

SHARED UNDERSTANDING IN ENGINEERING DESIGN CAPSTONE TEAMS

BY

ALEXANDER PAGANO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mechanical Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Doctoral Committee:

Professor Anthony Jacobi, Chair
Teaching Assistant Professor Molly H. Goldstein, Director of Research
Associate Professor Emma Mercier
Professor Matthew West

ABSTRACT

This study explores how Latent Semantic Analysis (LSA) can be used to understand the shared content of open-ended reflections (*RQ1*) as a means of identifying the typical elements of a collaborative design team's shared understanding of a design problem (*RQ2*). Undergraduate student design teams of 5-6 were recruited from a capstone design course in a Mechanical Engineering department at a large midwestern university in Spring 2022. Eight teams were selected for further study based on response rates. All students were asked to submit reflections throughout the semester, with one prompt repeating three times. Additionally, a survey was conducted to collect demographic information and additional context about the students' perceptions of their project. An emergent coding scheme was used to label utterances from responses to the repeated reflection prompt according to the type of content contained within each utterance ($n=664$). Utterances were also processed using LSA to approximate each utterance ($n=5162$) as a vector in the k -reduced semantic space. The sharedness of each utterance was computed as the cosine similarity between the vector representing the utterance and a composite vector representing the sum of utterances from teammates in response to the same reflection prompt.

The most common type of utterance found in design team reflections was Design Considerations. Design Considerations ($U=61337$, $p<0.001$) and Project Descriptions ($U=61337$, $p<0.001$) were found to have higher sharedness than the overall group of utterances using the Mann-Whitney U-Test. Process ($U=21128$, $p=0.010$), Perspective ($U=42216$, $p<0.001$) and Obstacles ($U=12961$, $p<0.001$) were found to have lower sharedness than the overall group of utterances using the Mann-Whitney U-Test. Three descriptive cases were explored representing

the approximate middle, lowest, and highest sharedness averages. Taken together, the descriptive cases demonstrated that sharedness varies dynamically over a project timeline, influenced by factors such as project clarity and team member specialization. Teams emphasized the importance of external representations such as block diagrams or rough prototypes in managing a shared understanding, not just among team members but also with stakeholders. The Team-Task-Design Mental Model Framework was used to examine these results, finding that utterances with high sharedness represented aspects of design-related mental models, whereas lower sharedness utterances represented aspects of task-related mental models. Future work will explore sensitivity analysis of LSA, validation of LSA sharedness against other mental models metrics, as well as a comprehensive ethnographic case study following teams more closely.

ACKNOWLEDGMENTS

To contribute to scientific research is to participate in a beautiful collective. By definition, peer-reviewed research depends on the efforts of communities of collaborators, working together to further human knowledge. While gratitude rains on the so-called “giants”, I would like to thank all participants who may call themselves scientists, not for the gravity of their contribution, but for their commitment to the discovery of objective truths through evidence-based inquiry. It seems this is no longer the norm in some, troublingly large, corners of society. Thank you to everyone who spends or has spent their time contributing to the endeavor of science, the references section represents only a small fraction of your influence on me and my work.

For me and for many others, the completion of a PhD represents the culmination of a lifetime goal. Regardless of the journey taken to get here, no one does it alone. The path that has brought me to this point has taken me to places I never expected. At each turn, there were people whose love, support, patience, curiosity, advice, and help buoyed me and made it possible to navigate what lay ahead. To each and every one of you, thank you, I would not be here without you.

I would particularly like to thank the wonderful teachers and professors that helped me learn enough about the world to carve out my own place in it. Your example has inspired me and set me on this path.

Thank you to my advisors, Dr. Molly H. Goldstein, and Dr. Anthony Jacobi. Your kindness, patience, experience, and guidance were invaluable and created opportunities for me to grow in ways that will serve me for the rest of my life. Tony, thank you especially for opening

doors to enable my reckless pursuit of curiosity and for your willingness to entertain my lofty ideas as well as your perspective to help ground them so they could be realized. Molly, thank you for sharing your expertise and for helping me become a better researcher. Also thank you for your boundless positivity and eagerness to help no matter how much time was needed.

The Siebel Center for Design gave me the opportunity to pursue my interests when others did not, and for that I am incredibly grateful. Thank you to Rachel Switzky for your commitment to grounding the center in evidence-based research and for creating countless opportunities for students like me to participate in that effort. Thank you also to Dr. Saad Shehab and Taylor Parks for being wonderful team members. Our countless talks have been equally enjoyable, enlightening, and productive.

To my whole family, but especially my mom, dad, and brother, I love you all so much. Your constant love and support kept me going. To Melody, you may never know it, but you made this possible.

Finally, thank you to Dr. Blake Johnson and all the students who participated in the study. Without you, none of this would exist.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	5
CHAPTER 3: METHODS	23
CHAPTER 4: RESULTS	56
CHAPTER 5: DISCUSSION	97
CHAPTER 6: CONCLUSION	122
REFERENCES	135
APPENDIX A: IRB LETTER.....	148
APPENDIX B: SHAREDNESS DATA BY TEAM	149

CHAPTER 1: INTRODUCTION

Introduction and Motivation

After showing the prompt, “I have an electric kettle that I use to make coffee in the morning, but it gets too hot – can you help me?” students write down responses to two questions, “what is the goal?” and “how do you define success?” This exercise is part of an introductory engineering design guest lecture I have been fortunate enough to give every semester for the last few years. Each time I’ve repeated the exercise, the result is the same. There is a group of students who commit to decreasing the water temperature, aiming for a specific target or accuracy. Another group interprets the prompt to mean that the kettle is a safety hazard, deciding that features are needed to prevent the user from being burned. Others assume a combination of safety and water temperature or focus on control or customization. The reliable lack of consensus serves to illustrate the importance of clarifying questions and seeking shared understanding in design contexts. Design problems are ill-structured and often emergent, revealing new questions and considerations over time as more information is gathered. The kettle example shows just how easy it is to reach wildly different conclusions when working with limited information. It is critical that design teams establish a shared understanding, not just among the team members, but also with the various stakeholders, because no amount of engineering optimization or efficiency will save a project that has misunderstood goals.

The motivation for this dissertation comes from my involvement with the Siebel Center for Design. In my role at SCD I have had the opportunity to work with many faculty members who teach design courses in different disciplines and who describe similar problems that their students face. Notably, students assume that all problems, including design, can be treated

similarly to textbook problems, and can be solved if the correct formula or approach is identified. This assumption leads to a solution-oriented approach primarily driven by a search for examples of similar solutions that may be co-opted. The tendency to skip research to start “cutting metal” is well-known in engineering education literature. However, this can lead to project failure in a number of ways which are all too common. First, the team does not properly understand the stakeholder’s needs, as shown in the electric kettle example. In this case, the team may develop a sophisticated solution only to discover when they turn it over to the client or stakeholders that it solves the wrong problem. Secondly, team members understand the problem differently from one another, resulting in a counterproductive tension pulling the project in different directions. Both of these problems are insidious and may not be realized until very late in the design process. While comparisons between novice and expert designers demonstrate that experts have a keener sense for the importance of understanding the challenge (Crismond & Adams, 2012), the existing research does not adequately address how successful design teams develop and manage a shared understanding throughout a design project. A more comprehensive theory of how design teams can successfully nurture a shared understanding of their objectives as they relate to real, unmet needs experienced by known stakeholders, would facilitate more effective engineering design problem solving across the board.

While the importance of a common goal or shared understanding may seem self-evident at first, it quickly becomes tangled by the complications of real-world design projects. What exactly does it mean to have a shared understanding or to be ‘on the same page’? What is shared, and with whom? How is this shared understanding developed and used? Surely there are instances within design where too much sharedness is detrimental, *e.g.* in cases of groupthink

during team brainstorming, or for teams with specialized roles where distinct skillsets are most useful. Conversely, there are undoubtedly times when more sharedness is better, for example, during requirements elicitation to ensure that the team correctly understands the stakeholder's needs. So then, while sharedness is important, it may be better to talk about maintaining the right levels of sharedness at the right times. What then is the "right level" of sharedness? How might this change over time? While progress has been made in the last 80 years, many of these questions remain unanswered. This dissertation offers contributions toward answering these questions.

Definitions and Explanations of Term Usage

Design – I tend to use both a noun and verb form of 'design' but avoid its use as a noun meaning the object of design, as in "the team showed me their design." This is because the transitive verb form, as in "they designed a bridge" requires the direct object to be specified, and this direct object could be the noun form of "design." This results in the confusing, but grammatically correct, "they designed the design." Since we often refer directly to the act of designing in design research communities, this presents a common lexical tripping hazard. To avoid this, I tend to not use "design" to mean the thing being designed, opting instead to be more descriptive or default to 'design object' or something similar for more general cases. However, I do use the verb forms to describe someone engaging in design work, as in "she will design that widget", and "our team designs collaboratively." I also like to use "design" to refer to the activity of designing, as in "the team was engaged in design." Or alternatively, to refer to the field of Design itself (note the capital), much like one refers to Math, or Science. This may be a less common way to use the word and might be a point of confusion among readers, however I

believe that an all-encompassing term is needed for the domain of Design despite the lack of a universal definition. The personal search for an answer to the question, “what is design?” is an important foundational step in design work and research.

Shared – In this work, I will predominately use shared to mean overlapping or common rather than distributed or complementary. This topic is discussed in Chapter 2, Mental Models as well as Chapter 3, Conceptual Framework.

Stakeholder – Stakeholders are all those players who are impacted by design decisions. Stakeholders include end-users as well as clients, but also others at various stages of the product’s development and lifecycle. For example, manufacturers are stakeholders impacted by the manufacturability of the design object, maintenance and service people are stakeholders impacted by the repairability of the design object. There may also be nonhuman “stakeholders” such as environmental and sustainability objectives, or others, e.g. dogs are stakeholders for the design of dog beds.

Need – The need is a design term of art used to refer to the problem faced by the stakeholders. Readers should not interpret “need” to be a value judgement of the urgency or importance of the request. A stakeholder can “need” a solution to their problem, even if their problem is not comparable to the human need for air and water. Needs can have different levels or priority and may generically refer to wants, desires, preferences, or other beliefs of the stakeholders. This term is sometimes clarified as *unmet needs*, *underlying needs*, or *stated needs*.

CHAPTER 2: LITERATURE REVIEW

Shared Understanding

The concept of a shared understanding amongst team members arises in a variety of literature contexts, “Joint Problem Space”, (Roschelle & Teasley, 1995), “Team Mental Models” (Rouse, Cannon-Bowers, & Salas, 1992), “Trading Zones” (Galison, 1997), communication across “Object Worlds” (Bucciarelli, 1994), “Cognitive convergence” (Teasley et al., 2008). While definitions and applications vary, these concepts describe a tendency for collaborating team members to seek common ground when communicating ideas and problem solving. This may be necessary when the individuals come from different backgrounds, as in Bucciarelli and Galison, or as knowledge is built collaboratively, as in Roschelle and Teasley, *et al.*, or when team members rely on each other to complete a task, as in Rouse *et al.*

Studies of engineering design practice show that despite a common tendency to describe “Engineering” as objective and technical, distinct from the social dimensions of engineering workplaces, engineering is a heterogeneous practice comprised of elements which are both human and technological (R. Stevens, Johri, & O’connor, 2015). The social and technical aspects cannot be separated (Faulkner, 2015; R. Stevens et al., 2015). In ethnographic studies of engineering design teams, Minneman and Bucciarelli both describe engineering as a social process in which designers negotiate a shared understanding (Bucciarelli, 1988, 1994, 2002; Minneman, 1991). As Roschelle & Teasley determined in an educational context, such a socially constructed shared understanding is mediated by a “Joint Problem Space” constructed in part by shared language (1995). The development of shared language may have broader implications for design, as Bucciarelli frames it in chapter six of *Designing Engineers*, “In this process of getting

the names right and developing shared meanings, they were constructing the objects of design as much as classifying a collection of clearly defined artifacts with reference to a set of given, clearly understood criteria.” (1994). Thus, shared language may offer a window into a design team’s shared understanding of the problem.

Mental Models

Origins of Mental Models Research

The idea that people create some internal representation of their world as a way to understand and move through it has been around for a very long time. Plato’s Allegory of the Cave, for example, draws the distinction between the representation of reality cast as a shadow on the wall of a cave and reality itself (Plato & Jowett, 1941). For the people in the cave, the shadow cast on the wall is all they know of the world, but people outside of the cave see the world differently from those inside it. Understanding depends on one’s perspective. The idea that each person has their own internal representations of the world persists, and came to take the name *mental models* in Craik’s, *The Nature of Explanation*, (Craik, 1943; Gottschling, 2009). Another early example of mental models is in Forrester’s *Industrial Dynamics* (1962). Forrester, considered to be a founder of system dynamics, used mental models to explain one’s often simplified internal representation of a complex system. In the 1980s the idea of mental models found wider use in the fields of cognitive psychology, and human factors and ergonomics by two books published in the same year. The first, by Johnson-Laird, used mental models as a way to explain human reasoning in the context of cognitive psychology (Johnson-Laird, 1983; Johnson-Laird, 1980). The second, by Stevens and Gentner, used mental models as a way to discuss human-machine interactions in the context of human factors and ergonomics (1983). Stevens and

Gentner described an individual's mental model of a piece of equipment which represents how each person understands the system to work and informs how they interact with it (1983). In all of these cases, mental models are considered to be a simplified representation of some external system that humans use to reason about and make decisions regarding that system. They are treated independently, each person possessing their own mental model, potentially comprising many models from different conceptual frames. Rouse and Morris offered the following definition for mental models, "Mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions (or expectations) of future system states" (1986). An important distinction is drawn between "mental models" and "knowledge", that mental models include more than just what is known, but also how knowledge is structured and related. Rouse and Morris suggest that mental models should be treated differently when they are employed in different disciplines such that the specific knowledge included and purpose for mental models in that context can be best understood (Rouse et al., 1992; Rouse & Morris, 1986).

Team Mental Models

Mental models were applied to teams in the 1990s. Early research on mental models in teams focused on operational teams, often in high-pressure scenarios such as military command and control, flight teams, or surgical teams (Klimoski & Mohammed, 1994; Mathieu, Goodwin, Heffner, Salas, & Cannon-Bowers, 2000; Rouse et al., 1992; Rouse & Morris, 1986; Xiao, Parker, & Manser, 2013). In these contexts, teams work through an established procedure to accomplish the task at hand. Operational teams may encounter unexpected scenarios and must respond creatively to resolve issues, but for the most part their tasks are characterized by a

known list of steps with established roles. Examples of domains in the team mental models literature include, flight crews (Cooke, Salas, Cannon-Bowers, & Stout, 2000; Mathieu et al., 2000; Weick & Roberts, 1993), power plant control crews (Waller, Gupta, & Giambatista, 2004), software development teams (Carley, 1997; Espinosa, Lerch, & Kraut, 2004), sport teams (Eccles & Tenenbaum, 2004; Webber, Chen, Payne, Marsh, & Zaccaro, 2000), management teams (Ensley & Pearce, 2001), long-term student groups (Peterson, Mitchell, Thompson, & Burr, 2000), and development teams (Akgün, Byrne, Keskin, Lynn, & Imamoglu, 2005; Davison & Blackman, 2005). With roots in cognitive psychology, many early notions of mental models employed cognitivist theories of learning and knowledge (Marshall, 2007). To drastically oversimplify, cognitivism focuses on the acquisition of knowledge and the internal representation and structure within an individual's mind (Ertmer & Newby, 2013). This individualistic conceptualization, with a focus on internal representations of an external, objective reality establishes, "strict dualisms between mind and body, thought and reality, individual and environment" (Marshall, 2007). Researchers familiar with teams and collaboration recognized that this cognitivist framework of mental models did not address the collective, situated, and socially constructed ways of knowing that exist in collaborative teams (Descombes & Schwartz, 2001; Ertmer & Newby, 2013; Marshall, 2007). As a result, mental models research began to accommodate more constructivist learning theories and principles. One example of this is the dual nature of team or shared mental models. Generally, team mental models refer not only to the individual mental models of the team members but also to, as Langan-Fox *et al.* said, "a synergistic functional aggregation of the teams mental functioning

representing similarity, overlap, and complementarity” (Badke-Schaub, Neumann, Lauche, & Mohammed, 2007; Langan-Fox, Anglim, & Wilson, 2004).

Similarity of mental models has long been recognized as an important metric (Klimoski & Mohammed, 1994; Rouse et al., 1992). In the context of operational teams, empirical studies have shown that similarity of mental models has a positive effect on team performance (Lim & Klein, 2006; Marks, Sabella, Burke, & Zaccaro, 2002; Mathieu, Rapp, Maynard, & Mangos, 2009; Smith-Jentsch, Mathieu, & Kraiger, 2005). While mental model elicitation methods and the methods used to determine similarity differ across studies, in a meta-analysis of twenty-three empirical studies on shared mental models, Dechurch and Mesmer-Magnus came to the conclusion that, “Across the elicitation methods, structure representation approaches, and representation of emergence used in past research, positive relations were observed between the similarity in mental models and team performance” (Dechurch & Mesmer-Magnus, 2010). For a more detailed review of the various methods used to elicit, represent, and compare mental models see Dechurch & Mesmer-Magnus, 2010. However, the importance of similarity also seems to vary by the type of content represented in the mental model. Several different types of mental models have been proposed as the components of the Mental Model in full. These include models of the Team, Task, Equipment, and Team Interactions (Rouse et al., 1992), among others such as a Transition Model (Smulders, 2007), and an Environment Model (Mohammed, Klimoski, & Rentsch, 2000). In general, these are considered to fall into two broader categories: Task-related and Team-related models (Badke-Schaub, Neumann, et al., 2007; Mathieu et al., 2000). Task-related mental models include knowledge of the task at hand, as well as content related to the technology, equipment, or tools in use. The equipment model might include

operating procedures, system limitations, potential failure points, or a physical understanding of where everything is and how it interacts. The task model might include the task procedures, likely scenarios, contingencies, suitable strategies, environmental constraints, etc. Team-related mental models include all content related to the team as individuals as well as the interactions between team members. The team model is focused on the teammates as individuals, describing what is known about the teammates' knowledge, skills, attitudes, preferences, tendencies etc. The team interaction model includes the knowledge about interactions between teammates and the properties of the team as a whole. It may include knowledge of the roles and responsibilities, sources of information, interaction patterns and the flow of information through the team, etc. While similarity of mental models has been generally shown to be related to team performance, the degree of this effect may vary for different types of mental models. For example, Bierhals *et al.* found that team and process models are related to successful team performance, but shared task models are not as critical due to specialization of team members (2007). Alternatively, Lim and Klein found that both task-related model and team-related model similarities were predictive of team performance (2006). This variation elucidates the importance of considering multiple mental model types.

There are several ways to interpret similarity and sharedness in the context of teams. For example, shared could mean overlapping sets, i.e., two independent mental models contain the same content such that the intersection between sets is shared. Alternatively, shared could mean distributed in the sense that content is divided between mental models. This is related to transactive memory, e.g. partners in a relationship dividing up the information that they are responsible for, so that the couple collectively has all of the knowledge they need but it is

distributed between the individuals (Wegner, 1987; Wegner, Erber, & Raymond, 1991). Additional distinctions have been drawn between sharedness as *identical*, *overlapping*, *equivalent*, *distributed*, and *complementary* (Cooke et al., 2000; Marshall, 2007; Rouse et al., 1992; Yi, Damen, & Toh, 2020). There is also the question of shared by whom - does sharedness necessarily involve all members of a team, or can things be shared between a subset of team members? This topic will be picked up in Chapter 3, in the Section, “Creating a Metric for Sharedness.”

No matter how it is defined, the similarity or sharedness of mental models says nothing of their accuracy. A team could have perfectly similar mental models which happen to be wrong in exactly the same ways, therefore similarity does not ensure accuracy. The premise of accuracy in mental models often hinges on the existence of some identifiable, correct version to serve as a baseline for comparison. For example, in operational teams accuracy may be measured as similarity to the mental model of an expert (Dechurch & Mesmer-Magnus, 2010; Lim & Klein, 2006) or to the average of several experts (Mathieu et al., 2000). However, even within teams working through a predetermined procedure, there may be choices which arise, unexpected or not, for which there are multiple valid options (Mathieu et al., 2009). Experts may disagree about which approach is best. This type of divergent decision-making is especially prominent in design (Badke-Schaub, Neumann, et al., 2007).

