient programs?</td>
</tr>
<tr>
<td>Behavioral Objectives</td>
<td>Tyler, Popham</td>
<td>Managers, psychologists</td>
<td>Prespecified objectives; quantified outcome variables</td>
<td>Behavioral objectives; achievement tests</td>
<td>Productivity, accountability</td>
<td>Are the students achieving the objectives? Is the teacher producing?</td>
</tr>
<tr>
<td>Decision Making</td>
<td>Stufflebeam, Alkin</td>
<td>Decision-makers, esp. administrators</td>
<td>General goals; criteria</td>
<td>Surveys, questionnaires, interviews, natural variation</td>
<td>Effectiveness quality control, control</td>
<td>Is the program effective? What parts are effective?</td>
</tr>
<tr>
<td>Goal Free</td>
<td>Scriven</td>
<td>Consumers</td>
<td>Consequences; criteria</td>
<td>Bias control; logical analysis, modus operandi</td>
<td>Consumer choice; social utility</td>
<td>What are all the effects?</td>
</tr>
<tr>
<td>Art Criticism</td>
<td>Eisner, Kelly, Consumers</td>
<td>Connoisseurs, critics</td>
<td>Consequences; criteria</td>
<td>Critical review</td>
<td>Improved Standards</td>
<td>Would a critic approve this program?</td>
</tr>
<tr>
<td>Accreditation</td>
<td>North Central Association</td>
<td>Teachers, public</td>
<td>Criteria, panel, procedures</td>
<td>Review by panel; self-study</td>
<td>Professional acceptance</td>
<td>How would professionals rate this program?</td>
</tr>
<tr>
<td>Adversary</td>
<td>Owens, Levine, Wolf</td>
<td>Jury</td>
<td>Procedures and judges</td>
<td>Quasi-legal procedures</td>
<td>Resolution</td>
<td>What are the arguments for and against the programs?</td>
</tr>
<tr>
<td>Transaction</td>
<td>Stake, Smith, MacDonald, Parlett - Hamilton</td>
<td>Client, Practitioners</td>
<td>Negotiations; activities</td>
<td>Case studies, interviews, observations</td>
<td>Understanding diversity</td>
<td>What does the program look like to different people?</td>
</tr>
</tbody>
</table>
Figure 2.3 Suggested components of evaluation (Wolansky, 1984, p. 10)

**Bloom’s Taxonomy**

In order to reduce the ambiguity associated with stating instructional objectives and translating these objectives into relevant test items, several researchers divided learning outcomes into cognitive, affective, and psychomotor domains. The well-known and most widespread technique for evaluating the quality of student performance in a systematic way is Bloom’s Taxonomy (Bloom et al., 1956). The most important feature of the Bloom’s Taxonomy is to provide six categories for classifying the cognitive domain which focuses on the broad processes that students respond to examination questions. The intention of
categories that are based on the level of complexity in the responding process is to be
hierarchical in terms of the intellectual demand required of the learner (Ebel & Frisbie, 1991;
Hannah et al., 1977):

1. **Knowledge**: the remembering of information, the dealing with specifics, less demanding
that comprehension, the relating of concepts or the translation of ideas from one form to
another, and the universals and abstractions in a field

2. **Comprehension**: the understanding of the material presented in the course, the explaining,
translating, interpreting to a new form or symbol

3. **Application**: ability to apply learning in new situations, using concepts, principles, rules,
thories, and laws to solve the new problems

4. **Analysis**: breaking course content into its component parts, understanding the
relationships between parts

5. **Synthesis**: putting parts together to form a new whole, production of a unique
communication, a plan, or proposed set of operations; derivation of a set of abstract
relations

6. **Evaluation**: the most demanding and requiring judgments using criteria remembered or
formulated by the learner, judgments in terms of internal evidence, or judgments in terms
of external criteria.

Since the middle 1950s, Bloom’s Taxonomy of the cognitive domain has been utilized
by test constructors because this domain seeks to measure learning outcomes from simple to
complex (Jacobs & Chase, 1992). Teachers use the cognitive domain most frequently to
write their objectives in terms that require simple remembering, or recall of information. This
enables students to realize the teacher’s intention and apply their knowledge in a test situation.

The affective domain focuses on assessment and is used widely in art appreciation,
attitude, and valuing of literature. This taxonomy has less influence than the cognitive domain
in ongoing testing and research. The five levels that are based on internalization are arranged
hierarchically in terms of increasing the level of involvement by the learner (Ebel & Frisbie, 1991; Hannah & Michaelis, 1977; Tittle et al., 1993).

1. **Receiving**: awareness, willingness to receive, controlled or selected attention

2. **Responding**: the responses by the learner that go beyond merely recognizing a phenomenon or concept

3. **Valuing**: acceptance of a value, preference for a value, commitment

4. **Organization**: conceptualization of value, organization of a value system

5. **Characterization by a value or personal judgment of a value, or value complex**: generalized set, characterization, the developing with the addition and integration of personal and social values.

The psychomotor domain was identified by educational psychologists. College examiners did not further develop this domain because they envisioned little or no need for it in terms of the assessment problems confronting them. The categories are related to gross and finely coordinated bodily movement, nonverbal communication, and speech behavior. The creative process in art, architecture, and design requires manipulative skills that are in the psychomotor domain, however, differences in individual abilities should also be considered.

**Evaluation Methods**

There are two steps for the learner during a learning process. According to Biggs and Collis (1982), the first step is to learn some data, such as facts, skills, concepts, or problem-solving strategies. Second is to use those skills, facts, or concepts in ways to explain what the learner learned and employed the skills, facts, or concepts to solve a problem or make a judgment. The second step uses a set of components such as knowledge or skills that may be
independent or integrated with each other. The authors note that, after a learner has been revealed this learning–application process, the educator needs to consider how much has been learned and how well the student has learned.

**Quantitative and qualitative evaluation**

Quantitative (how much) and qualitative (how well) aspects may be involved in evaluation. Quantitative evaluation is concerned mainly with what the teacher wants to measure to explain how much the students have learned; for example, the number of words a student can spell or the number of problems a student can solve correctly. Quantitative judgments estimate a numerical dimension related to frequency of occurrence, amount, monetary worth, or other aspects subject to quantifiable measurement. In fact, quantitative evaluation is used commonly in educational testing. On the other hand, qualitative evaluation involves how well the student performed, such as scoring an essay question. Qualitative judgments of a product or performance involve an estimate of the degree of the existing characteristics, the quality of a product, or the level of development. They usually are made in subjective ways that are rarely for the benefit of the student and are incorporated into the final grade by means of an equally private calculation (Biggs & Collis, 1982; Dave, 1971). A comparison between qualitative and quantitative methods in evaluation is shown in Table 2.2.

Hedrick (1994) pointed out that the terms of qualitative and quantitative evaluation “. . . often are used loosely to refer to a wide range of philosophies and methods” (p. 47), and provide a diagram which may help teachers and evaluators to distinguish between paradigm, design, and method (Table 2.3).
Table 2.2. Attributes of the qualitative and quantitative paradigms (Cook & Reichardt, 1979)

<table>
<thead>
<tr>
<th>Qualitative Paradigm</th>
<th>Quantitative Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advocates the use of qualitative methods.</td>
<td>Advocates the use of quantitative methods.</td>
</tr>
<tr>
<td>Phenomenologism and verstehen; concerned with understanding human behavior from the actor's own frame of reference.</td>
<td>Logical-positivism, seeks the facts of causes of social phenomena with little regard for the subjective states of individuals.</td>
</tr>
<tr>
<td>Naturalistic and uncontrolled observation.</td>
<td>Obtrusive and controlled measurement.</td>
</tr>
<tr>
<td>Subjective.</td>
<td>Objective.</td>
</tr>
<tr>
<td>Close to the data; the insider perspective.</td>
<td>Removed from the data; the outsider perspective.</td>
</tr>
<tr>
<td>Grounded, discovery-oriented exploratory, expansionist, descriptive, and inductive.</td>
<td>Ungrounded, verification-oriented, confirmatory, reductionist, inferential and hypothetical-deductive.</td>
</tr>
<tr>
<td>Valid; real, rich, and deep data.</td>
<td>Reliable, hard, and replicable data.</td>
</tr>
<tr>
<td>Ungeneralizable; single case studies.</td>
<td>Generalizable; multiple case studies.</td>
</tr>
<tr>
<td>Holistic.</td>
<td>Particularistic.</td>
</tr>
<tr>
<td>Assumes a dynamic reality.</td>
<td>Assumes a stable reality.</td>
</tr>
</tbody>
</table>
Table 2.3. Examples of paradigms, design approach, and methods (Hedrick, 1994)

<table>
<thead>
<tr>
<th>Term</th>
<th>Qualitative</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paradigm</td>
<td>Constructivist approach</td>
<td>Positivist (scientific method) approach</td>
</tr>
<tr>
<td>Design</td>
<td>Hermeneutic dialectic</td>
<td>Experimental</td>
</tr>
<tr>
<td></td>
<td>Pattern matching</td>
<td>Quasi-experimental</td>
</tr>
<tr>
<td></td>
<td>Case study</td>
<td>Representative samples</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case study</td>
</tr>
<tr>
<td>Method</td>
<td>In-person interviews (unstructured)</td>
<td>In-person interview (structured)</td>
</tr>
<tr>
<td></td>
<td>Focus groups</td>
<td>Questionnaires</td>
</tr>
<tr>
<td></td>
<td>Thick description</td>
<td>Observational recording</td>
</tr>
<tr>
<td></td>
<td>Observational recording</td>
<td>Administrative records</td>
</tr>
</tbody>
</table>

The term "paradigm," according to Hedrick (1994), refers to the philosophy or school of thought underlying the research approach. Positivist evaluation is based on the scientific method, which searches for the increasing likelihood of objective, unbiased answers to be proven successful in the physical sciences. The role of the evaluator in this approach is providing information to decision-makers. Thus, the responsibility of the evaluator is only to use the results neutrally and objectively. On the other hand, the constructivist approach searches to play an active role in program decision making. The aim of this approach is to change and empower all stakeholders.

The term "design" refers to "... how an approach handles issues of causal attribution and representativeness" (Hedrick, 1994, p. 48). Generally, quantitative study designs employ
experiments and quasi-experimental approaches. The sampling procedures and projection
methods are very important in this approach. Qualitative researchers, however, use the
scientific method of hypothesis testing with qualitative data.

The term "method" refers to how data collection turns out and what form the data
obtain. In other words, the method used should help to identify whether the data exist in a
qualitative or quantitative form. Qualitative methods often focus on techniques that add
perspectives, raise additional issues, and accumulate details, but less attention is focused on
data collection procedures. On the other hand, quantitative methods generally give emphasis
on obtaining specific items of information, and systematic approaches about people or places.

**Norm-referenced and criterion-referenced evaluation**

In order to accumulate evidence that students are learning and mastering essential
knowledge and skills, an evaluation of the performance of the students or learners should be
considered. Generally, there are two ways to judge learner performance. One is through
norm-referenced evaluation. Most school teachers commonly use this method. The term
"norm" refers to normal, usual, or average. Therefore, norm-referenced evaluation indicates
the comparison of a person's score with the average score of some relevant group of people.
However, there is a tendency for teachers to grade the observations and compare individuals' performances in terms of the relative position they hold in some class or known group. As the evaluator uses the norm-referenced judgment, it is focused on how well the student compares to others who are under the same or a similar instructional experience. With norm-referenced
evaluation, teachers or evaluators can make judgments quickly about the best student in a classroom or a particular group when the evaluation is expressed in simple ranking terms.

Another method is criterion-referenced evaluation which compares a person's score with scores that represent distinct levels of performance in some specific content area. In other words, criterion-referenced evaluation intends to assess how well the individual is performing in terms of a known standard or criterion. The criterion-referenced score shows the degree of proficiency achieved by an individual without reference to anyone else (Biggs & Collis, 1982; Ebel & Frisbel, 1991; Popham, 1972; Wolansky, 1985).

Criterion-referenced evaluation has two major characteristics. According to Wolf (1990), first, it employs a narrowly defined topic or skill because the domain measured is restricted to a very clear objective. Second, there is a predetermined standard that separates acceptable from unacceptable performance. When using a predetermined dividing line between acceptable and unsatisfactory performance, it is possible to obtain results from a criterion-referenced test showing that a particular percentage of learners has achieved an acceptable level of performance on a particular objective or skill. The expectation of obtaining such results probably is one of the major interests of criterion-referenced measurement. Biggs and Collis (1982) indicated that criterion-referenced evaluation provides the outline in advance because of the list of standards on what the student needs to do and what mistakes should be avoided.

On the other hand, the problem of using the criterion-referenced evaluation is that many of the objectives of instruction do not meet the specifications of a clear and narrowly
defined domain. Most educational objectives are remarkably complex—even specific objectives are complex—and require the use of several concepts, skills, and abilities.

Criterion-referenced evaluation defines domains narrowly, and the results are to be clearly interpretable. As a variance exists between the nature of educational objectives and the method of testing, educators should put the objectives first (Wolf, 1990).

**Summative and formative evaluation**

Summative evaluation is defined as “grading” by Biggs and Collis (1982). It is directed toward a general assessment of degree of achievement as compared to the larger outcomes. Summative evaluation is used primarily after the completion of a unit or block of instruction, or after the entire course is completed. Moreover, this method is utilized when the evaluator wants to reach a decision as to the adoption of a particular instructional treatment (Bloom et al., 1971; Popham, 1972). The essential characteristic of summative evaluation is to analyze the examination or test results with respect to the effectiveness of learning or instruction.

If the intention of examination is to provide feedback to students, then the evaluator should tell the students which specific objective samples are being assessed and which have been included and are being evaluated. For example, a letter C shows little information to the student whether, in some fashion or other, he or she did better than others or did not do as well as others. Thus, the scores or results of the examination should be conveyed to the students as who did reasonably well in the aspects being evaluated. In other words, the scores
Formative evaluation, on the other hand, is “... an ongoing process that is more remedial in intent” (Biggs & Collis, 1982, p. 7). It is part of an ongoing instructional process to determine the degree of mastery of a given learning task. Formative evaluation is conducted during the instructional process with a view to improve it. In other words, formative evaluation works with incomplete educational treatments, whereas summative evaluation judges a completed educational treatment (Bloom et al., 1971, 1981; Popham, 1972). Thus, formative evaluation should provide the kind of evidence that is desired during the instructional/learning processing and look for a useful method to report the evidence. In other words, the essence of formative evaluation is to tell students as soon as possible what progress they are making.

