INVESTIGATING DIFFERENCES IN FORMATIVE CRITIQUING BETWEEN INSTRUCTORS AND STUDENTS IN GRAPHIC DESIGN

by

Liwei Zhang

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Technology Leadership & Innovation

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Scott Bartholomew, Chair

Department of Technology Leadership & Innovation

Dr. Greg Strimel

Department of Technology Leadership & Innovation

Dr. Esteban Garcia-Bravo

Department of Computer Graphic Technology

Approved by:

Dr. Nathan Mentzer

Head of the Graduate Program

To my mom

ACKNOWLEDGMENTS

I want to express my deep gratitude to Dr. Scott Bartholomew, who has invested a lot of efforts in helping me through the entire process of my thesis work. Thank you for your hours meeting with me every week and being available every time I need your help. I really appreciate your support, guidance and encouragement throughout my studies.

I am particularly grateful for the assistance given by Dr. Esteban Garcia Bravo throughout the implementing of this thesis. Thank you for enabling me to conduct this study in your class. I could not have accomplished this without your support.

My sincere thanks are also extended to Dr. Greg Strimel, whose advice and assistance are always there when I need. Thank you for your insightful comments and hard questions that have enabled me to rethink my thesis from different perspectives.

I would also like to thank Dr. Todd Kelley and Jung Han for their generous help through my protocol analysis. Thank you, Jung, for your time and input being my second coder.

Last but not least, I want to appreciate Dr. Scott Bartholomew, Dr. Nathan Mentzer, and Dr. Patrick Connolly who have funded my master's study. Thank you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Zhang, Liwei. MS
Institution: Purdue University
Degree Received: May 2019
Title: Investigating Differences in Formative Critiquing between Instructors and Students in Graphic Design
Committee Chair: Dr. Scott Bartholomew

Critique is an essential skill of professional designers to communicate success and failure of a design with others. For graphic design educators, including critique in their pedagogical approaches enables students to improve both their design capability and critique skills. Adaptive Comparative Judgment (ACJ) is an innovative approach of assessment where students and instructors make comparisons between two designs and choose the better of the two. The purpose of this study was to investigate the differences between instructors' and students' critiquing practices. The data was collected through think-aloud protocol methods while both groups critiqued the same design projects.

The results indicate that it took students longer to finish the same amount of critiques as those completed by instructors. Students spent more time describing their personal feelings, evaluating each individual design, and looking for the right phrases to precisely express their thoughts on a design. Instructors, with more teaching experience, were able to complete the critique more quickly and justify their critique decisions more succinctly with efficient use of terminology and a reliance on their instincts.

CHAPTER 1. INTRODUCTION

1.1 Background

Critique is a form of assessment in design, in which individuals review design work and give verbal or written comments to designers (Soep, 2005). Critique is considered an important aspect of graphic design education as various forms of critiquing activities allow students to observe design work by others, as well as receive and provide feedback (Barrett, 2000; Dannels & Martin, 2008; Elkins, 2014; Hokanson, 2012; Motley, 2015). Besides receiving feedback from peers and/or instructors, students can also learn how to address design problems as they go through the process of giving feedback to others and as they are exposed to more solutions to a design problem by reviewing others' work (Bartholomew, Strimel, & Yoshikawa, 2018). Critique also helps students develop their abilities to communicate the success and failure of a design with others using professional language (Dannels, Gaffney, & Martin, 2008), which is considered an important skill if students are to become successful professionals in graphic design (Öztürk & Türkkan, 2006). For educators, including critique in graphic design pedagogies may not only help their students develop skills that are necessary in future careers (Motley, 2017), but also provide an opportunity to investigate students' understanding of aesthetic qualities and industry terminology and adjust their teaching accordingly (Soep, 2005).

1.1.1 Peer Critique

Students analyze and assess their peers' design projects and receive feedback from other students during *peer critique*. In such contexts, students can act as both designers and critics. Peer critique is widely used in many university-level studio-based graphic design classes (Barrett, 2000; Bartholomew, Zhang, Garcia-Bravo, Strimel, 2019; Cambre, Klemmer, & Kulkarni, 2018) as it

allows students to integrate what they learned from critiquing with their own design process (Klebesadel & Kornetsky, 2009).

1.1.2 Formative Critique

Formative critique occurs before the final version of a design has been submitted. This allows students time to rethink their design based on what they have learned from the critique session (Motley, 2017). Both instructors and students can act as critics in a formative critique process (Hokanson, 2012). In doing so, this process may be beneficial to students as the timing of feedback enables them to reflect and apply adjustments to their design project prior to final submission (Hattie & Timperley, 2007).

1.1.3 Adaptive Comparative Judgment

One approach to critique, which has gained traction recently in graphic design education, is adaptive comparative judgment (ACJ) (Bartholomew, et al., 2019; Potter et al., 2017). Unlike traditional forms of critique in graphic design where critics look at one design at a time and analyze the apparent successes and failures (Hokanson, 2012), ACJ involves individuals comparing two design projects and selecting which one, of the two displayed, is "better". For example, when using the online software *CompareAssess*, the ACJ process is facilitated as two design artifacts are shown side by side on the computer screen (see Figure 1.1). A viewer—often called a "judge" in ACJ settings—views the items and judges/selects the one that is better based on predetermined criteria. Judges can also leave written comments justifying their decision or feedback for each design through ACJ platforms such as *CompareAssess*. Although the ACJ approach is different than traditionally-used rubric-centered approaches, ACJ as an assessment approach has been found to be both valid and reliable in a variety of settings (Kimbell, 2012; Pollitt, 2004; 2012) including

creative and ill-structured design projects (Potter, et al., 2017; Bartholomew, Strimel, & Jackson, 2018; Bartholomew, Strimel, & Yoshikawa, 2018).

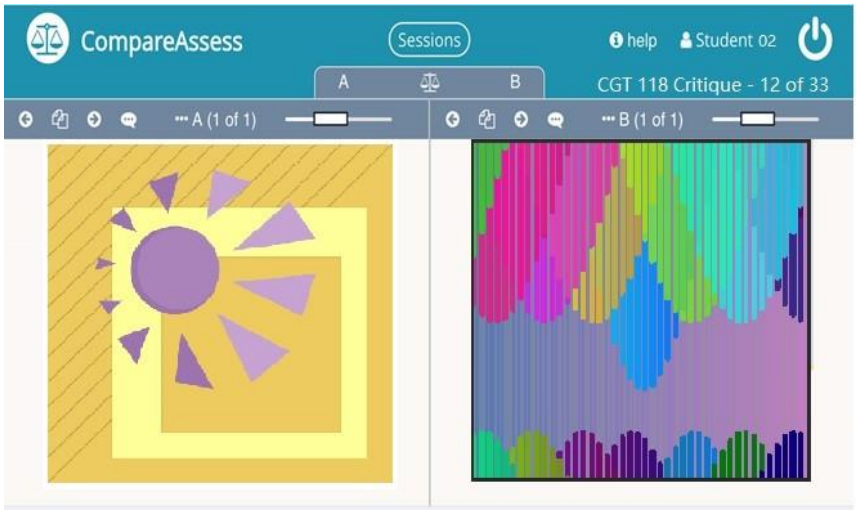


Figure 1.1 *CompareAssess* Interface

CompareAssess has been specifically used in visual design education as an efficient way to assist peer formative critique in visual design classes and has been linked with higher student satisfaction and achievement (Bartholomew, et al., 2019). However, in addition to the learning gains experienced by students using ACJ formatively, previous research with ACJ has also revealed differences in the ACJ results (rankings and parameter values of design artifacts and written feedback associating with each artifact generated by a group of judges) generated by students and instructors. These differences are potentially indicative of different perceptions between groups during design critique around quality (Strimel, Bartholomew, Purzer, Zhang, Yoshikawa, 2018; Bartholomew, Ruesch, Hartell, & Strimel, 2019). Relatedly, Hartell (2015) utilized think-aloud protocols (TAP) to specifically analyze teachers' perceptions while performing ACJ and found that teachers achieved a high agreement in criteria of successful design portfolios. However, Hartell (2015) did not study students to investigate the potential correlation between teachers' and students' perceptions. Therefore, this proposed research seeks to add to the

literature related to critique through ACJ through additional TAP study into the similarities and differences between students and instructors while performing graphic design critique through ACJ. An exploration of the similarities and differences in student and teacher ACJ experiences may yield useful results related to teaching, learning, and assessment in graphic design critique settings.

1.2 Significance

Formative critique, either by peers or teachers, is important in graphic design pedagogy as it allows students to improve their own design and, in the case of peer critique, may lead students to develop critiquing skills through providing and receiving feedback (Elkins, 2014). Both design capabilities and critiquing skills are essential to the development of future graphic designers (Motley, 2017).

Despite research indicating that critique is a necessary and useful pedagogy, results from other research indicates that students often critique differently from teachers (Barrett, 2000; Strimel et al., 2018). This disparity sometimes results in ineffective or even misleading feedback from peers that can negatively affect students' learning (Cambre, et al., 2018). A variety of reasons could lead to differences in peer and teacher critique, including differences between novice and expert designers (Barrett, 2000; Dannels & Martin, 2008), experience in critiquing (Motley, 2015), understanding of design criteria (Hicks, Pandey, Fraser, & Klemmer, 2016), design values held by individuals (Bartholomew, Yoshikawa, et al., 2018), and others not yet identified. The reasons, or rationale, behind critique decisions are often difficult for educators to understand as they do not manifest themselves during classroom practices (Elkins, 2015). As these potential differences between students and teachers are not widely known and understood, it becomes difficult to employ best practices while preparing students in graphic design settings. Research is needed to

identify the gap between how students and teachers critique thereby potentially identifying what educators might do to bridge the gap between student and teacher critiquing. It is important to understand the existing critique differences as an understanding of why students and instructors critique differently may lead to a refinement in teacher pedagogies, which facilitates the development of specific skills in students such as improved critique and increased alignment with best practices demonstrated by instructors.

1.3 Statement of Purpose

The purpose of the research is to investigate differences between instructors and students while performing critique in an introductory undergraduate level graphic design course at a public university in the Midwestern United States. This research involves conducting TAP with student and teacher groups while performing critique of student design projects through ACJ in order to investigate the similarities and differences in design critique processes between groups. The findings may potentially benefit educators in the specific field of study by outlining difference in critique approaches between the two groups. An understanding of these difference may further provide guidance for improving teaching pedagogy towards facilitating student development of critiquing skills.

1.4 Research Question

The research question that guided this study was:

What are the differences, if any exist, between the critiquing practices of instructors and students through adaptive comparative judgement for graphic design projects?

The following sub-questions were used to investigate the research question stated above:

1. What relationship, if any exists, between the ACJ results (ranks and parameter values) generated by students and instructors in an introductory graphic design course?
2. How long does it take for each student/instructor to make each judgment in ACJ?
3. What factors influence students'/instructors' judgment decisions?

1.5 Scope

This study was conducted in the context of an introductory graphic design course in a large Midwest university. Previous research with critique and ACJ had been done in this course and preliminary findings have demonstrated differences in ACJ rankings generated by students and instructors (Bartholomew, et al., 2019). To investigate the differences between students and instructors, three instructors who teach the class and assess student projects and three students who were enrolled in the course at the time of study, were recruited to participate. This study intentionally used a small number of participants due to the limited availability of instructors and students and the intensive nature of TAP.

Each participant performed design critique of student work in two separate ACJ sessions (one for instructors, one for students) using the *CompareAssess* software. *CompareAssess* was chosen as the software interface for facilitating the ACJ because 1) it has demonstrated the ability to facilitate ACJ and provide high-reliable assessment results, 2) in previous work the users have identified its user-friendly nature which facilitates judgments, 3) it can provide an optimized display of images, and 4) it has been used multiple times in similar contexts to this study and has demonstrated high feasibility for student and teacher adoption (Bartholomew, et al., 2019).

All participating students and instructors were asked to practice TAP while assessing the student design work through *CompareAssess*. Both the ACJ and the TAP processes were recorded

by desktop screen recorders which captured both audio statements from the participants and the screen display.

1.6 Assumptions

The assumptions of this study are:

1. All participants honestly express their true feelings and thoughts while think-aloud protocols and surveys are implemented.
2. Participants have never seen any of the design projects before the study – thereby the first time they compare a design project in *CompareAssess* is the first time they see the project – thus potentially lowering bias connected to previous exposure to student work.
3. Student participants have received the same instructions in class before the study.
4. Design projects, randomly selected from previous years for the ACJ sessions, are representatives of student design work.
5. The desktop screen recorder accurately records screen displays and sounds during the study. The use of the screen recorder was tested before the beginning of the study to make sure the settings can accurately capture the data needed in the study.
6. Knowing they are being recorded does not significantly affect participants' behaviors.
7. *CompareAssess* software functions properly with every login and password (prior to the study each login was checked, and the researcher assisted participants to log into *CompareAssess* during the study).
8. Judgments made by each participant in *CompareAssess* are comparable.

1.7 Limitations

The scope of this thesis is limited in the following ways:

1. The researcher has been involved in the prior research project that was conducted in the same course. This prior involvement could lead to a bias of the researcher because the pre-defined codes were generated during a previous project. The design of the data analysis process was utilized in an effort to mitigate this potential bias. This process involved the researcher generating a new list of thematic codes during the transcription of videos. The pre-defined codes from the previous study were only used as a reference and starting point for the new codes.
2. The student participants are from introductory graphic design course at college level. The students' learning experience in graphic design before college could influence the results obtained through the ACJ sessions. These experiences, which may have influenced the students, were identified in pre-surveys inquiring participants' background and experience in graphic design. The results from these pre-surveys were used to mitigate affects.
3. Background of instructors varies in terms of gender, age, years of graphic design education, and years-of-teaching. These differences may affect their critiquing, but this impact was taken into consideration during the data analysis stage.
4. The study only includes one project and a sample of 10 design artifacts due to the time frame of this study. With this data each participant made 34 judgments in about 10-40 minutes while thinking aloud.
5. The study only includes three instructors and three students; the number of participants limits the findings to a small group. However, this constraint is externally imposed as there are only three instructors for the identified course. Thus, the findings from this study may still be useful despite the small number of participants.

1.8 Delimitations

The following statements set up the boundaries of this thesis:

1. The study does not focus on informal graphic design education outside of schools.
2. The study does not look at written comments or feedback provided during ACJ.
3. The study does not focus on the effect on participants' design ability through the design critique processes; it only focuses on their performance during design critique.
4. The study only focuses on ACJ as the technique to assist design critique process, other forms or tools of design critique were not used.
5. This study does not focus on graphic design education at any other level than undergraduate.
6. This study only focuses on differences between instructors and students as they critique design work. The instructor participants are graduate teaching assistants and a professor in the computer graphic technology program. These instructor participants are the ones who give instructions in class and assign grades to undergraduate student design projects. Not all instructors may perform the same in critiquing as professional practicing graphic designers, however, this study does not focus on the differences between students and professionals and/or instructors and professionals.

1.9 Definitions

Adaptive Comparative Judgment (ACJ): an assessment approach that enables judges to choose the better one from two artifacts and generate rankings after repeated pairwise comparisons (Pollitt, 2012). Referred to as *adaptive* because after each artifact has been compared five times, an algorithm is applied to select pairs with similar number of wins and losses (Pollitt, 2012). ACJ has been found to be more reliable than traditional marking in design education (Kimbell, 2012).

CompareAssess: an online software that implemented ACJ to enable users to compare artifacts on the screen and leave comments to each artifact. *CompareAssess* also provides administrators the reliability coefficient of the judgment session, the rankings and parameter values of all artifacts compared, and statistics data (e.g. misfit values) for each item and judge (DigitalAssess, 2018).

Critic: in this thesis, a critic is a term that refers to the person who critically reviews a design artifact.

Critique: a form of assessment in graphic design, in which critics observe, reflect, and comment regarding what a design is about (*explorative critique*), or whether the design is good or not (*argumentative critique*). In educational environments, critique can be done by students (*peer critique*) or teachers (*teacher critique*). Critique can be completed during an assignment (*formative critique*) or after (*summative critique*) the final design is submitted (Soep, 2005; Motley, 2015).

Graphic Design: a professional job, a field of study, an innovative visual artifact, or the process of crafting creative visual content to solve visual communication problems (Motley, 2017).

Judge: in this thesis, a judge is a term that refers to the person who makes judgments in ACJ.

Think-Aloud Protocols (TAP): a research methodology that collects, compiles, and analyzes audio or video records of individual participants speaking aloud their own thoughts as they work on an assigned task or activity (Ericsson & Simon, 1993).

1.10 Summary

This chapter provided the background of the problem and research question to be addressed in this thesis. The chapter also described the significance and scope of the study and discussed the assumptions, limitations, and delimitations of the study.

The following chapter provides a literature review around key elements of the proposed study, including graphic design education, critique in graphic design education, adaptive comparative judgment, and think-aloud protocols.

CHAPTER 2. LITERTURE REVIEW

The purpose of this chapter is to review the literature around critique in graphic design education, adaptive comparative judgment, and think-aloud protocols in educational research. Additionally, current literature is cited which provides a rationale for studying adaptive comparative judgment in design critique and conducting think-aloud protocols in this study.

2.1 Graphic Design Education

Graphic design is a term which can refer to a professional job, a field of study, an innovative visual artifact, or the process of crafting creative visual content to solve visual communication problems (Motley, 2017). It requires learners to develop skills of creativity, visual communication and problem-solving – all skills which are difficult to teach but can often be developed through repetition and guided practice (Elkins, 2015). As a field of study, graphic design is also called visual communication, or visual design. Besides the technical skills of creating visual content, an expert designer is also capable of effectively communicating solutions through visual content as well as evaluating visual artifacts created by oneself and peers (Hokanson, 2012).