Team Mental Models in Design

The creative, emergent, and ill-structured nature of design causes several issues if “designing” is treated as the task in the existing team-related and task-related mental model framework. First, design is not procedural. The requisite steps are not known at the start. This

poses methodological issues as mental model assessment techniques may rely on prior knowledge of the relevant components at play to prepare tools for assessment (Badke-Schaub, Neumann, et al., 2007). For example, card sorting activities to elucidate the structure of one's mental model require a set of cards to be created beforehand. These cards must suitably map the contents of one's mental model so that their arrangement provides an accurate representation of the structure (Dechurch & Mesmer-Magnus, 2010). In design, the content of one's mental model may include any number of things which cannot be predicted at the start. Second, there may be many valid approaches to any design problem. Experts, if they exist at all, may disagree about the correct approach, making accuracy difficult to define in the usual way. Third, diversity is beneficial in design. Teams who think exactly alike are less able to fully explore the design space and may develop blind spots. This contradicts the assumption that teams who have more similar mental models are able to better perform collaborative tasks in the context of design. Finally, unlike operational teams, design teams often create something new. This is a complicated task that requires an understanding of how the system works as a whole, how the design solution works, its function, behaviors, and structures, as well as how the design solution fits into the overarching system. There are many existing models and system frameworks in engineering design that are intended to help wrangle this complexity, including the functions, behaviors, and structures of FBS (Gero & Kannengiesser, 2004), Suh's Axiomatic Design (1995), and Quality Function Deployment (Chan & Wu, 2002). Would these be considered task-related models, or are they distinct? While it may be technically accurate to say, as is common in the literature, that all design considerations fall under the task-related model because they relate to the overarching task of "designing", this lumps together one's understanding of the object of design with their

understanding of the design tasks that refine and develop it. The object informs the tasks just as much as the tasks shape the object. Lumping together the mental models of task and object would obscure the complicated interactions that happen as a team iteratively develops a design, uncovering new tasks and considerations as the design object is refined. This suggests that there are “design-related” mental models which could be distinguished from task-related and team-related models (see “Conceptual Framework”, in Chapter 3).

Design-specific conceptual frameworks of team mental models have precedent in the literature. Researchers identified the need to adapt mental models frameworks for use in different contexts relatively early (Rouse & Morris, 1986). Design-specific mental models also predate their use in design teams. Stevens and Gentner hint at a designer’s mental model of the system being designed, and indicated that this may differ from the mental model of the user (1983). In a 2007 special issue of *CoDesign*, Badke-Schaub *et al.* introduce the concept of team mental models in design, and propose a Team, Process, Task, Context, Competence framework (Badke-Schaub, Lauche, & Neumann, 2007; Badke-Schaub, Neumann, et al., 2007). This special issue of *CoDesign* contains eight papers in total including the lead paper by Badke-Schaub *et al.* as well as five comments and two empirical studies. Boos discussed optimal sharedness from a social psychology perspective, and highlighted that the degree of sharedness that may be most beneficial varies over time and across different tasks and design phases (2007). Marshall presented issues related to the cognitivist underpinnings of mental models, and suggested a practice-based approach that accommodates a situated understanding of cognition (2007). Schaub highlighted the relevance of task characteristics, and discussed differences between operational teams and design teams (2007). Goldschmidt emphasized the importance of visual

representations in the development of shared mental models (2007). Smulders focused on the importance of mental models for the communication and synchronization between teams, and argued for the inclusion of a transition model which accounts for the boundary interactions across teams (2007). Bierhals *et al.* measured the impact of shared mental models on performance in student and industry teams (2007). They reported that team-related mental models and process models were related to performance, with team-related models most impactful. However, task models were not predictive of team performance. They explain this as a result of the division of responsibilities for various subtasks (Bierhals et al., 2007). Kleinsmann *et al.* investigated influencing factors for shared understanding in multidisciplinary design teams, and identified barriers and enablers at three organizational levels corresponding to the actor, the project, and the company (Kleinsmann, Valkenburg, & Buijs, 2007).

Since their introduction to the design team research community, mental models have continued to be an oft-used construct among researchers (Borrego, Karlin, McNair, & Beddoes, 2013; Casakin & Badke-Schaub, 2017; Dong & Kleinsmann, 2016; Kim, 2019; Yi et al., 2020). In a retrospective review, Mohammed *et al.* described the state of art of team mental models research, outlining many best practices and future directions (2010). Notably, they advocate for the specification and measurement of multiple mental model content domains, rather than taskwork or teamwork alone. They also highlight the need for methods which can be used to measure the development of mental models over time. In a systematic review of mental models in engineering design teams, Kim demonstrates that while early findings continue to be robust in light of new results, the community remains spread across a variety of measurement and analysis techniques (2019). There are still significant opportunities for continued team mental models

research, especially in collaborative design contexts. Methods that allow for the exploration of the mental model development across multiple domains are especially needed. One promising approach involves the use of a natural language processing technique called Latent Semantic Analysis.

Latent Semantic Analysis

Introduction to Latent Semantic Analysis (LSA)

LSA is a mathematical and statistical method that is used to approximate written or transcribed utterances as numerical vectors (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Landauer, Foltz, & Laham, 1998). This enables the computation of the degree of similarity between two utterances using the cosine angle between vectors. The vector approximation is computed from the stochastic usage of terms in the context of a document (e.g. page, paragraph, sentence, utterance) and is therefore entirely informed by the set of documents, collectively called the corpus. No additional training input is needed, and all semantic meanings and associations are implied from the data itself. With a sufficiently large corpus, the latent semantic associations between terms used in context provide a robust approximation of the concepts discussed in each utterance. One way to think about this is that LSA computes a kind of average meaning for each term based on all of documents in which it appears (Landauer et al., 1998). Simultaneously, the meaning of all documents is approximated by a kind of average of all the words they contain (Landauer et al., 1998). In this way, each document is approximated in a reduced semantic space as a vector. Since all documents exist in this same semantic space, they can be compared directly to determine their similarity, and otherwise manipulated using vector operations. For example, the utterance, “the fruit basket contains apples, oranges, and bananas”

would be assigned a vector which expresses a connection between the terms “fruit basket”, “apples”, “oranges”, and “bananas.” Note here that terms need not be single words, but can be multiple words in sequence, generically called n-grams. If another utterance is added to the corpus, for example, “the transit system relies on buses” this utterance will be approximated with a vector that connects “transit system” and “buses.” These vectors also encode information about the semantic relationships between terms like “apples” and “buses” which do not appear together directly. These vectors can be used to determine the similarity of the utterances quantitatively.

LSA Process

The procedure for Latent Semantic Analysis is as follows, for more information see Dong, 2005, Dumais, 2004, and Landauer et al., 1998. First a term document matrix (TDM) is constructed such that each column of the matrix corresponds to a document, and each row corresponds to a term contained in the corpus. This matrix should be very large since there must be a row for every term used in the entire corpus, including all n-grams. This TDM is then populated with the number of times each term appears in each utterance. Since most terms will not appear in most utterances, this matrix will be very sparse. This very large, sparse matrix can be approximated in a consolidated form by using singular value decomposition (SVD) to reduce its dimensionality. The TDM is first inverted to get frequencies and normed to get the term-frequency, inverse-document-frequency matrix, $\mathbf{X}$. This matrix, $\mathbf{X}$, is $n \times m$, with n terms, and m documents. Then use SVD to decompose $\mathbf{X}$ into the matrix product, $\mathbf{X} = \mathbf{T}\mathbf{S}\mathbf{D}^T$, where $\mathbf{S}$ ($m \times m$) is the diagonal matrix of singular values, and $\mathbf{T}$ and $\mathbf{D}$ are the left and right orthonormal singular matrices respectively. The dimensionality of $\mathbf{X}$ can be reduced by taking only the k -most significant singular values of $\mathbf{S}$, replacing all others with 0. The resulting matrices, $\mathbf{X} \approx \tilde{\mathbf{X}} =$

$\mathbf{T}_k \mathbf{S}_k \mathbf{D}_k^T$, give the best least-squares approximation to $\mathbf{X}$ with k parameters, such that the rows in $\mathbf{T}_k$ are the term vectors in LSA space and the rows in $\mathbf{D}_k$, are the document vectors in LSA space (Dumais, 2004). The k -reduced matrix, $\tilde{\mathbf{X}}$ has dimension $m \times k$, each row-vector corresponds to a document expressed in k parameters. In this way, documents can be compared using the cosine angle between their k -reduced vector form in $\tilde{\mathbf{X}}$.

Fidelity of LSA to Human Results

LSA has been shown to preform about as well on the Test of English as a Foreign Language (TOEFL) as the average TOEFL test-taker (Landauer & Dumais, 1997; Landauer et al., 1998). When trained on the text of introductory psychology textbooks then tested using multiple-choice tests from the publishers of the books, LSA performed well above chance, and while it performed worse than college students in a related comparison using a psychology course exam, it did well enough to receive a passing grade (Landauer et al., 1998). Dumais identifies a number of ways that LSA may be similar to human responses including essay grading, semantic priming, textual coherence, similarity neighborhoods, word sense disambiguation, and analogical reasoning (Dumais, 2004).

LSA for Mental Models Research

LSA has also been applied to team mental models research, building upon earlier LSA work on design team communication (Dong, 2005; Dong, Hill, & Agogino, 2004; Hill, Dong, & Agogino, 2002; Song, Dong, & Agogino, 2003). Dong's 2005 article contains overviews of studies which use LSA to assess both synchronous, i.e. transcripts, and asynchronous communication, i.e. emails, missions statements, customer feedback, product testing and engineering memos, presentation materials and personal and group reflections (Dong, 2005). For

asynchronous communication, they measured semantic coherence and variation in semantic choice as metrics of congruent communication and reported that low variation in semantic choice and high semantic coherence were related to the quality of design outcomes (Dong, 2005; Hill et al., 2002). Synchronous communication data from the 1994 Delft Protocols Workshop (Cross, Christiaans, & Dorst, 1996) and the Bamberg Study of design thinking (Stempfle & Badke-Schaub, 2002) was analyzed using a running average semantic coherence of utterances that were temporally adjacent, showing that successful teams had patterns of conversation in which ideas built on one another, i.e. increasing coherence, “lead[ing] to the construction of the socially held representation” of the design problem (Dong, 2005). Dong also measured knowledge convergence in design teams by computing the cosine similarity between the group concept centroid and each speaker’s running average (2005). Notably, the resulting curves representing each speaker distinguish whose contributions are most similar to the group centroid, thus dominating discussion, and who may be *yes-men* or naysayers. The rate of convergence may also be significant, with teams converging too quickly falling victim to groupthink or cognitive priming, while teams that converge too slowly may face other problems. While the convergence rates corresponding to specific problem areas are not identified, there are notable differences between teams indicating the viability of this approach (Dong, 2005). In a 2013 paper, the connection to team mental models was further honed as the authors used LSA to measure sharedness and accuracy of team mental models (Dong, Kleinsmann, & Deken, 2013). They computed sharedness as the similarity, by cosine angle, between the average of each team member’s contributions up to a given point in time and the group’s concept centroid at that time (Dong et al., 2013). Accuracy was computed using the total room mean square error between

each team member's similarity to the team mental model (Dong et al., 2013). While these methods are interesting and potentially very useful, there are several possible problems. Firstly, similarity or sharedness computed with respect to the group's "concept centroid", defined as the average of all utterances from all speakers thus far, may be artificially high due to self-similarity as will be discussed in Chapter 3 in the section, "Creating a Metric for Sharedness." Secondly, because team coherence is computed as the similarity between the running concept centroid and the concept centroid at the end of the session, it will always converge to 100% coherence, which may improperly suggest that the team is converging around a similar understanding. Nonetheless, the "concept centroid", as an average of the team's contributions approximated use LSA, represents a notable contribution for mental models research, because it offers a simplified representation of the team's semantic understanding of the problem at hand. As such, it closely matches common definitions of team mental models, and may have utility as an analogue for them (Dong, 2005; Dong & Kleinsmann, 2016; Dong et al., 2013).

An LSA-based approach to mental models' measurement has a number of strengths. It does not require any knowledge about mental model content *a priori*. Since all of the processing is done after data collection and semantic meanings are derived from the corpus alone, LSA can be used where methods such as card sorting which require prior knowledge of mental model content are not applicable. LSA works with unstructured text-based data which can be sourced from transcripts or written documents; therefore, it can be used for think-aloud protocol, open-ended responses, reflections, or project reports. With a relatively small amount of initial preparation, a large amount of data can be processed quickly, allowing for data collection at a large scale. It is not necessary to pull participants away from the task to engage in some

measurement procedure, such as card sorting or concept mapping, therefore design can continue uninterrupted. As Dong argues, LSA produces a simplified representation of the information expressed in the text, which is similar to the definition of mental models as simplified representations of the world (Dong, 2005; Dong & Kleinsmann, 2016). Text-based data has long been used in mental models research and may be considered to represent a portion of the author's mental model at the time the text was created (Carley, 1997). While language is not the only means of communication or knowledge expression that we use, it is a major one that is ubiquitous and accessible. Of course, the internal, intangible mechanisms of cognition cannot be directly probed and measured, so researchers must rely on tangible expressions of them – studying textual data is one way this can be accomplished. Shared language has long been shown to be an expression of collaborative thinking, for example in the joint concept spaces described by Roschelle and Teasley (1995).

However, a reliance on language is not without weaknesses. Several authors have already highlighted the importance of external representations as a mediator of shared understanding, for example, sketches, prototypes and other so-called boundary objects (Bucciarelli, 1994, 2002; Goldschmidt, 2007; Smulders, 2007). Sometimes words are incapable of capturing a concept, thus there are likely aspects of the mental model which are not accessible through text-based analysis. For example, as a trumpet player I have an embodied sense for the combination of embouchure contortion, diaphragm pressure, and breath support required to play specific notes. In order to produce that note consistently, I must be able to reliably assume the same physical orientation in response to seeing a note written on a page. My “mental” model associated with playing the trumpet necessarily includes that embodied sense of what it feels like to play that

note. However, despite my best attempt to describe it in words, this description does little to communicate that feeling to someone who does not play trumpet (or someone who does). In this way, all attempts to measure mental models are limited as they cannot simultaneously capture all ways of knowing. Thus, LSA is limited to capture only those aspects of mental model which are expressed through language. Another important distinction is that semantic similarity is not equivalent to similarity as described in most mental models literature. While semantic similarity doesn't necessarily imply conceptual agreement, similarity of mental models does. However, semantic similarity may be indicative of shared mental models regardless of the difference in their meanings. Regardless, care should be taken when comparing similarity of mental models assessed through other measures to those from LSA. Given the complexity of the LSA algorithm and its contrast to the other mental model measurement approaches, researchers may have a hard time integrating LSA into their practices, which could make a high-quality systematic comparison between LSA findings and other mental models measurement approaches more difficult. Additionally, the sensitivity of LSA in the context of mental models is not well established in the literature. The number of semantic dimensions, k , is likely important, but the stability of conclusions drawn about mental models in light of changing parameters such as k has not been well documented. As of now, mental model structure is not well represented in studies using LSA. However, this is an opportunity for future research as the approximation of text-based utterances as vectors (i.e. position vectors in some semantic space) potentially enables the use of existing mathematical approaches - see Dong & Sarkar for an interesting use of SVD for a similar purpose (Dong & Sarkar, 2014).

Examples of LSA Usage

LSA has been used widely in a variety of different applications including information retrieval and filtering (Deerwester et al., 1990; Dumais, 2004), education (Landauer & Psotka, 2000), modeling human memory (Dumais, 2004; Landauer et al., 1998). In engineering design specifically, LSA has been used in a number of ways. Elena & Summers used LSA to assess the novelty of requirements generated by students with and without a lecture intervention (2019). Fu *et. al* measured the impact of introducing different example solutions on team convergence by using LSA to determine semantic similarity of written responses to an open-ended prompt. Participants answered the prompt four times over the course of a short design session, allowing for change over time to be measured (2010). Wood *et. al* used LSA to measure the similarity between student self-reports at several points during a collaborative engineering design problem solving session, and used these findings to conclude that interactions between team members as short as seven minutes lead to more similar mental models (2014). Goucher-Lambert *et. al* used LSA to identify pre-existing design concepts which are conceptually near or far from the designer's current design state, then presented these pre-existing concepts as inspirational stimuli (2020).

CHAPTER 3: METHODS

Conceptual Framework

While similar task-related and team-related mental models have been shown to improve team performance in operational team contexts, design teams cannot simply aim to increase mental model similarity or sharedness and expect better outcomes. Often in design projects, differing mental models are needed, for example during ideation or when team members are disciplinary experts in a specific content area. However, at other times, sharedness is critical, for example when eliciting stakeholder requirements, or when discussing interfaces between sub-assemblies. The degree of sharedness which is most beneficial changes over time and for different tasks within a design project. There are also unanswered questions as to what is shared. While the categories of team-related and task-related mental models work well for operational teams, they present challenges when applied to design directly. Researchers have long advocated for discipline specific frameworks for mental models, and design is no exception.

Team-Task-Design Mental Model Framework

I propose a separate category of “design-related” mental models be included when considering design teams. The resulting Team-Task-Design Mental Model Framework is shown in Figure 3.1. This framework would include an “Object Model” representing the understanding of the object of design, and inspired by Bucciarelli’s, “Object worlds” (1994), as well as a “Need Model” representing the understanding of the unmet need. Example contents of the team-related, task-related, and design-related mental models are outlined in Table 3.1.

Object and Need Models develop together as the team’s understanding of the Need expands and becomes clearer, and they respond to this new understanding of the need through

iterative development of the design object. At the end of the project, if object satisfies the need the Object and Need models as represented in Figure 3.1 may eventually overlap. This design development may be moderated by some established process, represented in a design-related “Process Model” as suggested by Badke-Schaub *et al.* (2007). Though, local process structure represented in the task-related model may be sufficient in many cases (Minneman, 1991). Distinguishing design-related models from task recontextualizes the task-related model to apply at a different timescale. Previously the assumption was that the task was “to design ____”, but this recontextualization instead focuses on subtasks within the activity of design which are guided by overarching, design-related considerations that span many tasks.