The purpose of formative evaluation is not to grade or certify the learner. It is to help both the teacher and the learner concentrate on the particular learning requirements toward mastery. When utilizing formative evaluation, the most fundamental action is the selection of a unit of learning. A unit contains subject matter to be learned over a given period of time. Therefore, it is very important to analyze the components of the unit for the purpose of using formative evaluation.

A comparison between summative and formative evaluation based on course, unit, and daily lessons at three instruction levels is shown in Table 2.4.
Table 2.4. A comparison between summative and formative evaluation (Ebel & Frisbie, 1991)

<table>
<thead>
<tr>
<th>Type of evaluation</th>
<th>Course</th>
<th>Unit</th>
<th>Daily Lesson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative evaluation</td>
<td>unit tests, projects,</td>
<td>quizzes, oral questioning</td>
<td>teacher questioning, student</td>
</tr>
<tr>
<td></td>
<td>papers, observation,</td>
<td>results, participation</td>
<td>questioning, quizzes,</td>
</tr>
<tr>
<td></td>
<td>participation patterns</td>
<td>records</td>
<td>activity observation, nonverbal</td>
</tr>
<tr>
<td>Summative evaluation</td>
<td>final examination</td>
<td>unit test, written project,</td>
<td>ordinarily not applicable</td>
</tr>
<tr>
<td></td>
<td>comprehensive project,</td>
<td>work product, presentation,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>research paper,</td>
<td>participation record,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>performance ratings</td>
<td>performance checklist</td>
<td></td>
</tr>
</tbody>
</table>

**Authentic evaluation**

Recent criticisms raise questions about evaluation: (a) traditional testing places emphasis on factual knowledge, thus it promote memorization instead of understanding; (b) tests direct students to believe that there is always one right answer; (c) when should one use multiple indicators of achievement; and (d) any single indicator, such as a percentile rank, limits the amount of information conveyed and is subject to error (Archbald & Newmann, 1988; Collins, 1990; Linn, 1991). Authentic evaluation of educational achievement has two main features: "First, all are viewed as alternatives to traditional multiple-choice, standardized
achievement tests; second, all refer to direct examination of student performance on significant tasks that are relevant to life outside of school” (Worthen, 1993, p. 445).

Authentic assessment can be called by several other terms such as performance assessment, appropriate assessment, alternative assessment, or direct assessment.

Authentic evaluations contain a variety of techniques such as written products, portfolios, check lists, teacher observations, and group projects. All forms of authentic assessment can be summarized numerically or put on a scale to meet federal requirements for comparable quantitative data. Today, authentic assessment is developed and used widely in the arts and apprenticeship systems (National Center for Fairtest, 1992).

Conventional assessment may be adequate if the goal is to monitor performance. However, if the aim is to improve performance across the board, tests should be composed of exemplary tasks, criteria, and standards. In authentic assessments, students have greater clarity about their obligations and are asked to master more engaging tasks. Thus, teachers can see assessment results as meaningful and useful for improving instruction (Wiggins, 1990).

Educational enterprises, having among their objectives the production of objects, will necessarily require product evaluations (Wolf, 1990). Products being scored or rated resemble essay questions in that the learner produces an object in his or her own style. The product can have varying degrees of correctness or completeness. For example, in a course in carpentry, a learner may be required to build a sawhorse according to a set of specifications. The product then can be judged on the basis of several criteria: stability, sturdiness, closeness
of fit of the joints, etc. Separate scores or ratings can be given for each criterion. In appraising learner products, usually it is not always possible to make judgments about the process in production.

**Grading**

Summative tests are used more often to assign grades than for any other purpose in most classroom evaluation. The intention of assigning grades is to categorize each student's learning in relation to that of other students. Traditionally, grades have been used to provide information to students and parents about performance, to certify the successful accomplishment of a grade or level of education, and to satisfy administrative record-keeping requirements. Bloom et al. (1971) pointed out that the most important concern before assigning grades is to understand clearly that assigning a grade is a judgmental, value-laden process.

Terwilliger (1989) indicated six propositions concerning the grading process:

1. Grading should be linked directly to the defined instructional goals that relate both to the content of instruction and to the cognitive complexity of the objective. In other words, the highest grades are assigned to those who achieve the most advanced outcomes.
2. All collected data for judging student achievement should be expressed in quantitative form, using a well-defined system of numeric ratings or check lists.
3. The evaluation data should be collected over a period of time.
4. A failing grade should reflect a categorical judgment that the student does not pass a certain level of requirement of outcome when compared to other students. The failing
grade also has a special significance to students in terms of their future educational options.

5. A clear evaluation plan should be prepared before evaluating. This plan should indicate the timing for data collection, conditions under which data collection takes place, and how the data are to be utilized in making summative judgments about students.

6. Teachers should consider an effective method of grading that is both practical and consistent with the particular classroom setting in which they work. Realistic expectations about evaluating student achievement may be accomplished through trial and error.

An overview of classroom evaluation and grade assignment is shown in Figure 2.4. In order to ensure fairness of grading, the basis on which grades are assigned should be made as specific as possible. This means that the grading process should not only be fair, but also it must be understood by those whom it affects. In other words, the student should be told whether or not the reported grade reflects effort, growth, or the degree to which the student is successful in attaining course objectives.

Teachers most commonly assign grades based on a scale of 0-100 (Bloom et al., 1971). Points may be assigned to each item on the test, then added to reach a total number of points up to 100. The raw score obtained is converted into a percentage, which is divide the number correct divided by the total number of points possible. For example, if the student gets 28 items correct out of 40, the percentage correct is 70. Percentages then are converted into grades—A, B, C, D, and F—based on the widely accepted convention: 100-90 = A, 89-80 = B, 79-70 = C, 69-60 = D, and below 60 = F.
Figure 2.4 Overview of classroom evaluation and grade assignment (Terwilliger, 1989)
Evaluation in Visual Arts/Design

Evaluation is important when assessing student achievement in visual arts/design. Three areas are discussed as follows: (1) measurement of creativity; (2) measurement of design; and (3) selecting criteria.

Measurement of creativity

The quality of teaching and the importance of the subject matter are determined through examination and testing. Students in the arts need evaluation (Waanders, 1986). However, the Council of Chief State School Officers indicated of testing at the state level in art assessment

...there are few acceptable models for assessment in the arts. Historically, there has been disagreement among arts educators regarding measurement of values attitudes and aesthetic understanding that, contrary to most academic subjects, are central to a comprehensive arts education (Wenner, 1986, p. 55).

Assessment of creativity involves mainly the evaluation of visual art. Educational decision-makers should use creativity assessment carefully. Generally, there are three approaches in creativity tests. The first is to test the person to identify a personality trait. The second is to test a product by comparing the work of one individual to that of others, using some set of judgmental criteria. The final approach is to test the process or cognitive style, to seek the means by which problems were solved or work was produced (Castiglione, 1986).

The definition of creativity, according to Castiglione (1986), was "...a sort of fuzzy set, blending at its edges with problem solving, talent, giftedness, and other related but
differing constructs” (p. 27). On the other hand, the expression of creativity often is constructed to mean a unique contribution by an individual in some respect.

Scoring a creative exercise never can be accomplished without some subjective judgment (Waamders, 1986). However, it is difficult to measure creativity because different people interpret things differently when they speak of creativity. The best approach to creativity measurement, therefore, may be to obtain an estimate of the consistency and the accuracy of measurement. According to Hocevar (1979), there are several means for estimating creativity:

- . . . tests of divergent thinking; attitude and interest inventories; personality inventories; biographical inventories; teacher nominations; peer nominations; supervisor rating; judgments of products; nomination of eminent persons; and self-reported creative activities and achievement. (p. 3)

The most important concern when one selects an appropriate approach for measuring creativity is to ask the goal of measurement regarding estimating potential for future development, measuring current level of performance, or predicting how individuals will perform in a particular setting. For example, if the goal of measurement is concerned with an individual’s current level of performance, then creativity should be evaluated as abilities.

**Measurement of design**

When rating students’ performance, the focus should be on several issues that assess the basis of their abilities in creating or producing a product. First, the performance of operations and characteristics of the product should be assessed accurately and objectively. Second, evidence of skill needed should be recorded to evaluate performance observable
directly from the end product. Third, the quality of the product should be determined clearly and criteria for judgment should be identified (Wolansky, 1985).

Jones (1970) indicated that importance in design measurements should relate closely to objectives and criteria. Moreover, in process-product-oriented manipulative skill tests, for the purposes of providing meaningful feedback to the student, Wolansky (1985) noted that "... is more convenient to develop separate instruments or checklist forms to rate procedures and later the product" (p. 40). A checklist provides specific guidance that help students to define good work and make judgments of their work.

Selecting criteria

The key to effective assessment is matching the assessment task to the intended student outcomes. When one judges the assessment task, criteria should be thought of simultaneously (Figure 2.5). Waanders (1986) pointed out that the first concern in arts evaluation must be to carefully define the criteria teachers use, the procedures teachers follow, and the way teachers communicate their judgments. In design evaluation, selecting criteria is one of the most difficult parts of an evaluation task that should translate aims and ideals into measurable realities. Jones (1970) stated that selecting criteria in design "... calls for both scientific precision and artistic flexibility. It is unlikely to be done well at the first attempt and one should expect to learn anew while searching for criteria that are appropriate to each new design" (p. 375).

The principle of selecting criteria is that of operationism. This principle depends on the assumption that no phenomenon exists unless the operations by which its aspect can be
detected by an observer can be specified in detail. Thus, Jones (1970) emphasized that without action there is no beauty in operational existence. This action indicates the presence of beauty regardless of the state of mind of the observer. In other words, "... if it can't be measured, it doesn't exist. . . . The thing to which it refers is incapable of measurement and is not objectively real" (p. 374).

Criteria are necessary because they help one judge complex human performance in a reliable, fair, and valid manner. Scoring criteria guide judgments and are made open to students, parents, and others (Herman et al., 1992). The advantages of using criteria for evaluation are discussed as follows:
1. Criteria clarify instructional goals. They define priority outcomes of the content, knowledge, or skills to be presented. Criteria also help teachers to define excellence and plan how to help students achieve it.

2. Criteria convey to students what constitutes excellence and how to evaluate their own work. They also help students see the perspectives of their teachers, their peers, and even the experts in the field.

3. Criteria provide guidelines that give a clear statement about good work. With clearly defined tasks, teachers or evaluators would come to more consistent, fair, and accurate judgments, than would be possible with vague or unstated tasks.

**An Overview of Graphic Design**

Graphic design is a well-conceived arrangement of graphic elements in visual communication. The elements include display type and illustrations. Display type indicates the message and text that should begin and flow in an orderly way. Illustrations must both attract and be compatible with the intended message. An effective design should be methodical and focused. A designer uses space, depth, balance, texture, and rhythm, among other elements, to orchestrate the presentation of message (Baird et al., 1993).

To achieve a successful graphic design work, all elements must seem to fit together to make a coherent whole. Zelanski (1984) explained that if all elements work together nicely, the whole will show the quality of design work as more than the sum of these elements.

Underlying all graphic design is finding the message content in visual form and making it experienced optimally. The function of words, either as criteria or as content, is crucial. This
means "... being sensitized to how visual form correlates with verbal meaning, and learning the power of the visual to create fundamental and swift impressions in ways that words cannot" (p. 11).

Furthermore, a design considered to be successful is a synthesis of all available information translated into words and images and projected in a dynamic form. The dynamic form depends on the designer's ability to blend the mainstream of visual communications with training, accumulated experience, and innate talent (Hurlburt, 1977).

**Principles of design**

Learning principles from specifics form are the basis for information. Information can be transferred from one situation to another. Principles are important to find an orderly way in chaotic situations. Hiebert (1992) explained that to learn the principle is "... to build an attitude of problem solving that allows seeing the larger context and formulating approaches that have an underlying, systematic structure requires sufficient isolation to allow the deep understanding of the relationship of constants and variables" (p. 9).

Baird et al. (1993) expressed a similar view by saying that a graphics professional depends on some basic design principles to make communication work. However, Zelanski (1984) stated an important reminder that these design principles are valuable as guides, but not as absolute rules to follow. According to Baird et al. (1993), the principles of design include balance, contrast, harmony, and rhythm.

*Balance* refers to creating equilibrium and proportion in a design. In other words, balance may help viewers see a design as a unified whole, and as having an existence that
supports various elements and does not appear heavier on one side than the other (Zelanski, 1984). The aim of good design is a balance that will be pleasing to viewers and will distribute visual weight appropriately throughout a layout (Baird et al., 1993).

There are two basic types of balance—symmetry and asymmetry. Symmetrical balance, according to Arntson (1993), is repeated identical shapes from left to right in mirrored positions on either side of a central vertical axis. Symmetrical balance is shown when figures are exactly alike in visual weight (Zelanski, 1984). Symmetrical design presents a quiet sense of order, stability, and tradition. It uses contrasts of value, texture, and shape to release boredom and introduce variety.

Asymmetrical balance, on the other hand, is shown when figures differ in visual weight. It is balanced through contrast to achieve equal visual weight among elements. In other words, symmetrical balance is presented through likeness while asymmetrical balance is shown through contrast (Arntson, 1993). Asymmetrical design has a greater sense of movement, instability, and relative weights. A good design depends on creating a carefully juggled balance of similarities and contrasts.

Contrast refers to a comparison of dissimilar elements, and helps to identify shapes and enhance visual variety in a composition. It can be shown not only by differences in size, shape, texture, and color, but also through position, direction, and spatial effects (Wong, 1987). Size is the most often used form of contrast in graphic design. The contrast between large and small can be distinct without overpowering the smaller elements so that they cannot contribute their share. The shape of objects produces a directional pull along the main
structural lines. Small complicated sets of contours have greater visual weight, than a large simple one.

Texture contrast indicates that a small, highly textured area will contrast with and balance a larger area of simple texture, applying especially in typographic design. Contrast change may be achieved through color. The brighter and more intense the color, the heavier is the color contrast. A large, dark color shape will be balanced by a small, bright color shape (Arntson, 1993).

_Harmony_ refers to unity. It may be defined as “... a logical and orderly arrangement of parts, with consonance, which is agreement or accord between parts. It is the syntax of the visual grammar” (Myer, 1989, p. 31). The harmony in a design should be the agreement created between visual elements throughout a visual field. This can be made by line, shape, color, space, and motion qualities. The Gestalt psychologists proposed that people perceive things as being harmony if they are coherent, that is, similar in most ways or similar in all ways (Myer, 1989). Baird et al. (1993) pointed out the following guidelines that can be used to achieve unity:

- **Watch proportions.** Avoid too much similarity in shapes, sizes, and emphasis. Monotony is as much a defect as great inequality in the distribution of elements.
- **Avoid clutter.** Give white space some room. Let type and illustration have their territory, but don't crowd them. Clutter is often the result of using conflicting typefaces.
- **Use contrast to attract attention and assert dominance.** But don't overdo this. Understatement can be a helpful ally in design.
- **Don't let visual elements float.** (p. 26)

_Rhythm_ refers to achieving an orderly but not boring representation through design elements such as line, shape, color, and texture. The viewer may see coherence in the rhythm
and follow its movement (Baird et al., 1993). Rhythm can be generated by controlling the
directions of and space between elements. It may be similar, contrasting, parallel, or radiating
(Wong, 1987). Rhythm may be represented through repetition of similar or varying elements
in a design because it represents a particular beat marking the movement of the viewer’s eye
through the work (Zelanski, 1984).