One of the primary goals of graphic design education is to prepare students for future careers in the graphic design industry (Motley, 2017). The environment of graphic design classrooms, therefore, often mimics the real working environment of professionals to help students develop skills and habits that will transfer from school to workplace (Öztürk & Türkkan, 2006). Professional training where students can learn by doing and mentoring, including internships and apprenticeships, are commonly used in graphic design education (Motley, 2017). Design studio, a technique where a group of students work on visual design projects while receiving personalized instruction through informal communications with peers and teachers (Neumann, 1988; Barrett,

2000; Utaberta, Hassanpour, Handryant, & Ani, 2013) has been increasingly used in graphic design education at college level (Hokanson, 2012).

2.2 Critique in Graphic Design Education

In the world of design and fine arts, critique is a specific form of assessment practice (Soep, 2005), which involves observation, reflection, articulation, and either written feedback or verbal comments of graphic design artifacts (Motley, 2015). Freedman (2003) described critique as "a form of social knowledge production done in the context of a cultural milieu of art program evaluation and student assessment" (p. 7). Barrett (2000) provided a definition of critique as "an opportunity for students to receive feedback from their teacher and peers regarding the aesthetic, stylistic, creative, and innovative aspects of their studio work" (p. 30), which emphasizes pre-determined standards as a key aspect of critiquing in educational settings. Standards that guide the critique process can be predetermined criteria that involved aesthetic elements, technical aspects of design, or the combination of both (Barrett, 2000; Motley, 2015).

One of the purposes of including critique in graphic design education is that learning how to critique can help students become better prepared for employment (Motley, 2017). Students need to develop critiquing skills to evaluate and discuss the success and failure of a design work - either their own work or work by others (Öztürk & Türkkan, 2006). In addition, research has shown that the more frequently students critique, the better they can perform in critiquing (Motley, 2015) – suggesting students may benefit from exposure to periodic critiquing experiences. Demonacos, Ellis, and Barber (2019) also suggested that doing peer evaluation activity can have a long-term influence on students' learning as students reported how much they valued the activity after its completion.

The process of critique and seeing others' designs and providing and receiving feedback can also help students in different ways (Bartholomew, Strimel, & Yoshikawa, 2018): 1) receiving feedback from peers and teachers can help students improve their own design work; 2) feedback from others enable students to see their own design from different perspectives; 3) understanding how one's own work is interpreted and judged by others is beneficial to learning (Barrett, 2000; Motley, 2015); and 4) providing feedback may give students opportunities to think of how to identify and fix problems in design (Motley, 2017).

2.2.1 Forms of Critique

There are a variety of forms of critique in graphic design classrooms; some examples include exploratory and argumentative critiques (Barrett, 2000), summative and formative critiques (Hokanson, 2012), and peer and teacher critiques (Motley, 2015). Table 2.1 categorizes critique approaches and identifies when to critique, who is providing the critique, and what is being critiqued.

Table 2.1 Forms of Critique

Category	**Critique Form**	**Brief Description**
What	Exploratory	What a design is about (Barrett, 2000)
	Argumentative	Whether a design is good or not (Barrett, 2000)
When	Summative	After final submission of design project, usually by teachers (Hokanson, 2012)
	Formative	Before final submission of design project, by teachers/students (Hokanson, 2012)
Who	Peer	By peers, usually formative (Motley, 2015)
	Teacher	By teachers, summative/formative (Motley, 2015)

2.2.1.1 Exploratory vs. Argumentative Critique

Exploratory critiques center on the *interpretation* of a visual artifact, where critics try to answer the question of what a design is about, while argumentative critiques emphasize an *evaluation* of whether a design is good or not (Barrett, 2000). These two terms originated from Smith's (1973) two forms of aesthetic criticism, 1) exploratory criticism consisting of critics' description, analysis, characterization, and interpretation, and 2) argumentative criticism involving critics' evaluation, argument, and defense. For example, in graphic design, exploratory critiques may include description of the design (e.g. "the designer filled the frame with rows of blocks"), while argumentative critiques may discuss how well an item fits the principles of design (e.g. "the repetition of block rows forms unity within the design").

2.2.1.2 Summative vs. Formative Critique

Summative critiques occur after a "final" version of a design has been submitted. Although both teachers and students may make summative critiques, teachers often perform as critics in summative critiques where they assign grades and make comments to student projects. This process is often accompanied by comments and feedback which are provided to students following the assessment (Motley, 2017). Conversely, formative critiques occur before the final version of a design has been submitted. Formative critiques may allow students time to rethink and refine their design, after receiving feedback from peers and/or teachers prior to final submission (Motley, 2017; Bartholomew, et al., 2019). Research has shown that formative critiques are more frequently used in graphic design educational environments than summative-only approaches (Motley, 2015).

2.2.1.3 Peer vs. Teacher Critique

During peer critique, students evaluate design work created by their peers and may also provide feedback – a process that can be challenging to some students who may feel it hard to

express honest thoughts on a design artifact created by a person they know personally; these difficulties can be especially pronounced when providing negative feedback (Hetland, 2013). While peer critique is widely used to promote critical thinking and develop students' skills in articulating effective feedback with disciplinary terminology, the feedback received from peer critique is often perceived to be less reliable and credible than teacher feedback (McCarthy, 2017). This perception may result in negative consequences, as students may be less likely to revise their design based on ineffective peer feedback; in other cases, this feedback may even be misleading and cause students to perform worse on the assigned work (Cambre, et al., 2018).

Differences between student and teacher performances in critique are usually reflected in what is perceived to be a lower quality in the feedback generated by students (Barrett, 2000). Several recent studies have been done to address how to improve quality of student peer feedback, including providing additional hints during student critiquing (Xu, Huang & Bailey, 2014; Kulkarni, Bernstein, & Klemmer, 2015), refining rubrics to better assist student peer critique (Yuan, Kuther, & Krause, 2016), letting students make pairwise comparisons while leaving feedback to each artifact (Cambre, et al., 2018), and collaborative critique learning through web-based technology (McCarthy, 2017). Differences in peer and teacher critique may originate from an assortment of reasons. As an example, researchers have found that novice and expert designers perform differently in critique (Barrett, 2000; Dannels & Martin, 2008) and students critique differently as they gain more experience (Motley, 2015). Further, critics' understanding of design criteria (Hicks, Pandey, Fraser, & Klemmer, 2016) and their design values (Bartholomew, Ruesch, et al., 2019) may also differ which could lead to different performance in critique. As a specific example, students may come from different cultural backgrounds and thereby value color choices differently.

Despite the differences noted above, additional research is needed to further investigate the gaps between teacher and peer performances in critique to more clearly identify *why* these differences may exist. A better understanding of these differences may allow educators to adjust or improve pedagogies to benefit students' development of critiquing skills in graphic design education and more closely align the critique outcomes from both groups. A closer alignment may result in better designing by students, increased trust in peer critique and feedback, and improved course outcomes.

2.3 Adaptive Comparative Judgment for Design Critique

In traditional forms of design critiques, a design artifact is often compared with a predetermined standard, such as outlined assignment requirements or criteria (Hokanson, 2012). These settings often require professional training for critics to ensure a complete understanding of the standard (Heldsinger & Humphry, 2010). Problems arise in traditional peer critique processes when students with little professional training may lack an understanding of design criteria or have difficulty identifying and describing evidence of a design success or failure (Cambre, et al., 2018). ACJ, however, as an approach to graphic design critique, has demonstrated potential for requiring fewer trainings for students to generate reliable assessment results (Heldsinger & Humphry, 2010) and thus may provide an alternative approach that is more reliable and feasible to conduct design critique (Bartholomew et al., 2019) through making pairwise comparison between two design artifacts based on the predetermined standard (Tarricone & Newhouse, 2016), in contrast to comparing a design artifact to some predetermined standard (e.g., rubric). Bartholomew et al. (2019) and Mortier et al. (2015) also found that students overall liked the experience of making comparative judgment.

ACJ is based on the Law of Comparative Judgment, a psychophysics principle originally published by Thurstone (1927), which states that human beings make decisions more easily when making comparative judgments compared to quality-based subjective decisions. For example, it may be hard for an individual to tell the weight of an object (e.g. a pencil, a piece of paper, etc.), but easier to tell if the piece of paper is lighter than the pencil. This law can be applied to both quantitative (e.g., weights) and qualitative comparisons (e.g., attitudes) (Thurstone, 1927).

Other researchers, building on Thurstone's work, have added to this body through work in several related areas. For example, Bradley & Terry (1952) showed that measurement of qualities is a more complex cognitive process compared to making qualitative comparisons which takes less time as it is based on immediate perceptions. Andrich (1988) developed a probabilistic model of comparisons in which the better item—as determined through a comparison—gains one point while the worse gains zero each time a comparison of pairs is made. Finally, Andrich & Luo (2003) showed that after repeated comparisons are made, items with more "points" (Andrich 1988) have higher possibility to be better than the ones with fewer points.

The first time that comparative judgment (CJ) was used as an educational assessment method was in 1993 by Pollitt and Murray (1996) on foreign language speaking proficiency. Thurstone's law was first introduced and tested in British education as an educational assessment tool, resulting in high reliability with high cost due to the numerous comparisons needed for judges to make (Pollitt, 2004). The idea of comparative judgment was then refined in use of educational assessment with the advance of computer technology and development of an adaptive algorithm to reduce the workload associated with judgments. The addition of a technology-based algorithm led to the concept of *adaptive* comparative judgment (ACJ).

ACJ is an adaptive version of CJ, which Pollitt (2012) explained in detail in his 2012 paper. To use ACJ for assessing student work, a judge compares two items on the screen, (e.g., two visual design projects), and chooses the better of the two items. Once the "winner" is chosen, another pair of items is randomly picked from the pool that contains all items to be compared. Initially each pair of items is randomly chosen but at the sixth round (a round of judgments is made when every item in the pool has been compared at least one time with another), the selection of pairs starts to be adaptive. In the adaptive stage each comparison consists of items chosen for display based on their "win-loss" record. The system updates the ranking order after each round so that the "winning" items move up and the "losing" items move down. Following round 12 of judgment the level of reliability is measured (Rangel-Smith, Lynch, 2018). In practice, ACJ requires considerable amounts of time, especially when there are many portfolios or scripts to assess. The current ACJ algorithm, used in *CompareAssess*, can generate satisfactory results with reliable rank order after 12 to 15 rounds (Rangel-Smith, Lynch, 2018) and higher reliability as more rounds of judgments are completed (Pollitt, 2012).

Results of each ACJ session generated by *CompareAssess* software include a rank and a parameter value for each item that has been compared, a misfit value for each item and each judge, and a reliability value for the session. The parameter values reflect how far apart between two adjacent ranked items. A misfit value of a judge is how much he or she judged differently from the other judges. A misfit value closer to the critical misfit means a particular judge was making judgments that did not align with other judges in the same session. The misfit value of an item indicates how much judges agree on the judgments made between this item and the other one. Close to or above the critical misfit means the artifact was reviewed controversially; some of the judges think it is good, some think it is not. The reliability value generated in ACJ is represented

as a Judgment Consistency Coefficient (JCC), which Pollitt (2015) defined as "the amount of consistency between all the judgments made" (Pollitt, 2015, p.1).

ACJ has been studied mostly in the UK (Kimbell, 2008; Pollitt, 2012; Jones, Swan, & Pollitt, 2015), Ireland (Seery, Canty, & Phelan, 2012; McMahon & Jones, 2015), Australia (Tarricone & Newhouse, 2016) and the US (Bartholomew & Yoshikawa, 2018; Bartholomew, Strimel, & Yoshikawa, 2018). ACJ publications have reached a wide range of educational areas, from design & technology (Kimbell, 2008; Seery, et al. 2012; Bartholomew, Strimel, & Yoshikawa, 2018) and English writing (Pollitt, 2012, Demonacos, et al., 2019), to chemistry (McMahon & Jones, 2015) and math (Jones et al., 2015). ACJ has also been shown to be a reliable and viable assessment tool for design portfolios (Tarricone & Newhouse, 2016) and design projects (Bartholomew, et al., 2019) in visual arts and design.

Specific rationale for utilizing ACJ for critique in graphic design education include: 1) creativity is an integrated part of design, which makes critique a difficult process because the assessment of creativity is challenging (Bartholomew, 2017). Holistic assessment—the basis of judgments in ACJ—fits well with critique while criterion-based assessment or analytical marking presents problems due to the inherent quality of creativity of design (Tarricone & Newhouse, 2016). 2) ACJ can produce higher reliability in assessment results despite the participation of novices as compared with traditional analytical marking (Heldsinger & Humphry, 2010; Kimbell, 2012). Relatedly, preliminary research has noted differences in ACJ rankings generated by instructors and students despite the reliability levels obtained by both groups (Bartholomew, Strimel, & Jackson, 2018; Strimel, et al., 2018). 3) Compared with group critique, ACJ allows each individual judge to reflect and express own thoughts freely in critique without interruption or influence by

others (Bartholomew, et al. 2019), while also retaining collaborative results and allowing for feasible investigation of each individual student/teacher's thinking.

2.4 Think-Aloud Protocols

Thinking-aloud is an activity in which participants, while working on an assigned task, speak out their thoughts as they occur, rather than thinking first and then expressing through speaking, writing, drawing, or other ways (Ericsson & Simon, 1993). Think-aloud protocols (TAP) are the records compiled during the activity and have been used in educational research resulting in high validity and reliability from a relatively small sample size (Park, 2014; Kelley, Capobianco, & Kaluf, 2014; Cowan, 2017). Similar to this proposed research, TAPs have also been used, as a research method, to analyze differences and similarities of design abilities between experts and novices (Ball, Ormerod, Morley, 2004; Perez, Fleming Johnson, & Emery, 1995). Hartell (2015) also utilized think-aloud protocols and analyzed the criteria instructors are thinking of while assessing design portfolios through ACJ at Swedish primary schools. Concurrent think-aloud protocols, a specific form of TAP which capture people's thinking and reasoning while simultaneously performing an assigned task, can more accurately reflect participant's thoughts than written reports after the task (Ericsson, 2003).

The procedures of TAP consist of 1) collecting protocols, 2) transcribing protocols into text, and 3) analyzing transcripts. The following sections provide the guidance for implementing TAP in this study as drawn from the literature and best practices around TAP.

2.4.1 Collecting Protocols

The collection of TAP data requires each participant to complete an assigned task in a quiet environment so the individual thinking aloud process can be clearly recorded. Participants are

asked to keep thinking aloud from the start to completion of the task. Despite a minimum amount of intervention, the researcher may sometimes need to quietly remind the participant to "keep thinking aloud" if any pause longer than a few seconds occurs in order to obtain accurate and complete concurrent protocols (Fonteyn, Kuipers, & Grobe, 1993). Another way is to offer an orientation session to instruct participants how to think aloud before the official start of data collection (Gibson, 1997) which will allow the participant to practice prior to engaging in TAP (Ericsson & Simon, 1993).

2.4.2 Transcription

Following verbal data collection, TAPs are transcribed by a researcher who is familiar with the domain knowledge and related terminology. Pauses and filler words, such as "um", "uh" and "ah", are identified and removed during transcription. Following transcription and conditioning the complete transcript is divided in to segments (e.g., by statements, by interview questions) to facilitate further analysis (Joseph & Patel, 1990).

2.4.3 Content Analysis

Content analysis (Cohen, Manion, & Morrison, 2018) is utilized to analyze the verbal reports from TAP. A similar approach was used in Hartell's (2015) related study and includes 1) breaking the text data into statements, 2) coding, and 3) identifying patterns, similarities and differences, and links, if any exist.

In the second step of coding a variety of approaches may be taken. For example, Hartell (2015) generated codes throughout the analysis process as patterns appeared while Kelley, et al. (2014) used pre-defined codes. Fonteyn, et al. (1993) suggested ways to improve findings from codes through validation by a group of experts or, if pre-defined, randomly selected portions of

transcript for code-checking and then comparing with original coding results to identify any discrepancies.

Coding generally requires a second rater, who is familiar with the content analysis method and domain knowledge, to code selected portions of transcripts to determine if percentage of inter-rater agreement of greater than 90% can be established (Saldaña, 2016), and Cohen's Kappa coefficient is greater than 0.41 (McHugh, 2012). Park (2014) suggested randomly selecting 20% of transcripts for the second rater to code and then comparing the results from both raters. If any discrepancies are identified between the two raters, further examinations should be done to determine the cause (e.g., rater's misunderstanding of specific codes, or ill-written definitions of thematic codes) (Fonteyn, et al., 1993).

2.5 Summary

This chapter reviewed the current literature related to the research topic (critique in graphic design education) and two tools (ACJ and TAP) used in this thesis. The following chapter describes the methods that were employed to identify differences between students and instructors performing design critique through ACJ.

CHAPTER 3. METHODS

This chapter discusses the specific data collection and analysis methods to be implemented for this study. It provides an outline of the context of study and an explanation of the sampling methods. This chapter also explains the data collection and analysis procedures to investigate the research questions.

3.1 Research Questions

The research question that guided this study was "What are the differences, if any exist, between the critiquing practices of instructors and students through adaptive comparative judgement for graphic design projects?"

The following sub-questions were used to investigate the research question stated above:

1. What is the relationship, if any exists, between the ACJ results (ranks and parameter values) generated by students and instructors in an introductory graphic design course?
2. How long does it take for each student/instructor to make each judgment in ACJ?
3. What are the factors that influence students'/instructors' judgment decisions?

3.2 Context of the Study

This research study was conducted in a large public research university located in the Midwest United States. In Fall 2017, there were 31,006 undergraduate students enrolled in the university with 63% identified themselves as White, 8% Asian, 5% Hispanic/Latino, 2% Black/African American, 2% multiracial, 0.1% American Indian and 2% unknown. The university also had 16% International students. There were 43% of female and 57% male students.