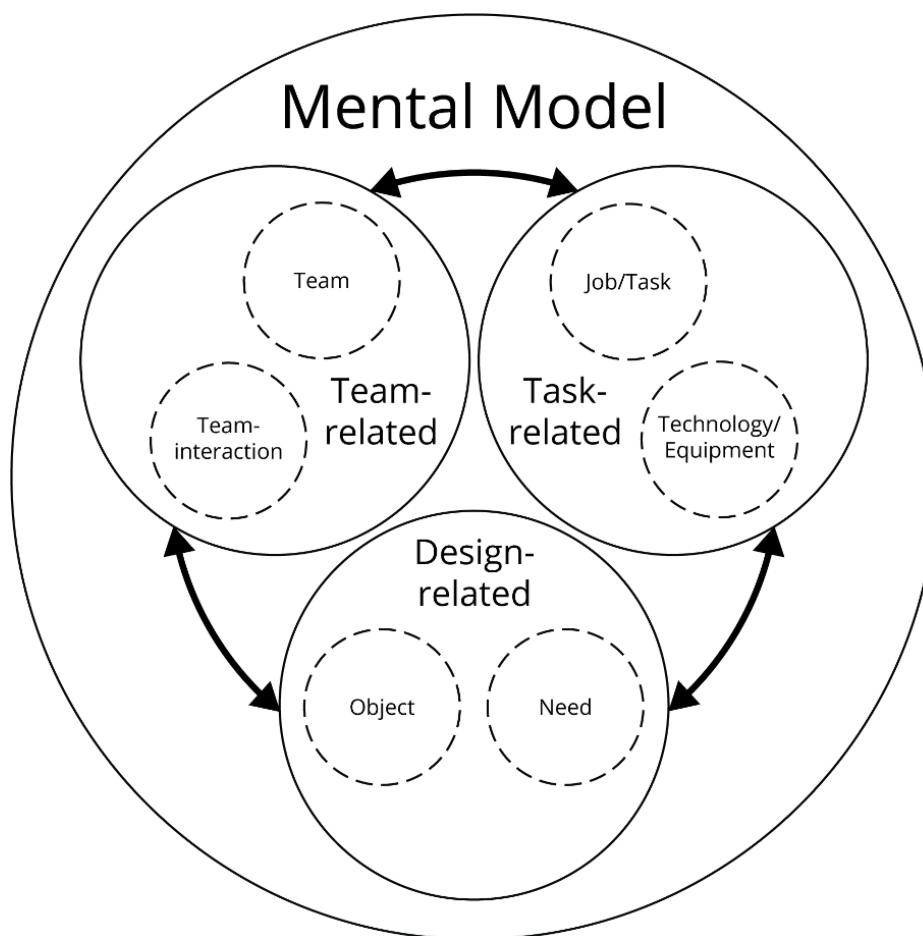


Figure 3.1 – Team-Task-Design Mental Model Framework

Table 3.1 - Mental Model Contents by Type		
Broader Model Category	Type of Model	Content
Design-related	Object	Functions (Gero & Kannengiesser, 2004)
		Behaviors
		Structure
		System Requirements
	Need	Stakeholder groups
		Stakeholder priorities
		Opportunity areas
		Stakeholder Requirements
		Unmet need
Task-related*	Technology/equipment	Equipment functions
		Operating Procedures
		System limitations
		Likely failures
	Job/Task	Task procedures
		Likely contingencies
		Likely scenarios
		Task strategies
		Environmental constraints
		Task component relationships
Team-related*	Team	Teammates' knowledge
		Teammates' skills
		Teammates' attitudes
		Teammates' preferences
		Teammates' tendencies
	Team interaction	Roles/responsibilities
		Information sources
		Interaction patterns
		Communication Channels
		Role interdependencies
Information flow		
*All entries under these headings are from prior work, see (Klimoski & Mohammed, 1994; Mathieu et al., 2000; Rouse et al., 1992)		

The main motivation for this shift is that there are clear tasks within engineering design that neatly fall under the existing task-related model paradigm, for example, a team tasked to create a CAD model of a motor mount as part of their design process. The potential contents of the mental models for this task can be considered using the examples outlined in Table 3.1 with

team-related and task-related entries from prior works cited in the table itself. The task-related model (see Figure 3.2) for the task of creating a CAD model of a motor mount would include an equipment/technology model covering the CAD software, its workflow, shortcuts, limitations, etc., as well as a task procedure outlining the steps of creating such a model, and the environmental constraints on the task such as deadlines, dependencies, or file compatibility requirements. Additionally, to design the motor mount the team must understand what motor is being used, how it is oriented, the interface between the motor mount and the rest of the assembly, etc. These considerations would exist within the Object model. Requirements that constrain the design may be part of the Need model, though the distinction between Object and Need is expected to become less clear as design progresses. Task-related considerations are specific to the task at hand, but design-related considerations are broader reaching, potentially connected to many tasks throughout design. This partially addresses the first issue mentioned previously, that design is ill-structured and emergent, not procedural, and therefore the components of the task-model cannot be easily predicted *a priori*. By instead considering shorter tasks individually within the broader context of design, we can treat different tasks differently. While design itself may be unpredictable, subtasks such as creating a CAD model may more closely resemble the procedural tasks faced by operational or procedural teams.

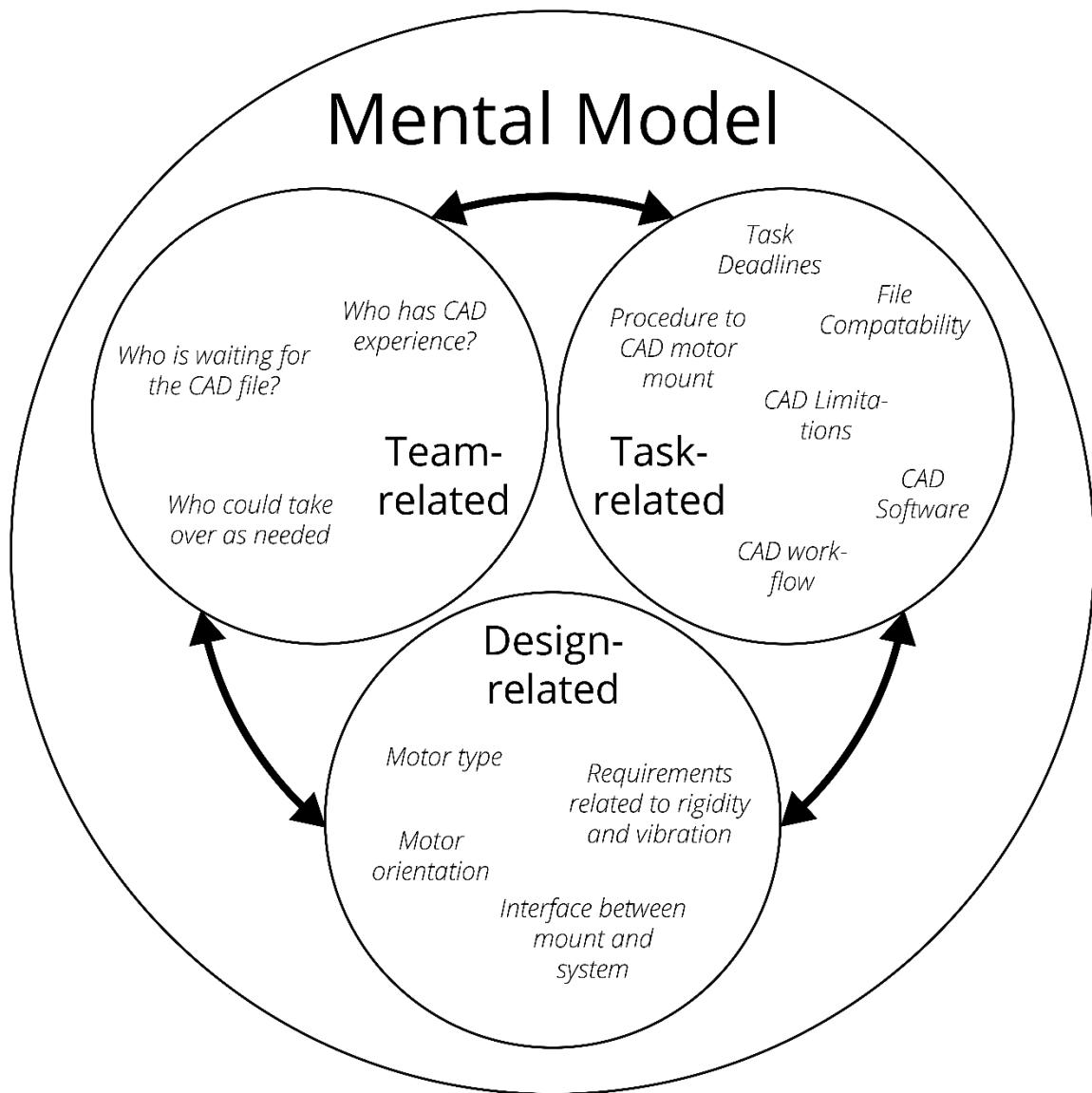


Figure 3.2 - Team-Task-Design Framework for Example CAD Task

The second issue, that mental model accuracy is difficult to define in design contexts, is also partially addressed by this recontextualization. Design includes both convergent and divergent processes, and while there would still be divergent design tasks which have many valid approaches and solutions, for many tasks, including things like creating a CAD model, there are reasonably identifiable “expert models.” By considering sub tasks individually, accuracy can be

defined separately for different tasks. This would allow for the assessment of accuracy for at least some task-related models in design contexts. For design-related models of Object and Need, accuracy is fairly simple to define. The accuracy of the Object model would be related to how closely it matches reality – how representative the model is of the real system being designed. For concepts within the Object model which are not yet realized, accuracy would be related to their technical accuracy according to the relevant scientific principles as well as their feasibility. While an inaccurate Task model might be a misunderstanding of the procedure or tools, an inaccurate Object model might be a misunderstanding of the system or an infeasible proposal. The accuracy of the Need model could be related to the similarity to that of the stakeholder experiencing the need. Since there may be several or many needs experienced by multiple stakeholders, the Need model would also include the relationships between all of these, including conflicts, priorities, etc. However, stakeholders may have incorrect understandings of their own needs, so it may be best to consider a theoretical ideal rather than relying on finding a perfect stakeholder. Perhaps this ideal could be the need as it is known when it is satisfied. Generally, distinguishing design-related mental models from task-related mental models may simplify the measurement of mental model accuracy in design teams.

Third, diversity is beneficial to design which contradicts the finding from past mental models literature that teams with more similar mental models are able to perform better. While there may be some aspects of design projects where sharedness is beneficial, there are certainly others where too much sharedness can lead to groupthink, resulting in poor exploration of the design space and potential blind spots. Therefore, some degree of difference between mental models is beneficial in at least some aspects of design. One way to identify areas where more

sharedness is most beneficial is to consider the *centrality* and *importance* of the model (Boos, 2007; Kameda, Ohtsubo, & Takezawa, 1997; Rouse et al., 1992). Centrality is the measure of how interrelated some element of the mental model is. In a network analogy, more central elements would have more connections. It follows logically that the necessity of similarity is dependent on centrality. The less central, the less crucial it is that the model is shared. Importance is a related metric but has its own distinctions. Even if a consideration is highly central, if it has low importance, then sharedness may be less crucial. Another way to consider these together may be impact or influence, with more impactful or influential considerations requiring more sharedness. Though this may raise additional questions about the breadth or depth of impact. Generally speaking, it should be expected that sharedness is more critical for core considerations with broad impact, as well as at interfaces and boundaries. Interfaces may be between incongruent mental models, for example as the team is meeting with stakeholders to elicit requirements, it is critical that the team and the stakeholder are able to cross the communication barrier to establish a shared understanding of the need. Interfaces may also be between assemblies or subassemblies within the Object model, for example, the team member creating the CAD model of the motor mount must have a shared understanding of the bolt pattern with the designer responsible for creating the matching component. This brings us to an interesting point about collaboration on engineering design teams. Teams tend to allocate subtasks to individuals or sub teams who specialize in different aspects of the design. This is most common in the design of complex systems incorporating multiple engineering disciplines, such as electromechanical devices. However, examples of real-world projects which do not require this kind of diverse expertise and specialization are increasingly rare. The consequence of

this is that often engineering designers, even those working on collaborative teams, are working independently at least some of the time. Their separate contributions must then be integrated into a cohesive whole. Therefore, while the considerations within one's Object world, as Bucciarelli describes them, are not required to be shared, sharedness is crucial at the interfaces or boundaries between Object worlds. In general, this would mean that task-related mental models might require relatively low sharedness due to the allocation of tasks to individuals or sub teams. Whole-team tasks such as group ideation sessions may require more sharedness, e.g. a shared understanding of the rules, procedures and protocols mediating the session. Sharedness in Object and Need models related to the whole system might be expected to be more critical as these whole-system considerations are highly central, however those related to individual components may be less shared. Following a typical Systems Engineering vee-model, this would mean that Object and Need models tend to be more central at the beginning of the design process, then as considerations get more specific, centrality decreases. Sharedness of team-related mental models is less predictable and may vary depending on the team and problem.

Study Design

Design is a central aspect of engineering practice which is often done in collaborative design teams where individual members understand the problem from different "object worlds" (Bucciarelli, 1994; Dym, Agogino, Eris, Frey, & Leifer, 2005). While prior team mental models research offers many useful insights, some findings and frameworks do not transfer directly to design teams due to the emergent, ill-structured nature of design. The Team-Task-Design Mental Model Framework may help to address some of these concerns by distinguishing design-related considerations which span multiple tasks as part of a separate model, distinct from, but

interacting with, the team-related and task-related mental models. As yet, more exploratory, and empirical work is needed to substantiate and refine the concept of design-related mental models. Nonetheless, it is assumed that teams benefit from some degree of shared understanding, even in design contexts. However, the content that is shared and degree of sharedness may vary over time and from task to task. Latent Semantic Analysis offers a promising way to measure aspects of mental models expressed in text, either through transcriptions or in written responses and project documents. Written reflections provide a useful venue for data collection as they are not tied to specific project milestones like some documentation such as project reports. Additionally, reflection prompts can be easily repeated to measure change over time and can be extended to cover projects of realistic lengths (i.e. 15-weeks) in ways that transcriptions often cannot. Teams may also directly benefit from reflections (Adams, Turns, & Atman, 2003; Turns, Sattler, Yasuhara, Borgford-Parnell, & Atman, 2014).

Research Questions and Study Design

To this end, this work seeks to answer the following research questions:

RQ1) In what ways can Latent Semantic Analysis be used to understand the shared content of open-ended reflections? What are the limitations of such analysis?

RQ2) What are the typical elements of a collaborative design team's shared understanding of a design problem?

To explore these questions, I developed a multi-case study (Yin, 2017) to investigate design teams in a Mechanical Engineering capstone design course at a large Midwestern University. The general structure of this study as it relates to the research questions is outlined in

Figure 3.3. Design teams working through their capstone projects as normal submitted individual responses to reflection prompts, with one prompt repeated at the beginning, middle and end of

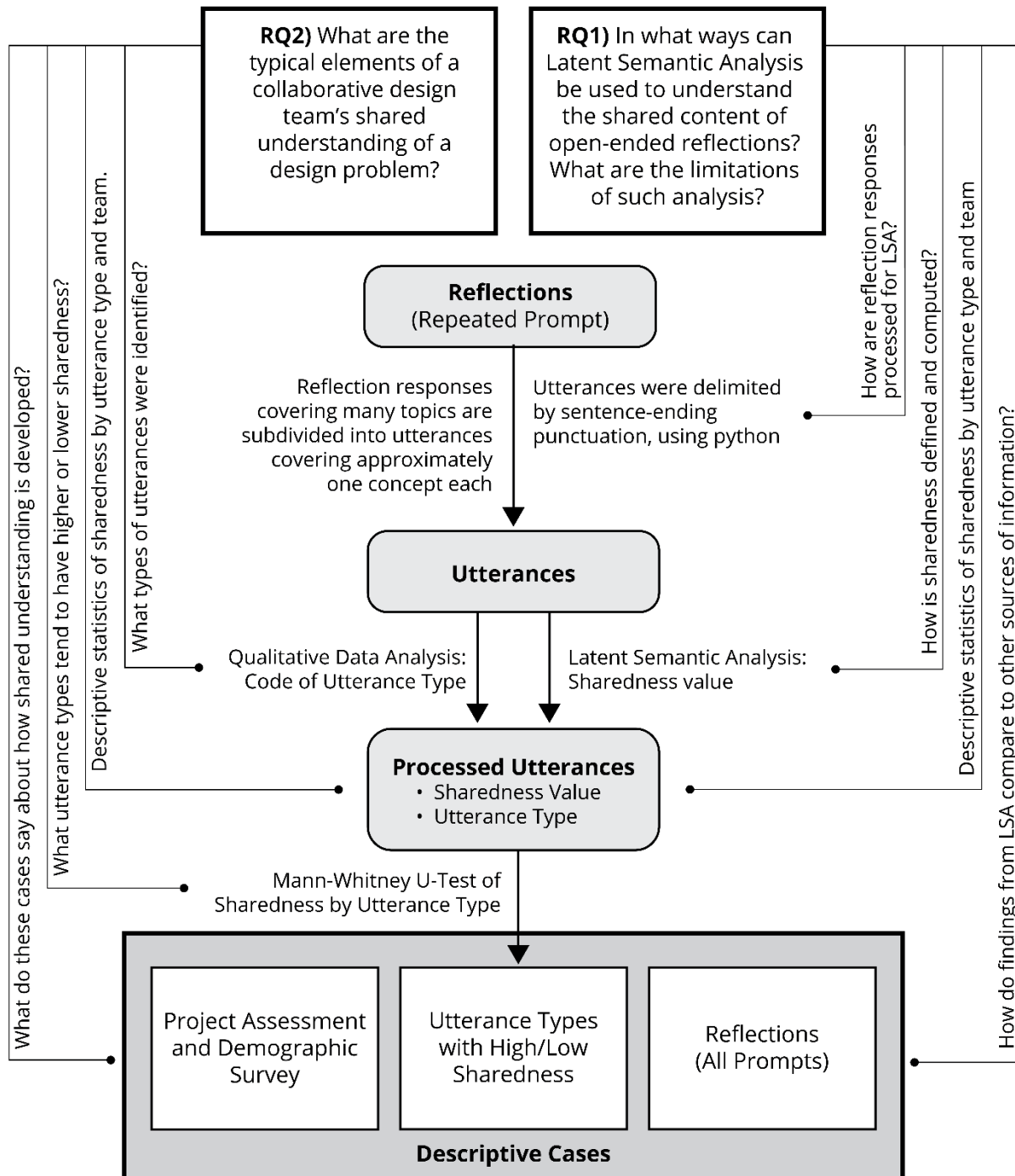


Figure 3.3 – Study Design

the semester (see Figure 3.4 for a timeline). This study explores these reflections, as well as responses to a project assessment and demographic survey. Reflection responses were divided into utterances, each roughly a sentence long, delimited by sentence ending punctuation. These utterances were then analyzed using LSA to compute their similarity to the group (aka *sharedness*) on a scale of -1 to 1. Utterances in response to the repeated prompt were also coded using an emergent coding scheme describing the type of content in each utterance. Then the Mann-Whitney U-Test was used to determine which utterance types tend to have high or low sharedness. Three teams were explored closer, pulling together sharedness findings from the repeated prompt, but also responses from the other reflection prompts as well as the survey to build descriptive cases. These descriptive cases explore how shared understanding is developed and the impact of project context as it relates to sharedness results from LSA.

Course Context

In the capstone course, teams of five to six students work together on design projects sponsored by industry, faculty, or others, including national labs. The course had a total enrollment of 175 students and there were 32 project teams. Note that large teams were required due to a limited number of available projects and abnormally high enrollment in the target semester. Participation in the study was optional, and all students completed the same work regardless of their participation in the research study. UIUC Institutional Review Board (IRB) approved this human subject research (#22770). Of the 175 students enrolled, 117 consented to participate in the study. The demographics of study participants are shown in Table 3.2.

The semester-long project was roughly divided into three phases with deliverables marking the completion of each phase. A timeline of these deliverables and project phases is

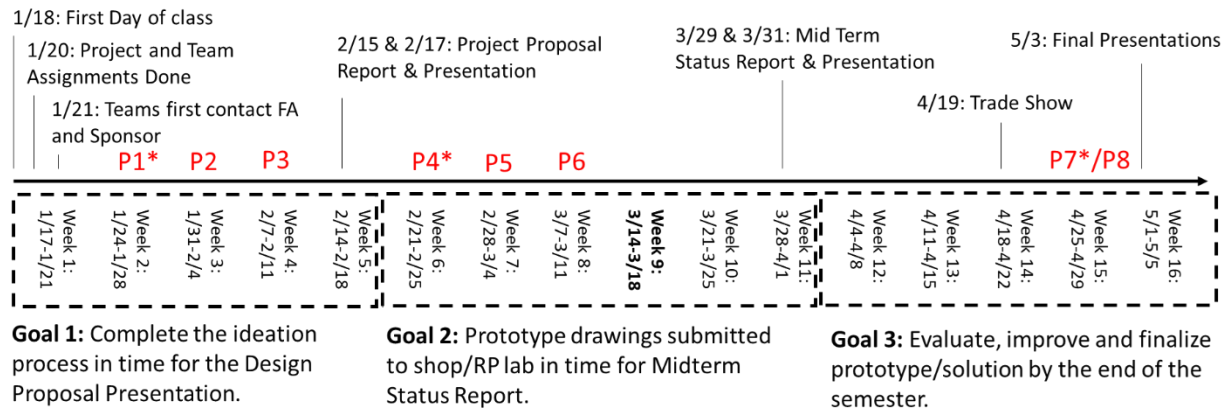


Figure 3.4 - Timeline of course activities and reflection prompts

shown in Figure 3.4. Goal 1 was to complete the ideation process in time for the Design Proposal Presentation. Goal 2 was to have prototype drawings submitted to shop/rapid prototyping (RP) lab in time for Midterm Status Report. Goal 3 was to evaluate, improve and finalize prototype/solution by the end of the semester. Final projects were demonstrated at a Trade Show event as well as in Final Presentations. In addition to weekly communication with the project sponsor, teams were also expected to meet regularly with a Faculty Advisor with relevant content expertise as well as a graduate TA assigned to their project. All enrolled students were asked to complete responses to reflection prompts throughout the semester.