Color

Color serves as a creative and expressive communication aid. It is a link—a carrier of
messages. In a broader sense, colors are a kind of code that can be easily understood and that
constitute an immediate and linear language. Favre and November (1979) observed that color
“... gives life to the visual message, it animates it, accentuates it, and makes it more
perceptible and of easy identification” (Favre & November, p. 13).

Color is not the property of objects or spaces. It is the sensation caused by certain
qualities of light that the eye recognizes and the brain interprets. The concept of color came
from Sir Isaac Newton. Newton discovered that when a beam of sunlight passes through a
glass prism, the white light breaks into a brilliant array of colored bands, later called the
“spectrum”.

Each color has three properties. First is hue, which is the name to identify color.
Second is value, which is the degree of lightness or darkness in a hue. A color can be
darkened by blending it with black and can be lightened by being mixed with a lighter hue of
the same color or by the addition of white. Third is saturation, which is a measure of a color's
purity and brightness (Arntson, 1993). The color blue can be high on the saturation scale but
lower on the value scale when compared. For instance, the color yellow is lighter than blue, but the quality of the blue can be more intense or saturated if it does not contain much gray.

Tastes in color change according to an individual’s age, sex, race, education, and cultural background. Wong (1987) believed that it is difficult to establish specific rules for creating effective color combinations. Color harmony is best defined as successful color combination. It can be achieved when one hue is dominant, all other hues are subordinated, or by placing things in balance. Myers (1989) indicated that color harmony is perceived under the hues, luminance (lightness or darkness), or the degree of purity presented in a sequence of logical and progressive steps. A simple rule was presented by Meyer for creating a good color scheme: Do not give any hues the same area, the same luminance, or the same degree of purity in a composition.

In fact, color harmony always can be expected by the designer, when selecting colors for a design work. Küppers (1981) indicated that the harmonic matching of color can be derived from clearly defined relations between the various hues. In addition, the basis of harmony “… will be found to consist of common features, connections, supplementary factors, or of degree of difference, distinctive features” (p. 144).

**Typographic design**

Typography is a greatly visual form of communication in graphic design. A successful typographic design work can make words sing or dance on the page and shape objects of beauty out of language. The major objective of the ten evaluated projects in the present study was the construction of a typographic poster. In order to develop a valid instrument, an
understanding of the elements of typographic design and typographic syntax and communication is necessary.

Ruedi Ruegg (1989) defined typography as covering

\[ \ldots \text{the entire spectrum of visual communication with script,} \ldots \]

We understand words as content in the first place, then as matters of form and technique. Typography in this sense is consequently limited to the purely visual, to the material on which the typographer must work in order to create an end-product that is placed before the reader. (p. 7)

In other words, typography is working on letters and words with logic order, priorities, sequences, and placement. Typography is an aesthetic intuition. It is combined with knowledge of materials, skill, and tools to make typography work (Burke, 1990).

Letters serve as parts of words and individual letters that are combined into new configurations. A word can be expressed as an idea, object, or event. In their book, *Typographic Design: Form and Communication*, Carter, Day, and Meggs (1993) indicated that “... signs are independent of the things they represent, yet by design they can be made to signify and reveal their meaning” (p. 47). Words are linked to form verbal sentences and typographic lines. The most important structural concerns of words and typographic lines are configuration and placement of lines of type. Lines of type can be arranged symmetrically or asymmetrically. The viewer must sense a clearly established relationship between individual lines of type and the surrounding space.

Dair (1967) used a whole cloth as an example to describe the relationship between the elements in typography:

Like the weaver, the typographer has the option of knitting his lines together tightly, or loosening them up by 'leading' to let the horizontal movement of his lines create a different textural effect. The texture of a type face derives from the distribution of
weight in each individual letter and the design of the letter itself. In the mass, each type face has its own textural and tactile individuality, as recognizable as the weave and feel of different type of cloth. (p. 35)

Moreover, Carter et al. (1993) pointed out the importance of understanding the language of typographic design and that one should learn typographic syntax and communication. The syntax is the connecting of typographic signs to form words and sentences on the page. In other words, typographic syntax is a process which arranges elements into a cohesive whole. Typographic space is the rhythmic and dimensional field in which typographic communication exists. This field consists of positive form (the typographic elements) and void (the spatial ground) upon which the elements are arranged.

The elements of typographic design such as letter, word, line, column and margin are used for a cohesive whole through the use of typographic space, visual hierarchy and grid systems. In addition, the relationship between form and meaning can expand and clarify content through the communicative use of visual form (Carter et al, 1993). The following figures indicate the major components of letterform construction (Fig. 2.6).

*Letter* typically functions as part of a word. By identifying the various components of individual letterforms, designers and typographers need to develop a fundamental and deep understanding of, and sensitivity to, the visual harmony and complexity of the alphabet (Carter et al., 1993). Typography starts with the letter and is the basic unit of all printed communication. A well-designed letter can create a focal point in a typographic design, with various sizes and unusual characteristics to draw the attention of the audience (Dair, 1967).
Fig. 2.6 The construction of letterform

*Word* is when the letter as a unit combines with a group of letters. Thus, words themselves become the unit in the larger relationship of the sentence or paragraph. They merge visually as part of the typographic line. The subtlety of the relationship between words in typographic design lies in the precise space that lets in just enough white space to distinguish one word and the next, but does not "chop up" the line into fragments. There are no firm rules for how large word space should be. Generally, it is based on the variety of each
Typeface and size. Type faces with a large x-height and large size need more space than small x-height and small size (Dair, 1967)

*Typeface* refers to the full range of characters on the printed page, and includes letters, numbers, and punctuation marks with the same style (Baird et al., 1993). Basically, variations of a typeface are based on the range in shape, size, personality, and historical development of the particular typeface (Burke, 1990). As a graphic designer, especially of typographic design, the ability to identify typefaces may help one to select the most appropriate typeface for a design. There are a variety of typefaces. Basically, five families of styles are classified (Carter et al., 1993; Graig, 1980): Old Style, Transitional, Modern, Egyptian, and Sans serif (see Fig. 2.7).

A design family is an integrated group with a geneal name such as Garamond or Times. The family consists of numerous fonts that vary in weight and style. Most type families contains at least two styles: roman and italic. The weight variations range from ultra light to ultra bold. Condensed and expanded versions are included also in some families (see Fig. 2.8).

*Type size or point size* refers to the body size of a letter. Picas are used to measure the line length. There are 12 points in one pica and 6 picas in one inch.

*Weight* is defined by the ratio between the relative width of the strokes of letterforms and their height. On the average, a letter of normal weight consists a stroke width of approximately 15% of height (Baird et al., 1993).
<table>
<thead>
<tr>
<th>Old Style</th>
<th>Transitional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caslon</td>
<td>Baskerville</td>
</tr>
<tr>
<td>Century Oldstyle</td>
<td>Bulmer</td>
</tr>
<tr>
<td>Galliard</td>
<td>Caledonia</td>
</tr>
<tr>
<td>Garamond</td>
<td>Electra</td>
</tr>
<tr>
<td>Goudy</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modern</th>
<th>Sans serif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bodoni</td>
<td>Avant Garde</td>
</tr>
<tr>
<td>Fenice</td>
<td>Eras</td>
</tr>
<tr>
<td>Firenze</td>
<td>Futura</td>
</tr>
<tr>
<td>Firmin Didot</td>
<td>Helvetica</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Egyptian</th>
<th><strong>Egyptienne</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lubalin Graph</td>
</tr>
</tbody>
</table>

Figure 2.7 Five basic styles of typeface

Leading of lines  Leading is the amount of white space between lines of type. Leading provides breathing room for the eyes as they can scan long passages of text. Theories concerning spacing between the lines can be considered when relating the x-height portion of the letter to its ascenders and descenders. Thus, the leading of lines varies based on the
design of the typeface. For example, when the line spacing is solid, letterforms should be selected with a large x-height and short ascenders and descenders. The leading space also can affect the readability of text. Too much space can reduce continuity, thereby decreasing readability. Figure 2.9 presents Helvetica typeface, which appears to be set tighter than the same size typeface with a small x-height, such as Courier. The space between the lines seems wider with Courier, giving it a light tonal value.

Although the rules of determining the leading space can induce one to make mistakes, there are no firm rules to determine how much leading should be selected. The decision should rely on good design judgment (Solomon, 1986).
A valid evaluation method can reflect the truly important outcomes of educational instruction. In visual art and communication, the roles and functions of evaluation are unclear. There is a lack of sufficient consideration to the method of evaluation that may be conducted.
Summary

In this chapter, the literature was examined regarding educational evaluation in general and historical development, role, and function in particular. Various evaluation models and the evaluation method based on the Bloom Taxonomy were reviewed for theoretical enhancement of the present research. Several methods of educational evaluation also were discussed.

A valid evaluation method can reflect the truly important outcomes of educational instruction. In visual art and communication, the roles and functions of evaluation are unclear. There is a lack of sufficient consideration to the method of evaluation that may be conducted. In fact, the evaluation method involved in subjective judgment of visual art and communication is too difficult to be cast as a single method, such as qualitative or quantitative, summative or formative, or authentic evaluation. The complexities of evaluation of visual art and communication are a challenge to those concerned about the essentials of design education.

The evaluated projects used in the present study were research posters which focused on well-known typefaces or typographers in the field. The principles of graphic design and the characteristics of typographic design were described in the literature review to help provide a rationale for selection of criteria in the instrument of the structured method. Clear objectives allow both the instructor and the student to evaluate progress. Well-defined and distinct criteria for evaluating student work provide guidelines that the student can use to understand how their work is to be evaluated or to understand the requirements of the
assignment. The instructor's evaluation is directed toward both student performance and the educational system as a whole.
CHAPTER III. METHODOLOGY

The methodology and procedures used to compare the structured and non-structured methods for the evaluation of graphic design projects are described in this chapter. Six sections are discussed as follows: Design of the Study; Population and Sample; Development of the Instrument; Reliability of the Instrument; Procedures of Data Collection; and Statistical Analysis of the Data.

Design of the Study

This study was a quasi-experimental design. Ary et al. (1990) defined experimental design as "... the conceptual framework within which the experiment is conducted" (p. 310). They described the functions of experimental design as follows: First, it establishes the environment for the comparisons required by the hypotheses of the experiment. Second, it allows the experimenter through statistical analysis of the data to make a meaningful interpretation of the results of the study. In order to conduct an experimental design, the use of randomization procedures and full experimental control are the goal of the study (true experimental design).

However, there are some situations which do not allow one to conduct a true experiment. Quasi-experimental designs may be a suitable method. Borg and Gall (1989) indicated that "... quasi-experimental designs are used when a random assignment of subjects to experimental and control groups is not possible" (p. 688).
Ten projects were used for evaluation in this study. The experiment was conducted mainly in a classroom or a design firm. It was impossible for the researcher to assign subjects randomly to a group. Thus, a quasi-experimental design was used for this study. In this experimental design, the evaluators who used the non-structured method were treated as the control group, and the evaluators who used the structured method were treated as the experimental group. The main purpose of this research was to investigate whether differences existed between these two evaluation methods.

**Population and Sample**

The evaluators involved in graphic design evaluation included college teachers, graphic designers, and college students in central Iowa who were the population in this study. The sample was selected for inclusion in the study from the population and is shown as Table 3.1.

Two kinds of samples were used in this study. First, student evaluators who had taken graphic design courses for three years and had obtained a homogeneous background were the criterion for selection. Thirty-two senior graphic design students at Iowa State University

<table>
<thead>
<tr>
<th>Evaluation method</th>
<th>Student Evaluators</th>
<th>Professional Evaluators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>10</td>
</tr>
</tbody>
</table>
were selected as the sample population. Sixteen students used the structured method and 16 students used the non-structured method for evaluating graphic design projects. Second, professional evaluators from central Iowa who had obtained professional knowledge in graphic design were selected. There were a total of 20 sample observations (10 for each method), by college faculty members, graphic designers, and graduate students who were majoring in the graphic design program.

**Development of the Instruments**

Two self-designed instruments were developed in this study. One was a simple form for the non-structured evaluation method. Another instrument was the evaluation matrix developed for the structured evaluation method. This section of the study will explain the development of the two evaluation forms, and includes an overview of the initial instruments, instrument development, validation by a panel of experts, and the pilot test.

**Overview of the initial instruments**

The initial instrument for the non-structured method consisted of one grading sheet for 10 projects, with 10 boxes on the sheet (see Appendix A). The total points for each project were 100, with a space provided for written comments for each project.

The evaluation matrix for the structured method was based on a literature review of the principles of design, typographical communication, and the researcher's experiences in graphic design evaluation. The inspiration for this matrix was derived from a grading sheet that was designed by the instructor of a graphic design studio course at Iowa State University.
The initial matrix included four parts: (1) Overall design and effectiveness, with four items (30 points); (2) Color, with two items (20 points); (3) Typographic hierarchy, with eight items (30 points); and (4) Image, with three items (20 points). The rationale for assigning weight for each part was based on the literature review and the project objectives (see Appendix B). An evaluation matrix was used for each project, with a total scale of 100 points, which was the same as the evaluation scale for the non-structured method.

**Validation by a panel of experts and the pilot test**

To assure content validity, all criteria used for the evaluation matrix were examined by a panel of eight experts from both academia and private practice. A list of the names and titles of the panel members is shown in Appendix C. In order to select the experts objectively, the panel members were selected based on the recommendation of the president of the Art Directors Association of Iowa. The criteria used for selecting the experts consisted of indicating persons on the list who were involved in graphic design teaching or professional practice with three or more years of experience. Then a letter was sent to each prospective panel member to request their assistance (see Appendix D).

Validity information is concerned with the degree to which the test is capable of achieving certain goals. There are several types of validity corresponding to different aims of testing. Content validity is especially important for achievement and proficiency measures, and for measures of adjustment or social behavior based on observation in selected situations (Isaac & Michael, 1990).
The purpose of content validity, according to Crocker and Algina (1986), is "... to assess whether the items adequately represent a performance domain or construct of specific interest" (p. 218). In other words, content validity is concerned with the extent to which an instrument measures what one thinks it is measuring (Ary et al., 1990).