The study took place in an introductory graphic design course where students create 2-dimensional full-color illustrations using computer software such as *Adobe Photoshop*, *Illustrator*, and *After Effects*. The course objectives included using shapes, colors, and other design elements to create digital images for communicating. The design of this course was inspired by the Bauhaus curriculum in Weimar, Germany. At the Bauhaus school, students' learning process starts by introducing the design principles and language, then letting students create and experiment. The design principles are intended to set the basis of a good design and, in this specific course, students were provided with eight items (see Figure 3.1) which include balance, contrast, repetition, unity, rhythm, economy, emphasis, and closure.

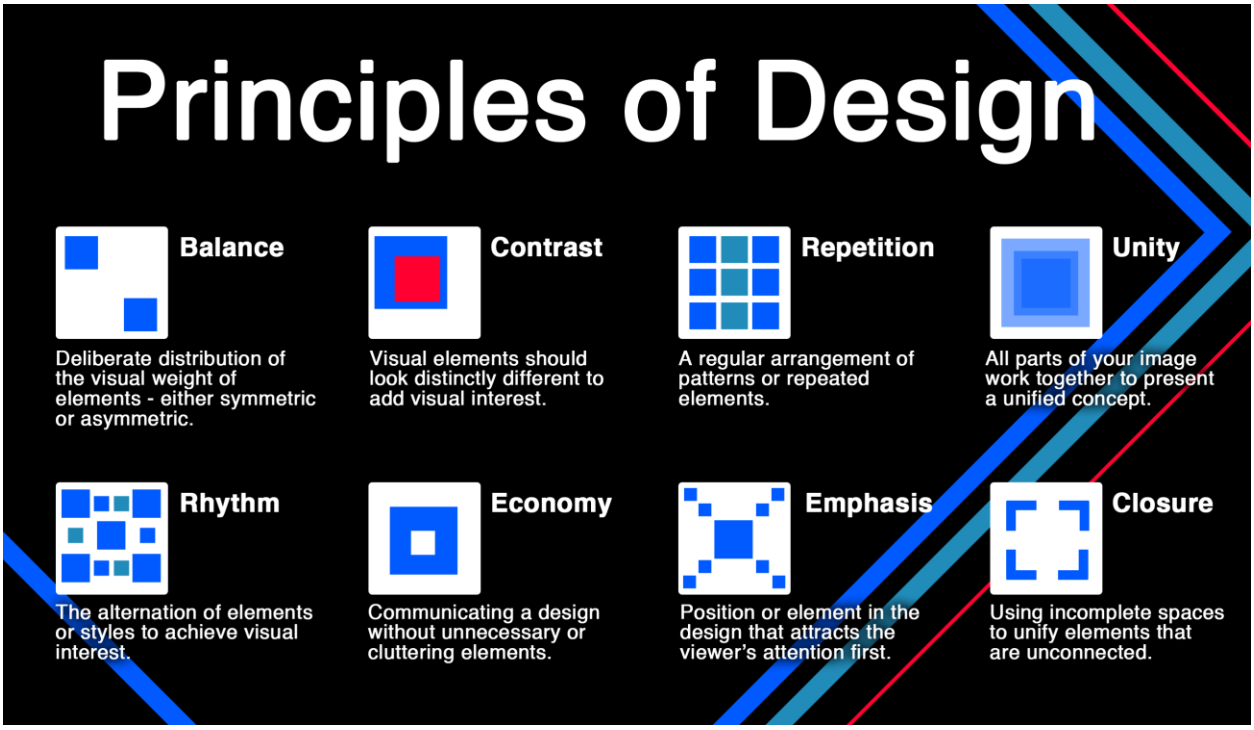


Figure 3.1 Principles of Design Handout, © Matt Wallace

At this university this introductory course was mandatory for all computer graphics technology undergraduate students and was usually provided in multiple class sections (each with approximately 50 students enrolled). In 2018, a survey was conducted with 108 students enrolled

in the course at the time showing 86% students in their freshman year, 9% in sophomore year, 4% in junior year, and 0.9% senior students. Among those surveyed, 50% were female and 50% were male students.

3.3 Data Collection

This section describes the data collection process of this study, including a description of pre-study surveys, the research subject recruiting procedures, participants' background and activities on the day of study, and the selection process of design projects used in this study. It also provided details about ACJ and TAP procedures.

3.3.1 Pre-Surveys

In this research, brief pre-surveys were distributed to instructors and students to investigate several traits related to their experience with design and teaching experience. The findings from these surveys were used to establish the maximum levels of comparability between subjects for analysis.

3.3.1.1 Instructor Survey

As the background of the three instructors may vary in terms of age, years of graphic design education, years-of-teaching, gender, and race—and these differences may affect their critiquing—the pre-survey questions were used to collect the demographic information, as well as specific graphic design and teaching experience, of instructors. The instructor survey questions are included as Appendix A.

3.3.1.2 Student Survey

As the student participants may have learning experiences in graphic design before college, which may influence the research results, the student surveys were aimed at inquiring regarding

their background and experience in graphic design. This pre-survey was distributed to all participating students to gain demographic information and help investigate student participants and their prior experience in graphic design. Student survey questions are listed as Appendix B.

3.3.2 Participants

Three instructors, and three students from the introductory graphic design course were recruited as participants in the study. The course instructors' roles were to provide instructions and answer questions in the studio-based classroom and assign grades to student design work based on the assignment criteria and course objectives.

3.3.2.1 Recruiting

Three instructors were selected based on criterion sampling – the recruited participants had to satisfy the following criteria: (a) the instructor was willing to volunteer for the study; (b) the instructor did not teach the class whose design projects were selected as items to be judged in the study; (c) the instructor had filled out and returned the assent form as required by university internal review board to participate in the study. Differences in demographics (i.e., age, year in school, years of graphic design education, gender, race, major) were identified through pre-study surveys and addressed in the analysis stage but did not influence the recruiting process due to the limited availability of instructors.

Three students were recruited from the three sections of class for participation in this study following these steps: (a) the researcher introduced the purpose and expectation of study to all students enrolled in the course; (b) students who were willing to volunteer for the study were asked to provide a one-hour-long time slot of availability on a sign-up sheet; (c) three students were selected based on their time availability and willingness to participate. Time availability was a key factor in the recruiting process because this study had to be completed within two weeks (while

students were working on their own design) prior to the project completion to avoid potential bias introduction following the completion of the assignment, grading, and student exposure to other's projects.

3.3.2.2 Instructors

All participating instructors were asked to complete a pre-study survey including demographic information and questions about their education background, teaching, and technical experience in graphic design. All three participating instructors were Caucasian with varied backgrounds in terms of gender, age, and years of learning and teaching experience in graphic design (see Table).

Instructor 01, a male graduate student, majored in Animation in college. He was 23 at the time of study and had studied graphic design for 3 years. In addition, he had worked at a start-up company to make logos and videos. However, this was his first semester teaching this course and had had no graphic design teaching experience before.

Instructor 02, a female graduate student, majored in User Experience Design in college. She was 23 at the time of study and had studied graphic design for 4 years. In addition, she had previous experience with graphic design in high school (yearbook design), interface design experience in college, and a marketing design internship. This was her third semester teaching this course, but she had no teaching experience other than this course.

Instructor 03 was a male assistant professor whose major was Animation in college. He was 39 at the time of study and had studied graphic design for 10 years. This was his first time teaching this course, but he had 7-years of experience teaching graphic design prior to the time of the study. Additionally, he had worked in the industry as a graphic designer for 5 years before

becoming a university professor. He also had developed new graphic design courses and taught at university level.

Table 3.1 Summary of Instructors' Information

	Instructor 01	**Instructor 02**	**Instructor 03**
Ethnicity	Caucasian	Caucasian	Caucasian
Gender	Male	Female	Male
Age	23	23	39
Major at college	Animation	User Experience Design	Animation
Graphic design experience	3 years	4 years	10 years
Teaching experience	1st semester	3rd semester	7 years

3.3.2.3 Students

All participating students were asked to complete a pre-study survey including demographic information and questions about their previous experience related to graphic design. The results are presented in Table 3.2. All three students were majoring in Computer Graphic Design and had some formal or informal learning experience in graphic design before college. However, for each student, this was their first time taking a graphic design course at university level.

Table 3.2 Summary of Students' Information

	Student 01	**Student 02**	**Student 03**
Ethnicity	Caucasian	Asian	Caucasian
Gender	Female	Female	Male
Year at college	Freshman	Sophomore	Sophomore
Major	Computer Graphic Design	Computer Graphic Design	Computer Graphic Design
Graphic design learning experience prior to college	learned video editing and website design in informal settings for four months before college	took a year-long graphic design course in high school where she learned Adobe Photoshop and Illustrator	took a graphic design course in high school for a year where he designed advertisement and short animations.

3.3.2.4 Days of Study

Following a demographic survey, ACJ trainings, and a thinking aloud practice task, each participant sat at a work table in a quiet room while making comparative judgments on graphic design projects and exercising think aloud protocols to justify their decisions. The entire TAP and ACJ process was recorded by a desktop screen recorder.

3.3.3 Design Project

The design project used during the critique by participants in this study was the first among five design projects in the course – each with increasing complexity. The design project is the first stage of graphic design learning process in the Bauhaus curriculum (Daichendt, 2010) and the learning objective of this project was to apply the design principles and create organized images using basic design elements (shapes and colors). It required students to create an abstract image that was aesthetically pleasing using geometric primitives like points, lines, shapes and curves, and at least six colors. The project also required students to create repetitions of shapes and color gradations (a graphic design technique of making gradual transition of colors). Figure 3.2 shows a sample student design.

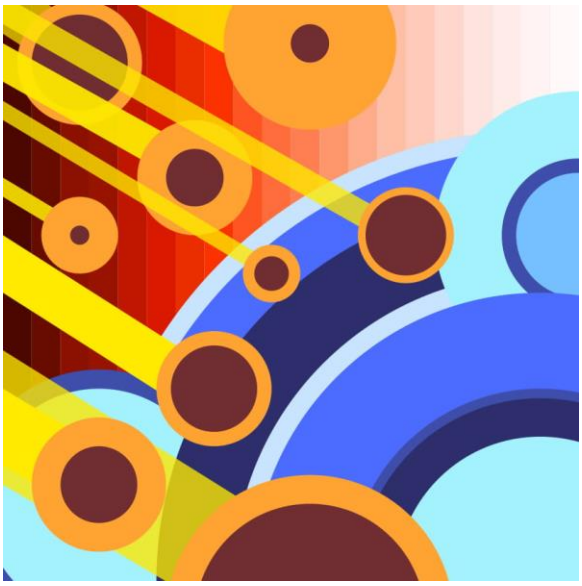


Figure 3.2 A Student Sample Design Project

Prior student submissions of their first project were collected from a previous year of this course – a step intentionally done to avoid any instructor or student recognition of designs. 50 projects were collected from a class section that none of the participating instructors had taught before and none of the students had been enrolled in. Out of those 50 projects, 10 design projects (see Appendix C) were selected through the following stratified sampling process:

1. Student designs were grouped into five groups by the grades received;

2. Two projects were randomly selected from each of the five groups;

3. The selected 10 projects were anonymized, and renamed with letters A-J.

4. The selected 10 design projects were used in the ACJ session to limit the ACJ comparisons required by each participating student and instructor.

3.3.4 ACJ Process

Prior to the study, the researcher uploaded the selected 10 design projects to *CompareAssess* and set up two ACJ sessions; both included the same 10 projects. The researcher

also created unique logins with fabricated email addresses for all participants to use through the study. Relatedly, all personal identifiers such as names, school accounts, or email addresses were not collected in the study and were not present on the student work assessed through *CompareAssess*.

All participants were trained prior to using *CompareAssess* to make judgments. The researcher provided the training and demonstration before they started their ACJ session. Participants were instructed to view pairs of designs and choose the better design in each case while thinking aloud. Instructors and students used *CompareAssess* to make comparative judgments with students in one ACJ session and instructors in another session. Students and instructors were put into two separate ACJ sessions as this facilitates the production of two rank orders for comparison in conjunction with the established research questions.

Each participant was expected to make 34 judgments and the ACJ process was estimated to take approximately 30 minutes. However, the actual time of completion varied with a range from 7.5 minutes to 41.4 minutes.

3.3.5 Thinking-Aloud Process

Each participant was sitting alone in a quiet room while making comparative judgments and thinking aloud. Each was provided a 10-minute practice task (e.g. to explain why they chose this university/major; describe their previous graphic design experience; or explain what they think is important to teach in a graphic design course) to help participants experience TAP and become comfortable with thinking aloud. This task was assigned after the ACJ training and prior to beginning the actual ACJ session so participants could become familiar with the process.

During the ACJ session screen recorders were used to capture the participant's voice and the screen display seen by each participant. No likenesses of participants were captured as part of this

study to protect the privacy and confidentiality of participants. Screen recorders only captured participants' voice and the screen they were viewing during each portion of the exercise.

The researcher originally intended to use both screen recorder and a point-of-view camera to capture what the participants were looking at. However, the video recordings captured by the point-of-view camera were discarded in favor of only using the screen recordings because 1) only a few gestures were noticed in field notes and screen recorder captured all the other information needed for analysis; 2) the POV camera ran out of battery one time during the process, so there is one incomplete video recording; 3) the analysis of recordings would be challenging because camera shook as participants moved their head.

3.4 Data Analysis

This section describes the data analysis process of this study, including how original videos recorded from the days of study were trimmed, how thematic codes were generated and refined from the transcribed videos, and the coding process.

3.4.1 Video Trimming

A desktop screen recorder was used to capture each participant's voice, mouse movements and images displayed on the screen while working through the ACJ session. The original screen capture videos included the thinking-aloud practice task and short segments at the end when the research turned the recorder off. Therefore, prior to data analysis, all videos were trimmed so each video recording included only the ACJ and TAP processes. The length of videos ranged from 7.5 minutes to 41.4 minutes.

3.4.2 Transcription & Code Generation

The researcher transcribed each of the six videos consisting of participants' verbal expressions made while critiquing. Each transcript was segmented by each judgment made and original thematic codes were generated as the researcher transcribed the videos. The codes were generated with reliance and influence from coding schemes established in Voet, et al. (2015) and Bartholomew, et al. (2019). Table 3.1 shows the list of codes produced by the researcher during this exercise with a representative example from the scripts for each of the identified themes. Initial codes were used in the first round of coding and secondary codes, also generated through similar approaches, were used in a second round of coding.

As an example of round 1 coding, a verbal expression stating that the judge prefers one project over another, or that the judge could not decide was coded as *Decision* or *Cannot decide*. An evaluative statement expressed as a positive, negative, or neutral remark regarding an item was coded as *Positive*, *Negative*, or *Neutral evaluation*. An informative statement that builds on a preference was code as a *Justification* and, in cases where the statement also included improvements, was coded as a *Suggestion*. An evaluation was distinguished from a justification because an evaluation builds on individual design project instead of a preference.

A neutral evaluation falls into exploratory form of critiques (see Table 2.1), which is often related to a description of what the design looked like to the judge. Positive and negative evaluations refer to times when the judge was analyzing the good and bad of a design without expressing any preference – these could be considered as either exploratory or argumentative critiques (Table 2.1). Lastly, a justification represents a time where the judge clearly expressed his or her arguments for whether a design is good or bad – this aligns with argumentative critiques as shown in Table 2.1.

For round 2 coding, items which were coded as Evaluation (positive, negative, or neutral), Justification, or Suggestion, were then coded again using sub-themes (see Table 3.1) which emerged through the coding process.

Table 3.3 Coding Scheme: Original Version

Category I	**Code**	**Example**
Decision	D	"Going with A"
Cannot decide	D.N	"I don't know. It's a hard choice"
Positive evaluation	E.P	"The color choices is like complimentary. The blue looks really popping compared to the pink and that looks really good."
Negative evaluation	E.N	"For me B felt a little too childish and especially like having a lime green color in the background."
Neutral evaluation	E	"Also it feels very like a game application from Minecraft."
Justification - informative	J.I	"I feel like for the creativity B has more appeal."
Justification - Suggestion	J.S	"If they change the background color which is purple to difference color which is more subtle, it would look more appealing."
Category II	**Code**	**Example**
Personal feelings	PF	"My brain can't accept it"
Requirements	RQ	"Just the use of lots of lots of shapes and more gradient looks well thought out."
Composition	CP	"It's definitely well composed together."
Creativity	CV	"I feel like for the creativity B has more appeal."
Design principles	DP	"B I think has more harmony"
Effort	EF	"it shows lots of effort by the student"

After this initial coding (both first and second round), the coding scheme was updated to better reflect the data; for example, code *Composition* from Category II, or CP was deleted because only 3 statements fit in this theme and the researcher found they could also fit into code theme *Design Principles* (DP). In addition, a new code theme was added to Category I as *Other* (O), for times when the participant rethought their judgment or made a statement which was not directly related to the current comparison. Additionally, a few definitions of codes were also clarified to finalize the coding scheme (see Table 3.2).

Table 3.4 Coding Scheme: Final Version

Category I	**Code**	**Example**
Decision	D	"Going with A"
Can't decide	D.N	"I don't know" "It's a hard choice"
Positive evaluation	E.P	"The color choices is like complimentary. The blue looks really popping compared to the pink and that looks really good."
Negative evaluation	E.N	"For me B felt a little too childish and especially like having a lime green color in the background."
Neutral evaluation	E	"Also it feels very like a game application from Minecraft."
Justification	J.I	"I feel like for the creativity B has more appeal."
Suggestion	J.S	"If they change the background color which is purple to difference color which is more subtle, it would look more appealing."
Rethink judgment	O	"This is the same situation as before."
Idle time	I	Blank
Category II	**Code**	**Example**
Design principles	DP	"B I think has more harmony" "Same compositional elements drawing your eyes to the center, so there is a movement."
Requirements	RQ	"Just the use of lots of lots of shapes and more gradient looks well thought out." "I find A to be more technically impressive and more aesthetically interesting"
Personal feelings	PF	"My brain can't accept it" "It's kind of peaceful. It makes my mind rest a little bit."
Effort	EF	"it shows lots of effort by the student"
Creativity	CV	"I haven't seen anybody in at least my class try to do something similar to this" "it gives me something that I'm not normally used to see."