Data Collection

Survey

A demographic and project assessment survey was used to collect information from participants about how the project went. This survey was sent to participants in week 14 of the semester. Survey questions can be seen in Table 3.3. Questions 1-9 used a five-point Likert response scale of Strongly Agree to Strongly Disagree. Questions 10-13 asked students to respond on a scale of 1-10, (with 1 low). Questions 14 and 15 asked students about their English

language usage as this analysis relies heavily on written English semantics. An independent t-test was used to determine if sharedness was dependent on possible dialectical variations in English speakers indicated by question 15. There was no significant difference, $t(2854) = 1.3$, $p = 0.2$, between the sharedness values of utterances from those who learned English in the United States ($n = 2312$, $mean = 0.187$, $std. dev. = 0.144$) compared to other countries ($n = 544$, $mean = 0.179$, $std. dev. = 0.142$). The final questions, 16-19 asked about demographics including grade expectations, social comfort, and gender and racial identity. The distributions of gender and racial identities in the course, as well as those for the teams selected for further analysis are shown in Table 3.2.

Table 3.2 - Basic demographic data for overall course total as well as teams selected for this study.		
Gender Identity	Course Total	Selected Teams
Female	29 (17%)	10 (29%)
Male	61 (35%)	13 (37%)
None of these/Other	0 (0%)	0 (0%)
Did not respond	85 (49%)	12 (34%)
Racial Identity	Course Total	Selected Teams
Asian or Pacific Islander	38 (22%)	5 (14%)
Black or African American	1 (1%)	1 (3%)
Hispanic or Latino	4 (2%)	1 (3%)
Native American of Alaskan Native	0 (0%)	0 (0%)
White or Caucasian	45 (26%)	15 (43%)
Multiracial or Biracial	0 (0%)	0 (0%)
A race/ethnicity not listed here	1 (1%)	0 (0%)
Prefer not to say	1 (1%)	1 (3%)
Did not respond	85 (49%)	12 (34%)

Table 3.3 - Demographic and Project Assessment Survey Questions	
Num	Question
Q1	The courses I've taken have prepared me for this project
Q2	This project required me to gain new knowledge, skills, and/or abilities
Q3	The sponsor provided my team with relevant background information
Q4	Project goals were stable throughout the semester
Q5	Sponsor expectations were clear at the start of the semester
Q6	This project had a clear starting point
Q7	This project required clarification before we were able to move forward
Q8	My understanding of the project has changed since the start of the semester
Q9	My team has a shared understanding of the project
Q10	On a scale of 1-10, evaluate the quality of communication with your sponsor
Q11	On a scale of 1-10, evaluate the quality of communication among your team
Q12	On a scale of 1-10, evaluate the quality of collaboration among your team
Q13	On a scale of 1-10, how familiar were you with your team members before the project
Q14	What language do you use most often with friends and/or family?
Q15	Where did you learn English (State/Province, Country)?
Q16	What grade do you expect to receive in [Design Course] (optional)
Q17	How comfortable are you in social situations? (optional)
Q18	Gender: Which of the following best describes you? (optional)
Q19	Race/Ethnicity: Which of the following best describes you? (optional)

Reflection Prompts

Students were asked to respond to reflection prompts throughout the semester as a normal course assignment. There were eight prompts, labeled P1-P8, with one prompt repeated three times during the semester as P1*, P4* and P7* (the asterisk indicates that this is the repeated prompt). These reflection prompts can be seen in Table 3.4. The sequence of prompts was intended to roughly follow the team's progression through the design project, so that relevant questions were asked at an appropriate time. There were several weeks where no reflection was required as the students had other deliverables due. While the prompt instructions suggested a minimum of 150 words to elicit complete responses, this limit was not enforced.

Table 3.4 - Reflection prompts listed by prompt ID. Note that prompts with an * are repeated.

ID	Prompt
P1*	(Minimum 150 words) Describe your current understanding of your design problem. Be sure to include answers to the following questions: What are the goals of the project? What challenges do you anticipate? How comfortable are you with this project? What questions and/or uncertainties do you have? If there is anything else you'd like to mention, please do.
P2	(Minimum 150 words) Consider any meetings you have had with your sponsor or faculty advisor. What were your expectations for those meetings? How did your experience compare to your expectations? Did they go well? Why or why not? What questions and/or uncertainties are you left with at this point? What, if anything, would you do differently next time? (If you have not yet met with your sponsor or faculty advisor, please explain the circumstances and how this may impact your work) If there is anything else you'd like to mention, please do.
P3	(Minimum 150 words) Consider what is decided and undecided at this point. Are there constraints or requirements for your design that cannot be changed? If so, what are these and why can't they be changed? How has your team approached making design decisions? What sources of information are used most when making decisions? Is there anything different about the things your team is decided on versus undecided on? Have you approached these things differently? Are you more confident with one or the other? If there is anything else you'd like to mention, please do.
P4*	<i>Repetition of P1 (See above)</i>
P5	(Minimum 150 words) What feedback have you gotten to this point? How has this feedback affected your understanding of the problem? How has this feedback affected your design? Are responding to all feedback equivalently? Why or why not? If there is anything else you'd like to mention, please do.
P6	(Minimum 150 words) Consider any diagrams, models, simulations, components and/or full prototypes you have created. What purpose did these serve in your design process? What aspects of your design did they explore? Were the methods that you used suitable? Why or why not? If there is anything else you'd like to mention, please do.
P7*	<i>Repetition of P1 (See above)</i>
P8	Briefly describe your overall experience with this project. How satisfied are you with this learning experience? How relevant is this experience to your future plans?

Team Selection and Overview

Eight teams of 32 were included in this analysis, GR, KL, LU, WZ, BK, ZH, CP and EK. Teams were selected based on the number of team members who submitted reflection responses at each instance of the repeated prompt, P1*, P4*, P7*. At least three responses to each instance of the prompt were required, otherwise the majority of team members would not be represented for that instance and comparison to the group average would have limited utility. Thus, only those teams with sufficient responses to these prompts were included in the analysis. However, utterances from all teams were included in the LSA corpus. Table 3.5 shows the project name corresponding to each Team ID and provides a brief description of project details. As many projects were industry sponsored, some details have been redacted to ensure confidentiality. Table 3.6 gives an overview of the members of each team with demographic details if provided by participants. Pseudonyms have been used and project names changed to protect the anonymity of the participants.

Table 3.5 - Project Names and Descriptions		
Team ID	Project Name	Project Details
GR	MEMS Project	This project is industry sponsored. The sponsor is located in another state. The project is centered around building a system to be used in the manufacture of silicon devices. The final deliverable is a full system design.
KL	Thermo Project 1	This project was sponsored by a university-affiliated group and was centered around the technical review of alternative energy sources. The final deliverable of this project was a feasibility study.
LU	Thermo Project 2	This project was sponsored by a university-affiliated group. In this project, the team was tasked with material characterization for applications associated with thermal energy. The final deliverable of this project was a technical report. This project recommended pre-requisite coursework.

Table 3.5 (cont.) - Project Names and Descriptions		
Team ID	Project Name	Project Details
WZ	Mass Customization Project	The project was industry sponsored. This project was centered around the use of 3D scanners and rapid prototyping for mass customization. The final deliverable for this project was a feasibility study, as well as proof of concept prototypes.
ZH	Vibration Damping	This project was industry sponsored. The sponsor was located in another state. The goal of this project is to find alternatives for an existing vibration damping system in a specific application. The final deliverable for this project was a technical report, and prototype.
BK	Instructional Project	This project was sponsored by a university-affiliated group. The team was tasked with developing instructional laboratory equipment and protocols for their use. The final deliverable for this project included a final bill of materials, as well as technical reports and proof-of-concept prototypes.
CP	Vehicle Control	This project was industry sponsored. At the start of the semester, a full project description was not available. Students were provided with details about the sponsor and a brief description of the project, which was centered on the use of a detector in a vehicle control system. Additional information was provided in the first few weeks of the project.
EK	Opto-mechanical	This project was sponsored by a national lab as part of an established capstone program. This project was a year-long project such that the students in Spring semester were continuing the work started by those in the Fall. The final deliverable for this project was a technical report and working prototype where applicable. The purpose of this project was to explore optomechanical systems for use in space.

Table 3.6 - Team Members by Project					
Team ID	Project Name	Team size (# in study)	Pseudonym	Demographic	# of Utts.
GR	MEMS Project	6 (6)	SP22_24707	Hispanic or Latino, Male	71
			SP22_01273	Black or African American, Male	37
			SP22_42298	Asian or Pacific Islander, Male	38
			SP22_43707	n/a	59
			SP22_66719	n/a	39
			SP22_95852	n/a	10
KL	Thermo Project 1	6 (4)	SP22_06499	White or Caucasian, Female	91
			SP22_43908	White or Caucasian, Female	49
			SP22_30927	White or Caucasian, Female	67
			SP22_93465	Asian or Pacific Islander, Male	48
LU	Thermo Project 2	5 (4)	SP22_75918	White or Caucasian, Male	77
			SP22_95447	White or Caucasian, Male	53
			SP22_26751	White or Caucasian, Male	57
			SP22_70689	Asian or Pacific Islander, Female	83
WZ	Mass Customization Project	6 (3)	SP22_48390	White or Caucasian, Male	68
			SP22_22009	n/a	23
			SP22_68263	n/a	81
ZH	Vibration Damping	5 (3)	SP22_81321	White or Caucasian, Male	42
			SP22_13278	White or Caucasian, Male	16
			SP22_74195	Asian or Pacific Islander Female	56
BK	Instructional Project	6 (4)	SP22_93282	Prefer not to say, Male	100
			SP22_80116	White or Caucasian, Female	36
			SP22_39708	n/a	35
			SP22_13510	n/a	46
CP	Vehicle Control	6 (5)	SP22_91567	White or Caucasian, Male	49
			SP22_72514	White or Caucasian, Male	59
			SP22_41617	n/a	31
			SP22_50384	n/a	72
			SP22_37767	n/a	35
EK	Opto-mechanical	6 (6)	SP22_98921	White or Caucasian, Female	53
			SP22_27826	White or Caucasian, Female	39
			SP22_04626	White or Caucasian, Female	47
			SP22_03268	Asian or Pacific Islander, Female	51
			SP22_65144	n/a	25
			SP22_18040	n/a	31

Data Processing and Analysis

Data processing and analysis was done using Python 3.10. Custom objects were used for data organization storage such that each participant in the study had a corresponding “folder” tagged with a number of attributes, including a procedurally generated pseudonym, as well as a team ID. Reflection responses and survey responses were imported and filed into their corresponding participant “folder.” These responses were themselves custom python objects, with their own attributes such as lists of utterances and their sharedness values. This custom python approach facilitated investigations student-by-student, team-by-team, or utterance-by-utterance while maintaining the connections between, for example, each utterance, its author, and the prompt it was in response to, and without requiring researchers to manage multiple spreadsheets. A requisite step of this process was subdivision of each reflection response into utterances. In this case, each utterance was roughly a sentence long and was delineated using a python script which subdivided reflection responses by sentence-ending punctuation (e.g., “.”, “?”, “!”). Note that special considerations may be necessary, for example for instances of periods in abbreviations or for different formatting such as bulleted lists. This method was chosen as a compromise between ease of implementation and the level of fidelity and accuracy attained. Ideally, each utterance would contain a single distinct semantic concept, though this is perhaps impractical. The total number of utterances from all teams was 5162 (126524 words), with 1774 utterances (38935 words) from the eight teams selected for analysis.

Utterance Coding Scheme

Utterances were coded using an emergent qualitative coding scheme representing their contents. The utterance codebook, with definitions for each code is shown in Table 3.7. The

process used for qualitative data analysis is as follows: First, all repeated prompt reflection responses were reviewed to identify early themes. Since the repeated prompt specifically asks for goals (project descriptions), questions, and challenges, codes covering these topics were proposed early. I then randomized utterances from Teams EK, GR, and CP in a spreadsheet and re-read through each utterance, applying, and developing relevant codes at the utterance level. In a preliminary check of interrater-reliability, 118 of the 647 utterances considered (18%) were coded by an additional PhD-credentialed researcher. The two coders agreed on 89.83% of utterances coded. Discrepancies between coders were not sufficient to alter any results reported as statistically significant, that is regardless of which researcher did the qualitative coding, the reported results did not change. With this confirmation, the preliminary coding scheme was used to code the remaining utterances.

Table 3.7 – Codebook for Utterance Type	
Code of Utterance Type	Description
Question	Any question asked. These are mainly clarifying questions about the project. They may be specific, or general
Expectation	Anticipated challenges, expected outcomes, speculations about what might occur
Process	Descriptions of what has been done, is currently being done, or what will be done to proceed through the design process
Project Description	Overview of the project, may include project goals or objectives
Perspective	The individual perspective, how they feel about the project, their team, their advisor, or TA
Obstacle	Obstacles which may impede progress. These are high-level challenges (e.g. advisor and sponsor disagree on scope, scheduling issues, supply chain issues), and should be concrete (potential or anticipated issues may be better coded as Expectations)
Design Consideration	A project-specific detail, should be content related (not process related)
Other	Utterance does not fit in the above categories or has unidentifiable meaning. Includes sentence fragments due to poor parsing of utterances, unless that fragment can be easily attributed to another code

Latent Semantic Analysis

The Latent Semantic Analysis procedure outlined in Chapter 2 was applied to a corpus constructed from all utterances (5162 utterances, 126524 words). Once again this analysis was completed in Python 3.10 using an implementation in the *scikit-learn* library (Pedregosa et al., 2011). A code segment illustrating the steps in this python implementation is shown in Figure 3.5. The first three lines import the required modules from the library. Line 5 creates a vectorizer object used to parse the corpus. This vectorizer has parameters for the n-gram range and stop words which can be used to fine tune the behavior of the LSA algorithm. For example, the n-gram “steering wheel” contains two words, though it refers to a singular artifact. To capture both the thing referred to by the n-gram “steering wheel” as well as the concepts of “steering” and “wheel” individually, then *ngram_range* = (1,2). This default range of (1,2) was used for the purpose of this study, but an analysis of the sensitivity to this parameter is planned for future work. Additionally, a list of stop words can be used to remove common words from the analysis. Words such as, “the”, “it”, “her”, “and”, “as”, etc. do not contribute significantly to the latent semantic meaning as assessed through LSA because of their ubiquity (Deerwester et al., 1990;

```
1 from sklearn.feature_extraction import text
2 from sklearn.feature_extraction.text import TfidfVectorizer
3 from sklearn.decomposition import TruncatedSVD
4
5 vectorizer = TfidfVectorizer(ngram_range=(1,2), stop_words='english')
6 lsa = TruncatedSVD(n_components=k)
7 vectorizer.fit(corpus)
8 X = vectorizer.transform(corpus)
9 X_kapprox = lsa.fit_transform(X)
```

Figure 3.5 - LSA Python Code Segment

Dong et al., 2004). These words can therefore be removed from consideration, so they do not contribute additional noise. In this analysis, I used the English language stop list included in the sci-kit learn implementation, the potential limitations of this are discussed later on. In line 6 of the code segment in Figure 3.5, The number of dimensions in the k -reduced semantic space computed through SVD is defined using the keyword *n_components*. The dimensionality of the semantic space determines the number of semantic concepts which are represented in the vector approximations of utterances. Values on the order of 10^2 are common, and while the choice of k value is important, a relatively wide window of usable values has been shown (Dong, 2005; Dumais, 2004; Landauer et al., 1998). A value of $k=200$ dimensions has been used in this study, and an analysis of the sensitivity to this parameter is planned for future work. In line 7, the vectorizer is used to tokenize the corpus, which is then transformed into the correct formatting for SVD in line 8. The SVD algorithm is applied in line 9, and the k -reduced matrix, $\tilde{\mathbf{X}}$, is returned as output and saved as *X_kapprox*. Each row-vector of this matrix corresponds to an utterance approximated in the reduced semantic space. Using these semantic-space utterance vectors, utterances can be compared pairwise to determine their semantic similarity using the cosine angle between vectors. Composite vectors can also be created as the vector sum of several utterances, for example, all utterances from teammates. In this way sharedness is computed for each utterance as the similarity (i.e. cosine angle) between that utterance and the composite vector representing all utterances from the speaker’s teammates in response to the same instance of the reflection prompt.

Statistical Analysis

In order to test if different types of utterances tended to have high or low sharedness, the Mann-Whitney U-Test (MWU) was used to compare the sharedness value distributions belonging to utterances with different types. This is similar to a one-sample T-test, used to determine if a subgroup falls significantly higher or lower on some metric than the group as a whole. For example, given a random sampling of people with known height and hair color, are redheads taller or shorter, on average? Unlike the Student T-test, MWU can be used to compare distributions of unequal sizes where the assumption that each follows a normal distribution cannot be assured. MWU tests the null hypothesis that a random sample drawn from each of two distributions, A and B, will have equal probability that $A > B$ and $B > A$ (Mann & Whitney, 1947). In other words, MWU can be used to compare distributions to see which is stochastically greater without use of the mean. For each team, the distribution of sharedness values for each utterance type code was compared to the overall distribution using the MWU test. This was done in both directions to check if the sharedness values of utterances of a specific utterance type were typically greater or less than those of the overall distribution.

Interpreting Sharedness Values

Pairwise Similarity between Utterances

As previously mentioned, LSA approximates utterances as vectors in the k —reduced semantic space. These vectors can be compared for similarity by the cosine angle between them, thus producing a measure of semantic similarity for the utterances themselves. Using this approach, semantic similarity ranges from -1 to 1. Utterances with semantic similarity of 1 are identical, and those with semantic similarity of -1 are semantic opposites. A semantic similarity

of 0 represents unrelated utterances. In LSA, the semantic meaning of words is defined by their

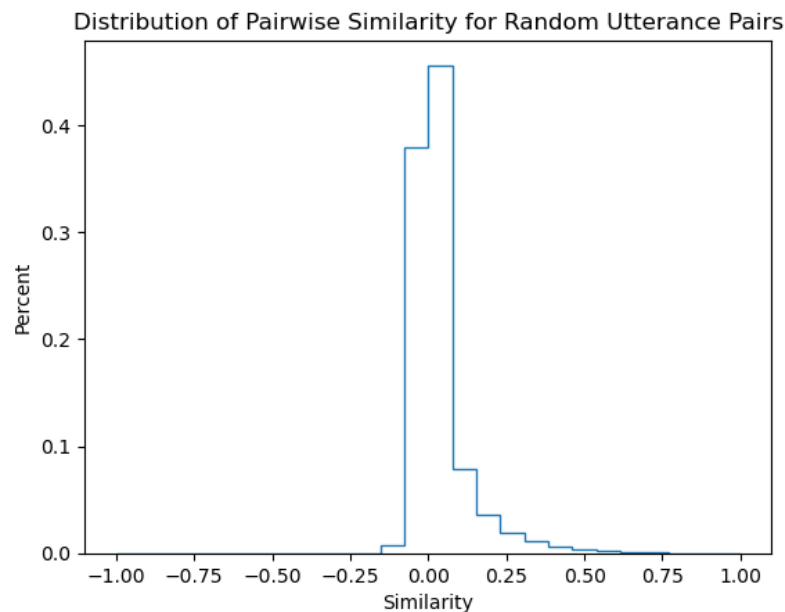


Figure 3.6 - Distribution of pairwise similarities for random utterance pairs

context of use within the corpus, and thus it is key that the corpus is sufficiently large to adequately represent word meanings. Terms which appear in utterances together are associated with one another. For example, a set of utterances about the weather will contain many terms related to the weather such as, hot, cold, rain, sunny, stormy, bright, dark, etc. When two utterances contain words which are associated in this way, their semantic similarity will be high. Because of this it is important to draw a distinction between similarity and agreement. The sentences, “today is a bright, sunny day” and “today is not a bright, sunny day” may have opposite meanings but they will likely be considered semantically similar according to LSA since they contain many of the same words. Likewise, the sentence, “today is dark and stormy” might also be considered semantically similar since the concept being discussed in all cases is the weather and the terms, “dark”, “stormy”, “sunny”, and “bright” are likely to appear in utterances discussing the weather, depending on the context and completeness of the corpus.