In order to achieve validity in the instrument, the appropriateness of the terminology used in the matrix was very important. Because of the experts' professional knowledge and experience, their comments and suggestions provided a great number of improvements for the instrument. The final instrument still included four subscores but the terminology was modified and an item was added to the revised instrument (making a total of 17 to 18 items).

The revised instrument was then pilot tested with a group of 10 persons who included teachers, graphic designers, graduate students, and undergraduate students majoring in graphic design. Minor modifications were made to the final evaluation matrix (see Appendix E).

Reliability of the Instrument

Reliability is an extremely important characteristic that indicates the quality of the instrument (Borg & Gall, 1989). The reliability of a measurement instrument may be defined as "... the degree of consistency with which it measures whatever it is measuring. This quality is essential in any kind of measurement" (Ary et al., 1990, p. 268).

However, there are many testing situations where there is only a single administration of one form of a test. How consistently examiners' perform on the test can be generalized to the domain of items on the single test form (Crocker & Algina, 1986). For estimating the
internal consistency of test scores from a single form, Cronbach's coefficient alpha is widely used in the Statistical Analysis System (SAS) package.

Borg and Gall (1989) reported that “... reliability coefficients vary between values of .00 and 1.00, with 1.00 indicating perfect reliability, which is never attained in practice, and .00 indicating no reliability” (p. 259). The reliability of the structured method was computed as .93 for all 18 items. According to Chase (1978), if reliability coefficients are equal or greater than 0.85, a high degree of reliability can be attained.

**Procedures of Data Collection**

Approval to administer the tests was obtained from the University Committee on the Use of Human Subjects in Research (see Appendix F). The procedures of the data collection were divided into two parts. One part of the data was collected from the student evaluators and another part was gathered from the professional evaluators. Each evaluator was provided with a statement of the project objectives and a cover letter explaining the purpose of the instrument and requesting their voluntary participation in the study (see Appendix G). Ten projects were demonstrated to the evaluators during a 50-minute testing period.

In order to gather reliable data, the participants among the professional evaluators were selected from different areas of graphic design, including academic, advertising, studio, and publication. The two evaluation instruments were assigned randomly to the participants. The total responses among the professional evaluators were 20, which was a return rate of 100%.
Student evaluators took the evaluation during a graphic design senior studio class held at Iowa State University. Students in two different sections of the studio class were randomly given the instruments. The responses of the students evaluators numbered 32, which was a response rate of 100%.

The collected evaluation forms first were examined for missing data and then entered into a computer. Data from each item were coded as shown in Table 3.2.

Table 3.2. Coding format

<table>
<thead>
<tr>
<th>Item</th>
<th>Column No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of evaluators</td>
<td>1</td>
</tr>
<tr>
<td>Type of methods</td>
<td>2</td>
</tr>
<tr>
<td>Rating score</td>
<td>3-4</td>
</tr>
<tr>
<td>Project no.</td>
<td>4-5</td>
</tr>
<tr>
<td>Code no.</td>
<td>7-9</td>
</tr>
<tr>
<td>Criterion item</td>
<td>10-45</td>
</tr>
</tbody>
</table>

Statistical Analysis of Data

The data obtained from each evaluator were analyzed by the Statistical Package for the Social Sciences (SPSS-X), which involved an examination of the demographic data and testing of the hypotheses. Mean scores were computed for all demographic variables in the
study and for the items related to each research hypothesis. Standard deviations were calculated for the demographic factors and for all the items. One way Analysis of variance (ANOVA) was used to analyze two variables. A reliability analysis (Cronbach's alpha) was conducted to establish the internal consistency of the instrument for the structured method.
CHAPTER IV. RESEARCH RESULTS, ANALYSIS, AND FINDINGS

The primary purpose of this study was to compare the use of the structured and non-structured methods of evaluating graphic design projects. The results and interpretation of the statistical analyses used to test the hypotheses of the study are presented in this chapter. This chapter is divided into four sections: (1) General Background of the Projects; (2) Results of Hypotheses Testing; (3) Comparison of Distribution of Rating Scores; and (4) Summary.

General Background of the Projects

In this research, ten research posters were evaluated by student and professional evaluators. Effective visual communication is the most important goal for any kind of graphic design. In order to understand the underlying concept behind the project from the designers' viewpoint, the background of each project will be described in this section. Due to the difficulty of maintaining clarity in the reproduction of each project, only selected text copy will be shown as it relates to the inspiration of the designer and the objective of the evaluation. In order to reproduce the colors with clarity, all the projects are shown in Appendix H.

The studio course was designed to emphasize the fundamental practice of graphic design for sophomore students. There were two objectives in this typographic project, wherein the students were to select only one project to complete. The first objective was to study an individual type family and learn its history and individual characteristics. Then the
student was to apply that knowledge to the design of the typographic poster. The purpose of this phase of the project was to promote an awareness and interest in the selected typeface.

The second objective was to study an individual typographer and to design a poster using the information studied. The purpose of the second objective was to promote an awareness and interest in the selected typographer.

The order of the projects was assigned randomly. Thus, there was no preference regarding the quality of the projects.

**Project 1 - Alexander Rodchenko (typographer)**

Rodchenko was an ardent Communist who delivered a spirit and eagerness to experiment to typography, montage, and photography. In 1923, Rochdchenko began to design magazines using contrasting bold, blocky type and hard-edge shapes against softer forms and edges of photomontages. The following is part of the text:

> We well know the power of propaganda. Nine tenths of the victories in the war and of our economic successes were due to the effectiveness and strength of our propaganda. . . Advertising is industrial and commercial propaganda . . . This weapon, commercial propaganda cannot be left in the hands of . . . foreign bourgeois elements.

**Project 2 - Gill Sans (typeface)**

Gill Sans was designed by Eric Gill, a British designer. In 1928, Gill Sans was first shown at a trade conference. A year later, the face then became the most popular typeface used in Britain and the United Kingdom. After World War II, Gill Sans was exported to the United States. This typeface employed the principles of proportion and shape from old-style
Roman designs, but was distributed with serif (a combined use of hairline and thickness). This combination makes Gill Sans one of the most beautiful, readable, and well-conceived typefaces of modern times (Brown, 1989; Haley, 1992). Part of the content from this project is as follows:

Gill’s sans serif eschewed those qualities of ornament and decoration implicit in serif construction, but retained enough colour and relief in the distribution of weight throughout each character and throughout the fount, to save the type from that wearing monotony normally inseparable from sans serif settings of any length. Gill’s is the most readable and legible of all modern sans serif designs, yet even his design has several limitation as a text type.

**Project 3 - Univers (typeface)**

In 1956, Adrian Frutiger who was born and educated in Switzerland, designed a family of over twenty variations on a sans serif face. Univers became the most important typeface and was basic to the post-war sans serifs because of its skillfully unobstrusive detailing. It eliminated everything but the essential forms of the letters and was carefully and delicately drawn. Moreover, Univers was designed geometrically, but was more complex than Futura (also designed by Frutiger). Part of the content from this project is as follows:

Adrian Frutiger’s typeface Univers reaches out into all fields of human activity, for which it provides a greatly extended palette of typographic variations. Limited to regular, italic, and bold in traditional typography-Univers extended those standards seven fold. The san-serif typeface has been known for the unity of it’s 21 fonts. Each font designed for the purpose of establishing continuity with the other. All 21 fonts have the same x-height and ascender and descender lengths forming a whole which can be used together with complete harmony. Univers has been called a universal typeface and has revolutionized typography’s realm of diversity.
Project 4 - Neville Brody (typographer)

Neville Brody was born and grew up in North London. He started in A-Level Art, which was very much from a Fine Art viewpoint. Then Brody changed to a three-year graphics program of courses at the London College of Printing. His tutors damned his work as uncommercial. Part of text is shown below:

In 1988, Neville Brody is the best known British graphic designer of his generation. His record cover designs have been highly regarded but most of all his work on magazines has transformed the way in which designers and readers approach the medium.

Project 5 - Goudy Old Style (typeface)

Goudy Old Style was designed by Fredric Goudy, a self-taught typographer from Bloomington, Illinois. Goudy Old Style was the most successful and satisfying of Goudy types (Carter, 1987). This type face is perceived of short ascenders and descenders as a design flaw. All the characters have very open counters and the width of the capitals is considerably broad, giving the face a classical standing. The following is a sample of the text:

Fredric Goudy felt that good manners in typography were as influential as more self-assertive types. His design of the conservative Goudy Old Style in 1915 was 1 of more than 123 type faces he developed. Plagued by fires that destroyed his studios, Goudy's old-fashioned work ethic and zest for life led him to become one of the most prolific typographers to date.

Project 6 - Frank Armstrong (typographer)

Frank Armstrong has become a leader in contemporary typographic design. He has been successful in combining both science and structure with art, resulting in very appealing designs. The following is a sample of the text:
Armstrong has designed a number of posters for musical events, in which he creates dramatic compositions that elegantly express the information. Armstrong had discovered there to be a unique similarity between music communication and typographic communication. He developed a unique ability in translating music into a typographic sense.

**Project 7 - William Caslon (typographer)**

In 1720, William Caslon built a type foundry in London. He also produced a line of distinct English that was excellently suited to text in mass. Although the characteristics of each letter look awkward, Caslon’s designs show a vivid, unaffected quality, and yield an even color in spite of the contrast. A portion of the text on the poster work is shown as follows:

Most type critics and historians contend that, given sufficient talent, it is relatively easy to create beautiful typefaces, but that it is altogether more difficult to produce a type of high utilitarian value. Caslon was able to do both. For over 200 years, Caslon was the type of choice among printers and typographers. It was used to set nearly every form of printed material from fine books to high-pressure advertising, to the most mundane ephemera.

**Project 8 - Frederic Goudy (typographer)**

The background of Frederic Goudy was discussed in Project 5. Part of the text in this project is shown as follows:

Goudy’s achievements are even more remarkable in that he was self-taught, making his first designs at the age of 30, and manufacturing his own type after 60. As a designer, Frederic Goudy displayed originality and great technical skill. As a printer, he developed a distinct personal style. First and foremost, Goudy realized that type design is not the rendering of individual letters, but the creation of the most versatile form of visual communication.
Project 9 - David Carson (typographer)

Due a lack of information about David Carson, only the quotation from the project is presented.

If the work is just centered around stylistic or surface considerations, with no attention to concept, I think people will tire of it much quicker. But a lot of this newer work will never infiltrate the mainstream anyway. It will just be a matter of degrees of acceptability. It’s not going to be a complete turn around with graphic design in this country. There are just more avenues where it is beginning to open up.

Project 10 - Bookman (typeface)

Bookman was designed from Oldstyle Antique, seen initially in the 1850s. The American Type Founders released the face when the merger of several foundries placed many versions under its control. The characteristics of this type face are large and open, with a high x-height, open counters, and an overall wide stance. Part of the content from this project is as follows:

Benguiat's goal was to design a typeface family with a clear resemblance to previous Bookmans, but which was distinctly different and more versatile. The large x-height and moderate contrast in stroke weight make it highly readable and legible under less than ideal reading conditions.

Results of Hypotheses Testing

_Hypothesis 1:_ There are no differences between structured and non-structured rating scores in evaluating graphic design projects.

\[ H_0: \mu_1 = \mu_2 \text{ and} \]
\[ H_a: \mu_1 \neq \mu_2 \]

Where: \( \mu_1 \) is the mean rating score of non-structured method
\( \mu_2 \) is the mean rating score of structured method
The purpose of this hypothesis was to determine whether there are significant differences between structured and non-structured mean rating scores in ten graphic design projects evaluated by student and professional evaluators. In order to draw a valid conclusion, the ten graphic design projects were tested as a group and then each project was tested individually.

The one-way ANOVA was the first procedure used to test differences among the variances between structured and non-structured methods. If the p-value of the F test is < $\alpha = .05$, then the value of the separate variance estimate will be used to test the null hypothesis. If the p-value of the F test is > $\alpha = .05$, then the value of the pooled variance estimate will be used to test the null hypothesis. The results for the overall group were shown in Table 4.1.

From the data provided in Table 4.1, since the probability value of the F test = .000 < $\alpha = .05$, the value of the separate variance estimate was used. The probability value of the

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>Pooled Var. Est. t-value</th>
<th>Pooled Var. Est. t-prob</th>
<th>Separate Var. Est. t-value</th>
<th>Separate Var. Est. t-prob</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-structured</td>
<td>260</td>
<td>78.85</td>
<td>12.48</td>
<td>1.92</td>
<td>.000</td>
<td>4.87</td>
<td>.000</td>
<td>4.83</td>
<td>.000</td>
</tr>
<tr>
<td>Structured</td>
<td>247</td>
<td>72.34</td>
<td>17.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
t test = .000 < \alpha = .05, thus, null Hypothesis 1 was rejected. It was concluded that the mean rating scores by student and professional evaluators were significantly different between structured and non-structured evaluation methods.

Since the result of null Hypothesis 1 was rejected based on the overall group of projects, the individual projects were then tested to determine whether there were significant differences between structured and non-structured methods. The same procedure was used to test the 10 projects. The results are presented in Table 4.2. Three projects (Project 4, 6, and 9) had significant differences in mean rating scores between structured and non-structured methods by student and professional evaluators.

In order to compare the distribution of the rating scores between evaluators and evaluation methods, the Pearson's correlation method was used. When comparing the non-structured and structured methods, the results from the Pearson's correlation method yielded a value of $r = -.21193$. This means that lower rating scores were found for the structured method while higher rating scores were found for the non-structured method.