3.4.3 Coding

The overall coding process was facilitated through *NVivo software*. In addition to the coding completed by the researcher another researcher also coded 25% of the videos (25 judgments by students, and 25 by instructors) independently, through *NVivo* as a check on inter-rater reliability.

The videos were segmented based on the results from transcription coding. Multiple statements, that were in sequence but assigned with the same code, were combined into one segment. Video segments without participants' voice (usually typified through a short pause or

transition between judgments) were marked as idle time. This second round of coding used the final coding scheme as shown in Table 3.2.

3.5 Summary

This chapter provided an overview of the methods that were used in this study to collect and analyze data. It described the context where the study was conducted, namely an introductory graphic design course at university level. The data collection process included preparation days when the researcher recruited participants and selected design projects to be critiqued and multiple days of study when participants took the survey and completed 34 comparative judgements while thinking aloud. The data analysis procedures included trimming videos, transcribing videos into scripts, generating codes, and coding. The following chapter presents the findings of this study.

CHAPTER 4. RESULTS

This chapter serves the purpose of presenting the results from the ACJ sessions and the thematic coding and comparing between students' and instructors' groups. The chapter is guided by the research question and sub questions listed in Chapter 2.

4.1 ACJ Results

To address the first research sub question, "what is the relationship, if any exists, between the ACJ results (ranks and parameter values) generated by students and instructors in an introductory graphic design course?", the researcher compared the ACJ ranks and parameter values between students and instructors through statistical correlation tests.

4.1.1 ACJ Ranks and Correlation

The two ACJ sessions resulted in two lists of rankings for the 10 design projects. Table 4.1 shows the comparisons between ranks generated by students and instructors - sorted by project.

Table 4.1 ACJ Ranks

Project	Student Rank	Instructor Rank
A	1	2
B	4	1
C	3	7
D	2	3
E	6	4
F	7	6
G	10	8
H	5	9
I	8	10
J	9	5

Figure 4.1 shows the correlation between student and instructor ranks generated by ACJ. There is considerable scatter with three design projects demonstrating a disparity of at least one rank (e.g., the one ranked 1st place by students is 2nd place by instructors), three with a disparity of two ranks, one with a disparity of three ranks, and three with a disparity of four ranks. The researcher also conducted a Spearman's correlation test and found no significant correlation between student and instructor rank $r_s = .564$, $p = .09 > .05$.

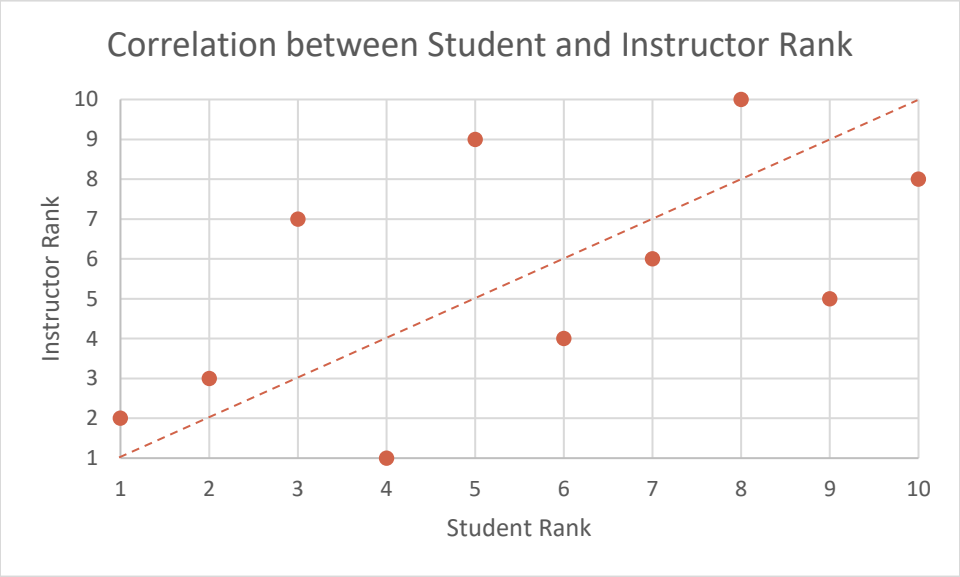


Figure 4.1 Correlation between Ranks by Two Groups

4.1.2 Parameter Values and Correlation

The two ACJ sessions also resulted in two lists of parameter values for the 10 design projects. Table 4.2 shows the comparisons between parameter values generated by students and instructors - sorted by project names. The results confirmed that the ACJ results generated by students and instructors did not align. For example, there was a large jump between the parameter value of Project A (4.04 as 1st place) and that of the 2nd ranked project (1.36), indicating that students overwhelmingly voted Project A to be the best of 10 design projects. Alternatively, instructors

voted Project A as the 2nd best with a parameter value (.75) - far behind the 1st place project (B) which had a parameter value of 3.11.

Table 4.2 Parameter Values

Project	Student PV	Instructor PV
A	4.04	0.75
B	0.98	3.11
C	1.13	-0.39
D	1.36	0.28
E	-0.49	0.23
F	-0.56	-0.23
G	-2.74	-0.64
H	0.04	-0.98
I	-1.65	-2.12
J	-2.1	0

Figure 4.2 shows the correlation between student and instructor parameter values generated by ACJ, which also demonstrated considerable scatter. The researcher also conducted a Pearson's correlation test and found no significant correlation between student and instructor parameter values $r_p = .465$, $p = .176 > .05$. The test results aligned with results from Spearman's correlation test between ACJ ranks – both indicating that the ACJ results generated by students and instructors were different and had no significant correlation.

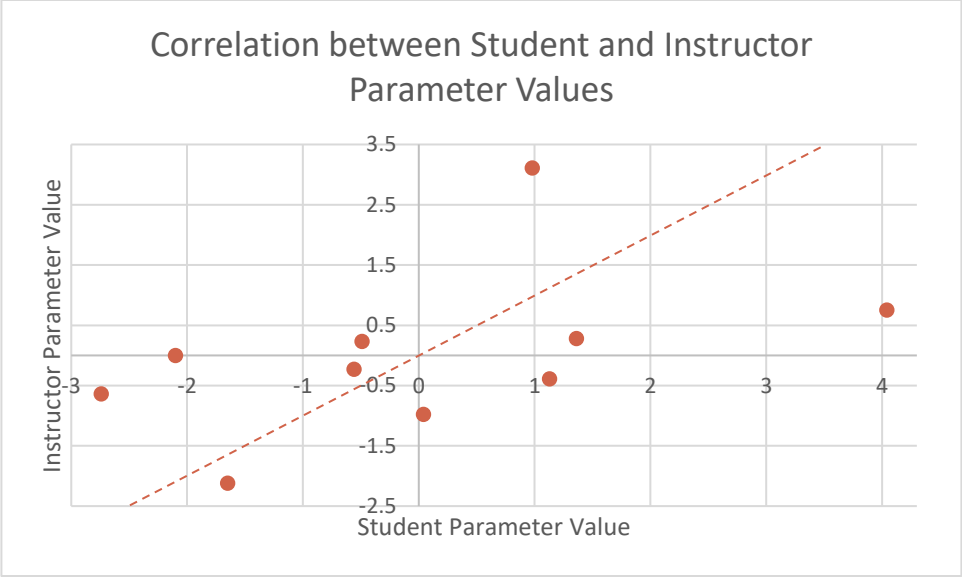


Figure 4.2 Correlation between Student and Instructor Parameter Values

4.1.3 Reliability and Misfit Values

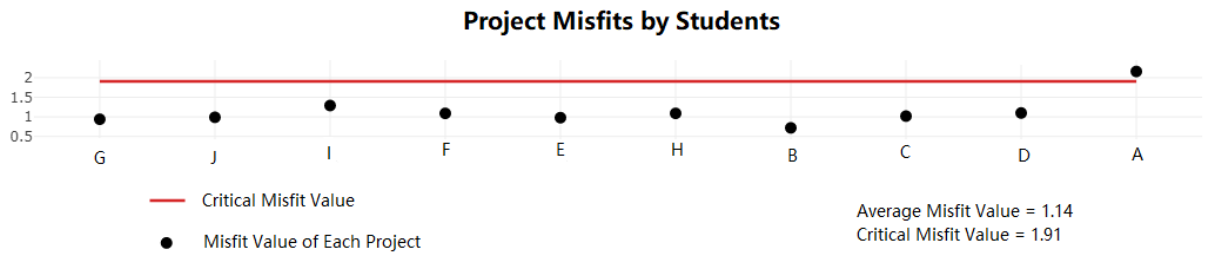
In addition to the rank order of all design artifacts, the *CompareAssess* software provided a reliability of each session and a misfit value for each judge and artifact. The reliability level of students' session was 0.83, and that of instructors' session is 0.80. Both were deemed to be sufficiently reliable for analysis (Rangel-Smith, Lynch, 2018) and higher than common scores obtained through a traditional rubric-based assessment (Demonacos, et al., 2019).

Table 4.3 Judge Misfits

Student Judge	**Misfit**	**Critical Misfit**
S01	1.61	2.26
S02	0.31	2.26
S03	0.95	2.26
Instructor Judge	**Misfit**	**Critical Misfit**
I01	0.98	1.26
I02	0.93	1.26
I03	1.15	1.26

Judge misfits from both ACJ sessions are shown in Table 4.3. No judge misfits are above the critical misfits, indicating that judges were making consistent judgments with the other judges in the same session. Misfit values of design projects are shown in Figure 4.3. Project A has a misfit value greater than the critical misfit (see Figure 4.3a). This means students' judgments on this design project were not completely consistent across all judges – an important finding as this project ended up as the highest ranked. Project H and Project C both have relatively high misfit values (see Figure 4.3b), reflecting levels of disagreement among the instructors on these two projects.

A)



B)

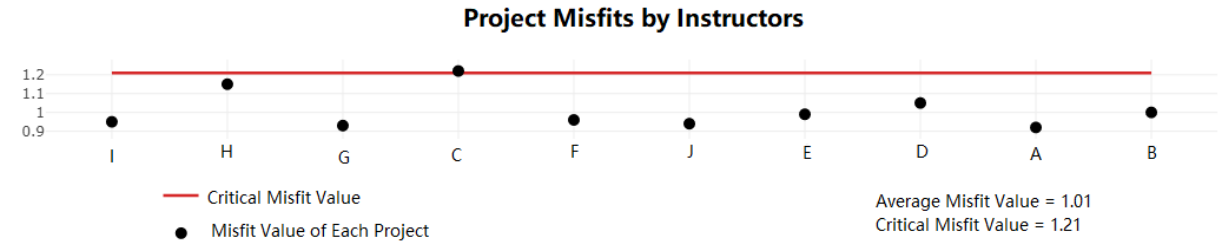


Figure 4.3 Project Misfits

4.2 Time Spent

On average, students spent about one minute (60.4s) making each judgment in ACJ while thinking aloud while the instructors spent less than half minute (28.1s). Therefore, students spent,

on average, more than twice the time instructors did in making one judgment. Further, it was noted that instructors with more years of teaching experience spent less time per judgment than others (see Table 4.4).

Table 4.4 Time Spent in ACJ

Judge	Time Spent in ACJ (min)	Time per Judgment (s)	Learning/Teaching Experience
S03	41.4	73.0	first semester
S01	32.4	57.2	first semester
S02	28.8	50.9	first semester
I01	26.9	47.4	1 semester
I02	13.5	23.9	3 semesters
I03	7.5	13.2	7 years

4.3 Coding Results

4.3.1 Inter-Rater Agreement

In order to check for inter-rater agreement a coding comparison query was utilized on 25% of the protocols in NVivo. The percentage of agreement between raters was calculated to be 93.42% and Cohen's Kappa coefficient to be 0.66 (combined average), which indicated a satisfactory level of agreement between coders (McHugh, 2012).

4.3.2 Category I

To further explore how students and instructors approached their critiquing decisions, the research conducted analysis on the thematic coding results from Category I codes (see Table 3.2 for details of each code). This section reports what students and instructors were thinking during their critiquing time based on the coding results.

4.3.2.1 Students' Group

Overall, students spent a large portion of their time in evaluation with both positive and negative remarks. For example, Student 01 spent 29.71% of her time in positive evaluation, Student 02 spent 22.58% of the time in positive evaluation, and Student 03 spent 24.41% of the time in positive evaluation (percentage of time coverage of each code was shown in Figure 4.4).

Of note, 29.26% of Student 02's critiquing time was spent in justifying a decision while this task took only 11.94% of Student 01's time and 7.19% of Student 03's time. These findings align with the field notes taken by the researcher on the day of study, which indicated that Student 01 and Student 03 were likely to evaluate both designs one by one and then decide which one was a better piece. Student 02, however, expressed her preference for one design at the beginning and then focused on this piece and justified why she thought it was a better design or what she thought made the other piece a worse design.

Student 03 spent 33.23% of his time focused on things that he did not like about the designs – much more than his peers (15.93% by Student 01 and 10.67% by Student 02). Another noticeable difference was that Student 02 did not reflect or remark back to previous critiques while making subsequent critiques, however Student 01 made five reflection statements and Student 03 made 14.

A)

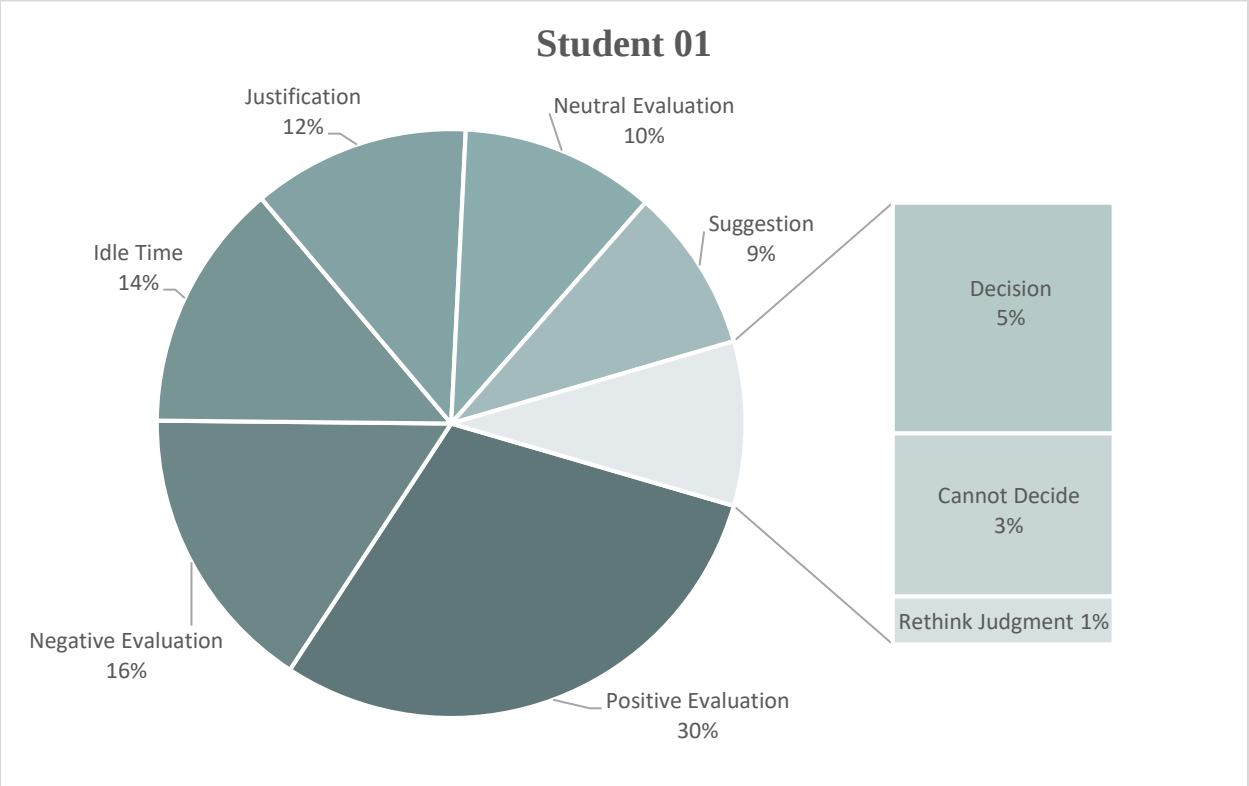


Figure 4.4 Percentage of Time: Students' Group

Figure 4.4 continued

B)