Thus, it will be important to establish a sense for the range of similarity values associated with the corpus used in this study. From 664 utterances, I generated 220116 utterance pairs, the distribution of their similarity values is shown in Figure 3.6. The mean pairwise similarity was 0.034, with a standard deviation of 0.090. The median was 0.009. Note that for most utterance pairs, their semantic similarity was very low, near zero. Thus, most utterance pairs are not semantically related, they cover different, unrelated topics. However, there is a portion of utterance pairs with higher semantic similarity as shown by the tail on the right side of the distribution in Figure 3.6. These utterances may belong to members of the same team, or speakers who faced similar situations during the design process. Examples of utterance pairs with varying levels of semantic similarity are shown in Table 3.8 and discussed below.

Table 3.8 - Examples of pairwise similarity between random utterances			
Ex.	Pairwise Similarity	Utterance 1	Utterance 2
A	0.998	<i>As of now, I am about 50% comfortable with this project</i>	<i>I am very comfortable with this project</i>
B	0.828	<i>Our optical alignment structure consists of a deployable baffle and lens mount</i>	<i>The goals of the project are to create an optical alignment structure with two lenses that has a hands off, deployable baffle and can view debris at a distance of 2 meters away</i>
C	0.700	<i>As I understand it, our problem is to limit [slip] of a [vehicle], but also to follow a planned path</i>	<i>Our current design problem is creating and refining an algorithm to allow a [vehicle] to follow a path in the most accurate way by minimizing the effects of [slip]</i>
D	0.199	<i>It was feedback that largely changed our paper which is difficult because we had discussed our plan at length prior to submitting the paper</i>	<i>Also, we have had to make last-minute modifications to our test plan due to scheduling changes with our original testing facility</i>

Table 3.8 (cont.) - Examples of pairwise similarity between random utterances			
Ex.	Pairwise Similarity	Utterance 1	Utterance 2
E	0.000	<i>A major challenge in the project will be humidity control, since as far as we know, the microscope cannot be exposed to humidity, so the testing stage must contain the humidity</i>	<i>The production process could then be more rapidly produced for high-volume orders</i>
F	-0.168	<i>For example, utilizing steam tunnel connections so that we wouldn't need more digging for new piping</i>	<i>However, after speaking with our sponsor, fifty meters seems to be in line with what is typically required in real-world applications</i>
G	-0.188	<i>Our current task is to redesign the [redacted] laboratory</i>	<i>How accurate these are will greatly influence the accuracy and precision of these custom earbuds</i>

Example A illustrates the difference between semantic similarity and agreement. These two utterances have extremely high semantic similarity (0.998) since they both contain descriptions of the speaker's confidence level with the project. However, Utterance 1 expresses a lower level of confidence than Utterance 2, so they are not in direct agreement with one another. Example B shows utterances which are a little more complicated. Both refer to the details of an "optical alignment structure", alluding to a "deployable baffle." The details of these utterances are different, and they have differing levels of completeness. However, they are still referring to the same underlying concepts, and therefore have a high semantic similarity (0.828). Example C continues this trend, both utterances are referring to the same topic, but use different, yet related, descriptors (see "limit [slip]" vs "minimizing the effects of [slip]"). Example D demonstrates a looser connection. Both utterances seem to be discussing some form of last-minute change in plan, however the circumstances and nature of the changes appear to be different. Example E demonstrates semantically unrelated utterances, with a similarity of 0. Examples F and G are interesting as they both have negative pairwise similarity. Negative similarity indicates opposing

concepts, which is a difficult concept to imagine. As previously demonstrated, this semantic opposition is distinct from disagreement, and the reader should not rely on their normal understanding of opposites (e.g. light and dark, hot and cold, big and small) in these cases. The distribution of pairwise similarity values shows that there is a relatively low number of negative pairwise similarities in the sample, and thus they are more a curiosity than an obstacle to interpreting LSA results.

Of course, little is learned about teams by looking at random pairs of utterances. So, I shift now to the similarities between team members as a way to create a metric for sharedness. However, since cosine angle compares vectors pairwise, some agreeable definition of a team composite vector must be established. This way each utterance can be compared to the team composite using the cosine angle between them. The team composite could be defined in a number of ways, several are discussed in the following section.

Creating a Metric for Sharedness

So far, I have shown how two utterances can be compared pairwise by the cosine angle of their vectors in the k -reduced semantic space. By extension, we can compute the similarity of an utterance vector to some team composite vector, to be defined, and use this as a measure of sharedness. Then by looking at the sharedness value of each utterance, the utterances which are most shared (*i.e.*, those with the highest sharedness values) can be further investigated to gain insights into the contents of a team's shared understanding. How then, should the team composite be defined to best represent each utterance's degree of sharedness?

There are several ways to go about this. The first would be the team centroid, as used by Dong, defined as the average of all document vectors from a team (2005). This team "concept

centroid”, is described by Dong to be a meaningful representation of the team’s mental model (Dong, 2005; Dong et al., 2013). Alternatively, take the vector sum of all vectors from the team to get a resultant vector. Note that the average vector (aka the concept centroid) and resultant vector point in the same direction, so the cosine angle between any vector and either the average or resultant defined in this way will be the same. Conceptually, the “average” of written statements is not a very clear concept or one that can be easily represented practically. However, the sum of those statements is perhaps more easily understood, especially due to the “bag of words” approach of LSA which is agnostic to word order or utterance order within a larger context. For example, imagine taking each reflection response from a team and dividing it into utterances as previously described, writing each on its own slip of paper, then placing these slips in a jar. The jar containing all slips of paper from the team represents the team composite vector, defined as the sum of all utterances from the team. Note that in the case of a team with one member who contributed very few utterances, and another who contributed many. The composite vector will be dominated by the many contributions of the first team member, with little impact from the second. This is the first way of defining sharedness that will be considered – similarity to a team composite taken as the sum of all semantic content vectors from that team. Let $\vec{a}_i$ be the vector representing the i^{th} utterance in student A’s response to the prompt. Then let $\vec{A}$ be the sum of all content vectors representing utterances from student A, such that $\vec{A}$ represents the semantic content of the entire response from student A. Define $\vec{B}, \vec{C}, \vec{D}, \vec{E}$ etc. in the same way to represent the responses from student B, C, D, E etc. respectively so that the entire team is captured.

$$\vec{A} = \sum_i \vec{a}_i \quad \text{Eq. 1}$$

The team composite is the sum of the vectors representing each student's response, e.g. $(\vec{A} + \vec{B} + \vec{C} + \vec{D} + \vec{E})$. Then, sharedness is calculated as the cosine angle between the utterance, $\vec{a}_i$, and this team composite as previously explained.

$$S_1 = \frac{\vec{a}_i \cdot (\vec{A} + \vec{B} + \vec{C} + \vec{D} + \vec{E})}{\|\vec{a}_i\| \|(\vec{A} + \vec{B} + \vec{C} + \vec{D} + \vec{E})\|} \quad \text{Eq. 2}$$

This option, S_1 , is conceptually clear and easy to calculate, however it has some theoretical limitations. Notably, the utterance being compared is included in both sides of the comparison. That is, the slip we are comparing to the contents of the jar is also in the jar. Consider the extreme case where all but one slip is blank. Using this first method to compute the sharedness value of the one non-blank slip, the sum of all slips in the jar is equal to the non-blank slip alone. Thus, the S_1 would be 1, suggesting that the team has an identical understanding of the concept represented on that one slip. However, in reality this example would represent a scenario where one student responded to the question with a single utterance, and no other team member responded at all. Surely, it would be incorrect to say that that team has an identical understanding based on that observed behavior. Indeed, including all utterances in the team composite ensures there will always be some degree of self-similarity, and this will artificially increase the resulting sharedness value. Thus, this self-similarity must be removed to get a suitable metric for sharedness. The simplest way to go about this would be to exclude the utterance from the team composite, or to express it more simply, subtract it. In other words, when comparing the utterance on the slip to the contents of the whole jar, the copy of that slip in the jar is first

removed. This is the second way to define the team composite and therefore the calculation of sharedness.

$$S_2 = \frac{\vec{a}_i \cdot ((\vec{A} - \vec{a}_i) + \vec{B} + \vec{C} + \vec{D} + \vec{E})}{\|\vec{a}_i\| \left\| ((\vec{A} - \vec{a}_i) + \vec{B} + \vec{C} + \vec{D} + \vec{E}) \right\|} \quad \text{Eq. 3}$$

While theoretically, this makes sense as a logical next step given the issues discussed with the first approach, there are still some conceptual issues with S_2 . Notably, we are still comparing an utterance from student A to a composite including the other utterances from student A. Imagine once again the extreme case that only one student responds to the prompt, however this time they respond in several utterances each containing related but slightly different semantic content. Each of these utterances would have a high sharedness value since the contributions from that one student are related to each other. Yet again, this scenario where all but one team member remains silent would be falsely marked as indicating a high degree of shared understanding amongst the team members. The sensible next step seems to be to remove student A from the team composite when considering the sharedness of contributions from student A. This is the third approach that will be considered.

$$S_3 = \frac{\vec{a}_i \cdot (\vec{B} + \vec{C} + \vec{D} + \vec{E})}{\|\vec{a}_i\| \left\| (\vec{B} + \vec{C} + \vec{D} + \vec{E}) \right\|} \quad \text{Eq. 4}$$

This third approach shifts the definition slightly away from ‘team composite’ as not all members of the team are represented by this sum. Thus, it may be better described as a measure of the extent to which semantic content is similar to one’s teammates rather than similar to a team composite. All three approaches have conceptual strengths and weaknesses, and while it may be subjective, I believe that the first and third approaches are easier to understand in

practice than the second. Rather than deciding an approach based on their definitions, I will apply all three and see their implications for analyzing and interpreting the data.

Figure 3.7 shows the distributions of sharedness values computed from methods S_1 (blue), S_2 (orange), and S_3 (green). The utterances in this sample belong to the eight selected teams described previously and are in response to the repeated prompt. Distributions are represented by histograms indicating the percent of utterances that have a sharedness within each bin. All three distributions are predominantly positive, though some instances of negative sharedness are present in sharedness metrics, S_2 and S_3 . If the utterances considered were random, unconnected statements the distribution of sharedness would be centered around zero, indicating most statements are unrelated, with some positive and negative sharedness utterances occurring by chance. However, a baseline level of similarity is expected for team members responding to the same prompt. First, they are working on the same underlying problem so the subject of their responses should be related, and second, they are responding to the same questions so they should be prompted to provide similar types of information. While two teams may respond to the same prompt differently than one another, their responses should be expected to be more similar than the responses to separate questions, for example. Therefore, the overall positivity of all these metrics of sharedness when calculated for this sample is a good sign for their validity, but not sufficient.

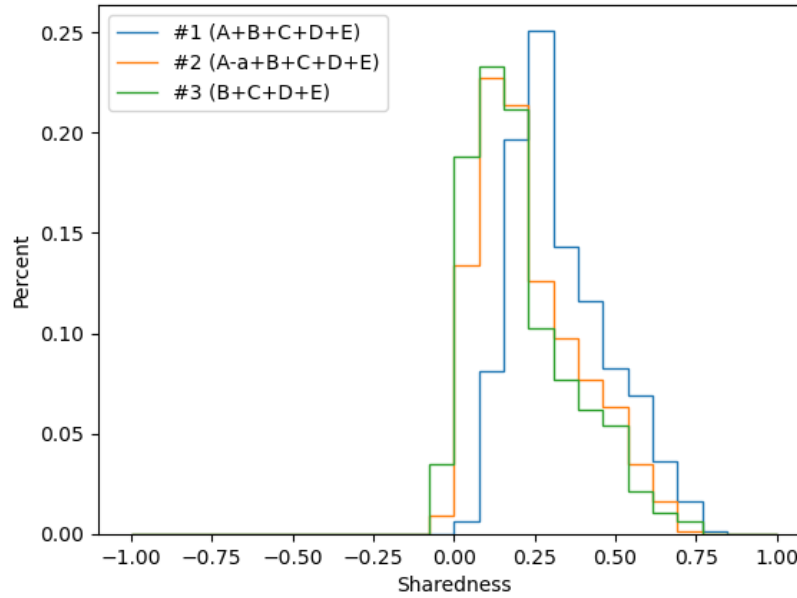


Figure 3.7 - Comparing distributions for different sharedness metrics

Also note that S_1 tends to be higher than sharedness metrics, S_2 and S_3 , which can also see this in their means, $\bar{S}_1 = 0.3296$, $\bar{S}_2 = 0.2351$, $\bar{S}_3 = 0.2045$. This is likely due to the direct self-similarity artificially increasing the S_1 values. To this end, S_2 would likely be higher than S_3 due to the implicit similarity of all utterances from a single student. This effect is present, but small in this sample. There is also an interesting difference in the shapes of the distributions, with S_1 being more symmetrical about its mean than S_2 or S_3 which are more positively skewed. As discussed above, this sample represents comparisons between team member's responses to the same prompt and should therefore tend to have positive sharedness. Additionally, the set of things related to any given concept, positively or negatively, is much smaller than the set of things which are unrelated. Therefore, I anticipate that the number of utterances with sharedness closer to 0 will be much higher than the number of utterances with very high (closer to 1) or very low (closer to -1) sharedness. By this logic, the distribution of sharedness values for this sample

should be positively skewed, with a mean above zero, with many observations of sharedness below the mean approaching 0, and some negative sharedness observations. Both S_2 and S_3 are good matches for this theoretical distribution, though S_3 is the more intuitive approach and will be used hereafter whenever “sharedness” is mentioned.

CHAPTER 4: RESULTS

Sharedness Descriptive Statistics

The distribution of sharedness (S_3) values for utterances in response to the repeated prompt is shown in Figure 4.1. Descriptive statistics for this distribution are provided in Table 4.1. The average (mean) sharedness value was 0.2045, with a standard deviation of 0.1575. Average sharedness is notably higher than the average pairwise similarity between random utterances (0.0342) as well as the average sharedness with other teams (0.0889), which indicates that when utterances are compared within the context of a team, similarity increases. This is likely because team members are working on the same project and are, therefore, referring to the same artifacts, players, and circumstances in their responses. Recall that the reflection prompt being answered in this sample is: *“Describe your current understanding of your design problem. Be sure to include answers to the following questions: What are the goals of the project? What*

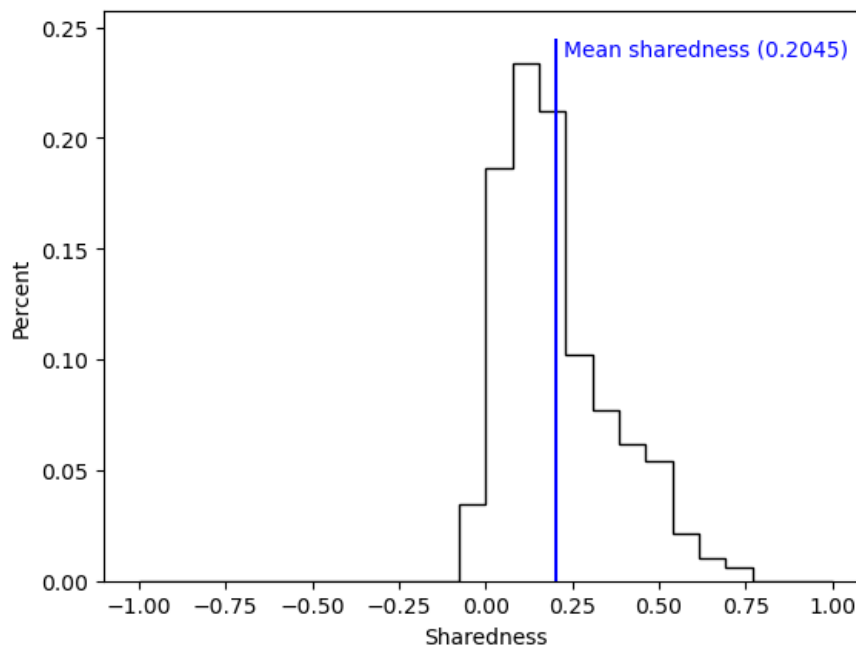


Figure 4.1 - Distribution of sharedness values, S_3

challenges do you anticipate? How comfortable are you with this project? What questions and/or uncertainties do you have? If there is anything else you'd like to mention, please do."

The distribution is positively skewed, with sample skewness calculated using the Fisher-Pearson coefficient of 0.9119. The maximum and minimum sharedness values were 0.7129 and -0.0499, respectively. The majority (59.5%) of utterances have low sharedness (< 0.2), whereas 40.5% have high sharedness, falling between 0.2 and 0.7129. The contents of this upper 40% are the most relevant to RQ2, as these utterances contain the information which is most shared amongst team members. Is there a particular type of content which tends to be more shared? Or alternatively, what patterns and trends can be identified in the utterances which have the highest sharedness values in the sample? To answer these questions, each utterance was labeled with a descriptive qualitative code representing the content of the utterance. The sharedness values of different utterance types were then compared to identify which tend to have high or low sharedness.

Table 4.1 - Descriptive Statistics for the Distribution of Sharedness Values	
Maximum	0.7129
Minimum	-0.0499
Mean	0.2047
Median	0.1694
Standard Deviation	0.1575
Variance	0.0248
Sample Skewness (Fisher-Pearson)	0.9119

Overview of Sharedness by Utterance Type

The coding scheme was discussed in detail in Chapter 3, see Table 3.7. The distribution of sharedness values for each of these utterance types is shown in the plot in Figure 4.2. The counts, medians, means and standard deviations can be found in Table 4.2. Roughly half of all

utterances were either Design Considerations (23.2%) or Perspectives (24.5%). Then in decreasing order, Project Descriptions (15.2%), Process utterances (11.4%), Expectations

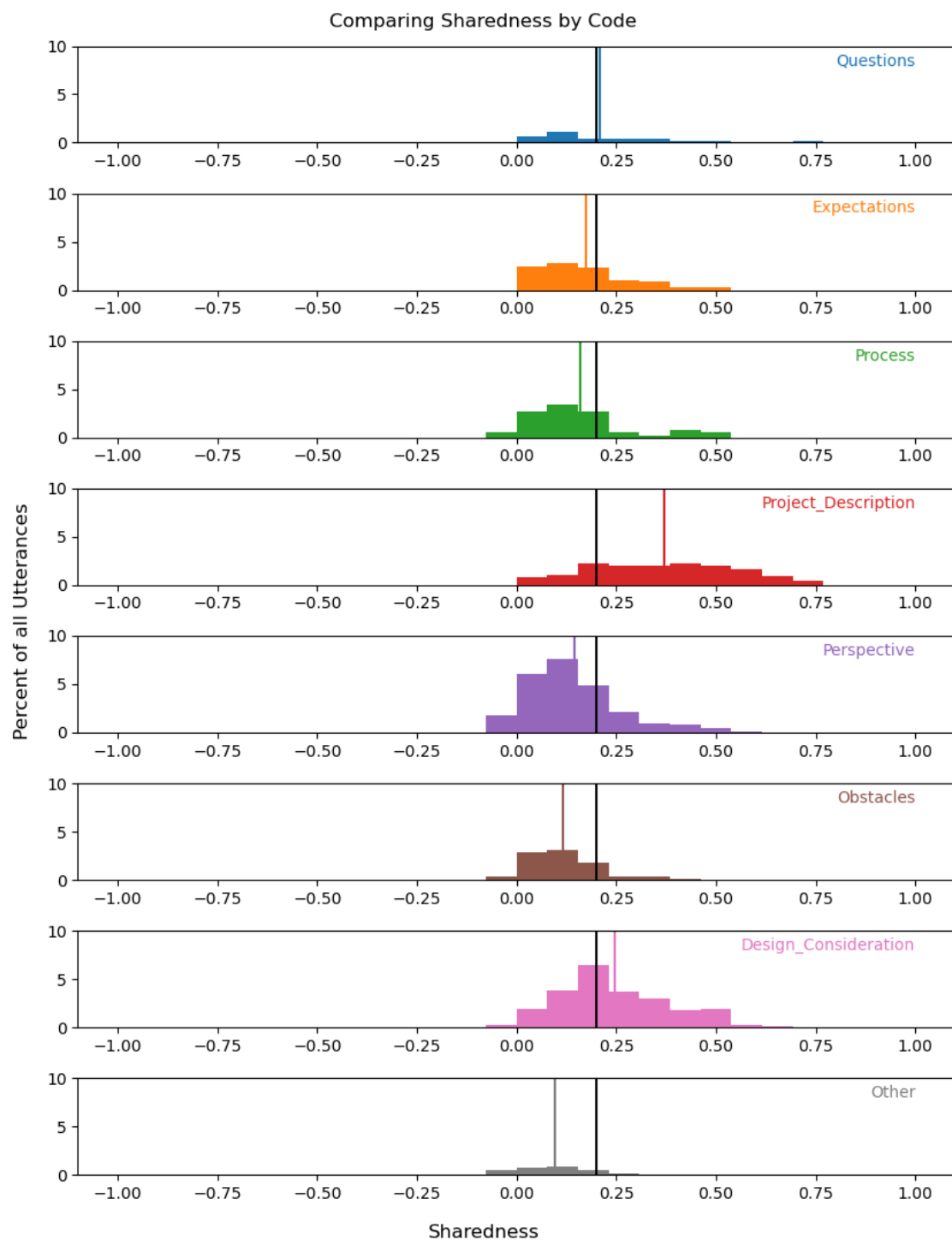


Figure 4.2 - Distribution of Sharedness Values by Utterance Type

(9.9%), Obstacles (8.9%), Questions (3.5%) and Other (3.3%). As a preliminary step toward determining which, if any, utterance types tend to have high/low sharedness, the mean sharedness for each utterance type can be compared to the overall mean, 0.205.