**Hypothesis 2:** There are no differences in project ratings when evaluated by professional and student evaluators using the structured method.

\[ H_0: \mu_1 = \mu_2 \text{ and} \]
\[ H_1: \mu_1 \neq \mu_2 \]

Where: $\mu_1$ is the mean rating score of student evaluators
$\mu_2$ is the mean rating score of professional evaluators

The purpose of this hypothesis was to detect whether there were differences in mean rating scores between student and professional evaluators using the structured method. The
Table 4.2. Test for differences between structured and non-structured rating scores in each project by student and professional evaluators combined

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est. t-value</th>
<th>Separate Var. Est. t-value</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>83.31</td>
<td>8.09</td>
<td>2.84</td>
<td>.012</td>
<td>.156</td>
<td>1.43</td>
<td>.162</td>
</tr>
<tr>
<td>Structured</td>
<td>25</td>
<td>78.8</td>
<td>13.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>82.73</td>
<td>8.78</td>
<td>3.00</td>
<td>.009</td>
<td>.072</td>
<td>1.8</td>
<td>.080</td>
</tr>
<tr>
<td>Structured</td>
<td>24</td>
<td>76.33</td>
<td>15.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>77.00</td>
<td>13.76</td>
<td>1.52</td>
<td>.305</td>
<td>.109</td>
<td>1.62</td>
<td>.111</td>
</tr>
<tr>
<td>Structured</td>
<td>25</td>
<td>69.96</td>
<td>16.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>70.04</td>
<td>12.94</td>
<td>2.13</td>
<td>.065</td>
<td>.006</td>
<td>2.85</td>
<td>.007</td>
</tr>
<tr>
<td>Structured</td>
<td>25</td>
<td>57.08</td>
<td>16.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>25</td>
<td>78.68</td>
<td>10.76</td>
<td>1.93</td>
<td>.116</td>
<td>.89</td>
<td>.378</td>
<td>.88</td>
</tr>
<tr>
<td>Structured</td>
<td>24</td>
<td>57.08</td>
<td>18.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>80.63</td>
<td>12.61</td>
<td>2.48</td>
<td>.026</td>
<td>.020</td>
<td>2.37</td>
<td>.023</td>
</tr>
<tr>
<td>Structured</td>
<td>25</td>
<td>69.60</td>
<td>19.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>73.96</td>
<td>9.94</td>
<td>1.68</td>
<td>.210</td>
<td>.59</td>
<td>.556</td>
<td>.59</td>
</tr>
<tr>
<td>Structured</td>
<td>24</td>
<td>72.04</td>
<td>12.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>87.00</td>
<td>9.39</td>
<td>1.54</td>
<td>.293</td>
<td>.30</td>
<td>.767</td>
<td>.30</td>
</tr>
<tr>
<td>Structured</td>
<td>25</td>
<td>86.12</td>
<td>11.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>75.23</td>
<td>16.93</td>
<td>1.29</td>
<td>.527</td>
<td>.221</td>
<td>.031</td>
<td>.221</td>
</tr>
<tr>
<td>Structured</td>
<td>25</td>
<td>64.00</td>
<td>19.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>26</td>
<td>79.81</td>
<td>11.43</td>
<td>1.10</td>
<td>.813</td>
<td>1.66</td>
<td>.103</td>
<td>1.66</td>
</tr>
<tr>
<td>Structured</td>
<td>25</td>
<td>74.36</td>
<td>11.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
procedure used to test the null hypothesis was the same as the procedure for testing Hypothesis 1. The results for the overall group are presented in Table 4.3.

From the data provided in Table 4.3, since the probability value of the F test $= .077 > \alpha = .05$ and the probability value of the pooled variance estimate was $.000 < \alpha = .05$, null Hypothesis 2 was rejected. It was concluded that there were significant differences in the rating scores between student and professional evaluators when using the structured method.

Since the result of null Hypothesis 2 was rejected, based on the overall group of projects, the individual projects were then tested to determine whether there were significant differences between student and professional evaluators when using the structured method. The same procedure was used to test the 10 projects. The individual results of the 10 projects are presented in Table 4.4. There were two projects (Project 2 and Project 7) where significant differences were found in the mean rating scores between student and professional evaluators when using the structured method.

Table 4.3. Test of differences in rating scores between student and professional evaluators using the structured method

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est.</th>
<th>Separate Var. Est.</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stu. Evaluator</td>
<td>159</td>
<td>69.69</td>
<td>15.92</td>
<td>.077</td>
<td>.000</td>
<td>4.83</td>
<td>.000</td>
<td>Rejected</td>
</tr>
<tr>
<td>Pro. Evaluator</td>
<td>88</td>
<td>77.12</td>
<td>18.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4. Results of individual projects testing

<table>
<thead>
<tr>
<th>Project</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est.</th>
<th>Separate Var. Est.</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project 2</td>
<td>73.43</td>
<td>17.56</td>
<td>2.09</td>
<td>.036</td>
<td>4.20</td>
<td>.000</td>
<td>Rejected</td>
</tr>
<tr>
<td>Project 7</td>
<td>71.89</td>
<td>16.92</td>
<td>2.15</td>
<td>.033</td>
<td>4.25</td>
<td>.000</td>
<td>Rejected</td>
</tr>
</tbody>
</table>
Table 4.4 Test of the ten projects for differences in rating scores between student and professional evaluators using the structured method

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est. t-value</th>
<th>t-prob</th>
<th>Separate Var. Est. t-value</th>
<th>t-prob</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 1.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>78.56</td>
<td>14.08</td>
<td>1.06</td>
<td>0.983</td>
<td>-.11</td>
<td>0.911</td>
<td>-.11</td>
<td>.911 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>79.22</td>
<td>13.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 2.</td>
<td>Stu. Evaluators</td>
<td>15</td>
<td>71.80</td>
<td>17.41</td>
<td>9.36</td>
<td>0.003</td>
<td>-2.00</td>
<td>0.057</td>
<td>-2.48</td>
<td>.023 Rejected</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>83.89</td>
<td>5.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 3.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>65.69</td>
<td>15.73</td>
<td>1.20</td>
<td>0.721</td>
<td>-1.75</td>
<td>0.094</td>
<td>-1.70</td>
<td>.109 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>77.56</td>
<td>17.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 4.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>56.06</td>
<td>13.79</td>
<td>3.73</td>
<td>0.027</td>
<td>-.35</td>
<td>0.728</td>
<td>-.30</td>
<td>.773 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>58.89</td>
<td>26.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 5.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>71.56</td>
<td>15.29</td>
<td>1.75</td>
<td>0.464</td>
<td>-1.86</td>
<td>0.077</td>
<td>-2.04</td>
<td>.056 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>8</td>
<td>83.00</td>
<td>11.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 6.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>67.38</td>
<td>15.72</td>
<td>2.8</td>
<td>0.082</td>
<td>-.74</td>
<td>0.467</td>
<td>-.64</td>
<td>.533 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>73.56</td>
<td>26.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 7.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>67.81</td>
<td>12.25</td>
<td>1.50</td>
<td>0.603</td>
<td>-2.53</td>
<td>0.019</td>
<td>-2.71</td>
<td>.015 Rejected</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>8</td>
<td>80.50</td>
<td>9.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 8.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>84.56</td>
<td>12.68</td>
<td>1.75</td>
<td>0.428</td>
<td>-.89</td>
<td>.384</td>
<td>-.96</td>
<td>.347 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>88.89</td>
<td>9.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 9.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>61.75</td>
<td>12.93</td>
<td>4.61</td>
<td>0.011</td>
<td>-2.77</td>
<td>.448</td>
<td>-.64</td>
<td>.538 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>68.00</td>
<td>27.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 10.</td>
<td>Stu. Evaluators</td>
<td>16</td>
<td>71.93</td>
<td>12.86</td>
<td>1.86</td>
<td>0.377</td>
<td>-1.37</td>
<td>.184</td>
<td>-1.50</td>
<td>.149 Retained</td>
</tr>
<tr>
<td></td>
<td>Pro. Evaluators</td>
<td>9</td>
<td>78.67</td>
<td>9.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In addition, when comparing the distribution of the rating scores between professional and student evaluators using the structured method, the result of the Pearson’s correlation method yielded a value of \( r = .2056 \). This suggests that, on average, when using the structured method, the professional evaluators gave higher scores than did the student evaluators.

**Hypothesis 3:** There are no differences in project ratings when evaluated by professional and student evaluators using the non-structured method.

\[
H_0: \mu_1 = \mu_2 \quad \text{and} \\
H_a: \mu_1 \neq \mu_2
\]

Where: \( \mu_1 \) is the mean rating score of student evaluators \\
\( \mu_2 \) is the mean rating score of professional evaluators

The purpose of this hypothesis was to determine whether there are significant differences in the mean rating scores between student and professional evaluators when using the non-structured method. The procedure to test this null hypothesis was the same as the procedure for testing Hypothesis 1. The results for the overall group are shown in Table 4.5.

From the data provided in Table 4.5, since the probability value of the F test = .001 < \( \alpha = .05 \) and the probability value of the separate variance estimate was .982 > \( \alpha = .05 \); therefore, the test of null Hypothesis 3 was retained. It was concluded that there were no significant differences in the mean rating scores in using the non-structured method between student and professional evaluators.
Table 4.5. Test of differences in rating score between student and professional evaluators using the non-structured method

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est. t-value</th>
<th>t-prob</th>
<th>Separate Var. Est. t-value</th>
<th>t-prob</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stu. Evaluator</td>
<td>160</td>
<td>78.83</td>
<td>10.91</td>
<td>1.84</td>
<td>.001</td>
<td>-.02</td>
<td>.981</td>
<td>-.02</td>
<td>.982</td>
<td>Retained</td>
</tr>
<tr>
<td>Pro. Evaluator</td>
<td>100</td>
<td>78.87</td>
<td>14.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Null Hypothesis 3 was retained based on the aggregate analysis of the group of 10 projects together, but tests of individual projects are still relevant to determine whether there were any significant differences between student and professional evaluators for any given project using the non-structured method. The same procedure was used to test each of the 10 projects. The individual results for the 10 projects are presented in Table 4.6. There were no significant differences found in each project in the mean rating scores between student and professional evaluators using the non-structured method.

When using the Pearson Product-Moment correlation to compare the distribution of rating scores between professional and student evaluators using the non-structured method, the results yielded a value of $r = .0015$. This indicates that, on average, the professional evaluators gave higher scores than did student evaluators when using the non-structured method.
Table 4.6. Test of ten projects for differences in rating scores between student and professional evaluators using the non-structured method

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est. t-value</th>
<th>t-prob</th>
<th>Separate Var. Est. t-value</th>
<th>t-prob</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project 1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>81.88</td>
<td>8.84</td>
<td>1.84</td>
<td>.358</td>
<td>-1.15</td>
<td>.262</td>
<td>-1.23</td>
<td>.230</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>85.60</td>
<td>6.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>82.44</td>
<td>7.31</td>
<td>2.33</td>
<td>.142</td>
<td>-.21</td>
<td>.834</td>
<td>-.19</td>
<td>.851</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>83.21</td>
<td>11.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>76.25</td>
<td>13.85</td>
<td>1.06</td>
<td>.883</td>
<td>-.35</td>
<td>.733</td>
<td>-.34</td>
<td>.736</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>78.20</td>
<td>14.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>68.00</td>
<td>12.01</td>
<td>1.42</td>
<td>.523</td>
<td>-1.02</td>
<td>.319</td>
<td>-.97</td>
<td>.344</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>73.30</td>
<td>14.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>76.13</td>
<td>9.91</td>
<td>1.31</td>
<td>.626</td>
<td>-1.49</td>
<td>.151</td>
<td>-1.44</td>
<td>.166</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>82.50</td>
<td>11.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>82.06</td>
<td>8.56</td>
<td>4.35</td>
<td>.010</td>
<td>.76</td>
<td>.453</td>
<td>.64</td>
<td>.534</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>78.20</td>
<td>17.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>74.31</td>
<td>7.13</td>
<td>3.71</td>
<td>.025</td>
<td>.22</td>
<td>.825</td>
<td>.19</td>
<td>.849</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>73.40</td>
<td>13.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>88.13</td>
<td>7.11</td>
<td>3.07</td>
<td>.053</td>
<td>.77</td>
<td>.451</td>
<td>.68</td>
<td>.511</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>85.20</td>
<td>12.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>77.19</td>
<td>12.17</td>
<td>3.59</td>
<td>.028</td>
<td>.74</td>
<td>.467</td>
<td>.64</td>
<td>.532</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>72.10</td>
<td>23.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Project 10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
<tr>
<td>Stu. Evaluators</td>
<td>16</td>
<td>81.56</td>
<td>8.61</td>
<td>3.04</td>
<td>.055</td>
<td>.99</td>
<td>.332</td>
<td>.88</td>
<td>.398</td>
<td></td>
</tr>
<tr>
<td>Pro. Evaluators</td>
<td>10</td>
<td>77.00</td>
<td>15.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retained</td>
</tr>
</tbody>
</table>
Hypothesis 4: There are no differences between the structured and non-structured method in rating scores when evaluated by student evaluators.

\[ H_0: \mu_1 = \mu_2 \] and 
\[ H_1: \mu_1 \neq \mu_2 \]

Where: \( \mu_1 \) is the mean rating score of non-structured method 
\( \mu_2 \) is the mean rating score of structured method

The purpose of Hypothesis 4 was to detect whether there are differences between structured and non-structured mean rating scores by student evaluators. The procedure to test the null hypothesis was the same as the procedure for testing Hypothesis 1. The results for the overall group are shown in Table 4.7.

From the data provided in Table 4.7, the probability value of the F test = .000 < \( \alpha = .05 \) and the probability value of the separate variance estimate was .000 < \( \alpha = .05 \); therefore, null Hypothesis 4 was rejected. It was concluded that the mean rating scores were significantly different between structured and non-structured methods evaluated by student evaluators.

Table 4.7. Test for differences between structured and non-structured in rating scores by student evaluators

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est.</th>
<th>Separate Var. Est.</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-structured</td>
<td>160</td>
<td>78.83</td>
<td>10.91</td>
<td>2.13</td>
<td>.000</td>
<td>5.98</td>
<td>.000</td>
<td>.97</td>
</tr>
<tr>
<td>Structured</td>
<td>159</td>
<td>69.69</td>
<td>15.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Since the result of null Hypothesis 4 was rejected based on the overall group of projects, the individual projects were then tested to determine whether there were any significant differences between structured and non-structured methods by student evaluators. The same procedure was used to test the 10 projects. The individual results of the 10 projects are presented in Table 4.8. There were 5 projects (Project 2, 4, 6, 9 and 10) by student evaluators where significant differences were found in the mean rating scores between structured and non-structured method.

A comparison using the Pearson correlation of the distribution of the rating scores between structured and non-structured methods by student evaluators yielded a value of $r = -0.3185$. This indicates that, among student evaluators, higher scores were given in the non-structured method than the structured method.

**Hypothesis 5:** There are no differences between the structured and non-structured method in rating scores when evaluated by professional evaluators.