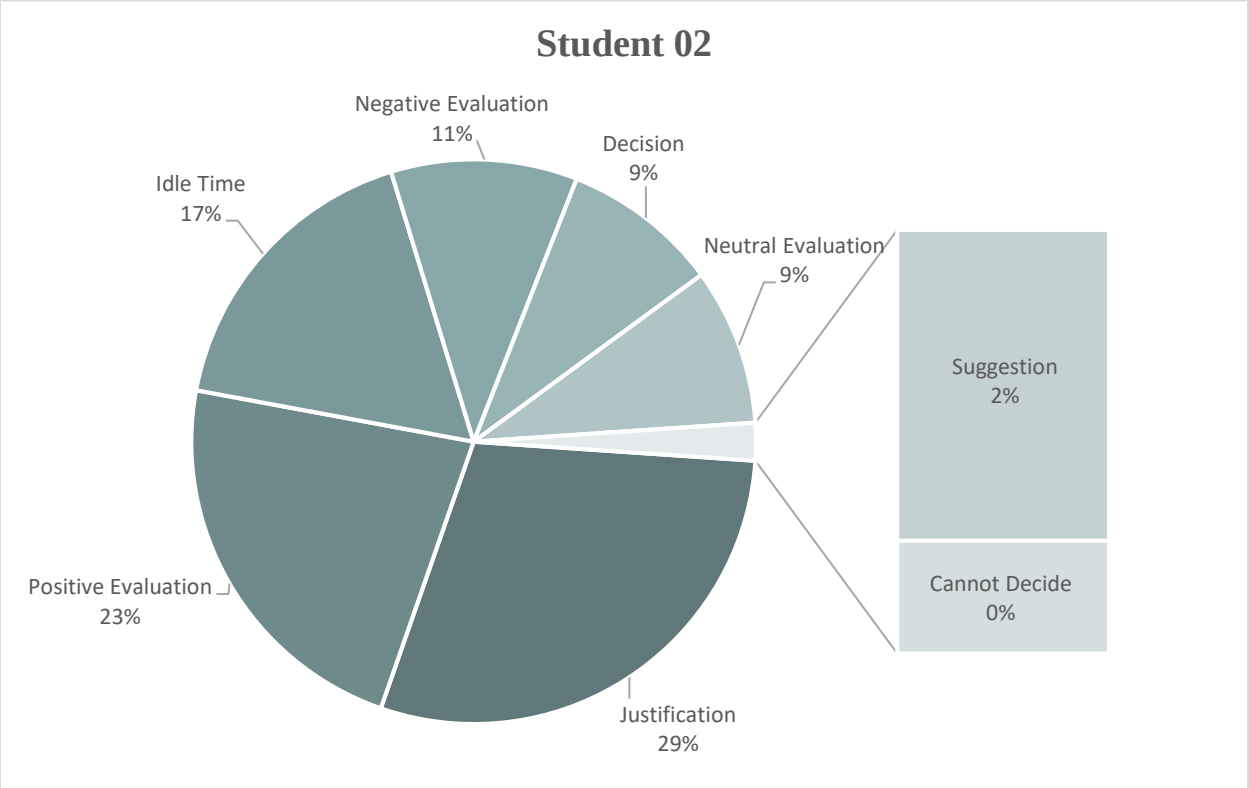


Figure 4.4 continued

C)

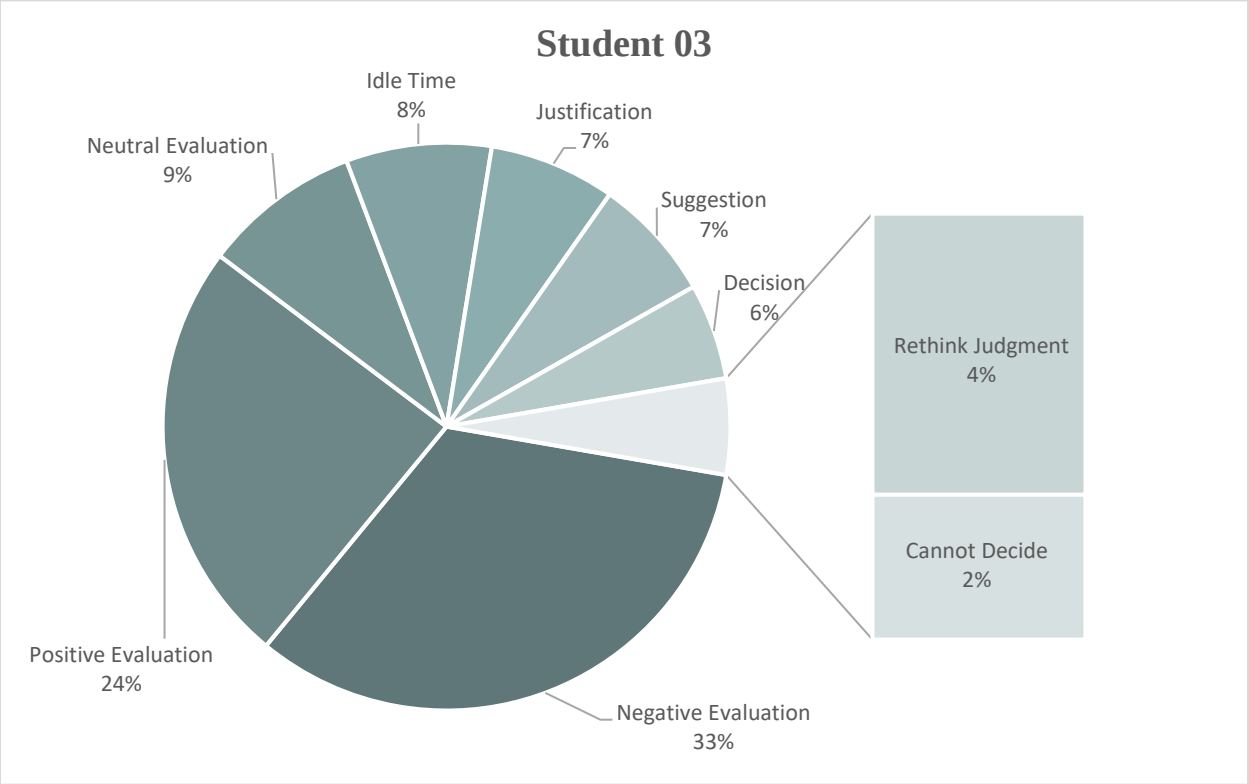


Table 4.5 presents the percentage of time in each coded area, the number of times each code was noted, as well as the average length in seconds of students speaking out one statement, or multiple statements expressing a similar meaning. One of the most time-consuming codes was Suggestion (J.S), which is the suggestion provided to make the design a better piece. For example, the average time spent for Student 01 to think aloud one suggestion (e.g., “I would change the colors. It could even be the same colors but I will make them stronger or more shiny in order to catch more people’s attention.”) was 11 seconds. In contrast, the average time for the student to state a decision, which was usually one short sentence (e.g., “between these two, I’m choosing A”), was only 3.2 seconds.

Table 4.5 Category I Coding Summary: Students' Group

Codes	Percentage of Time	Number of Codes	Average Length (s)
S01			
Decision (D)	4.72%	29	3.2
Cannot Decide (D.N)	3.33%	14	4.6
Neutral Evaluation (E)	10.68%	25	8.3
Negative Evaluation (E.N)	15.93%	33	9.4
Positive Evaluation (E.P)	29.71%	53	10.9
Justification (J.I)	11.94%	33	7.0
Suggestion (J.S)	9.01%	16	11.0
Rethink Judgment (O)	0.98%	5	3.8
Idle Time (I)	13.69%	33	8.1
S02			
Decision (D)	8.99%	36	4.3
Cannot Decide (D.N)	0.57%	2	4.9
Neutral Evaluation (E)	8.99%	18	8.6
Negative Evaluation (E.N)	10.67%	18	10.3
Positive Evaluation (E.P)	22.58%	34	11.5
Justification (J.I)	29.26%	49	10.3
Suggestion (J.S)	1.58%	2	13.6
Idle Time (I)	17.35%	33	9.1
S03			
Decision (D)	5.47%	41	3.3
Cannot Decide (D.N)	1.86%	8	5.8
Neutral Evaluation (E)	9.00%	15	14.9
Negative Evaluation (E.N)	33.23%	58	14.2
Positive Evaluation (E.P)	24.31%	61	9.9
Justification (J.I)	7.19%	19	9.4
Suggestion (J.S)	7.05%	21	8.3
Rethink Judgment (O)	3.60%	14	6.4
Idle Time (I)	8.29%	33	6.2

4.3.2.2 Instructors' Group

Figure 4.5 shows the results from the coding of the instructor's time thinking and critiquing. There were some similarities within the instructors' group identified by the researcher. The top two categories were the same for all three instructors: 1) justification of their decisions and 2) idle time. This aligns with researcher's field notes which noted that instructors tended to think aloud

in terms of which one of the two projects was better and then either justify their decision or quickly move on by clicking to select the better piece.

A)

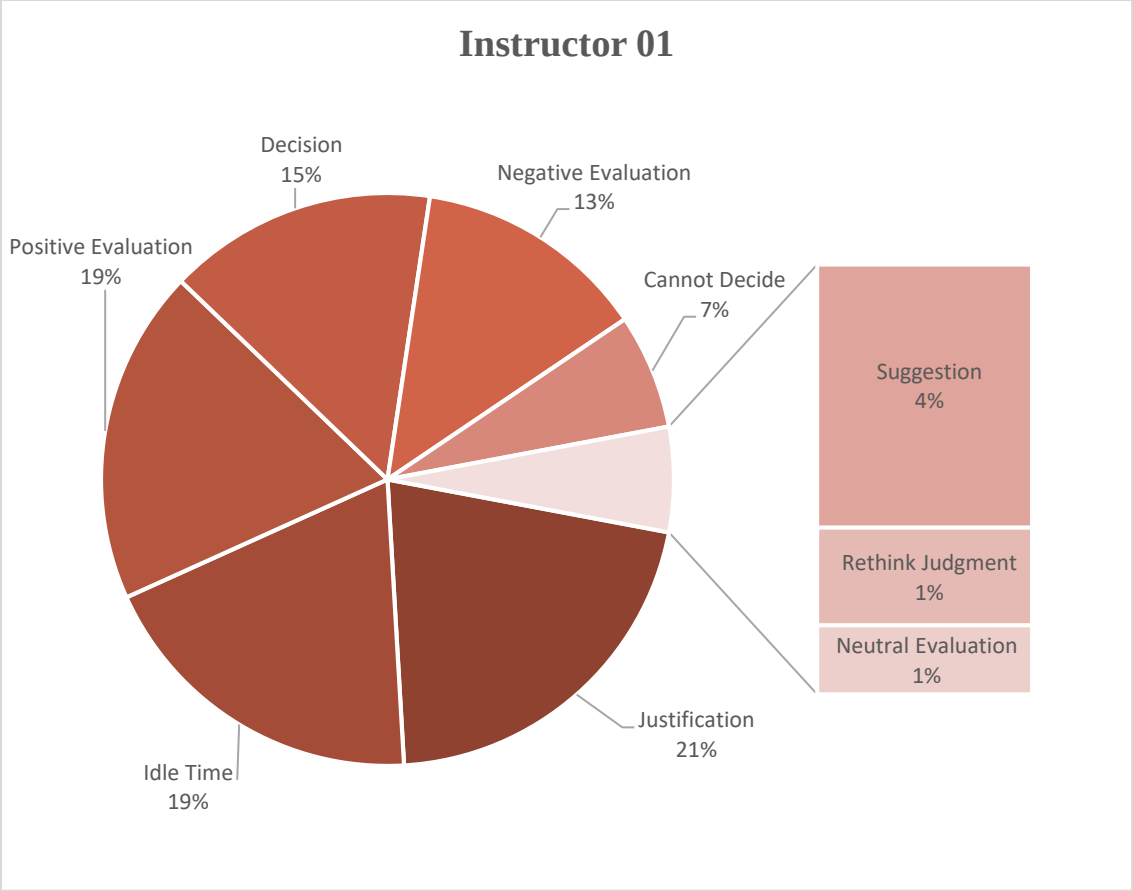


Figure 4.5 Percentage of Time: Instructors' Group

Figure 4.5 continued

B)

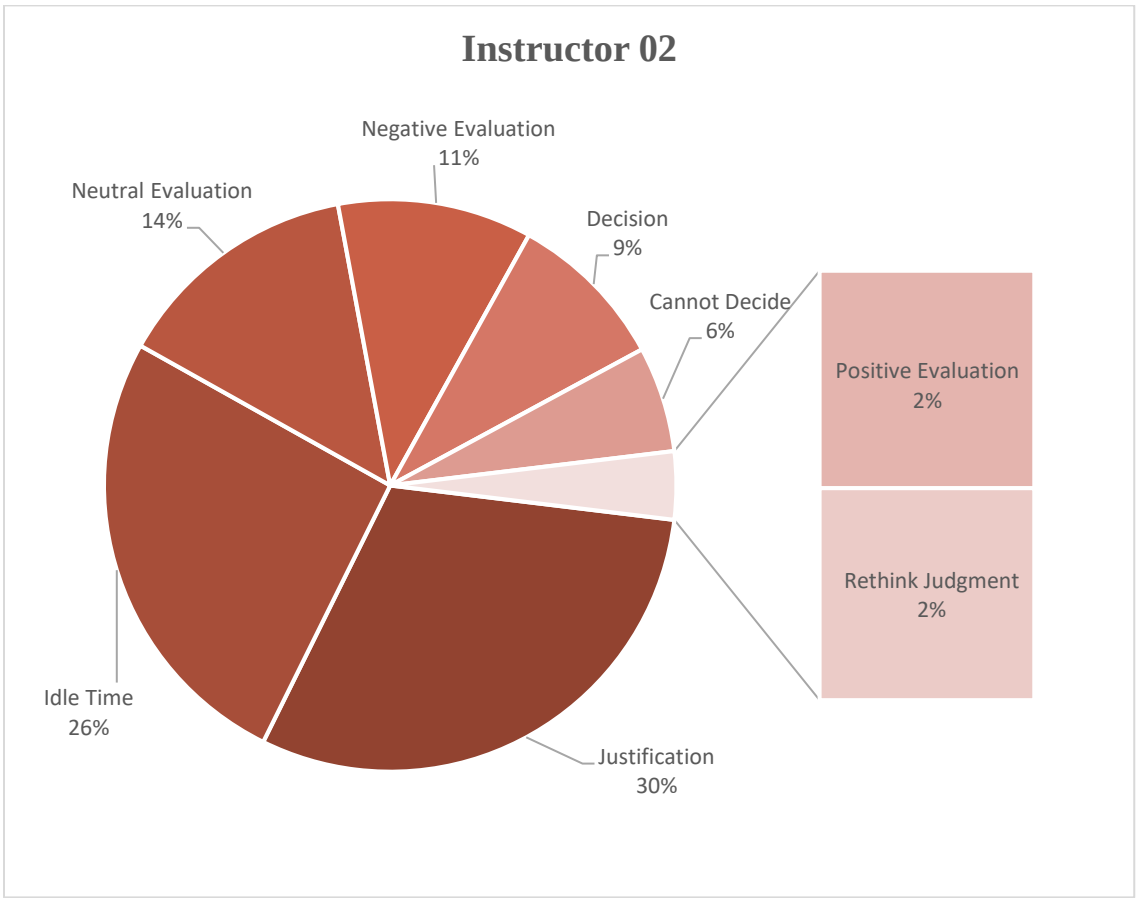


Figure 4.5 continued

C)

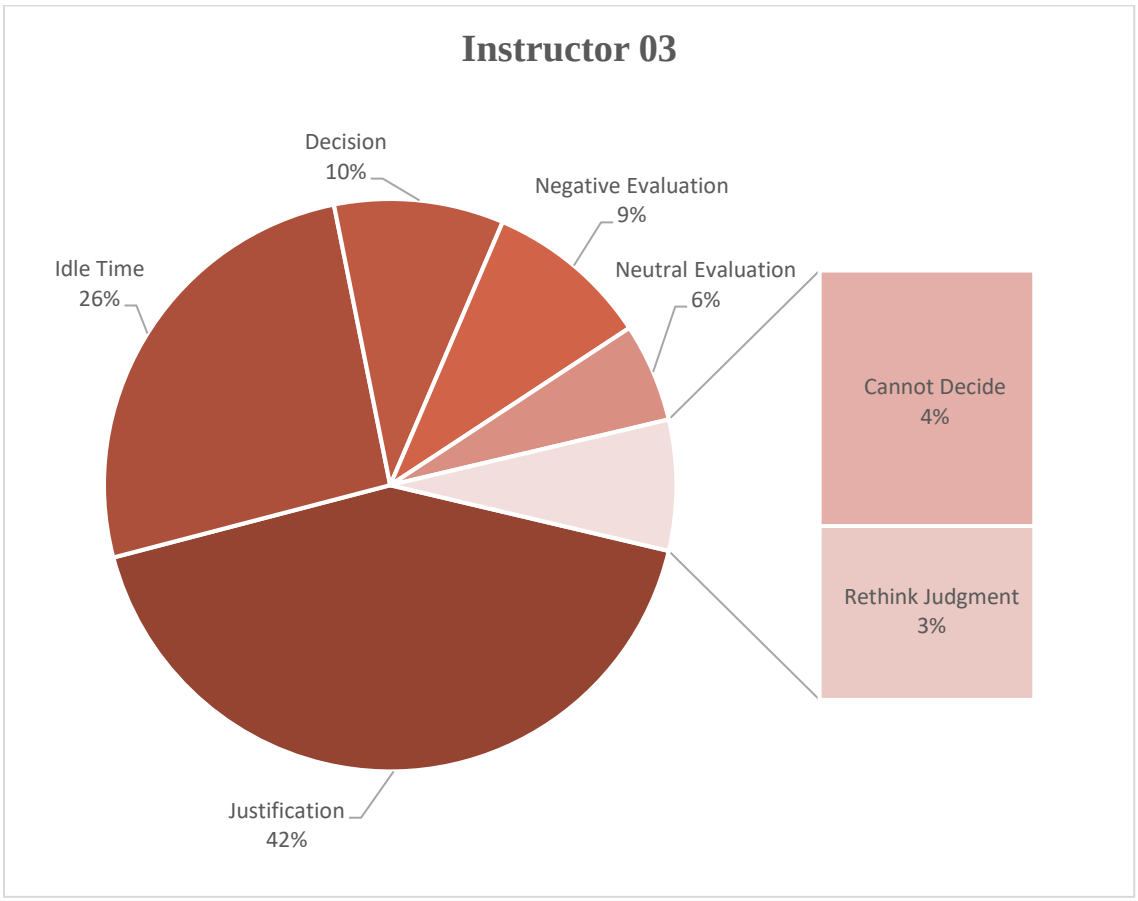


Table 4.6 Category I Coding Summary: Instructors' Group

Codes	Percentage of Time	Number of Codes	Average Length (s)
I01			
Decision (D)	15.20%	59	4.2
Cannot Decide (D.N)	6.49%	16	6.5
Neutral Evaluation (E)	0.95%	4	3.8
Negative Evaluation (E.N)	13.17%	22	9.6
Positive Evaluation (E.P)	18.93%	31	9.8
Justification (J.I)	21.12%	38	9.0
Suggestion (J.S)	3.62%	5	11.7
Rethink Judgment (O)	1.35%	2	10.8
Idle Time (I)	19.17%	33	9.4
I02			
Decision (D)	9.06%	34	2.2
Cannot Decide (D.N)	5.97%	12	4.0
Neutral Evaluation (E)	13.95%	14	8.1
Negative Evaluation (E.N)	10.99%	15	5.9
Positive Evaluation (E.P)	1.94%	2	7.9
Justification (J.I)	30.40%	45	5.5
Rethink Judgment (O)	1.89%	5	3.1
Idle Time (I)	25.80%	33	6.3
I03			
Decision (D)	9.58%	14	3.1
Cannot Decide (D.N)	4.39%	5	3.9
Neutral Evaluation (E)	5.57%	5	5.0
Negative Evaluation (E.N)	9.31%	8	5.2
Justification (J.I)	42.24%	34	5.6
Rethink Judgment (O)	2.99%	3	4.5
Idle Time (I)	25.91%	33	3.5

Table 4.6 presents the percentage of time coverage, number of each code, as well as the average length in seconds of instructors making one statement, or multiple statements expressing the same meaning. For example, Instructor 03 averaged between 3.1-5.6 seconds to make one statement which corresponded roughly to one full sentence.