Table 4.2 – Comparing Sharedness Descriptive Statistics by Utterance type				
Utterance type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	23 (3.5%)	0.175	0.208	0.160
Expectations	66 (9.9%)	0.152	0.175	0.118
Process	76 (11.4%)	0.123	0.161	0.138
Project Description	101 (15.2%)	0.377	0.370	0.177
Perspective	163 (24.5%)	0.151	0.144	0.117
Obstacles	59 (8.9%)	0.105	0.115	0.090
Design Consideration	154 (23.2%)	0.212	0.246	0.139
Other	22 (3.3%)	0.081	0.087	0.086
<i>All Combined</i>	664	0.169	0.205	0.157

Additionally, the mean of each utterance type's sharedness distribution is shown in its corresponding color in Figure 4.2. A black vertical line has been drawn at 0.2 to serve as a visual reference of the overall mean. Project Descriptions and Design Considerations have means of 0.370 and 0.246 respectively and are the two categories which fall above the overall mean (~0.2). Questions (0.208) falls very close to the mean, while all others are below 0.2. The medians tell a similar story, though Design Considerations falls a little lower in this case. This suggests that utterances coded as Project Description or Design Consideration may tend to have higher sharedness than all other categories. However, the extent to which this effect is significant is unclear from the means and medians alone. The Mann-Whitney U-Test can be used to determine if these effects are significant by comparing the distribution of sharedness values for each utterance type to the overall distribution of sharedness values. This test will allow us to determine if any utterance types have stochastically higher or lower sharedness than usual. A

Mann-Whitney U-test was performed to evaluate whether sharedness differed by utterance type. The results are shown in Table 4.3.

Design Consideration and Project Description coded utterances were found to have stochastically higher sharedness than the overall distribution. Conversely, utterances coded as Process, Perspective, Obstacles, and Other all have stochastically lower sharedness. Questions and Expectations were not significantly different from the overall distribution and therefore don't tend toward either extreme. These results suggest that the elements of reflections which were most shared amongst team members in the sample corresponded to descriptions of the overall project goals as well as the detailed design considerations. Alternatively, those utterances which described individual perspectives, descriptions of process, and obstacles tended to be less shared. To the extent that reflection responses capture the mental models of participants, this analysis tells us that in general, the teams in the sample developed a shared understanding of the project goals and the details of their design but may not have shared a common perception of how things were going, including the obstacles they faced, nor a shared, explicitly represented, understanding of their process.

Table 4.3 – Mann-Whitney U-Test by Utterance type		
Utterance type	Significant Direction	Test Result
Questions	No difference	Not significant
Expectations	No difference	Not significant
Process	Low	U=21128, p=0.010
Project Description	High	U=51123, p<0.001
Perspective	Low	U=42216, p<0.001
Obstacles	Low	U=12961, p<0.001
Design Consideration	High	U=61337, p<0.001
Other	Low	U=3831, p<0.001

As another example of this, recall the 40% of utterances which have sharedness values above the mean. These high sharedness utterances are those to the right of the black line in

Figure 4.2. There are some utterances from each utterance type that fall above average, but most are Project Description (30.9%) and Design Consideration (30.5%) utterances. Then in descending order, Perspective (14.9%), Process (8.8%), Expectations (7.3%), Obstacles (3.8%), Questions (3.4%), and Other (0.4%). The high percentages of Project Description and Design Consideration utterances among the top 40% sharedness values further supports the claim that the teams in the sample have developed some shared understanding of the project goals and the details of their designs.

Finally, consider two example utterances with high sharedness and the utterances from their teammates, shown in Table 4.4 and Table 4.5. Recall that the sharedness value is the similarity between an utterance and a composite which represents all the utterances from the speaker's teammates in response to the same prompt. As such, the utterances represented by the composite are also provided explicitly. This allows us to manually search for similar content between the sample utterance and the rest of the team's responses.

These example utterances are the two highest sharedness utterances in the sample. Both are Project Descriptions, and similar words and phrases are clearly shown in the teammate's responses, demonstrating that the semantic content of these utterances is shared by the team. One possible explanation for this is that teams were given a description of their project at the beginning of the semester. These sponsor-provided descriptions are confidential and unfortunately cannot be reproduced here, however, there are many examples of phrases and descriptions that the students borrow and use in their responses. For example, "silicon device extraction" and "minimize manual labor" appear directly in a project brief. However, these initial sponsor project descriptions are not the only source of information that students coalesce around.

Interestingly, the team responsible for the utterances in Table 4.5, Team CP, was not provided with a detailed project brief at the start of the semester. Therefore, the team must be converging around a shared understanding developed elsewhere. The circumstances surrounding different project teams will certainly impact the means by and extent to which each team shares a common understanding of their project, and I will explore these in more detail in the following sections.

Table 4.4 - High Sharedness Utterance Example 1 (Team GR)	
Example Utterance	<i>The primary goal for this project is to develop a silicon device extraction process requiring minimal manual labor, minimizing damage to the devices, and optimizing efficiency in transfer yield and rates (sharedness = 0.7129)</i>
Teammate ID	Teammate's Utterances
SP22_01273	• <i>The project that our company is sponsoring is looking to expand their business through the creation of a silicon device extractor</i> • <i>At the very least, the company sponsor is requesting a proof of concept with the extraction process, and after that there is an ask for a localization system that can both identify the position of a [sensor] in the wafer as well as index the [sensor] as it is extracted from the wafer and placed into a wafer tray</i> • <i>Throughout the extraction process, the company sponsor would also appreciate some force readings related to the application upon the [sensor] and the wafer stand</i> • <i>Some challenges involve the precision needed as well as the speed and acceleration in the z - motion</i> • <i>The mask design of the wafer decreases the necessary force required to separate the [sensor] from the wafer, and some FEA will be done to characterize the force required for extraction.</i>
SP22_42298	• <i>I think I have a very solid understanding of the design problem associated with our project</i> • <i>The goal of this project is to design a system that can automate the extraction process of the silicon gauges that [sponsor] use to make their [sensor]</i> • <i>It is crucial that the extraction has high quality, since we do not want the devices to have excess or less material than the desired shape</i> • <i>The ultimate goal is also to increase the speed at which the extraction is performed, but that is slightly outside the scope of our part in this project</i> • <i>I am very comfortable with the needs of this project and the solutions that we have prepared for our sponsor</i> • <i>The main worry I have ahead lie is whether our extraction method can actually meet the target yield rate of over [redacted]%, as the first 30 tests we have run seem to show that the devices break very easily during extraction</i> • <i>Part of this is due to imperfections in the nozzle alignment to the devices, and part of the reason is due to using tweezers to move the device to the tray after extraction.</i>

Table 4.5 - High Sharedness Utterance Example 2 (Team CP)	
Example Utterance	<i>The main goal of our project is to design a control algorithm that accounts for these slip conditions and makes the [vehicle] go in a straight line (sharedness =0.698)</i>
Teammate ID	Teammate's Utterances
SP22_91567	• <i>As I understand it, our problem is to limit track slip of a [vehicle], but also to follow a planned path</i> • <i>Therefore, we have two problems: detecting slip and localizing our robot</i> • <i>I anticipate the localization to be challenging, as this is an ongoing problem that has been studied as long as robotics has been studied</i> • <i>In addition, localization becomes harder with smaller vehicles because the accuracy of the positioning sensors is much more important</i> • <i>The scale of the project also makes measuring ground speed more difficult, as a smaller vehicle at low speeds means we need to be more accurate than a full-sized vehicle</i> • <i>I am comfortable with the concepts involved in the project, but not necessarily the tools and timeframe we have to solve it</i> • <i>Finally, I am a bit concerned that the scope of the project is not fully agreed upon between the sponsor and all team members.</i>
SP22_37767	• <i>The goal of our project is to get a small scale [vehicle] to follow a path with heavy slip conditions within [redacted] max error</i> • <i>The major challenges relate to 3 main areas: sensors, control system, and ground truthing</i> • <i>The sensor selection portion of our project has been completed and will only be revisited if we experience significant problems</i> • <i>The ground truthing system we have researched to implement seems relatively straight forward -- with a near out of the box solution</i> • <i>That said, it still requires some software and logistical changes to work for our use case</i> • <i>That brings us to our major challenge point -- which is the control system</i> • <i>There is no agreed upon -- or even publicly known -- solutions to making a path finding robot that's goal is to reduce slipping [redacted]</i> • <i>Solving for our own plant model even seems a large difficulty, as slip and its force responses are not well encompassed in a pure physics model</i> • <i>Our main questions relate to this, such as: what is the best signal to base our control loop around -- deviation from path, a dynamically adjusted heading pointing back to path, something else</i> • <i>All of these questions we will be bringing up to [advisor] and control system specialist this week, and as we move through out project.</i>
SP22_41617	• <i>My current understanding of our design problem is that we have to make a [vehicle] travel in a straight line while there is gravel on the floor which causes [slipping]</i> • <i>The challenges I anticipate are that it our computer model won't reflect the real model accurately and that hardware parts will break and take a long time for replacement parts to ship.</i>
SP22_72514	• <i>Our current design problem is creating and refining an algorithm to allow a [vehicle] to follow a path in the most accurate way by minimizing the effects of [slip]</i> • <i>The main goal of the project is for the vehicle to travel along with the path, with at no point more than [redacted] of deviation from the said path</i> • <i>Additionally, it should travel within [redacted]% of its maximum speed and be able to move autonomously for up to [redacted]</i> • <i>I anticipate calculating and accommodating for slip to be a major hurdle</i> • <i>I'm fairly comfortable with the project, but I do think once our hardware has arrived I'll feel a lot more comfortable with how to work to find a solution to the problem</i> • <i>My largest uncertainty is how to interpret data provided to us by the sensors, primarily the IMU</i> • <i>Hopefully, once I see a few example outputs of that data, I can better conceptualize on how we'll capitalize on that information.</i>

Overview of Sharedness by Team

Table 4.6 shows the mean sharedness by team for each utterance type. Recall that mean sharedness for all utterances in the sample was ~0.2. For each team, the mean sharedness for all coded utterances combined is shown in the bottom row of Table 4.6. Interestingly, the variation in the overall mean sharedness between teams is relatively small, the highest from Team EK (0.234) and the lowest from Team WZ (0.158), with a standard deviation of just 0.032. This suggests that all teams in the sample achieved some level of baseline shared understanding indicated by an average sharedness greater than zero, though no teams landed remarkably high or low within the context of this sample. However, by considering specific utterance types, more differences between teams appear.

Table 4.6 - Overview of mean sharedness by team									
	<i>All combined</i>	Team BK	Team CP	Team EK	Team GR	Team KL	Team LU	Team WZ	Team ZH
Questions	0.208	0.269	0.414	0.278	0.200	0.162	0.099	0.119	n/a
Expectations	0.175	0.204	0.214	0.143	0.130	0.203	0.181	0.104	0.293
Process	0.161	0.164	0.100	0.182	0.172	0.145	0.231	0.087	0.179
Project Description	0.370	0.295	0.401	0.465	0.440	0.286	0.273	0.257	0.458
Perspective	0.144	0.159	0.121	0.162	0.154	0.124	0.103	0.194	0.126
Obstacles	0.115	0.162	0.100	0.109	0.126	0.089	0.120	0.079	0.152
Design Consideration	0.246	0.285	0.256	0.278	0.306	0.173	0.251	0.156	0.219
Other	0.087	0.126	0.099	0.178	0.025	0.094	0.032	0.078	n/a
<i>All combined</i>	0.205	0.223	0.224	0.234	0.226	0.163	0.179	0.158	0.226

The utterance type with the overall mean highest sharedness was Team EK's Project Description utterances (0.465), while the lowest was Team GR's Other (0.025). If the Other code is ignored, the lowest was Team WZ's Obstacles utterances (0.087). For most teams, Project Description was the type of utterance with the highest sharedness, except Team CP who had higher sharedness in Questions utterances. However, Team CP only had one utterance coded as

Question, so this may reasonably be an outlier. While we can gain some insight into the types of utterances which were more or less shared by considering which utterance types have average sharedness above or below the overall mean, investigation of means does not explain differences in the distributions of sharedness values, such as sample size, variance, etc. Thus, I used the one-sided Mann-Whitney U-test to compare each group of utterances of a specific type to the overall group to determine which have high or low sharedness.

For each team and each utterance type, the group of sharedness values of utterances belonging to that specific utterance type was compared to the sharedness values of all utterances from all teams using a Mann-Whitney U-test. This allows us to determine if a group of utterances belonging to a certain utterance type has stochastically higher or lower sharedness than “normal.” Where “normal” here is given by the distribution of all observed sharedness values. Table 4.7 provides an overview of the significant results from these one-sided Mann-Whitney U-tests. A cell marked “low” indicates that the utterances belonging to the corresponding utterance type had significantly *lower* sharedness than the overall group of all utterances. Conversely, a cell marked “high” indicates that the utterances belonging to the corresponding utterance type had significantly *higher* sharedness than the overall group of all utterances. Empty cells indicate no significant result, meaning that the sharedness values for utterances from that utterance type and team were not distinguishable from the overall distribution. The results for all teams, including p-values can be found in Appendix B.

Table 4.7 - Overview of significant results from one-sided Mann-Whitney U-Tests.									
	<i>All combined</i>	Team BK	Team CP	Team EK	Team GR	Team KL	Team LU	Team WZ	Team ZH
Questions									
Expectations								low	
Process	low		low					low	
Project Description	high	high	high	high	high	high		high	high
Perspective	low		low			low	low		low
Obstacles	low		low	low		low		low	
Design Consideration	high	high	high	high	high		high		
Other	low						low		

There were no teams with a significant difference between the sharedness values of their Questions utterances and the sharedness of the overall group of utterances. This is to be expected from the previous chapter's results looking at all teams together as Questions utterances were not found to have significantly different sharedness from the overall group. We might expect Expectations utterances to be similar to Questions based on previous discussions, however there is a slightly different story. One team, Team WZ, has low-sharedness for Expectations utterances. That is, when we consider all Expectations utterances from all teams together, we see no significant difference in sharedness compared to the group of all utterances from all teams. However, when considering just the Expectations utterances from Team WZ, we see that this group of utterances has lower sharedness. Thus, in general, Expectations utterances tend to have roughly average sharedness, but Team WZ demonstrates unusually low sharedness. Notably, all teams except Team LU had high sharedness Project Description utterances, in line the with previous results. Project Description utterances had the highest mean sharedness, indicating that these utterances represent the content which is most commonly shared amongst team members (as represented by the semantic similarity of utterances measured through LSA). It is interesting

that all teams except LU demonstrate this trend. The other utterance type which tended to have high-sharedness as indicated by the previous chapter, Design Consideration utterances, had five of eight teams which followed this trend. Teams KL, WZ, and ZH had Design Consideration utterance sharedness values which were not significantly different from the overall group and are therefore lower than expected from the previous chapter. In the following sections, I will explore the details further on a team-by-team basis.

Descriptive Cases

This section provides an in-depth examination of several design teams in the form of an exploratory, multi-case study. As Yin explains, case study is a useful empirical method to “investigate a contemporary phenomenon in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident” (2017). Investigating several cases in more detail will facilitate an exploration of how and to what extent each team developed a shared understanding. This will be presented in light of sharedness results previously reported in order to address the research questions. I’ll first explain the reasoning behind the selection of the teams for case study review, then will provide an overview of the sources of data that will be considered in this chapter. Each case will then be presented in its own section.

Thus far, I have shown that LSA can be used to provide a measure of sharedness at the utterance level, and that utterances containing different content, e.g. those with different codes, such as Project Description versus Obstacles, tend to have different sharedness values. While some, such as Project Description utterances, tend to have high sharedness compared to the overall group, others tend to have lower sharedness, but not all teams exemplify these trends to

the same extent. The following section will investigate examples of high and low sharedness as it manifests in different project contexts. Cases were selected to represent the range of sharedness values in the sample, with teams selected for overall highest average sharedness, and overall lowest average sharedness. Additionally, a group in the middle will also be explored, chosen due to interesting circumstances which will be explained shortly. The overall average sharedness values for each team are shown in Table 4.6. Team WZ had the lowest overall average sharedness at 0.158. Team EK had the highest, at 0.234. Team CP, whose overall average sharedness was fourth highest at 0.224, will be included also, as this team's experience provides some additional insight into the effect of unclear goals at the start of the project.

The multi-case study is constructed from several sources of data. Firstly, to substantiate the findings so far, I will present utterances previously reported from the sharedness data, which cover only responses to the repeated reflection prompt (P1*, P4*, P7* in Table 3.4). By providing quotes from students in context, I aim to illustrate examples of high and low sharedness and how these might relate to the circumstances described in the reflection responses themselves. As the repeated prompt was posed three times, at the beginning, middle, and end of the semester, comparing between these instances will provide insight into how the team's shared understanding changed over time. Secondly, I will go beyond this group of repeated reflections to consider responses to the other prompts asked throughout the semester. These are shown in Table 3.4 and target various aspects of the design process. As such, they will provide additional context about the team's project circumstances and approach. Note that the sharedness values of these utterances are not included in the previously reported data as they address prompts other

than the repeated prompt. Likewise, these utterances were not part of the subgroup which was assigned descriptive codes.

Finally, answers to the survey questions will also be included (see Table 3.3). A summary of responses to Likert and 1-10 scale questions will be provided for each team. Respondents also had the opportunity to explain their answers in an open-ended response field. These explanations will be included when informative to the reader. It should be noted that while quotes may be partially redacted or otherwise blinded for reasons of confidentiality, care has been taken to preserve the original meaning, messaging, or intent. The synthesis of data presented is intended to provide an accurate representation of each team's experience, to the best of the author's knowledge and ability.

Case 1: Vehicle Control Team (CP)

Team CP had 6 students in total, 5 consented to be included in the study. Two responded to demographic questions. This team included SP22_91567 (White or Caucasian, Male) and SP22_72514 (White or Caucasian, Male). Two additional students, SP22_41617 and SP22_50384, responded to the survey but did not answer demographic questions. This project was sponsored by a Fortune 500 company. At the start of the semester, a full project description was not available. Students were provided with details about the sponsor and a brief description of the project, which centered on the use of a detector in a vehicle control system. Despite this unusual start, this team had the fourth highest overall average sharedness, 0.224, as shown in Table 4.8.