$H_0: \mu_1 = \mu_2$ and $H_a: \mu_1 \neq \mu_2$

Where: $\mu_1$ is the rating score mean of non-structured method $\mu_2$ is the rating score mean of structured method

The purpose of this hypothesis was to determine whether there are significant differences between mean structured and non-structured rating scores by professional evaluators. The procedure to test the null hypothesis was the same as the procedure for testing Hypothesis 1. The results for the overall group are shown in Table 4.9.
Table 4.8. Test for differences between structured and non-structured rating scores in each projects by student evaluators

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est.</th>
<th>Separate Var. Est.</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project 1. Non-structure</td>
<td>16</td>
<td>81.88</td>
<td>8.84</td>
<td>2.54</td>
<td>.081</td>
<td>.80</td>
<td>.432</td>
<td>.80</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>78.56</td>
<td>14.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 2. Non-structure</td>
<td>16</td>
<td>82.44</td>
<td>7.31</td>
<td>5.67</td>
<td>.002</td>
<td>2.24</td>
<td>.033</td>
<td>2.19</td>
</tr>
<tr>
<td>Structured</td>
<td>15</td>
<td>71.80</td>
<td>17.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 3. Non-structure</td>
<td>16</td>
<td>76.25</td>
<td>13.85</td>
<td>1.29</td>
<td>.628</td>
<td>2.02</td>
<td>.053</td>
<td>2.02</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>65.69</td>
<td>15.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 4. Non-structure</td>
<td>16</td>
<td>68.00</td>
<td>12.01</td>
<td>1.32</td>
<td>.600</td>
<td>2.61</td>
<td>.014</td>
<td>2.61</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>56.06</td>
<td>13.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 5. Non-structure</td>
<td>16</td>
<td>76.13</td>
<td>9.91</td>
<td>2.38</td>
<td>.113</td>
<td>.98</td>
<td>.335</td>
<td>.99</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>71.56</td>
<td>15.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 6. Non-structure</td>
<td>16</td>
<td>82.06</td>
<td>8.56</td>
<td>3.37</td>
<td>.021</td>
<td>3.36</td>
<td>.002</td>
<td>3.30</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>67.38</td>
<td>15.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 7. Non-structure</td>
<td>16</td>
<td>74.31</td>
<td>7.13</td>
<td>2.95</td>
<td>.044</td>
<td>1.87</td>
<td>.077</td>
<td>1.83</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>67.81</td>
<td>12.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 8. Non-structure</td>
<td>16</td>
<td>88.13</td>
<td>7.11</td>
<td>3.18</td>
<td>.032</td>
<td>.98</td>
<td>.335</td>
<td>.98</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>84.56</td>
<td>12.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>61.75</td>
<td>12.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 10. Non-structure</td>
<td>16</td>
<td>78.83</td>
<td>10.91</td>
<td>2.23</td>
<td>.131</td>
<td>2.49</td>
<td>.019</td>
<td>2.49</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>69.69</td>
<td>15.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.9. Test for differences between structured and non-structured rating scores by professional evaluators

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est. t-value</th>
<th>t-prob</th>
<th>Separate Var. Est. t-value</th>
<th>t-prob</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-structured</td>
<td>100</td>
<td>78.71</td>
<td>14.71</td>
<td>1.62</td>
<td>.020</td>
<td>.72</td>
<td>.473</td>
<td>.71</td>
<td>.480</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>88</td>
<td>77.12</td>
<td>18.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the data examined in Table 4.9, the probability value of the F test = .020 < α = .05 and the probability value of the separated variance estimate was .480 > α = .05; therefore, null Hypothesis 5 was retained. It was concluded that the rating scores by professional evaluators were not significantly different between structured and non-structured methods.

Null Hypothesis 5 was retained based on the aggregate analysis of the group of 10 projects together, but tests of individual projects are still relevant, to determine whether there were any significant differences between structured and non-structured methods by professional evaluators for any given project. The same procedure was used to test each of the 10 projects. The individual results for the 10 projects are presented in Table 4.10. The results show that, for the ten projects, there were no significant differences found in the rating scores by professional evaluators between structured and non-structured methods.

When comparing the distribution of the rating scores between structured and non-structured methods by professional evaluators, the Pearson’s correlation method showed a
Table 4.10. Test for differences between structured and non-structured rating scores in each project by professional evaluators

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>F Ratio</th>
<th>F-prob</th>
<th>Pooled Var. Est. t-value</th>
<th>t-prob</th>
<th>Separate Var. Est. t-value</th>
<th>t-prob</th>
<th>Results of null Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structure</td>
<td>10</td>
<td>85.60</td>
<td>6.15</td>
<td>4.42</td>
<td>.040</td>
<td>1.32</td>
<td>.205</td>
<td>1.27</td>
<td>.229</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>79.22</td>
<td>13.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structure</td>
<td>10</td>
<td>83.20</td>
<td>11.16</td>
<td>3.85</td>
<td>.071</td>
<td>-1.7</td>
<td>.870</td>
<td>-1.17</td>
<td>.866</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>83.89</td>
<td>5.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structure</td>
<td>10</td>
<td>78.20</td>
<td>14.27</td>
<td>1.46</td>
<td>.581</td>
<td>.09</td>
<td>.930</td>
<td>.09</td>
<td>.931</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>77.56</td>
<td>17.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>10</td>
<td>73.30</td>
<td>14.34</td>
<td>3.45</td>
<td>.083</td>
<td>1.49</td>
<td>.154</td>
<td>1.45</td>
<td>.174</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>58.89</td>
<td>26.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>10</td>
<td>82.50</td>
<td>11.36</td>
<td>1.04</td>
<td>.937</td>
<td>-0.9</td>
<td>.928</td>
<td>-0.9</td>
<td>.928</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>8</td>
<td>83.00</td>
<td>11.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>10</td>
<td>78.20</td>
<td>17.85</td>
<td>2.17</td>
<td>.269</td>
<td>.45</td>
<td>.655</td>
<td>.45</td>
<td>.663</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>73.56</td>
<td>26.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>10</td>
<td>73.40</td>
<td>13.75</td>
<td>1.89</td>
<td>.412</td>
<td>-1.22</td>
<td>.239</td>
<td>-1.27</td>
<td>.223</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>8</td>
<td>80.50</td>
<td>9.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>10</td>
<td>85.20</td>
<td>12.46</td>
<td>1.69</td>
<td>.471</td>
<td>-.72</td>
<td>.483</td>
<td>-.73</td>
<td>.477</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>88.89</td>
<td>9.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>10</td>
<td>72.10</td>
<td>23.06</td>
<td>1.45</td>
<td>.590</td>
<td>.35</td>
<td>.730</td>
<td>.35</td>
<td>.733</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>68.00</td>
<td>27.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project 10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-structured</td>
<td>10</td>
<td>77.00</td>
<td>15.01</td>
<td>2.54</td>
<td>.204</td>
<td>-.29</td>
<td>.778</td>
<td>-.29</td>
<td>.774</td>
<td>Retained</td>
</tr>
<tr>
<td>Structured</td>
<td>9</td>
<td>78.67</td>
<td>9.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
value of $r = -.0526$. This denotes that higher scores were given in the non-structured method than the structured method among professional evaluators.

**Hypothesis 6:** There is no correlation between each subscore and total score when evaluated by student and professional evaluators using the structured method.

$H_0$: $\rho = 0$ and $H_1$: $\rho \neq 0$

Where: $\rho$ is the correlation coefficient between each subscore and total score

The purpose of this hypothesis was to detect whether there is a significant correlation between each subscore (Overall Design, Color Analysis, Typographic Analysis/Syntax, and Image Analysis) and total score. The Pearson’s correlation method was used to test this hypothesis, and the results are shown in Table 4.11.

The correlation coefficients between each subscore and the total score shown in Table 4.11 were: $r = .9215, .8332, .8802,$ and $.8479,$ with $p = .000 < .05$ in each case. Therefore, null Hypothesis 6 was rejected. It was concluded that there was a significant correlation between the subscore and total score. Since the values of $r = .9215, .8332, .8802,$ and $.8479 > 0$, there is a positive relationship between the subscore and total score.

Table 4.11. Correlation coefficients between subscores and total score

<table>
<thead>
<tr>
<th>Total score</th>
<th>Subscore 1 (Overall Design)</th>
<th>Subscore 2 (Color Analysis)</th>
<th>Subscore 3 (Typographic Analysis)</th>
<th>Subscore 4 (Image Analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>.9215**</td>
<td>.8332**</td>
<td>.8802**</td>
<td>.8479**</td>
</tr>
</tbody>
</table>

** significance level < .01
Hypothesis 7: There is no correlation between each subscore and total score when evaluated by student evaluators using the structured method.

\[ H_0: \rho = 0 \text{ and} \]
\[ H_1: \rho \neq 0 \]

Where: \( \rho \) is the correlation coefficient between each subscore and total score

The purpose of this hypothesis was to detect whether there is a significant correlation between each subscore (Overall Design, Color Analysis, Typographic Analysis/Syntax, and Image Analysis) and total score by student evaluators. The Pearson's correlation method was employed to test this hypothesis, and the results are presented in Table 4.12.

As shown in Table 4.12, the correlation coefficients between each subscore and total score were \( r = .9215, .8332, .8802, \) and \( .8479 \), with \( p = .000 < .05 \) in each case. Therefore, null Hypothesis 7 was rejected. It was concluded that there was a significant correlation between subscore and total score by student evaluators. Since the values of \( r = .9175, .8135, .8717, \) and \( .8530 > 0 \), there is a positive relationship between subscore and total score.

Table 4.12. The correlation coefficients between subscores and total score by student evaluators

<table>
<thead>
<tr>
<th>Total score (Overall Design)</th>
<th>Subscore 1 (Color Analysis)</th>
<th>Subscore 2 (Typographic Analysis)</th>
<th>Subscore 3 (Image Analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>.9175**</td>
<td>.8135**</td>
<td>.8717**</td>
</tr>
</tbody>
</table>

** significance level < .01
Hypothesis 8: There is no correlation between each subscore and total score when evaluated by professional evaluators using the structured method.

\[ H_0: \rho = 0 \quad \text{and} \quad H_a: \rho \neq 0 \]

Where: \( \rho \) is the correlation coefficient between each subscore and total score.

This hypothesis was intended to detect whether there is a significant correlation between each subscore (Overall Design, Color Analysis, Typographic Analysis/Syntax, and Image Analysis) and total score by professional evaluators. The Pearson's correlation method was used to test this hypothesis, and the results are presented in Table 4.13.

The correlation coefficients between each subscore and total score by professional evaluators as shown in Table 4.13 were: \( r = .9290, .8536, .8797, \text{and} .8372, \) with \( p = .000 < .05 \). Thus, null Hypothesis 8 was rejected. It was concluded that there was a significant correlation between subscore and total score by professional evaluators. Since the values of \( r = .9290, .8536, .8797, \text{and} .8372 > 0 \), there is a positive relationship between the subscore and total score.

Table 4.13. The correlation coefficients between subscores and total score by professional evaluators

<table>
<thead>
<tr>
<th>Total score</th>
<th>Subscore 1 (Overall Design)</th>
<th>Subscore 2 (Color Analysis)</th>
<th>Subscore 3 (Typographic Analysis)</th>
<th>Subscore 4 (Image Analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>.9290**</td>
<td>.8536**</td>
<td>.8797**</td>
<td>.8372**</td>
</tr>
</tbody>
</table>

** significance level < .01
Comparison of Rating Scores

Presented in this section are the mean rating scores of the ten projects evaluated by student and professional evaluators using the structured and non-structured methods. Means and standard deviations were calculated for each project. The purpose of this section is to show a comparison of the means of the rating scores of each project based on the different methods and the different evaluators.

As shown in Table 4.14 and Figure 4.1, the means of the rating scores of the 10 projects evaluated by student evaluators using structured and non-structured method, ranged from a low of 56.06 (project 4/structured) to a high of 88.13 (project 8/non-structured). In addition, the means of the rating scores using the non-structured method are all higher than for the structured method. This suggests that student evaluators gave higher scores when using the non-structured method. Similarly, lower standard deviations were shown by the non-structured method while higher standard deviations were shown by the structured method.

As evaluated by professional evaluators, the means of the rating scores of the 10 projects, shown in Table 4.14 and Figure 4.2, using structured and non-structured methods, ranged from a low of 58.89 (project 3 & 4/structured) to a high of 88.89 (project 8/structured). The means of the rating scores using non-structured method all higher than structured method. This indicates that professional evaluators gave higher scores when using the non-structured method. Similarly, most lower standard deviations were shown by the
Table 4.14. Means and standard deviations of rating scores using structured and non-structured methods by student and professional evaluators

<table>
<thead>
<tr>
<th>Category</th>
<th>Student Evaluators</th>
<th>Professional Evaluators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Project 1.</td>
<td>16</td>
<td>81.87</td>
</tr>
<tr>
<td>Non-structure</td>
<td></td>
<td>78.56</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>78.56</td>
</tr>
<tr>
<td>Project 2.</td>
<td>16</td>
<td>82.44</td>
</tr>
<tr>
<td>Non-structure</td>
<td>15</td>
<td>71.80</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>71.80</td>
</tr>
<tr>
<td>Project 3.</td>
<td>16</td>
<td>76.25</td>
</tr>
<tr>
<td>Non-structure</td>
<td></td>
<td>65.68</td>
</tr>
<tr>
<td>Structured</td>
<td></td>
<td>65.68</td>
</tr>
<tr>
<td>Project 4.</td>
<td>16</td>
<td>68.00</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>56.06</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>56.06</td>
</tr>
<tr>
<td>Project 5.</td>
<td>15</td>
<td>76.13</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>71.56</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>71.56</td>
</tr>
<tr>
<td>Project 6.</td>
<td>16</td>
<td>82.05</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>67.38</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>67.38</td>
</tr>
<tr>
<td>Project 7.</td>
<td>16</td>
<td>74.31</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>67.81</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>67.81</td>
</tr>
<tr>
<td>Project 8.</td>
<td>16</td>
<td>88.13</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>84.56</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>84.56</td>
</tr>
<tr>
<td>Project 9.</td>
<td>16</td>
<td>77.19</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>61.75</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>61.75</td>
</tr>
<tr>
<td>Project 10.</td>
<td>16</td>
<td>81.56</td>
</tr>
<tr>
<td>Non-structured</td>
<td>16</td>
<td>71.94</td>
</tr>
<tr>
<td>Structured</td>
<td>16</td>
<td>71.94</td>
</tr>
</tbody>
</table>
Figure 4.1 A comparison of mean scores between structured and non-structured methods by student evaluators

Figure 4.2 A comparison of the mean rating scores between structured and non-structured methods by professional evaluators
non-structured method while most higher standard deviations were shown by the structured method.

Summary of Hypothesis Testing

There were eight hypotheses tested, with six rejected in whole or in part and two retained. A summary of results is shown as follows:

*Hypothesis 1 (Rejected in part):* Significant differences were found in the mean rating scores between structured and non-structured method for three of ten graphic design projects that were evaluated by student and professional evaluators combined.

*Hypothesis 2 (Rejected in part):* Significant differences were found in the mean rating scores between student and professional evaluators for two of ten graphic design projects using the structured method.