Another item of note is that there were 34 judgments made by each instructor, but Instructor 01 made 59 statements of decision while Instructor 02 made 34 and Instructor 03 made 14.

Instructor 01 repeatedly expressed his preference between two design projects and, sometimes while deciding between the two, he would reach a decision and then change his mind. Conversely, Instructor 03 traditionally followed a much-more streamlined approach as he clicked the one he thought was better and moved on—often without thinking aloud his decisions.

4.3.3 Category II

Codes from Category II were derived from the rationale behind ACJ evaluation, justification, or suggestions. Figure 4.6 presents a visual of the percentages of time spent by participants in thinking aloud by the different factors that influenced their decisions.

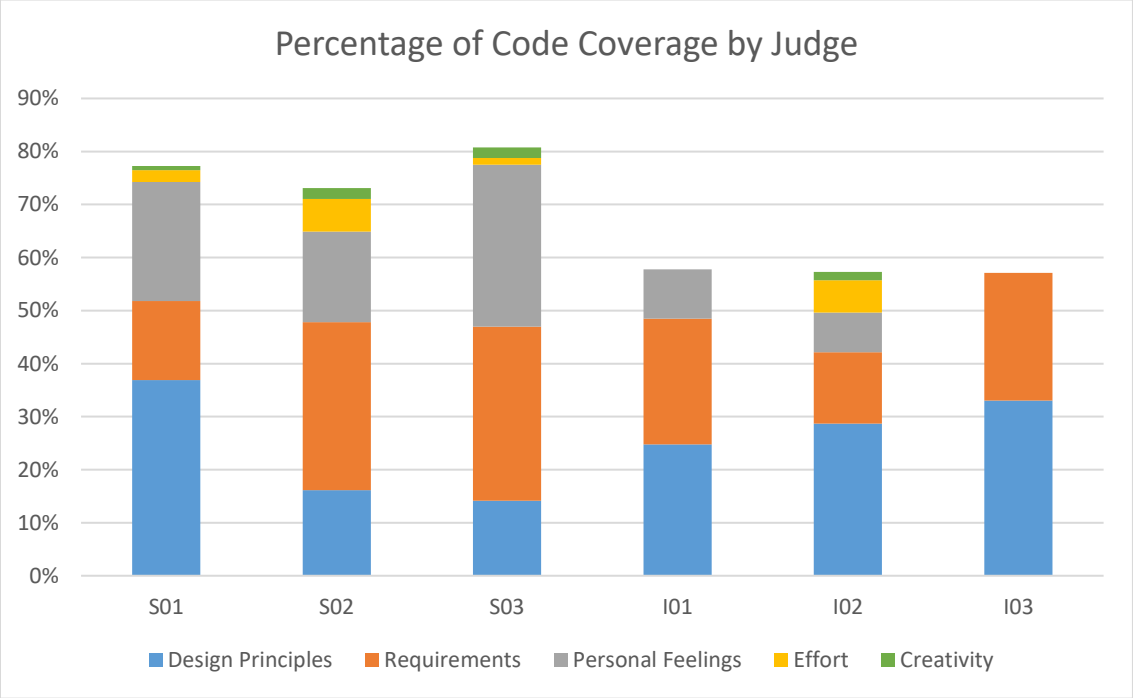


Figure 4.6 Percentage of Code Coverage by Judge

Design principles and project requirements were the two main factors which influenced students' and instructors' choices, while creativity was the least influential factor. For students, their personal feelings were taken more into consideration while they were critiquing design

projects than the instructors who did not generally voice their personal feelings during critiques - especially Instructor 03 who only used design principles and project requirements to determine the better one between two design projects.

Instructors spent less time thinking aloud rationale behind their judgment decisions compared with their students. The average percentage of time spent in evaluation, justification or suggestion was 77% for students whereas it was 57% for instructors.

4.4 Cases

Three cases were identified from the ACJ results which included designs that were ranked very differently in the two ACJ sessions. The researcher revisited the videos, codes, and analysis for these three to further explore the research questions related to the similarities and differences in critiquing approaches for students and instructors.

4.4.1 Case 01: Project C

Project C was ranked as 3rd place by students but ranked as 7th by instructors. It was first noted that the misfit value (1.22) of Project C by instructors was higher than critical misfit (1.21), which indicates that the three instructors did not have significant agreement on this design project. The researcher further investigated what made instructors rank this design lower. When the researcher looked at the comparative judgments associated with this project the researcher noticed several intriguing comments related to judgments with this project. A few representative comments are included here:

"I think maybe because this shape in the middle is a little too symbolic. Kind of looks like any eye."

"[Project C] just has too many competing elements."

"[Project C] seems to have more work put into it. But it just doesn't fully come together."

Although Instructor 02 recognized the effort designer put into Project C, she thought it to be less of a unified design and less simplistic comparing to the other designs.

"The colors are too intense in [Project C]."

"I'm seeing the same options multiple times and I'm second guessing myself whether I'm being consistent in my judgments."

Instructor 03 held a negative opinion on the color use in Project C at the beginning. Then in the later comparisons, he tried to stay consistent with his own judgments, so he kept not choosing it due to "less evidence of a deliberate use of the principles of composition".

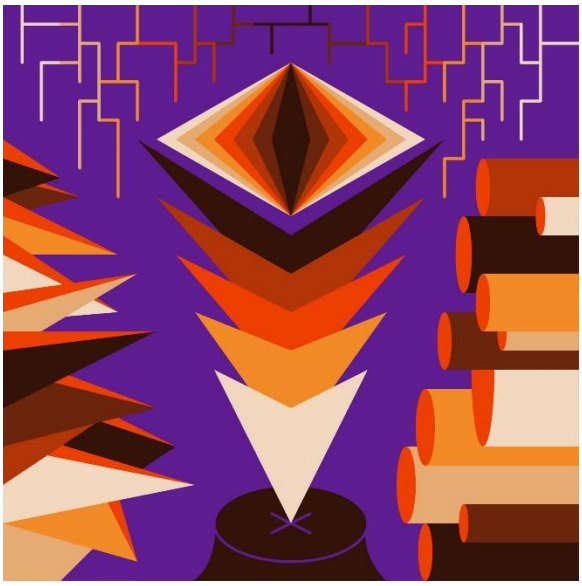


Figure 4.7 Project C ranked 3rd by students but 7th by instructors

Students, however, held positive thoughts towards Project C in terms of its use of design principles (symmetry and some variations) and color use. Here are some representative examples of students' quotes for Project C:

"It has some symmetry to it if you set a line in the middle of the image. But they also added some variations. For examples, the cylinders on the right and the pyramids on the left."

"The orange really brings out the depth perception ... Also, the yellow and purple are complementary colors and it's working really well together."

4.4.2 Case 02: Project H

Project H was ranked as 9th by instructors who agreed that "the colors don't match." Students seemed to agree as they also mentioned color use when they chose the competing design as the

better one over Project H. Since students ranked Project H in the middle (5th) out of 10, the researcher looked specifically at what students were saying positively about it. Students' quotes of Project H are listed here:

> *"It looks like … if you were doing an art piece in a modern building. I liked the corners; the corners are pretty cool with this like triangles and squares kind of thing following the lines. It gives kind of movement towards the center … It doesn't make me tired and I could look at this for many many minutes and I wouldn't be tired."*

Student 01 liked Project H a lot because of her personal feelings and a sense of movement when shapes and colors are put together. One time when she found it hard to decide between Project H and another design, she voted for Project H because of the creativity and effort put into it.

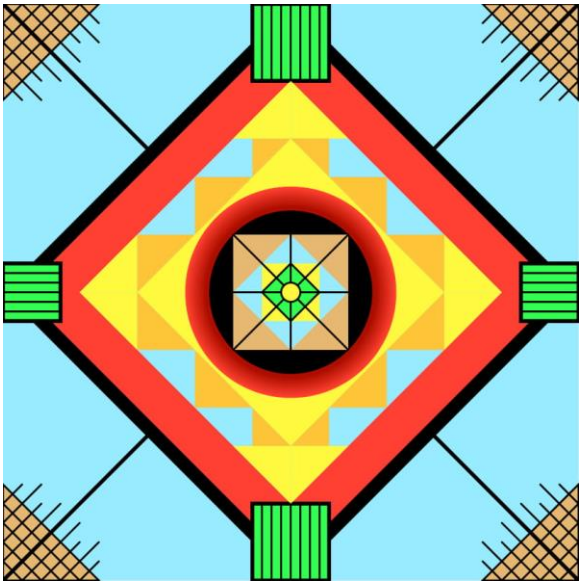


Figure 4.8 Project H ranked 5th by students but 9th by instructors

Student 02 thought Project H had better color choice with some creativity and effort into it. However, she only chose Project H once as the better design out of two, suggesting that overall she would rank it lower than the 5th place ranking from the overall session.

"Everything here is pretty standard, but it's designed well, and it still looks nice."

Student 03's comments suggest an approach to Project H based on symmetry as a point of "good and bad of a design". He made comments revolving around the project having "four or even eight-way symmetry".

"Symmetry makes a design look nice, but same time it is not thinking outside of the box when a lot of people in class are doing this sort of symmetrical design."

Further, Student 03 mentioned color choice in Project H as a problem at the first glance because of the use of green, but stated later that "it's not overbearing at all. It's used pretty minimally."

4.4.3 Case 03: Project J

Project J was ranked as 9th (second lowest) by students but ranked as 5th by instructors. Students disliked Project J because it looked random to them, especially the use of shapes.

"Looks like someone put random shapes, color them also randomly and submitted it."

"The shapes might just be a little bit of more conflict with four triangles."

"There's too much loose space on the top bottom right and left and I don't have a specific place that my eyes are caught attention to."

The researcher then looked specifically at what instructors liked about it. Instructors' quotes about Project J are listed here.

"I like these bright colors in the middle. I think it matches well with the kind of damper colors in the background."

Instructor 01 thought the color composition of Project J is good, although he criticized the tension caused by the green shapes touching the edges. Instructor 02 also said she liked the color of Project J. However, neither of them chose Project J when viewing it compared with others.

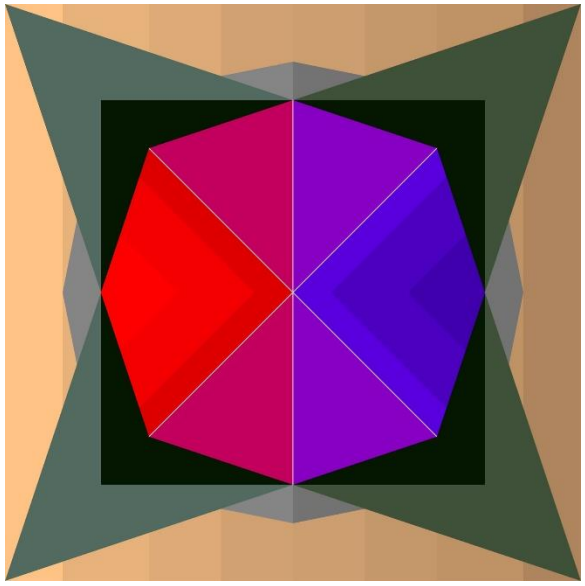


Figure 4.9 Project J ranked 9th by students but 5th by instructors

Instructor 03 selected Project J the first time he saw it, but he did not provide any reasons.

"Don't know why, but I prefer the composition of [Project J]."

In later comparisons, he thought aloud his negative opinions about Project J, but stuck to his decision because of the color treatment and he wanted his judgments to be consistent.

"Neither of these examples are very impressive. I think they both using formal symmetry, both lacking a bit of color structure, but [Project J] is slightly better."

"I'm seeing the same options multiple times and I'm second guessing myself whether I'm being consistent in my judgments. I think in this case the composition and color treatment of [Project J] is more pleasing."

4.5 Summary

This chapter has presented the results from the data analysis. Included here are a few key findings of the study:

1. There was no significate correlation found between the ACJ ranks and parameter values generated by students and instructors.
2. Students tended to spend considerable amount of time in each judgment – often describing their personal feelings of both designs, while instructors rarely did so.
3. Students utilized more time evaluating the good and bad of each design and then deciding which one was better; however, instructors with more teaching experience appeared to use experience and instinct to make decisions more quickly and with less discussion.

CHAPTER 5. DISCUSSION & CONCLUSIONS

This chapter includes a discussion around the findings and related observations made by the researcher while completing this study. It also provides a summary of each chapter, concluding remarks around this study, and some potential directions for future work.

5.1 Critiquing Practices of Instructors and Students

The research question guiding this study was: "What are the differences, if any exist, between the critiquing practices of instructors and students through adaptive comparative judgement for graphic design projects?" Overall findings from this research indicate that 1) students spent a considerable portion of their time evaluating each individual design project before deciding on a better design while instructors tended to quickly justify their preferences after a decision; and 2) there were differences in the ACJ results produced by instructors and students which related to design principles, preferences, and values.

While investigating the relationship between the ACJ results (ranks and parameter values) generated by students and instructors, the researcher found no significant correlation between the ACJ ranks and parameter values generated by students and instructors. Further, this disparity was more noticeable between the two groups' parameter values than the ranks alone. The two main factors—identified through the analysis performed in this research—demonstrated that while both students' and instructors' judgment decisions were based on the principles of design and project requirements, the students' judgment decisions were greatly influenced by their personal feelings while instructors did not rely on their personal feelings to make judgments.

Further differences were identified related to the amount of time utilized for the students and instructors to make each judgment in ACJ. Generally, it took instructor less time to make each

judgment in ACJ than students did. Further, instructors who had had more teaching experience spent less time completing each judgment than the new instructor with little teaching experience who performed similar as students did.

In addition to the overall findings noted here, several insights gleaned through this study and the accompanying data collection and analysis, will be shared in the following sections.

5.2 Instructor and Student Roles

The results from this study suggest that instructors spent less time completing the critiquing session and that instructors with more experience spent less time making each judgment (see Table 4.5) than those with less experience. Given that, for both students and instructors, ACJ was a new approach of making critiques compared with traditional rubric-based methods, it appears that the different roles of students and instructors could be a part of the potential causes which resulted in the differences in time spent. Instructors become more trained in summative assessment as they gain more teaching experience. Observations by the researcher indicated that instructor participants in this study seemed to be familiar with comparing and determining which item of student work is better than another – even though they had not used ACJ previously. Further, it was noted in the researcher's field notes that instructors, especially Instructor 03, worked to complete the critiquing activity as quickly as possible even when they allocated the same amount of time (an hour time-slot) as students did to participate in the study. Instructors 02 and 03 finished each judgment quickly and, once they had finished one, they quickly moved on to the next one. However, Instructor 01 spent 47.4s to complete each judgment - similar to the time spent of 50.9s by Student 02 and close to the average time of 60.37s by the students' group. Looking at Instructor 01's think-aloud protocols, it appears that he acted like an "experienced student" while making judgments – a finding that may align with this being his first time as an instructor of the course.

Additionally, the researcher noted that the instructors' decisions overall often lacked constructive feedback - at least when they were not specifically asked to provide it (in this study all participants were asked to think aloud which design project they thought is better and why but were not explicitly directed to provide any type of feedback). Instructors tended to move between pairings without providing feedback—as the students were more apt to do—and they appeared to have more capability in using terminology efficiently to explain and to justify. As a result, students overall provided more suggestions than instructors did in this study. Specifically, the three students thought aloud 16, 2, and 21 pieces of constructive feedback while making 34 comparative judgments on 10 design projects, while one of the instructors gave five suggestions and the other two gave none.

Table 5.1 Students' and Instructors' Performance in Evaluation

Judge	Number of Negative Comments	Number of Positive Comments	Percentage of Time in Evaluation	Learning/Teaching Experience
S03	58	61	66.54%	first semester
S01	33	53	56.33%	first semester
S02	18	34	42.25%	first semester
I01	22	31	33.05%	1 semester
I02	15	2	26.88%	3 semesters
I03	8	0	14.88%	7 years

Table 5.1 presents the number of negative and positive evaluation comments by each participant and the percentage of their time spent in evaluation (negative, positive, and neutral). All three students and Instructor 01 spoke out more positive comments on design projects than negative comments – especially Student 01 and 02 whose positive comments were almost doubly represented when compared with the negatives. However, instructors 02 and 03 rarely spoke positive thoughts about a design in their critiquing practices with two and zero positive comments respectively. It was also noted that instructors with more teaching experience spent less time

evaluating each design; for example, Instructor 03 spent only 14.88% of his time in evaluation while the average percentage is 24.94% for instructors' groups and 55.04% for students' group.

Overall instructors with more teaching experience tended to complete the critique quicker, made less suggestions, spoke fewer positive things about students' design, and spent less time evaluating each design. This may have been caused by their experience, their "gut feelings" about design, their expertise, or something else. However, the researcher noted that this may also be a problem in terms of the role of instructors – were they providing enough feedback to students and were they able to articulate that feedback in a way that students could use?