Team CP submitted 86 utterances in response to the repeated prompt throughout the semester. The highest percentage of these were Design Considerations at 33.7% which is higher

than the overall percent of these types of utterances (23.2%, see Table 4.2). The team's highest sharedness utterance type was Project Description at 0.401, though these utterances had a relatively high standard deviation of 0.218.

Table 4.8 - Team CP Sharedness Statistics by Utterance type				
Utterance type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	1 (1.2%)	0.414	0.414	
Expectations	11 (12.8%)	0.180	0.214	0.120
Process	7 (8.1%)	0.096	0.100	0.084
Project Description	14 (16.3%)	0.399	0.401	0.218
Perspective	16 (18.6%)	0.103	0.121	0.095
Obstacles	6 (7.0%)	0.089	0.100	0.072
Design Consideration	29 (33.7%)	0.238	0.256	0.122
Other	2 (2.3%)	0.099	0.099	0.095
All Combined	86 (100.0%)	0.185	0.224	0.167

As before, Mann-Whitney U-tests are used to determine which utterance types have stochastically greater or lower sharedness than the overall group (see Table 4.9). For Team CP, Project Description ($p < 0.001$) and Design Consideration ($p = 0.009$) utterances had significantly higher sharedness than the overall group. While Process ($p = 0.037$), Perspective ($p = 0.018$), and Obstacles ($p = 0.044$) utterances had significantly lower sharedness than the overall group. The distribution of sharedness values for each utterance type is shown in Figure 4.3. Notably, there is a cluster of Project Description utterances with sharedness above 0.5, which is quite high, but also utterances with sharedness between 0 and ~0.4.

Table 4.9 - Mann Whitney U-Test by Utterance type – Team CP		
Utterance type	Less than	Greater than
Questions	U=579.5 p=0.902	U=579.5 p=0.099
Expectations	U=4003.5 p=0.708	U=4003.5 p=0.292
Process	U=1414.5 p=0.037	U=1414.5 p=0.963
Project Description	U=7040.0 p>0.999	U=7040.0 p<0.001
Perspective	U=3682.0 p=0.018	U=3682.0 p=0.982
Obstacles	U=1187.0 p=0.044	U=1187.0 p=0.956
Design Consideration	U=12113.5 p=0.991	U=12113.5 p=0.009
Other	U=411.0 p=0.176	U=411.0 p=0.825

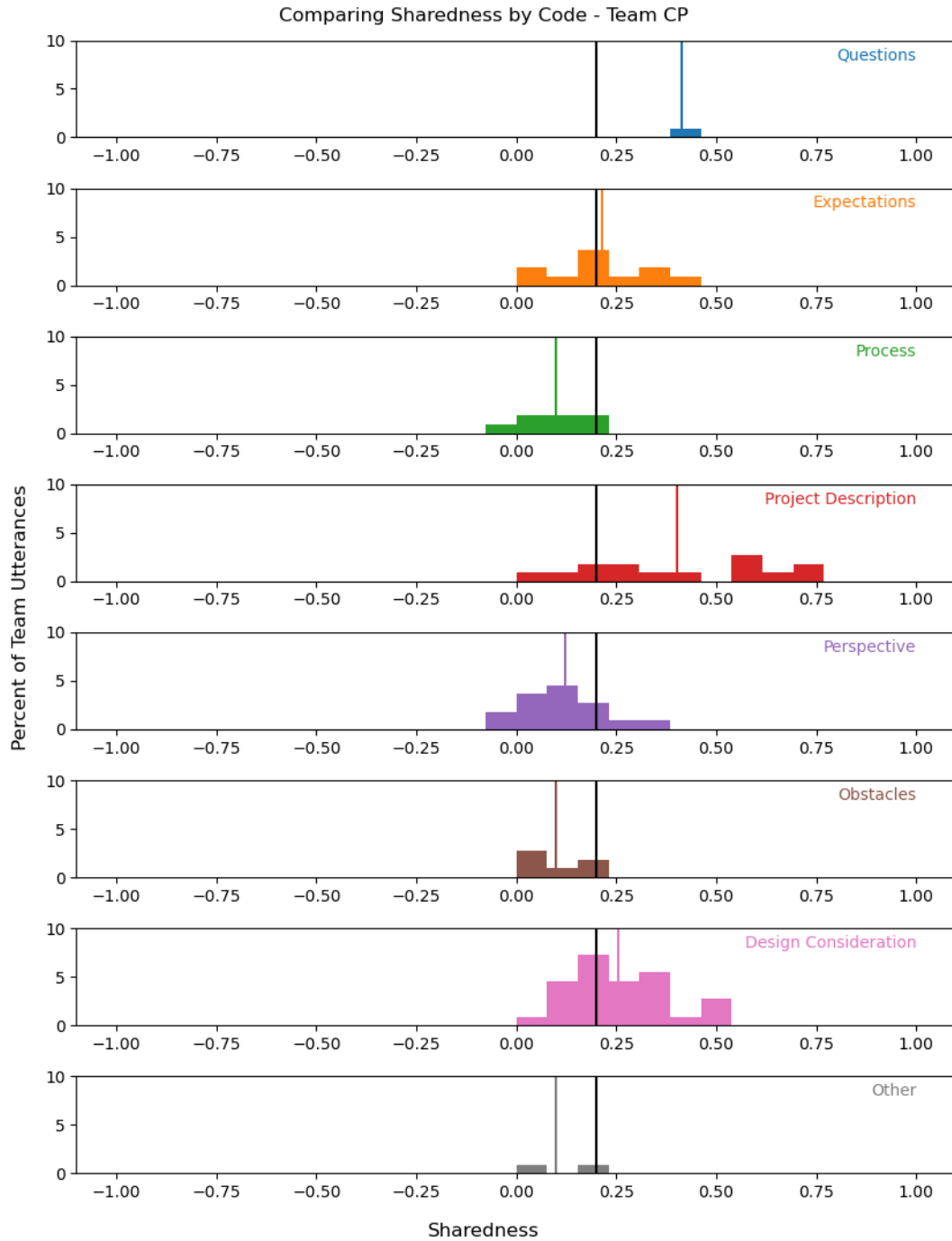


Figure 4.3 - Histograms showing the distributions of sharedness values for each code (Team CP).

Responses to the first iteration of the repeated prompt were submitted during the second week of the semester. Teams were assigned during the first week, and each team was required to

reach out to schedule a meeting with their faculty advisor and project sponsor. Team CP had limited information available due to an incomplete project description at the start of the semester (*“Our current understanding of our project is limited since we don't have all the details from our sponsor”*, Obstacles, sharedness=0.207, SP22_50384). They describe the problem similarly, but this is likely because their understanding of the problem is based on the limited information provided by the sponsor (*“The goal of the project is to analyze discrepancies between a ground true speed sensor and the vehicle's speed”*, Project Description, sharedness=0.432, SP22_72514), (*“My current understanding of the design problem is that a sensor must be installed [redacted to preserve anonymity] such that it is able to accurately and reliably measure ground speed”*, Project Description, sharedness=0.280, SP22_91567), (*“our goal is to develop a way to accurately detect the ground speed of a vehicle”*, Project Description, sharedness=0.367, SP22_50384). However, they cite what they believe to be relevant experience, (*“I feel fairly comfortable with this project despite the lack of a detailed description”*, Perspective, sharedness=0.109, SP22_72514), (*“I am relatively comfortable with this project because it involves many topics that we've learned in previous courses such as signal processing and dynamics”*, Perspective, sharedness=0.202, SP22_50384). Though this experience is not always confidence boosting, (*“I have worked on similar projects, which proved very difficult and took much longer than a single semester”*, Perspective, sharedness=-0.026, SP22_91567), (*“These often required algorithm development that I am not comfortable with”*, Perspective, sharedness=0.207, SP22_91567). The average sharedness of all utterances from this team in response to this first iteration of the repeated prompt is 0.166.

In week five, students submitted deliverables for the Project Proposal Report and Presentation. Responses to the second iteration of the repeated prompt (P4* in Table 3.4) were submitted in week six. The team seems to have a more complete understanding of their project, and the average sharedness has increased to 0.281. Notably the goal seems to have changed from detecting ground speed to a control algorithm, (*“Our current design problem is creating and refining an algorithm to allow a [vehicle] to follow a path in the most accurate way by minimizing the effects of slip [redacted]”*, Project Description, sharedness=0.590, SP22_72514), (*“The main goal of our project is to design a control algorithm [...]”*, Project Description, sharedness=0.698, SP22_50384). Additionally, some specific requirements appear in several students’ responses, (*“at no point more than [redacted] of deviation from the said path.”*, Project Description, sharedness=0.664, SP22_72514), (*“Additionally, it should travel within [redacted]% of its maximum speed and be able to move autonomously for up to [redacted]”*, Project Description, sharedness=0.113, SP22_72514), (*“The goal of our project is to get a small scale [vehicle] to follow a path with heavy slip conditions within [redacted] max error”*, Project Description, sharedness=0.693, SP22_37767). The team identified a few different potential challenges, (*“The challenges I anticipate are that it our computer model won't reflect the real model accurately and that hardware parts will break and take a long time for replacement parts to ship.”* Expectations, sharedness=0.180, SP22_41617), (*“The main concern is that our model will not account for real-life conditions and will thus be inaccurate”*, Expectations, sharedness=0.351, SP22_50384), (*“My largest uncertainty is how to interpret data provided to us by the sensors, primarily the IMU”*, Design Consideration, sharedness=0.156, SP22_72514), (*“I anticipate the localization to be challenging, as this is an ongoing problem that has been*

studied as long as robotics has been studied”, Expectations, sharedness=0.149, SP22_91567), (*“That brings us to our major challenge point -- which is the control system”*, Design Consideration, sharedness=0.172, SP22_37767).

The final iteration of the repeated prompt (P7* in Table 3.4) took place in week fifteen of the semester. The previous week, teams shared their work with other teams, FAs, and sponsors, in a “Trade Show.” At this point, the team has progressed to a more detailed understanding of their project. While they seem to agree that the goal of the project is to control a vehicle, accounting for loss of traction, they explain the goals slightly differently than one another. Average sharedness has decreased to 0.202, (*“The goal of the project is to design an algorithm that minimizes the effect of slip on a [vehicle] over a variety of different terrains by not allowing the vehicle to go off course by a specific threshold and maintaining a speed within a predetermined range, as well as provide the accompanying sensor suite”* Project Description, sharedness=0.540, SP22_72514), (*“As I understand it, the goal of my project is to implement a traction control system on a small-scale RC vehicle”*, Project Description, sharedness=0.156, SP22_91567), (*“Right now, our project is to design an algorithm that allows a [vehicle] to autonomously account for [slip] and go in a straight path or follow a desired trajectory.”* Design Consideration, sharedness=0.485, SP22_50384) (*“[Slip] is important because [slip] is a barrier to autonomy”*, Design Consideration, sharedness=0.121, SP22_50384), (*“My understanding of the design problem are that our thing is like a [vehicle] that slips because of bad soil conditions”* Project Description, sharedness=0.236, SP22_41617). The team seems to draw a distinction between the algorithm and the physical implementation, including the sensor suite (*“In addition*

to developing the algorithm, we have to build and test it on a physical prototype with hardware and sensors” Design Consideration, sharedness=0.217, SP22_50384).

The team’s average sharedness increases from week two (0.166) to week six (0.281), then drops to week fifteen (0.202). However, the Project Description utterances’ average sharedness moves from 0.359 in week two, to 0.449 in week six, then to 0.310 in week fifteen. The relatively high Project Description sharedness at the start of the semester seems to be informed by the limited information provided by the sponsor. Since the team is working with a small amount of information, their descriptions of the project are similar. However, the overall average sharedness in week two is relatively low (0.166), which suggests that the team is pushing beyond their shared understanding to pull in connections from their existing mental models and prior experiences. These perspectives are unique to the individual and tend to have low sharedness. This can be seen in the low sharedness Perspective utterances shown in week two. Once the team has had the opportunity to clarify their project goals with the sponsor, there is a corresponding increase in sharedness, especially in Project Description utterances. In week six, several students include design specifications or requirements. After this, sharedness decreases toward the end of the semester. While there are still high sharedness utterances in week fifteen, the responses are more individualized.

Further insights into the team’s experience are given in responses to the survey and other prompts. A summary of Team CP’s multiple choice survey responses is given in Figure 4.4. The team felt strongly that the sponsor did not provide them with relevant background information, and that sponsor expectations were not clear at the start of the semester. We can see this in an explanation from SP22_91567, “*the sponsor continuously expanded[sic] the scope of the project from ground speed detection to traction control to following a straight, planned path to planning and following a complex course over different surfaces.*” Also, the team rated the quality of communication with their sponsor very low, with an average rating of 1.5 out of 10. They agreed or strongly agreed that the project required clarification before they were able to move forward, and that their understanding has changed since the start of the semester. There was some disagreement about whether they had a shared understanding, which is elucidated by this explanation from SP22_91567, “*The team is now on the same page, but we were not for a long time as our problem statement was very broad and each member interpreted it differently.*”

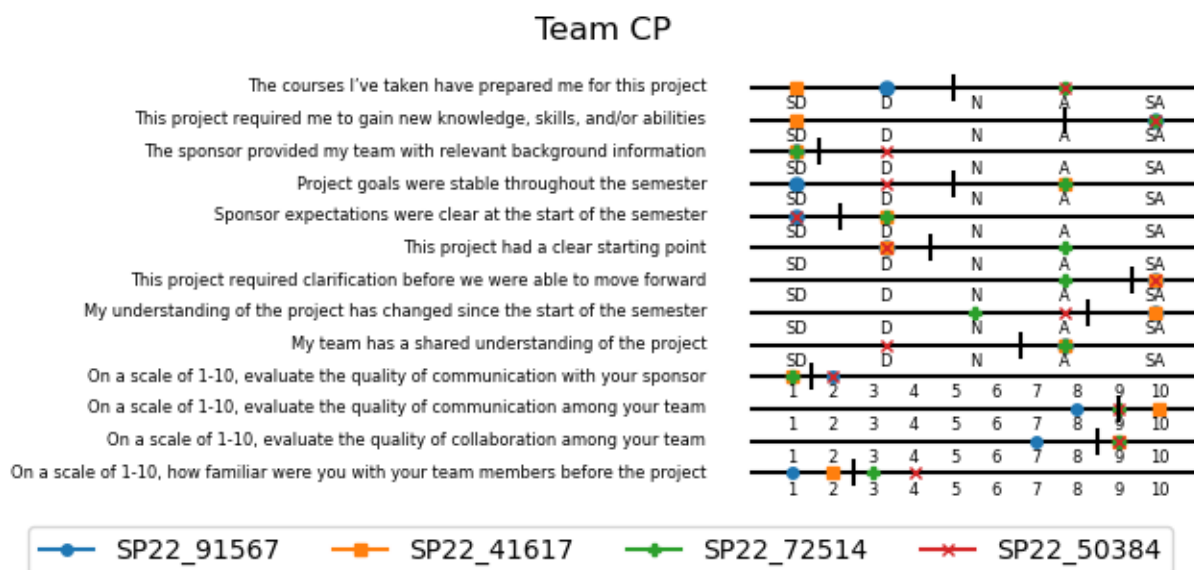


Figure 4.4 - Summary of Team CP’s multiple-choice survey responses

In their responses to reflection prompt P2, the team provides some insight into their meetings with the sponsor, faculty advisor, and teaching assistant. These reflections seem to describe how the team's understanding of their project changed as new information was introduced. They report a change in project scope, (*"In our sponsor meeting, we were prepared with questions, research, etc However, when discussing our project with our sponsor, the project was much different than we had anticipated"*, sharedness=0.514, SP22_91567), (*"That said, the scope of the project was much different than we expected [...]"*, sharedness=0.197, SP22_37767). Additionally, they described how their expectations for their collaborators shifted as a result of these meetings. That is, they gained a new understanding of who was responsible for what aspects of the project, or who they should bring certain kinds of questions. This is reminiscent of the "team interaction knowledge" and "team knowledge" elements of a shared mental model reported in literature (Johnson et al., 2007), (*"I expected to receive more vision and advice regarding the path and deadlines for our project"*, sharedness=0.237, SP22_72514), (*"I think the FA meeting differed from these expectations a bit while the TA meeting was more aligned"*, sharedness=0.248, SP22_72514), (*"Both went well, but the FA meeting presented itself as much more of a technical resource for our project while the TA was the aforementioned guidance fort the class structure"*, sharedness=0.255, SP22_72514), (*"Both still went well, but my opinions and expectations are certainly restructured moving forward"*, sharedness=0.098, SP22_72514). Notably, the team indicated an intention to utilize this newfound knowledge about their team to better target questions to their advisor's and sponsor's expertise, (*"Next time I think I would try to show up with more precise questions to ask rather than the more open ended outline we had"*, sharedness=0.325, SP22_72514), (*"Next time, we will try to make more efficient*

use of our sponsor's time by preparing and asking more targeted questions regarding the project", sharedness=0.335, SP22_50384)

In the responses to reflection prompt P6, one student goes into detail about how the team is using a block diagram to, “keep team members on the same page.” (*"So far, we have constructed a working block diagram of our plant and controller systems which are continually updated to reflect changes", sharedness=0.065, SP22_91567), ("This serves not only to keep all team members on the same page, but it is also a good presentation tool that we can reference at any point and assume that it is up to date", sharedness=0.034, SP22_91567), ("It is also used in each TA and FA meeting to discuss which parts of the system are in progress that week", sharedness=0.024, SP22_91567), ("This block diagram also drives our simulation: a Simulink model of our vehicle(plant) which enables us to design and test different control algorithms independently of our physical vehicle", sharedness=0.619, SP22_91567), ("This is useful for two reasons", sharedness=-0.013, SP22_91567), ("The first is that this enables the parallel development of hardware and software, and second, this enables us to illustrate our progress without physical demonstrations, which may not always be convenient", sharedness=0.080, SP22_91567), ("Finally, it has proved useful to have a platform to work on the system model that can be accessed remotely by all team members", sharedness=0.244, SP22_91567). The same student gives some insight into the team’s decision making process in their response to P3 in week 4 (*"So far, we have taken simple votes for most of the decisions, with members presenting various solutions based on their experience and comfort level with the technology and/or solving similar problems", sharedness=0.109, SP22_91567), ("Then, most of the decisions have been based heavily on both sponsor suggestions and on the team’s experience", sharedness=0.184,**

SP22_91567) ("*These decisions are being made with a Pugh matrix rather than a simple vote and are going much more slowly as research is done*", sharedness=0.087, SP22_91567)

Case 2: Mass Customization Team (WZ)

Team WZ (Mass Customization Project) had 6 students total, 3 consented. One responded to demographic questions. This team included SP22_48390 (White or Caucasian, Male). Another student, SP22_22009, responded to the survey but did not answer demographic questions. This project was centered around the use of 3D scanners and rapid prototyping for mass customization. The project was industry sponsored, though the sponsor and faculty advisor were one and the same. The final deliverable for this project was a feasibility study, as well as proof of concept prototypes.

Table 4.10 - Team WZ Sharedness Statistics by Utterance type				
Utterance type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	6 (7.8%)	0.085	0.119	0.086
Expectations	9 (11.7%)	0.096	0.104	0.050
Process	6 (7.8%)	0.056	0.087	0.097
Project Description	11 (14.3%)	0.222	0.257	0.116
Perspective	19 (24.7%)	0.175	0.194	0.181
Obstacles	4 (5.2%)	0.064	0.079	0.040
Design Consideration	19 (24.7%)	0.189	0.156	0.111
Other	3 (3.9%)	0.098	0.078	0.055
All Combined	77 (100.0%)	0.132	0.158	0.134

Team WZ's utterances had a mostly typical distribution of utterance types, especially among those utterance types with the highest percentages (see Table 4.10). The three utterance types representing the highest percent of utterances were Design Considerations and Perspectives tied at 24.7%, then Project Descriptions at 14.3%. Despite a relatively standard breakdown of utterance types, Team WZ had the lowest overall average sharedness, 0.158, of all teams in the

sample (overall average was 0.205). The highest average sharedness utterance type for this team was only 0.257, for Project Description utterances. This is lower than the average sharedness of all team's Project Description utterances, 0.370.