*Hypothesis 3 (Retained):* No significant differences were found in the mean rating scores using the non-structured method when evaluated by professional and student evaluators.

*Hypothesis 4 (Rejected in part):* Significant differences were found in the mean rating scores between the structured and non-structured method for five of ten projects that were evaluated by student evaluators.

*Hypothesis 5 (Retained):* No significant differences were found in the mean rating scores between structured and non-structured methods when the ten projects were evaluated by professional evaluators.
Hypothesis 6 (Rejected): A significant correlation was found between each subscore and total score when the ten projects were evaluated by student and professional evaluators using the structured method.

Hypothesis 7 (Rejected): A significant correlation was found between each subscore and total score when the ten projects were evaluated by student evaluators using the structured method.

Hypothesis 8 (Rejected): A significant correlation was found between each subscore and total score when evaluated by professional evaluators using the structured method.

Findings

When the ten graphic design projects were evaluated by student and professional evaluators, significant differences were found between structured and non-structured methods. The means of the rating scores using the non-structured method were all higher than the means of the rating scores using the structured method. In addition, the standard deviations of the non-structured method were smaller than the standard deviations of the structured method.

The highest rating score from student evaluators was given to the project 8 by both methods, among professional evaluators, rating scores were the highest in the structured method and second highest in the non-structured method for this project. In the same way, the lowest rating score from student evaluators was given to the project 4 by both methods; among professional evaluators, rating scores were the lowest in the structured method for this
project. This may suggest that a consistency exists for using both structured and non-structured methods by both evaluation groups, especially for the structured method.

There were significant differences found in mean rating scores between student and professional evaluators using the structured method to evaluate projects. The means of the rating scores using the structured method indicate that professional evaluators gave higher scores than did student evaluators. The same results were found when both evaluators used the non-structured method.

No significant differences were found in the rating scores between structured and non-structured methods when evaluated by professional evaluators. However, there were significant differences found in the rating scores between structured and non-structured methods when evaluated by student evaluators. Moreover, when comparing the distribution of the rating scores between structured and non-structured methods, student and professional evaluators all gave higher scores when using the non-structured method.

The relationship between subscores and total score were found to be all highly correlated by both student and professional evaluators.

Summary

The results of the statistical analysis used in testing the hypotheses and the findings of the study are presented in this chapter. The characteristics and background of the ten projects, based on different objectives relating to understanding the effectiveness of design were explained. One-way ANOVA was used in Hypotheses 1 to 5, to test for differences between structured and non-structured evaluation methods, and between student and
professional evaluators. The results of the F-test were examined first, then either a pooled or separate variance estimate was selected, based on the value of F-probability. The Pearson Product-Moment correlation was used in Hypotheses 6, 7 and 8 to test the relationship between subscores and total score in the structured method by student and professional evaluators, respectively.

The results of hypothesis testing were presented. Out of eight hypotheses tested, two were retained and six were rejected wholly or in part. Significant differences were found between structured and non-structured methods by student and professional evaluators. Significant differences were found on mean rating scores when using the structured method by student and professional evaluators. Significant differences were found between structured and non-structured methods by student evaluators. Significant correlations were also found between subscores and total score when student and professional evaluators, and student or professional evaluators separately utilized the structured method.

When comparing the mean scores between evaluation methods and evaluators, the Pearson’s correlation method provided information showing distinctions between rating scores. Most of the lower scores were given when using the structured method. When using either the structured method or the non-structured method, on average, professional evaluators gave higher scores than student evaluators did. Most student or professional evaluators gave lower scores when using the structured method than when using the non-structured method.
CHAPTER V. SUMMARY, DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Within the preceding four chapters of this research, the introduction, review of literature, methodology, and statistical analysis and findings were presented. The purpose of this chapter is to present a summary of the results of the preceding chapters, provide a discussion, draw conclusions based on the research questions, and make recommendations for practice and research.

Summary

Evaluation is a process of gathering information and interpreting information to help evaluators make valid judgments to meet goals and objectives. The teacher needs evaluation as a guide to measure the effectiveness of teaching and to collect evidence indicating where the instruction might be improved. In addition, the student needs to know how well he or she is accomplishing the course objectives.

This experimental research design was conducted with 32 senior graphic design students as student evaluators and 20 graphic design professionals as professional evaluators. Two self-designed instruments were developed and used to evaluate 10 posters emphasizing typographic design. A simple form was used to evaluate the non-structured method whereas an evaluation matrix was used for the structured method.

The results of this study showed that there were significant differences in the mean rating scores between structured and non-structured methods in evaluating graphic design.
projects by student and professional evaluators, significant differences between student and professional evaluators using the structured method, but no significant differences between student and professional evaluators using the non-structured method, and no significant differences between structured and non-structured methods evaluated by professional evaluators. However, significant differences were found between the structured and non-structured method evaluated by student evaluators, and there were significant correlations between subscores and total score evaluated by student or professional evaluators using the evaluation matrix in the structured method.

**Discussion**

The purpose of this study was to compare the judgments of student and professional evaluators when evaluating graphic design projects using structured and non-structured evaluation methods. Several factors can be discussed from the results of the statistical analyses:

1. When comparing the two evaluation methods, statistically significant differences were found between structured and non-structured methods used by students and professional evaluators. Based on the concept of totality of the design project which represents that design, the non-structured method is viewed as a whole with an interrelationship among individual parts. On the other hand, the structured method is based on the concept of a separated whole, divided into individual parts, which provides a systematic method to evaluate a design project. Due to a difference in the methods of evaluation, the results
may indicate that both methods have characteristics with different capabilities to evaluate graphic design.

2. There were no significant differences found in the mean rating scores using the non-structured method when evaluated by student and professional evaluators. However, there were significant differences found in the mean rating scores using the structured method when evaluated by student and professional evaluators. Evaluation using the non-structured method is based on employing one measuring score for one project. On the other hand, the structured method of evaluation is based on a score for each project that is obtained by employing an 18-item evaluation matrix. Furthermore, the standard deviation in the structured method, in most cases, presented a larger number than the non-structured. This may suggest that the rating score obtained by using the non-structured method is more consistent than the rating score obtained by using the structured method.

3. There were no significant differences found between use of structured and non-structured methods by professional evaluators. However, significant differences were found between structured and non-structured methods by student evaluators. Generally, the background of professional evaluators would imply more practical experience than that of student evaluators. Professional evaluators may also have more confidence than student evaluators as well as a better understanding of the finer qualities of good graphic design. Thus, when professional evaluators rate design projects, their results may be more consistent, despite the use of two different evaluation methods. On the other
hand, student evaluators have less experience than professionals. When using the structured method, student evaluators may not be familiar with the evaluation matrix and they may make different judgments when using two different methods to evaluate the same projects. This may explain the inconsistency of the results when using two different evaluation methods by student evaluators.

4. Significant correlations were found between each subscore and total score when the ten projects were evaluated by student and professional evaluators using the structured method. This means that Overall Design, Color Analysis, Typographic Analysis/Syntax, or Image Analysis are all highly correlated to the total score. This may indicate that the rating score of any one of the criterion is an important part when evaluating graphic design projects.

5. On the returned non-structured forms in this study, 23 of 26 evaluators (88%) wrote comments (optional) in order to explain their reasons for the rating score. The content of most comments was similar to the criteria listed in the structured method (see Appendix I). On the other hand, only 6 out of 26 evaluators (23%) using the structured method wrote comments. The content of these comments focused only on the suggestion of the method of design expression. This denotes that the structured method provides more specific information than the non-structured method for evaluators.
Conclusions

In this section an attempt is made to respond to the research questions that guided this study.

1. Do the evaluations differ between the structured and non-structured methods completed by both student and professional evaluators?

There were significant differences found in the rating scores between the structured and non-structured method when the graphic design projects were evaluated by student and professional evaluators. When comparing the means of the rating scores, the non-structured method was given higher rating scores than the structured method. Moreover, the standard deviations of the non-structured method were smaller than the structured method.

2. Do the evaluations differ between professional and students evaluators using the structured or non-structured method?

No significant differences were found in the mean rating scores between student and professional evaluators using the non-structured method. However, there were significant differences found in the mean rating scores between student and professional evaluators using the structured method. When comparing the means of the rating scores, both student and professional evaluators gave higher scores on the non-structured method and lower scores on the structured method. In addition, professional evaluators always gave higher scores than did student evaluators when using either the structured or non-structured method.

3. Do the evaluations completed by student or professional evaluators differ between structured or non-structured methods?

There were no significant differences found in the rating scores between structured and non-structured methods when evaluated by professional evaluators. However, there were
significant differences found in the rating scores between structured and non-structured methods when evaluated by student evaluators.

4. What is the degree of relationship of each subscore to the total score in the evaluation matrix of the structured evaluation method when evaluations are completed by student evaluators or professional evaluators?

Each subscore and total score were found to be significantly correlated by both student and professional evaluators.

Recommendations

In this section, several recommendations of the study are brought forward for practice. The purpose is to provide suggestions for teachers and students of visual communication, and potential areas for further study.

Recommendations for teachers and students

Evaluation is a vital part of any educational activity. There were significant differences found between structured and non-structured methods. Thus, it is strongly suggested that the structured method for evaluating graphic design project should be used in evaluation. When guidelines indicating what constitutes quality work are clearly stated, it helps teachers to communicate to students how to evaluate their own work. Moreover, since the criteria are made public to all students, the scores can be determined in a fair and objective manner, based on whether the production has met each criterion or the quality of the production. Thus, teachers are able to evaluate design work accurately and consistently without bias.
Using the structured method can help inexperienced teachers improve their effectiveness and to establish confidence and experience in instruction. The students know what guidelines they should consider when working on the projects, and the teacher clearly understands what criteria students should meet. Overall, effective use of the structured method will reduce misunderstanding and arguments about the fairness of grading.

On the returned non-structured forms in this study, 88% of evaluators wrote comments (optional) in order to explain their reasons for the rating score. The content of most comments was similar to the criteria listed in the structured method. On the other hand, only 23% evaluators wrote comments when using the structured method. The content of these comments focused only on the suggestion of the method of design expression. This significantly addresses the fact that most evaluators desire to communicate their rationale for evaluation, rather than to award just a rating score.

The findings also have shown that significant differences existed between the structured and non-structured methods by the student evaluators. As discussed in the preceding section, the student evaluators may lack experience in using the structured method. When working on the design projects, self-evaluation is an important part in the design process. Using the structured method to evaluate one’s own project nevertheless will help a student not only to achieve the requirements set by the teacher, but also develop skill in self-evaluation. Furthermore, it will enhance the judgment of the quality of design work.
The role of evaluation is to provide valid evidence to help both teachers and students achieve instructional goals. The structured method provides a systematic evaluation method that may better capture meaningful instructional outcomes.

**Recommendations for further study**

The recommendations for further study are based on the findings and conclusions of the study.

1. This study found significant differences between the structured and non-structured method. Since there is little research in the systematic evaluation method in visual communication, more detailed studies should be conducted using the structured method.

2. It was found that, when comparing the means of the rating scores between the non-structured and structured method, most scores by the non-structured method were higher. Further investigation should be made to ascertain whether the whole of a work is greater than the sum of its parts.

3. The results revealed that, when using either the structured or the non-structured method, on average, professional evaluators gave higher scores than did student evaluators. Since professionals should have obtained more experience and critical judgment experience than students, further investigation of the perception of judgment would prove worthwhile.
REFERENCES


ACKNOWLEDGEMENTS

So many people have impacted my life to make this culmination of my doctoral research possible. First, I thank God for giving me the talents and ability to realize my dreams. I thank Him for a wonderful husband, Jerry, and my son, Jonathan. I have been able to accomplish so much because they have walked along beside me. Jerry, you helped me through many stressful times and we occasionally managed to have fun, too. Jonathan, remember that someday soon, I will be there for you, too, just as you supported and encouraged dad and me!

I am forever grateful for my mother and father who instilled in all their children, the desire to realize our dreams, and that belief in oneself can make anything possible. To my eldest brother, thank you for your emotional support. You started me on the long road to develop skill in art. I remember your words: "Life is very long and you have to make every step count." To my other brothers and sister, thanks for being there for all your support in so many ways. So many times I felt as if you were here, even when we in different parts of the world.

My experiences at Iowa State University will always be remembered. I am especially thankful to my major professor, Dr. John Riley, for his guidance and direction throughout my doctoral program. I am grateful for his commitment and his belief in me. Many thanks to the other members of my committee: Dr. Larry Bradshaw who shared a deep interest in graphic communication and technology; Dr. Mack Shelley who gave me much confidence and support in the statistical analyses to complete the written phase of this research; Dr. Roger Smith who
gave me many ideas through our discussions about research; and Dr. Bill Gillette who encouraged me ever since I was in his audio/visual communications class.

I am grateful to Dr. John Dugger, chairperson, and the faculty of the Department of Industrial Education and Technology for support and guidance through my doctoral program. I am also indebted to several faculty members in the Department of Art and Design who provided their expertise and helped me to formulate my research design and objectives. In particular, my graduate classes during my M.F.A. program were invaluable to the completion of this research.

I am grateful to Rashid Bax who helped me during the research phase by giving me a lot of valuable ideas and advice. To my editor and friend, Pat Hahn, thank you for providing your home, computer system, and assistance to complete this dissertation. I enjoyed your good humor as we worked together.

To the many friends I made at Iowa State University and Ames, thank you for all the good times we shared. Over a seven-year period, we all helped one another to reach our goals to graduate and start a new life in our chosen field. I have so many friends who are scattered throughout the globe and I look forward to visit you again someday.
APPENDIX A. INITIAL INSTRUMENTS
<table>
<thead>
<tr>
<th>Number of Projects:</th>
<th>Points</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall Design &amp; Effectiveness (30 points)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniqueness of visual form (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness of Visual Communication (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Color (20 points)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness of Color (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrelationship of Color Quality (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Typographic Hierarchy (30 points)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Headline</strong></td>
<td>Appropriateness of Size (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appropriateness of Weight (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of Word Spacing (4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appropriateness of Typefaces (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Body Text</strong></td>
<td>Appropriateness of Size (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appropriateness of Weight (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of Leading (4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appropriateness of Typefaces (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Image (20 points)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness of Image(s) (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness of Size (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness of Placement (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Research Poster

<table>
<thead>
<tr>
<th>Number of Projects</th>
<th>Points (1-100)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B. STATEMENT OF PROJECT OBJECTIVE
OBJECTIVE:
Select one of the following objectives:

(1) Study an individual type family in close detail. Learn its history and individual characteristics and apply this knowledge to the design of a typographic poster.

(2) Study an individual typographer. Learn of the person’s history and contributions to the field of typography.