For example, one of the instructors (03), with the most graphic design experience, reflected about critiquing with ACJ and the thinking aloud experience and expressed his concerns when he had a hard time translating his instincts into verbal expressions.

> *"When I am seeing the design, my eyes are inspecting the elements, switching between reviewing the individual parts (shapes, line work, crafting details), recognizing the part/whole relations (layout, overall color scheme, visual flow), and comparing these between the two examples.*
>
> *Concurrently I am having emotional reactions to these inspections and about this process. This may be as simple as positive or negative reactions to various elements. The reflection or recognition is where I am noticing that this emotional reaction links to my experience, my learned understanding of how design works (or doesn't, or should, or could work). It's like I'm automatically checking or confirming.*
>
> *When challenged to deliver a verbal protocol of the process, … I have to interrupt this instinctual process and translate those instincts into a communication element that can be understood."*

5.3 ACJ as a Learning Tool

Another finding from the study that became apparent to the researcher was the potential for ACJ to be used as a learning tool to assist student learning in critiquing graphic design.

5.3.1 Learning through Extended Reviews

The researcher noted that students changed their opinions on one design as they worked through the ACJ pairings and repeatedly viewed it. Making comparative judgments appears to have helped shape their idea of what makes a good design and changed their understanding of the principles of design. For example, Student 01 held negative feelings on the Project G at the beginning.

"[Project G] scares me a little bit. It reminds me of ... the old video game machine ... like Pacman stuff like the little avatars or deleted square it would pursue like spaceships."

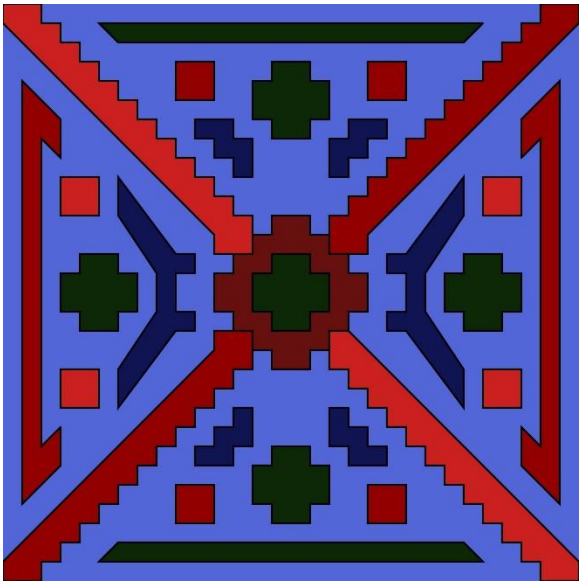


Figure 5.1 Project G ranked 10th by students

Then, when she saw Project G the second time, she started to have mixed feelings about it because of the symmetrical distribution of shapes. The third time and in subsequent comparisons, she found it to be more aesthetically pleasing.

> *"Although I have mixed feelings about the shapes and everything, it's for my brain more aesthetically thought. It's you see like the artist put the shapes in a specific place. It is in a way organized."*

The fifth time she critiqued on Project G, she realized that she had changed her opinions on it, and she was able to provide some constructive suggestions.

> *"[Project G] I thought in the beginning was too organized, but now I'm thinking it's well distributed. But it could be better distributed. Maybe move the lines further apart and make the rectangles in the middle a bit smaller."*

This, and other examples, suggest to the researcher that ACJ may be valuable as a learning tool instead of simply an assessment tool in the context graphic design education. Students, when using ACJ as a learning tool, may be able to build up their understanding of design principles, what make(s) a good design, and how to improve a design. Further, ACJ could also potentially be used in teacher training, as a learning tool for new instructors to gain more critiquing experience, and in building their understanding of project/assignment criteria.

5.3.2 Proceed with Caution

If ACJ was to be utilized as a learning tool in the critiquing practices, the ACJ ranks by students would not be vitally important in terms of whether they are correlated with ranks generated by instructors, or whether they can be considered as reliable assessment results. However, caution is needed when students act in providing and receiving written comments from peer critiques through ACJ. It has been found that students generated different ACJ results from instructors; meanwhile they may also provide substantially-different comments, which could lead to questions on the quality and value of peer feedback. Peer feedback that is of low quality or considered as "not credible" by students could be misleading (Cambre, et al., 2018) and not helpful to students' revision on their design projects. Therefore, it appears that students would need to be critical while evaluating peer feedback and use this feedback "wisely" to refine their design projects. Students' likelihood of changing their point-of-view on a design project during critiques may improve the quality of peer feedback because ACJ platforms (e.g. *CompareAssess*) allow students to amend their written comments if they see a project more than once.

5.3.3 Limitations

In this study, the researcher only investigated the experiences of three students, three instructors, and a total of 10 design projects. Due to the limited number of participants in this study, the ACJ rank of one design was heavily influenced by one judge's preferences. This limitation was highlighted in the included case study of the three design projects ranked very differently between two groups. For example, Project J would rank much lower and Project C would rank higher if Instructor 03 did not try to keep his own judgments consistent from start to finish. Likewise, Project H would rank much lower if Student 02 were the sole judge.

Further, related to use in contemporary classrooms, a larger number of design projects would also require more judgments to make by each judge. In this case it may be an issue for instructors as judges – one instructor may have to make hundreds of judgments to get a rank order of student work for one class. However, if students were the judges, this could also become easier – every student could assist with the necessary reviews and would also receive a chance to review more projects than they may have in traditionally-used critiquing activities.

5.4 Suggestions

Based on the findings of this study and the included references from related studies, the researcher identified several suggestions for consideration related to instructors and educators and how they may facilitate students' learning through critiques.

5.4.1 Themes of Critiquing

The coding themes, generated in this study, should be considered by instructors, educators, and future researchers in graphic design education. This study found that students spent significantly more time evaluating each design project rather than simply comparing the two projects and justifying their preferences for which was better. In practice, instructors may notice this difference in students' and, if desired, could work to train students to emphasize certain traits, approaches, or aspects. The researcher suggests providing additional instructions to students so they can focus on comparisons—if that is desired by instructors—more quickly rather than evaluating each individual project for merit. In addition, instructors may want to inform their students that pure compliments may not assist their peers as much as feedback with specific areas of improvement identified. Even though positive comments might encourage students and build

up their confidence, students may want to be more critical and it emphasize constructive feedback which may be more valuable in critiques.

The lack of suggestions by instructors was discussed in Section 5.1. There was no doubt to the researcher that generally these instructors were capable of providing constructive feedback to students and that they viewed it as part of their roles to do so as a means of promoting student learning. However, the findings of this study suggest that instructors with more years of teaching experience may need to be more aware that while they are capable of completing the critique task quicker, they must emphasize providing enough feedback or information to their students to distill down and pass on their own understanding and learning. This finding extends to teacher education where new teachers should be likewise instructed relative to their critiquing. Critiquing formatively as an instructor should not only be critical but should also provide suggestions that are viable for student growth.

The researcher also suggests encouraging students to utilize professional language in critiques instead of expressing personal feelings that are often more subjective for amateur judges. The following sections discusses why the researcher believes this is important and how to teach students to use terminology.

5.4.2 Teaching Professional Language

This study found that it took students almost double the length of time instructors needed to complete the same amount of critiques through ACJ. Where did students spend the other half of their critiquing time? Evaluating each design project, expressing their personal feelings on a design, and looking for the exact phrases or terminology to explain and justify their thoughts on a design appeared to take up significant amounts of time. An instructor would say “no focal point” using just a few words, while Student 03 expressed the same meaning with a paragraph:

"There's nothing there. Everything blends into itself and there's nothing that's emphasized or stands out. It's just kind of all there. It doesn't really catch your eye. The only thing that would really catch your eye is that if it's amongst a bunch of pictures like this it would stand out for not standing out."

It was clear to the researcher that students were struggling with the use of terminology even though, in the case of this study, students were provided with a handout explaining design principles (see Figure 3.1) as part of the essential course resources. The principles of design were explained in class prior to the start of this study. However, none of the student or instructor participants had the handout by their side as they were critiquing, nor any documents that could remind them of using graphic design terminology or critique language. The researcher suggests that instructors should provide some guidance before, during, or after critiquing through ACJ related to using the associated syntax.

5.5 Summary of Thesis

5.5.1 Chapter One: Introduction

Chapter One introduced the research topic of this study – critique in graphic design education. ACJ, as a form of assessment, has been utilized in formative critiquing of graphic design, but previous research has suggested that students critique differently from instructors through ACJ. This chapter explained what problem is investigated in this study and why this problem is important. Understanding the differences between students' and instructors' critiques is important to understand potentially-useful refinements in teaching pedagogies that may facilitate student

learning and critiquing in graphic design education. This chapter also discussed the scope, assumptions, limitations, and delimitations of this study.

5.5.2 Chapter Two: Literature Review

Chapter Two presented a review of literature relevant to this thesis study. It discussed critique, how critique aligns with goals of graphic design education, ACJ, how ACJ has been used in design learning and formative assessment, and how to implement TAP as suggested in related studies.

5.5.3 Chapter Three: Methods

In Chapter Three, the researcher discussed the method and design of this thesis study, the research questions guiding the research, the study context, the participants and how they were recruited and involved in this study, the procedures of using surveys, ACJ and TAP to collect data, and the data analysis process – code generation and video coding.

5.5.4 Chapter Four: Results

The results of comparisons between students' and instructors' group from ACJ sessions and coding process were presented in Chapter Four. This chapter was divided by four sections: 1) the results of ACJ, including ranks and parameter values, 2) time spent by each judge, 3) the results from coding, and 4) three case studies of graphics that were ranked differently by students and instructors.

5.5.5 Chapter Five: Discussion & Conclusions

Chapter Five discussed potential takeaways for instructors, teacher trainers and researchers from this thesis study. As part of the conclusions, it summarizes each chapter in this paper,

provides a final conclusion and suggestions for future work, and identifies potential directions for future research studies based on the findings of this thesis (section 5.6).

5.6 Conclusions

In conclusion, one of the reasons students and instructors from this introductory graphic design course critique differently from each other is the differences in role of instructors and role of students. With ACJ as a learning tool, students appeared to develop their understanding of aesthetical qualities while critiquing. While the students were able to keep the principles of design and project requirements in mind while critiquing, they spent considerable time describing their personal feelings of a design and struggled with finding the exact words or terminology to explain and to justify. To improve students' critiquing skills may require additional effort from instructors in teaching professional language used in graphic design and design critiques. Conversely, for the instructors, their experience and expertise appeared to expedite their decision-making process. In turn the researcher also noted that this may have worked to block them from communicating the rationale behind their critique. Instructors, especially experienced ones, may need to pay special attention to this "expert blind spot" as they critique student work and find ways to consciously share their thoughts with their students.

5.7 Future Work

The results from this study may help refine the current pedagogical practices and course structure of graphic design courses to enable future students to have better learning experiences and empower instructors to become better teachers. For example, the findings of this research suggest that ACJ, if used as a learning tool in students' critiquing activities, may foster the development of critiquing skills and an understanding of how design principles can be applied to

make an aesthetically pleasing image. Some potential ways of teaching students to use terminology in critiques were provided, but additional teaching strategies needs to be developed, tested and refined. It also suggests that instructors may need practice in verbalizing the rationale behind their formative critiques so they can best convey their experience and skills to their students. This can be applied to professional development for both experienced instructors and new instructors.

Future studies need to be done to further investigate critique differences within larger groups and with more projects; for instance, research can be done to include multiple projects from a course. The differences within the groups could be explored (this study did not address these because of the limited number of research subjects involved); for example, students with different levels of learning experience (e.g., freshman vs. senior students), instructors who teach similar introductory graphic design course at a different university (or from different countries), and so forth. Similar methodology could also be applied in different educational settings (courses/subjects/grade levels); for example, research can be done to use design products, design portfolios, or technical writing pieces where ACJ has been tested but has not yet been studied about how students and their instructors would approach similar or different assessment results through ACJ.

APPENDIX A. INSTRUCTOR SURVEY QUESTIONS

What is your gender?

- ○ Male
- ○ Female

What is your age?

What is your ethnicity?

- ○ White
- ○ Black or African American
- ○ American Indian or Alaska Native
- ○ Asian
- ○ Native Hawaiian or Pacific Islander
- ○ Other

What is your major at college?

How long have you been studying graphic design?

How long have you been teaching this course?

Have you taught any graphic design course other than this course?

- ○ Yes
- ○ No

If so, how long?

Please briefly describe if any graphic design experience outside of Purdue (think of technical experience like projects, internships and jobs, as well as teaching experience, high school).

APPENDIX B. STUDENT SURVEY QUESTIONS

What is your gender?

- ○ Male
- ○ Female

What is your ethnicity?

- ○ White
- ○ Black or African American
- ○ American Indian or Alaska Native
- ○ Asian
- ○ Native Hawaiian or Pacific Islander
- ○ Other

What is your school classification?

- ○ Freshman
- ○ Sophomore
- ○ Junior
- ○ Senior

What is your major?

How long have you been studying graphic design?

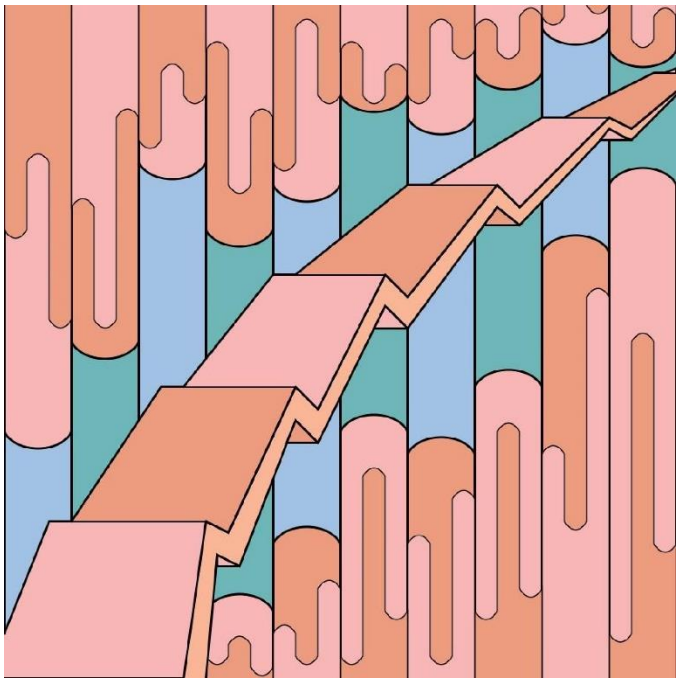
Have you taken any graphic design course other than this course?

- ○ Yes
- ○ No

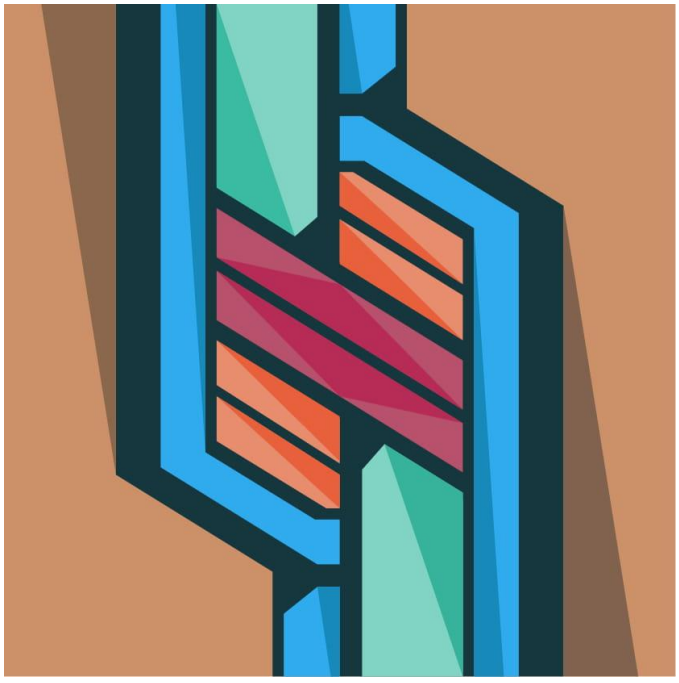
If so, how long?

Please briefly describe if any graphic design experience outside of Purdue.

APPENDIX C. DESIGN PROJECTS

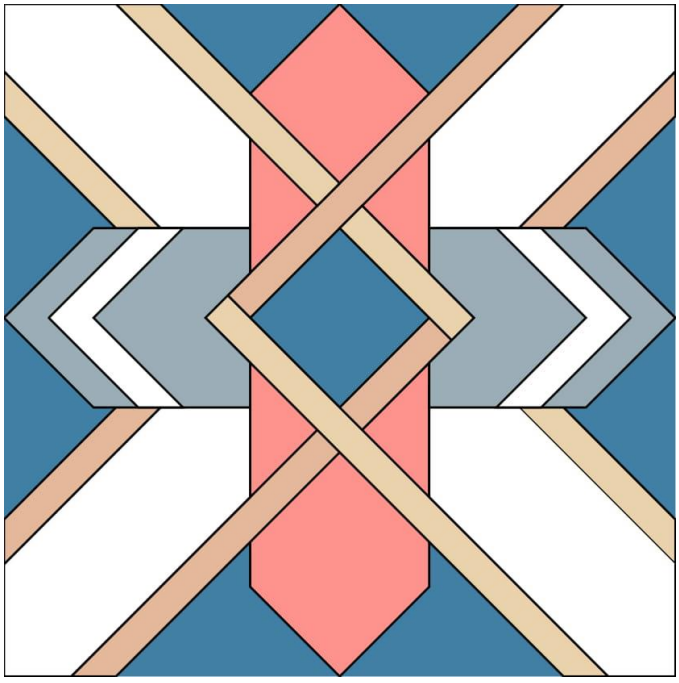


Project A



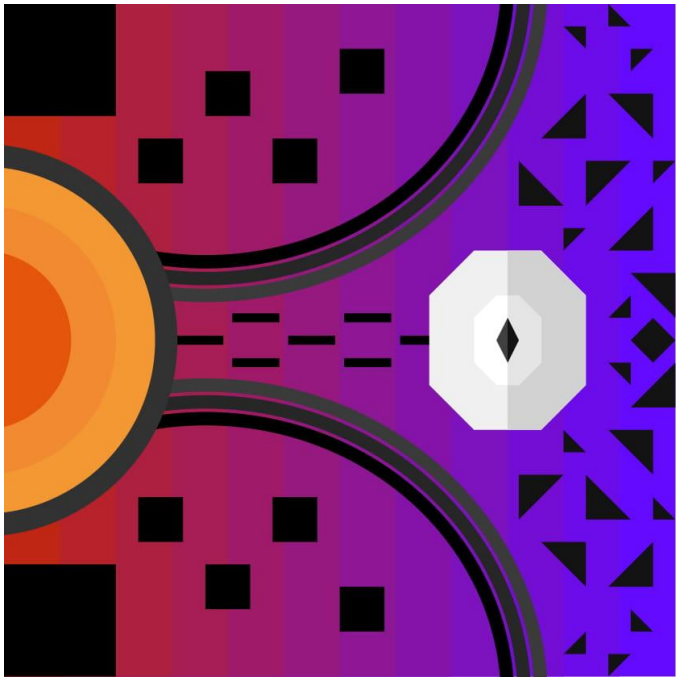
Project B

Project C

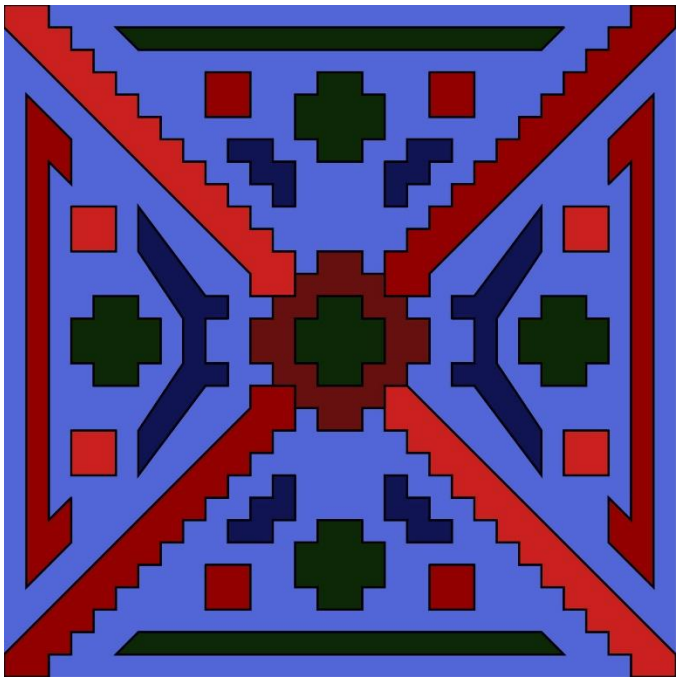


Project D

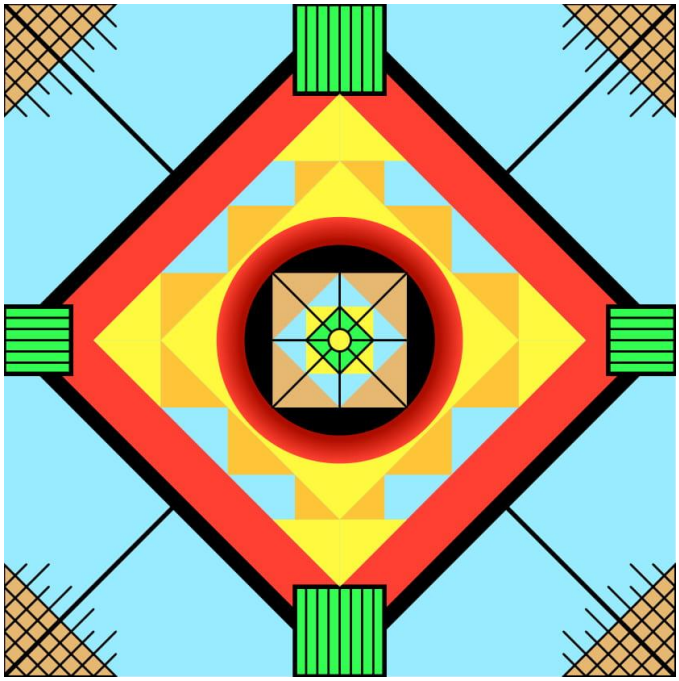
Project E



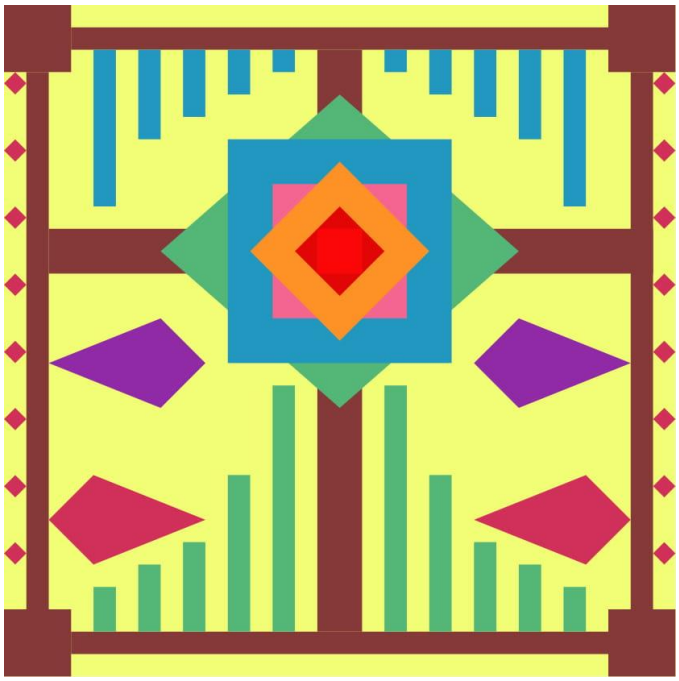
Project F



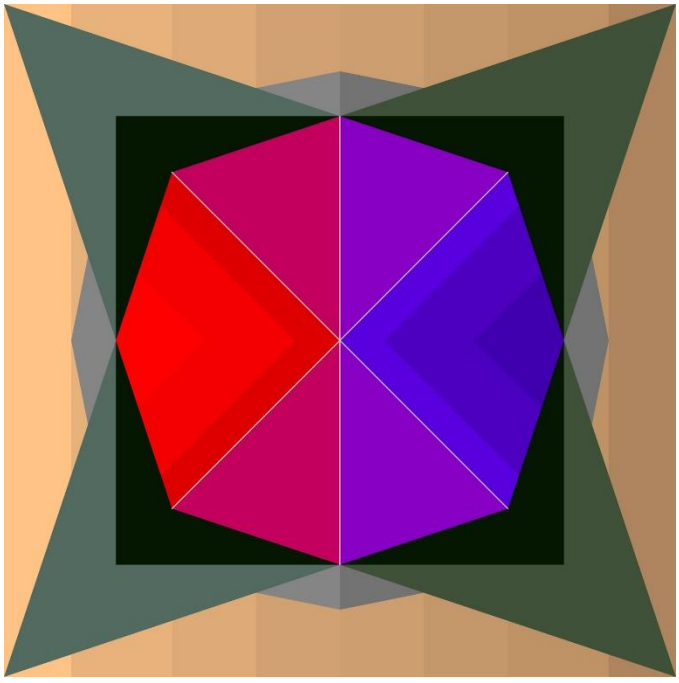
Project G



Project H



Project I



Project J

REFERENCES

Andrich, D. (1988). *Rasch models for measurement*. Sage Publications.

Andrich, D., & Luo, G. (2003). Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. *Journal of applied measurement*, *4*(3), 205-221.

Ball, L. J., Ormerod, T. C., & Morley, N. J. (2004). Spontaneous analogising in engineering design: a comparative analysis of experts and novices. *Design studies*, *25*(5), 495-508.

Barrett, T. (2000). Studio critiques of student art: As they are, as they could be with mentoring. *Theory into Practice*, *39*(1), 29-35.

Bartholomew, S. R. (2017). Assessing Open-Ended Design Problems. *Technology and Engineering Teacher*, 76(6), 13-17.

Bartholomew, S. R., Ruesch, E. Y., Hartell, E., & Strimel, G. J. (2019). Identifying design values across countries through adaptive comparative judgment. *International Journal of Technology and Design Education*, 1-27.

Bartholomew, S. R., Strimel, G. J., & Jackson, A. (2018). A comparison of traditional and adaptive comparative judgment assessment techniques for freshman engineering design projects. *International Journal of Engineering Education*, *34*(1), 20-33.

Bartholomew, S. R., Strimel, G. J., & Yoshikawa, E. (2018). Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *International Journal of Technology and Design Education*, 1-23.

Bartholomew, S. R., Zhang, L., Garcia-Bravo, E., & Strimel, G. S. (2018). *A Tool for Formative Assessment and Learning in a Graphics Design Course: Adaptive Comparative Judgment.* (Under Review).

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*(3/4), 324-345.

Cambre, J., Klemmer, S., & Kulkarni, C. (2018, April). Juxtapeer: Comparative Peer Review Yields Higher Quality Feedback and Promotes Deeper Reflection. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 294). ACM.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). New York: Routledge.

Cowan, J. (2017). The potential of cognitive think-aloud protocols for educational action-research. *Active Learning in Higher Education*, 1-14.

Daichendt, G. J. (2010). The bauhaus artist-teacher: Walter gropius's philosophy of art education. *Teaching Artist Journal, 8*(3), 157-164.

Dannels, D., Gaffney, A. H., & Martin, K. N. (2008). Beyond content, deeper than delivery: What critique feedback reveals about communication expectations in design education. *International Journal for the Scholarship of teaching and Learning*, *2*(2), 12.

Dannels, D. P., & Martin, K. N. (2008). Critiquing critiques: A genre analysis of feedback across novice to expert design studios. *Journal of Business and Technical Communication*, *22*(2), 135-159.

Demonacos, C., Ellis, S., & Barber, J. (2019). Student peer assessment using adaptive comparative judgment: grading accuracy versus quality of feedback. *Practitioner Research in Higher Education Journal*, *12*(1), 50-60.

DigitalAssess. (2018). Retrieved October 10, 2018, from http://digitalassess.com/about

Elkins, J. (2014). *Art critiques a guide* (3rd ed.). Washington, DC: New Academia Publishing.

Elkins, J. (2015). *Why art cannot be taught: A handbook for art students*. University of Illinois Press.

Ericsson, A. (2003). Valid and non-reactive verbalization of thoughts during performance of tasks. *Journal of Consciousness Studies*, *10*(9-10), 1-18.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data, Rev*. the MIT Press.

Freedman, K. (2003). *Teaching visual culture: Curriculum, aesthetics, and the social life of art*. Teachers College Press.

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative health research*, *3*(4), 430-441.

Gao, S., Krogstie, J., & Siau, K. (2011). Developing an instrument to measure the adoption of mobile services. *Mobile Information Systems*, *7*(1), 45-67.

Gibson, B. (1997). Talking the Test: Using Verbal Report Data in Looking at the Processing of Cloze Tasks. *Edinburgh Working Papers In Applied Linguistics*, *8*, 54-62.

Hartell, E., & Skogh, I. B. (2015). Criteria for Success: A study of primary technology teachers' assessment of digital portfolios. *Australasian Journal of Technology Education*, *2*(1).

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Heldsinger, S., & Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher, 37*(2), 1-19.

Hetland, L. (2013). *Studio thinking 2: The real benefits of visual arts education*. Teachers College Press.

Hicks, C. M., Pandey, V., Fraser, C. A., & Klemmer, S. (2016, May). Framing feedback: Choosing review environment features that support high quality peer assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 458-469). ACM.

Hokanson, B. (2012). The design critique as a model for distributed learning. In *The next generation of distance education* (pp. 71-83). Springer, Boston, MA.

Jones, I., & Alcock, L. (2013). Peer assessment without assessment criteria. *Studies in Higher Education, 39*(10), 1774-1787.

Jones, I., Swan, M., & Pollitt, A. (2015). Assessing mathematical problem solving using comparative judgement. *International Journal of Science and Mathematics Education, 13*(1), 151–177.

Joseph, G. M., & Patel, V. L. (1990). Domain knowledge and hypothesis genenation in diagnostic reasoning. *Medical Decision Making, 10*(1), 31-44.

Kelley, T. R., Capobianco, B. M., & Kaluf, K. J. (2015). Concurrent think-aloud protocols to assess elementary design students. *International Journal of Technology and Design Education*, *25*(4), 521-540.

Kimbell, R. (2008). E-assessment in project e-scape. *Design and Technology Education: An International Journal*, *12*(2), 66-76.

Kimbell, R. (2012). Evolving project e-scape for national assessment. *International Journal of Technology & Design Education*, 22, 135-155.

Klebesadel, H., & Kornetsky, L. (2009). Critique as signature pedagogy in the arts. In *Exploring signature pedagogies: Approaches to teaching disciplinary habits of mind* (pp. 99-120). Sterling, VA: Stylus Publishing.

Kulkarni, C. E., Bernstein, M. S., & Klemmer, S. R. (2015, March). PeerStudio: rapid peer feedback emphasizes revision and improves performance. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 75-84). ACM.

McCarthy, J. (2017). Enhancing feedback in higher education: Students' attitudes towards online and in-class formative assessment feedback models. *Active Learning in Higher Education*, *18*(2), 127-141.

McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.

McMahon, S., & Jones, I. (2015). A comparative judgement approach to teacher assessment. *Assessment in Education: Principles, Policy and Practice*, *22*(3), 368–389.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Mortier, A. V., Lesterhuis, M., Vlerick, P., & De Maeyer, S. (2015). Comparative judgment within online assessment: exploring students feedback reactions. *Communications in Computer and Information Science, 571,* (pp. 69-79).

Motley, P. (2015). Learning—to and from—the Visual Critique Process. *New Directions for Teaching and Learning*, *2015*(141), 77-86.

Motley, P. (2017). Critique and process: Signature pedagogies in the graphic design classroom. *Arts and Humanities in Higher Education, 16*(3), 229-240.

Neumann Jr, R. K. (1988). A Preliminary Inquiry into the Art of Critique. *Hastings Lj*, *40*, 725.

Öztürk, M. N., & Türkkan, E. E. (2006). The Design Studio as Teaching/Learning Medium—A Process-Based Approach. *International Journal of Art & Design Education*, *25*(1), 96-104.

Park, E. (2014). Learners' first exposure to written L2 input: Insights from learners' spontaneous think-alouds. In Z. Han & R. Rast (Eds.), *First Exposure to a Second Language: Learners' Initial Input Processing* (pp. 40-63). Cambridge: Cambridge University Press.

Perez, R. S., Johnson, J. F., & Emery, C. D. (1995). Instructional design expertise: A cognitive model of design. *Instructional Science*, *23*(5-6), 321-349.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied statistics*, 126-135.

Pollitt, A (2004, June). *Let's stop marking exams*. Paper presented at the International Association for Educational Assessment conference, Philadelphia.

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy and Practice*, *19*(3), 281–300.

Pollitt, A. (2015). On 'Reliability' bias in ACJ. *Cambridge Exam Research*. Retrieved from https://www.researchgate.net/publication/283318012_On_'Reliability'_bias_in_ACJ.

Pollitt, A., & Murray, N. L. (1996). What raters really pay attention to. *Studies in language testing*, *3*, 74-91.

Potter, T., Englund, L., Charbonneau, J., MacLean, M. T., Newell, J., & Roll, I. (2017). ComPAIR: A New Online Tool Using Adaptive Comparative Judgement to Support Learning with Peer Feedback. *Teaching & Learning Inquiry*, 5(2), 89-113.

Rangel-Smith, C, Lynch, D. (2018). Addressing the issue of bias in the measurement of reliability in the method of Adaptive Comparative Judgment. In *2018 PATT36 International Conference: Research and Practice in Technology Education: Perspectives on Human Capacity and Development. Athlone, Ireland. June 18–21, 2018.*

Saldaña, J. (2016). The coding manual for qualitative researchers (3rd ed.). Sage Publications.

Seery, N., Canty, D., & Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education*, *22*(2), 205–226.

Smith, R. A. (1973). Teaching aesthetic criticism in the schools. *Journal of Aesthetic Education*, *7*(1), 38-49.

Soep, E. (2005). Critique: Where art meets assessment. *Phi Delta Kappan*, *87*(1), 38-63.

Strimel, G. S., Bartholomew, S. R., Purzer, S., Zhang, L., & Yoshikawa, E. (2018). *Informing Engineering Design Through Adaptive Comparative Judgment.* (Under Review).

Tarricone, P., & Newhouse, C. P. (2016). Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability. *International Journal of Educational Technology in Higher Education*, *13*(1). https://doi.org/DOI 10.1186/s41239-016-0018-x

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286.

Utaberta, N., Hassanpour, B., Handryant, A. N., & Ani, A. I. C. (2013). Upgrading education architecture by redefining critique session in design studio. *Procedia-Social and Behavioral Sciences, 102*, 42-47.

Voet, M., Gielen, M., Boelens, R., & De Wever, B. (2018). Using feedback requests to actively involve assessees in peer assessment: effects on the assessor's feedback content and assessee's agreement with feedback. *European Journal of Psychology of Education*, *33*(1), 145-164.

Xu, A., Huang, S. W., & Bailey, B. (2014, February). Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 1433-1444). ACM.

Yuan, A., Luther, K., Krause, M., Vennix, S. I., Dow, S. P., & Hartmann, B. (2016, February). Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1005-1017). ACM.