PROBLEM:
For objective (1):
*Design a poster using the typeface you have studied. The poster’s intention is to promote awareness and interest in the typeface.*

The following may be included on the poster:
Name of typeface, name of designer, characteristics of the face, example of the numerals and a small block of pertinent text.

Design of the poster should reflect the feeling of the typeface and possibly of the time period in which it was popular.

For objective (2):
*Design a poster for this typographer using the information you have studied. The poster’s intended use is to promote awareness and interest of this person.*

The following may be included on the poster:
Name of typographer, the image of the person, an example of the typographer’s work as image and an appropriate amount of written text that reflects your research information.

Design the poster to reflect the feeling of the typographer’s style and possibly of the time period in which it was popular. Do not design it as the designer would have. There is a difference.

PRESENTATION:
The dimensions are 11”x17” horizontal or vertical. You may use color. You may use solid geometric shapes, rules and screens, but they should be used sparingly.
APPENDIX C. LIST OF PANEL MEMBERS
## Instrument Validation Panel

<table>
<thead>
<tr>
<th>NAME</th>
<th>POSITION/TITLE</th>
<th>SCHOOL/COMPANY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Messer, Randy</td>
<td>Art Director</td>
<td>Perfection Learning Corp.</td>
</tr>
<tr>
<td>2. Baer, Roger</td>
<td>Associate Professor</td>
<td>Iowa State University</td>
</tr>
<tr>
<td>3. Ure, Cheri</td>
<td>Temporary Assistant Professor</td>
<td>Iowa State University</td>
</tr>
<tr>
<td>4. Mauck, Kent</td>
<td>President</td>
<td>Mauck + Associates</td>
</tr>
<tr>
<td>5. Stiles, Kelly</td>
<td>Senior Art Director</td>
<td>Pattee Design</td>
</tr>
<tr>
<td>6. Cooper Smith, Sally</td>
<td>Art Director</td>
<td>Cooper Smith &amp; Co.</td>
</tr>
<tr>
<td>7. Fontain, Lisa</td>
<td>Associate Professor</td>
<td>Iowa State University</td>
</tr>
</tbody>
</table>
Dear [Name],

I am a graduate student completing my doctoral research under the guidance of Dr. John N. Riley in the Department of Industrial Education and Technology at Iowa State University. Presently, I am conducting a research study about the comparison of structured (using an evaluation matrix) and non-structured (traditional) methods for the evaluation of graphic design projects.

[Name], your name was recommended by Jody Tramontina, President of the Art Directors Association of Iowa. Based on your professional knowledge and experience in graphic design, I am asking your assistance in this study which intends to establish a more reliable evaluation matrix for this structured method. Ten graphic design professionals are being contacted to participate as panel members. This phase of the study will be completed during the month of August.

Please complete the enclosed Agreement of Participate form and return it in the enclosed, stamped envelope. As a participant, you will receive an evaluation matrix at two different times. Your participation in this study is voluntary and any information that you provide will be kept strictly confidential. If you desire, a copy of the results will be provided after the conclusion of this study.

I hope that you will be able to participate in this research. If you have any questions or concerns regarding this study, please contact me at the above address or at 128 B University Village, Ames, IA 50010. My phone and fax numbers are (515) 292-7451.

Sincerely yours,

Jen Yen  
Doctoral Candidate

Dr. John N. Riley  
Professor & Major Advisor
August 1, 1994

[Name]
[Address]
[City, State, Zip Code]

Dear [Name],

Thank you for agreeing to participate in my research study. Enclosed is a first draft of the evaluation matrix which is based on the literature review and the objective of the project. A description of the project is printed on the gray sheet.

I am asking your feedback for each criterion, including the appropriateness of terms and point value for each criterion. Additional criteria and comments for the evaluation matrix are encouraged. The revised matrix may be faxed to me for your convenience.

If you have any questions or concerns regarding this study, please contact me at the above address or write to me at 128 B University Village, Ames, IA 50010. My phone and fax numbers are (515) 292-7451.

Sincerely,

Jen Yen
Doctoral Candidate

Dr. John N. Riley
Professor & Major Advisor
APPENDIX E. REVISED INSTRUMENTS
<table>
<thead>
<tr>
<th>Title of Project: <em>Research Poster</em></th>
<th>Points</th>
<th>Total Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Project:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Overall Design
- Appropriateness of visual form (10 points)
- Unity of components (10 points)
- Visual hierarchy/organization (10 points)

### Color Analysis
- Appropriateness of color (10 points)
- Interrelationship of color (10 points)

### Typographic Analysis / Syntax
- Interrelationship of headline/display type and body text (3 points)

#### Display Type/
Headline
- Appropriateness of size (3 points)
- Appropriateness of weight (3 points)
- Use of letter spacing (3 points)
- Appropriateness of placement (3 points)

#### Body Text
- Appropriateness of size (3 points)
- Appropriateness of weight (3 points)
- Use of leading (3 points)
- Appropriateness of line length (3 points)
- Appropriateness of placement (3 points)

### Image Analysis
- Appropriateness of image(s) (10 points)
- Appropriateness of size (5 points)
- Appropriateness of placement (5 points)

### Comments:


<table>
<thead>
<tr>
<th>Title of Project: <em>Research Poster</em></th>
<th>Total Points (1-100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Project:</td>
<td></td>
</tr>
</tbody>
</table>

Comments:
APPENDIX F. HUMAN SUBJECTS APPROVAL FORM
Information for Review of Research Involving Human Subjects

Iowa State University

(Please type and use the attached instructions for completing this form)

1. Title of Project: A comparison of structured and non-structured methods for the evaluation of graphic design projects.

2. I agree to provide the proper surveillance of this project to insure that the rights and welfare of the human subjects are protected. I will report any adverse reactions to the committee. Additions to or changes in research procedures after the project has been approved will be submitted to the committee for review. I agree to request renewal of approval for any project continuing more than one year.

   Jen Yan
   Typed Name of Principal Investigator
   Date: 8/11/94
   Signature of Principal Investigator

3. Signatures of other investigators

   John N. Riley
   Date: 8/11/94
   Relationship to Principal Investigator: Major Professor

4. Principal Investigator(s) (check all that apply)
   - Faculty
   - Staff
   - Graduate Student
   - Undergraduate Student

5. Project (check all that apply)
   - Research
   - Thesis or dissertation
   - Class project
   - Independent Study (490, 590, Honors project)

6. Number of subjects (complete all that apply)
   - 20 # Adults, non-students
   - 20 # ISU student
   - # minors under 14
   - # minors 14 - 17
   - Other (explain)

7. Brief description of proposed research involving human subjects: (See instructions, Item 7. Use an additional page if needed.)

   see attachments

(Please do not send research, thesis, or dissertation proposals.)

8. Informed Consent:  
   - Signed informed consent will be obtained. (Attach a copy of your form.)
   - Modified informed consent will be obtained. (See instructions, item 8.)
   - Not applicable to this project.
9. Confidentiality of Data: Describe below the methods to be used to ensure the confidentiality of data obtained. (See instructions, item 9.)

Individual responses will not be coded for personal identifiers. Only group data will be reported.

10. What risks or discomfort will be part of the study? Will subjects in the research be placed at risk or incur discomfort? Describe any risks to the subjects and precautions that will be taken to minimize them. (The concept of risk goes beyond physical risk and includes risks to subjects’ dignity and self-respect as well as psychological or emotional risk. See instructions, item 10.)

No discomfort or risk is expected. Participation is voluntary.

11. CHECK ALL of the following that apply to your research:

☐ A. Medical clearance necessary before subjects can participate
☐ B. Samples (Blood, tissue, etc.) from subjects
☐ C. Administration of substances (foods, drugs, etc.) to subjects
☐ D. Physical exercise or conditioning for subjects
☐ E. Deception of subjects
☐ F. Subjects under 14 years of age and/or ☐ Subjects 14 - 17 years of age
☐ G. Subjects in institutions (nursing homes, prisons, etc.)
☐ H. Research must be approved by another institution or agency (Attach letters of approval)

If you checked any of the items in 11, please complete the following in the space below (include any attachments):

Items A - D Describe the procedures and note the safety precautions being taken.

Item E Describe how subjects will be deceived; justify the deception; indicate the debriefing procedure, including the timing and information to be presented to subjects.

Item F For subjects under the age of 14, indicate how informed consent from parents or legally authorized representatives as well as from subjects will be obtained.

Items G & H Specify the agency or institution that must approve the project. If subjects in any outside agency or institution are involved, approval must be obtained prior to beginning the research, and the letter of approval should be filed.
Checklist for Attachments and Time Schedule

The following are attached (please check):

12. [ ] Letter or written statement to subjects indicating clearly:
   a) purpose of the research
   b) the use of any identifier codes (names, #’s), how they will be used, and when they will be
      removed (see Item 17)
   c) an estimate of time needed for participation in the research and the place
   d) if applicable, location of the research activity
   e) how you will ensure confidentiality
   f) in a longitudinal study, note when and how you will contact subjects later
   g) participation is voluntary; nonparticipation will not affect evaluations of the subject

13. [ ] Consent form (if applicable)

14. [ ] Letter of approval for research from cooperating organizations or institutions (if applicable)

15. [ ] Data-gathering instruments

16. Anticipated dates for contact with subjects:

<table>
<thead>
<tr>
<th>First Contact</th>
<th>Last Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/24/94</td>
<td>9/6/94</td>
</tr>
</tbody>
</table>

17. If applicable: anticipated date that identifiers will be removed from completed survey instruments and/or audio or visual tapes will be erased:

   12/31/94

18. Signature of Departmental Executive Officer

   [Signature]

   Date: [Signature]

   Department or Administrative Unit: [Signature]

19. Decision of the University Human Subjects Review Committee:

   [ ] Project Approved
   [ ] Project Not Approved
   [ ] No Action Required

   [Signature]

   Date: [Signature]

   Name of Committee Chairperson: [Signature]

   Date: [Signature]

   Signature of Committee Chairperson: [Signature]
APPENDIX G. COVER LETTERS FOR EVALUATORS
September 1, 1994

Dear Professional Evaluator:

I am conducting a research study in the area of evaluation of graphic design projects to complete the dissertation requirement in my program of study at Iowa State University. Your cooperation is being sought to gather data about the comparison of structured and non-structured methods for the evaluation of graphic design projects.

I am asking graphic design professionals in the central Iowa to participate in this study so that information can be assessed to determine an appropriate evaluation method to employ in college classrooms. As a graphic design professional, the information you supply will be used to develop a method to enhance student self-evaluation in the design process.

Please complete the enclosed evaluation form. It will take approximately 45 minutes to complete. Your participation is voluntary and any information that is provided will be kept strictly confidential. All data will be analyzed and reported as group data only. Your experience and knowledge in typographical projects is beneficial to the success of this research.

I greatly appreciate your assistance in this study. If you have any questions, please feel free to discuss them with me.

Sincerely,

Jen Yen
Doctoral Candidate

John N. Riley
Professor & Major Advisor
September 2, 1994

Dear Student Evaluator:

I am conducting a research study in the area of evaluation of graphic design projects to complete the dissertation requirement in my program of study at Iowa State University. Your cooperation is being sought to gather data about the comparison of structured and non-structured methods for the evaluation of graphic design projects.

I am asking senior graphic design students at Iowa State University to participate in this study so that information can be assessed to determine an appropriate evaluation method to employ in college classrooms. As a graphic designer, the information you supply will be used to develop a method to enhance student self-evaluation in the design process.

Please complete the enclosed evaluation form. It will take approximately 45 minutes to complete. Your participation is voluntary and any information that is provided will be kept strictly confidential. All data will be analyzed and reported as group data only. Your experience and knowledge in typographical projects is beneficial to the success of this research.

I greatly appreciate your assistance in this study. If you have any questions, please feel free to discuss them with me.

Sincerely,

Jen Yen
Doctoral Candidate

John N. Riley
Professor & Major Advisor
APPENDIX H. TEN GRAPHIC DESIGN PROJECTS
Goudy Old Style

Project 5

Project 6

Project 7
Frederic Goudy

Project 8

Bookman

Project 9

Project 10
APPENDIX I. SELECTED COMMENTS FROM EVALUATORS
Selected Comments From Evaluators in Structured and Non-structured Method

<table>
<thead>
<tr>
<th>Non-structured Method</th>
<th>Structured Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories are based on the evaluation matrix of the structured method</td>
<td>* good sense of typography</td>
</tr>
<tr>
<td></td>
<td>* great use of tension in this poster is</td>
</tr>
<tr>
<td></td>
<td>* this piece is static</td>
</tr>
<tr>
<td><strong>Overall Design</strong></td>
<td>* I really enjoy the screened back image</td>
</tr>
<tr>
<td>* good composition</td>
<td>* this whole piece is way too digitized, it can done more effectively in other ways</td>
</tr>
<tr>
<td>* overall good composition</td>
<td>* a good sense of type sensibility here, but special concept needs some work</td>
</tr>
<tr>
<td>* meets objectives</td>
<td></td>
</tr>
<tr>
<td>* nice design</td>
<td></td>
</tr>
<tr>
<td>* great layout</td>
<td></td>
</tr>
<tr>
<td>* overall format lacks unity</td>
<td></td>
</tr>
<tr>
<td><strong>Color Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>* color choices are not reflecting of the time period</td>
<td></td>
</tr>
<tr>
<td>* bright contrasting colors</td>
<td></td>
</tr>
<tr>
<td>* great choice of colors</td>
<td></td>
</tr>
<tr>
<td>* colors don’t really reveal Brody type of style</td>
<td></td>
</tr>
<tr>
<td>* the color are terrible</td>
<td></td>
</tr>
<tr>
<td><strong>Typography Analysis/Syntax</strong></td>
<td></td>
</tr>
<tr>
<td>* text is a little too small</td>
<td></td>
</tr>
<tr>
<td>* excellent size contrast</td>
<td></td>
</tr>
<tr>
<td>* text is hard to read</td>
<td></td>
</tr>
<tr>
<td>* type usage is a bit over simplified</td>
<td></td>
</tr>
<tr>
<td>* body copy too small for a poster</td>
<td></td>
</tr>
<tr>
<td>* some of the type isn’t very readable</td>
<td></td>
</tr>
<tr>
<td>* type to read needs to be large</td>
<td></td>
</tr>
<tr>
<td><strong>Image Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>* good use of images</td>
<td></td>
</tr>
<tr>
<td>* don’t understand the use of the image</td>
<td></td>
</tr>
<tr>
<td>* love the subtle image of Goudy</td>
<td></td>
</tr>
<tr>
<td>* strong images</td>
<td></td>
</tr>
<tr>
<td>* images reflect designers style and time period</td>
<td></td>
</tr>
</tbody>
</table>