In fact, Team WZ's average sharedness per utterance type was lower for all types of utterances. Despite the fact that Team WZ's Project Description utterances had lower average sharedness than the overall average Project Description utterance, Team WZ's, Project Description utterances had significantly higher sharedness than the overall group of all utterances regardless of utterance type or team ($p=0.046$). Expectations ($p=0.021$), Process ($p=0.015$), and Obstacles ($p=0.031$) utterances had significantly lower sharedness than the overall group (see Table 4.11). The distribution of sharedness values for each utterance type is shown in Figure 4.5.

Table 4.11 - Mann Whitney U-Test by Utterance type – Team WZ		
Utterance type	Less than	Greater than
Questions	U=1321.0 $p=0.078$	U=1321.0 $p=0.923$
Expectations	<i>U=1811.5 $p=0.021$</i>	U=1811.5 $p=0.979$
Process	<i>U=963.0 $p=0.015$</i>	U=963.0 $p=0.985$
Project Description	U=4735.5 $p=0.954$	<i>U=4735.5 $p=0.046$</i>
Perspective	U=5800.5 $p=0.275$	U=5800.5 $p=0.725$
Obstacles	<i>U=611.0 $p=0.031$</i>	U=611.0 $p=0.969$
Design Consideration	U=5605.5 $p=0.204$	U=5605.5 $p=0.796$
Other	U=489.5 $p=0.064$	U=489.5 $p=0.936$

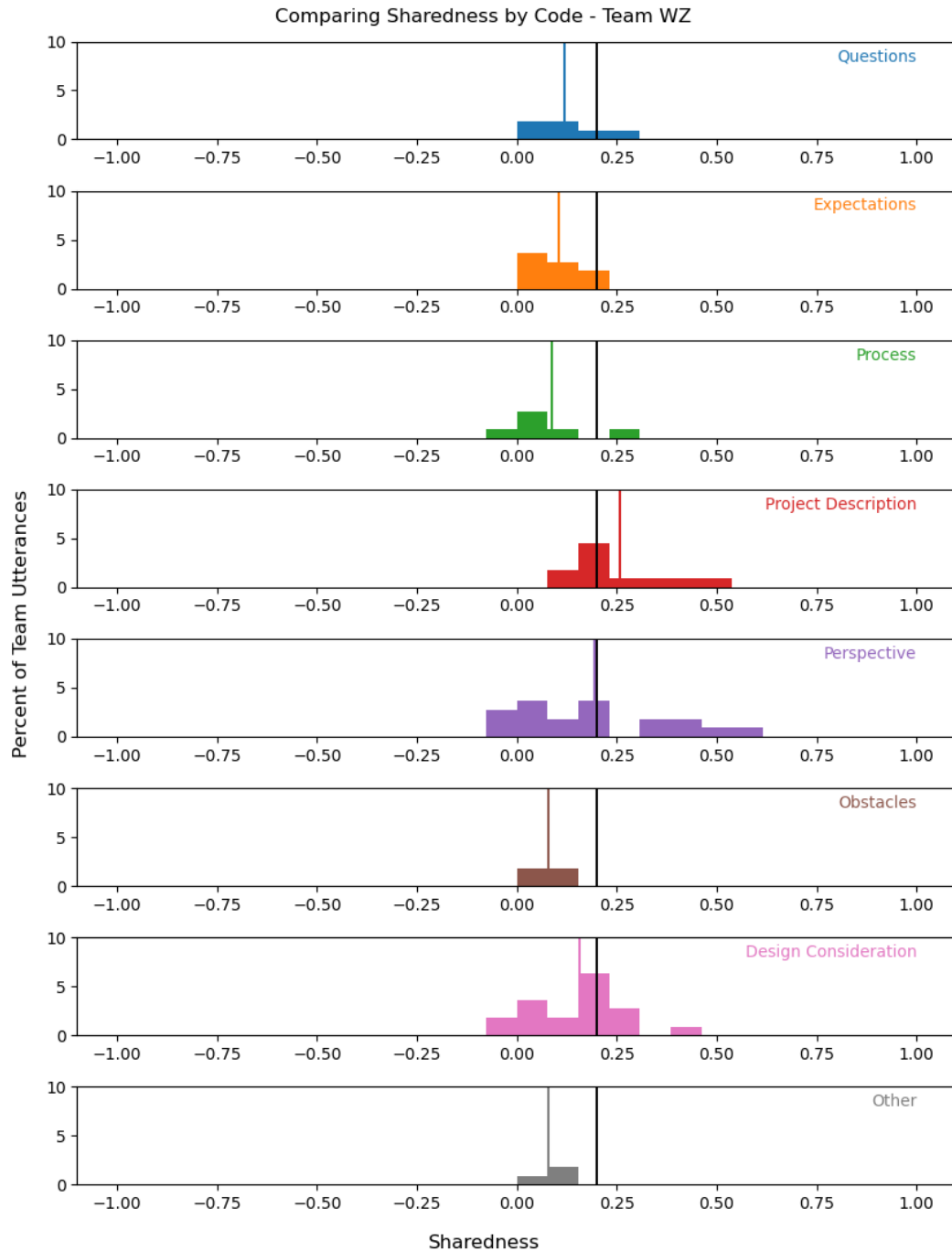


Figure 4.5 - Histograms showing the distributions of sharedness values for each code (Team WZ).

In response to the first iteration of the repeated prompt, P1*, Team WZ's overall average sharedness was 0.149. Average sharedness of Project Description utterances, 0.258. The team offers related, but distinct explanations of what the project goals are, including focuses on

comfort, (*"Our project needs us to find solution to the discomfort caused by the [redacted] when worn for long periods of time"*, Project Description, sharedness=0.143, SP22_22009), python-compatible scanning technology (*"My current understanding of the design problem is that we will be creating a python or related code that works with a smart phone scanning technology interface"*, Project Description, sharedness=0.248, SP22_48390), and a combination of scanning and mass production through 3D printing (*"By utilizing a 3D scanning technology with a Python-based plugin, customized, 3-D printed models of [redacted] will be mass produced by scanning [redacted]"*, Project Description, sharedness=0.384, SP22_68263).

In response to the second iteration of the repeated prompt, P4*, overall average sharedness dropped slightly to 0.128. Average sharedness of Project Description utterances was 0.200. Similarly to the first iteration, the team appears to describe the goal in related but distinct ways, (*"The goals of the project are to digitalize the customisation of [redacted]"*, Project Description, sharedness=0.217, SP22_22009), (*"Our project aims to make a design that makes an impression of [redacted] using 3D scanning software feasible for mass customization"*, Project Description, sharedness=0.159, SP22_48390), (*"The goals of this project are to create an [redacted] customized to each customer with the ability to be created in high volumes"*, Project Description, sharedness=0.227, SP22_68263). These project description utterances provide general overviews of the problem with varying focus, but do not provide much detail.

In response to the third and final iteration of the repeated prompt, P7*, overall average sharedness was 0.210. Average sharedness of Project Description utterances was 0.289. The team's descriptions of the project are still vague in some cases, (*"The goal of the project was to digitalize the process of creating customised [redacted]"*, Project Description, sharedness=0.222,

SP22_22009), (*"Our project involves automating the customization of personal consumer [redacted]"*, Project Description, sharedness=0.404, SP22_68263), and more detailed in others, (*"The main goal of this project was to evaluate the effectiveness of digital manufacturing for customized [redacted], but also general customization to a lesser degree"*, Project Description, sharedness=0.206, SP22_48390), (*"Specifically, the goals were to create a computer plugin that can automate the majority of the printing process, to find biocompatible materials, to make and evaluate prototypes, and to find a good 3D scanning application"*, Project Description, sharedness=0.501, SP22_48390). Several of these project description utterances include a new focus on “automation” that wasn’t explicitly discussed in prior weeks. However, some team members use the term differently, (*"Users will be able to scan [redacted] and receive [redacted] that involves very little automation on [sponsor’s] part"*, Project Description, sharedness=0.113, SP22_68263).

Only two members of the team responded to survey questions, and neither provided additional information in the optional open-ended response questions (see Figure 4.6). However, these two team members responded positively to the majority of questions. They agreed that their team had a shared understanding of the project, but agreed or had no opinion as to whether their understanding had changed since the start of the semester. The team reports that they collaborated reasonably well (average score of 8.5 out of 10) and rated their communication both amongst team members as well as with the project sponsor highly (average score 9 out of 10). This team’s Faculty Advisor was also their project sponsor so there were fewer stakeholders to

Team WZ

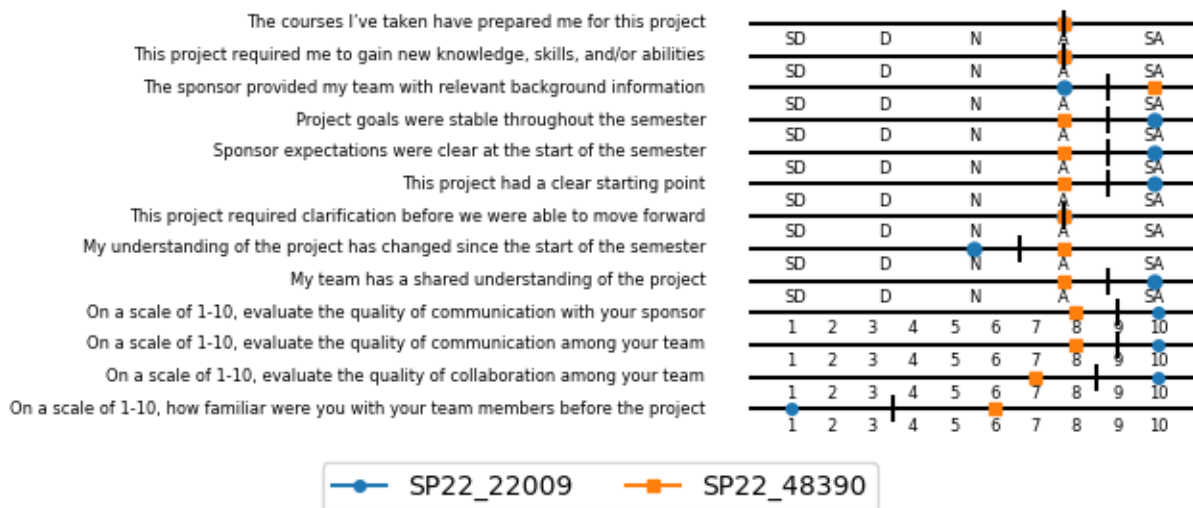


Figure 4.6 - Summary of Team WZ's multiple-choice survey responses

accommodate for this project, and this team did not have to reconcile differences between various stakeholders.

Throughout the semester, the team noted the importance of the coding aspects of their project. However, this presented unique challenges, as only a few members of the team had relevant experience, (*"The challenges we faced were putting the computer plugin together, especially in a timely manner"*, Obstacles, sharedness=0.142, SP22_48390) (*"This, being the main portion of our project, was the most difficult to complete when only one-two of us were able to understand what was needed to get the plugin working"*, Obstacles, sharedness=0.082, SP22_48390), (*"With prior experience in Python automation, I have a bit of knowledge of what is expected of this group"*, Perspective, sharedness=0.203, SP22_68263), (*"I want to learn more in the coding aspect as I have the least experience in it."*, Perspective, sharedness=0.128, SP22_22009), (*"I'm not comfortable with the coding aspects, but the other deliverables seem to be up my area of expertise"*, Perspective, sharedness=0.175, SP22_48390)

This theme appeared again at the very end of the semester, with students indicating that they had different areas of responsibility within the group (*"I didn't work as much with the rest of the project qualifications, mostly report writing and the above mentioned portions of the project"*, sharedness=0.125, SP22_48390), (*"From the group bonding experience and working with people of various backgrounds, I was able to figure out who needs to do what and when"*, sharedness=0.012, SP22_68263), (*"It was important to know who could be depended on when a task needed to be done"*, sharedness=-0.019, SP22_68263). Several utterances also mentioned learning about coding in a group, (*"I was able to learn a lot more about plugins and coding in groups, something I have never really experienced"*, sharedness=0.126, SP22_68263), (*"It was interesting to see how other people code and I wanted to gauge or improve my coding."*, sharedness=-0.040, SP22_68263).

Team WZ's average sharedness dips from 0.149 in week two, to 0.128 in week six, then increases to 0.210 in week fifteen. Similarly, the Project Description utterances' average sharedness decreases from 0.258 in week two, to 0.200 in week six, then increases to 0.289 in week fifteen. This team had the lowest overall average sharedness, as well as the lowest average sharedness in each utterance type. Their project required a lot of coding, and only a limited number of team members had the requisite programming experience. Other members of the team mentioned that they helped the team in other ways, like report writing. Low sharedness overall could be indicative of this specialization.

Case 3: Optomechanical Team (EK)

Team EK (Optomechanical Project) had 6 students total, 6 consented. Of these, four responded to demographic questions. This team included SP22_98921, SP22_27826, and

SP22_04626 (White or Caucasian, Female), and SP22_03268 (Asian or Pacific Islander, Female). This project was sponsored by a national lab as part of an established capstone program where the same project prompt is given to capstone teams at a number of schools. This project was a year-long project, the students in the Spring semester had the option of continuing the work started by those in the Fall. The final deliverable for this project was a technical report and working prototype where applicable. The purpose of this project was to explore optomechanical systems for use in space.

With six team members participating in the study, Team EK had a total of 110 utterances submitted in response to the repeated prompt throughout the semester (see Table 4.12). Of these, the highest percent were Perspective utterances (26.4%), followed by Design Considerations (23.6%), then Project Descriptions (12.7%). This distribution of utterance types is typical of the teams in the sample. Team EK had the highest overall sharedness of any team with an overall average sharedness of 0.234. Their highest sharedness utterance type was Project Description with an average sharedness of 0.465, which is higher than the overall average Project Description sharedness for all teams, 0.370.

Table 4.12 - Team EK Sharedness Statistics by Utterance type				
Utterance type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	9 (8.2%)	0.308	0.278	0.200
Expectations	10 (9.1%)	0.136	0.143	0.092
Process	11 (10.0%)	0.111	0.182	0.151
Project Description	14 (12.7%)	0.474	0.465	0.141
Perspective	29 (26.4%)	0.142	0.162	0.103
Obstacles	8 (7.3%)	0.128	0.109	0.059
Design Consideration	26 (23.6%)	0.260	0.278	0.167
Other	3 (2.7%)	0.141	0.178	0.090
All Combined	110 (100.0%)	0.180	0.234	0.173

Once again, Mann-Whitney U-tests can be used to compare each group of coded utterances to the overall group to determine which have high or low sharedness (see Table 4.13). For Team EK, Project Description ($p < 0.001$) and Design Consideration ($p = 0.011$) utterances had significantly higher sharedness than the overall group. Obstacles utterances had significantly lower sharedness than the overall group ($p = 0.045$).

Table 4.13 - Mann Whitney U-Test by Utterance type – Team EK		
Utterance type	Less than	Greater than
Questions	U=3561.5 p=0.839	U=3561.5 p=0.161
Expectations	U=2621.0 p=0.127	U=2621.0 p=0.874
Process	U=3234.5 p=0.258	U=3234.5 p=0.743
Project Description	U=8152.0 p>0.999	U=8152.0 p<0.001
Perspective	U=8513.5 p=0.146	U=8513.5 p=0.855
Obstacles	U=1729.0 p=0.045	U=1729.0 p=0.955
Design Consideration	U=10929.0 p=0.989	U=10929.0 p=0.011
Other	U=960.5 p=0.458	U=960.5 p=0.543

The distributions of sharedness values for each utterance type are shown in Figure 4.7.

Note that there are relatively few low sharedness Project Description utterances. Compare to Design Considerations which covers a similar range of sharedness values but has more utterances with lower sharedness than Project Description. Similar to the overall distribution from all teams combined, the Perspective utterances from Team EK seem to have more tightly clustered sharedness than other utterance types.

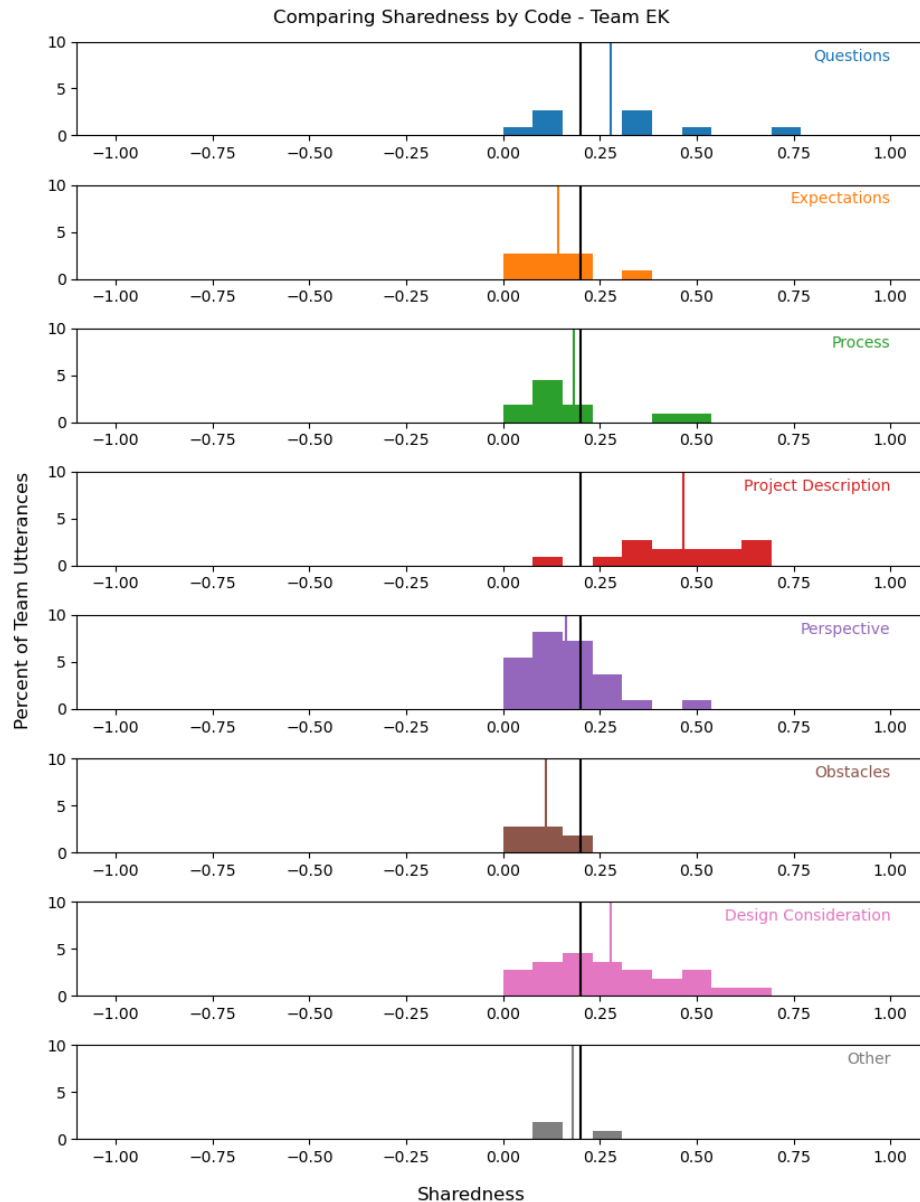


Figure 4.7 - Histograms showing the distributions of sharedness values for each code (Team EK).

In response to the first iteration of the prompt, P1*, Team EK's overall average sharedness was 0.298. However, average sharedness of Project Description utterances was much higher at 0.517. Most descriptions appear to center around the goal of viewing space debris, ("[...] looking at the materials we currently have, the goal of the project seems to be creating a

prototype of a telescope-like structure that would view space debris", Project Description, sharedness=0.629, SP22_03268), (*"The goals are to view space debris and it brings up questions of what is space debris, how will we validate that it works in a zero gravity vacuum, what material would be best to sustain -40 to 70 degree temperatures while keeping our product under 5 pounds, how and when will we acquire lenses, mirrors, and sensors"*, Project Description, sharedness=0.592, SP22_98921). Others focus on the structure itself, (*"I think the ultimate goal of this project is to prototype this structure with those design constraints and do some experimental testing or computational testing to make sure the structure can withstand vibrations"*, Project Description, sharedness=0.442, SP22_04626). While the team refers to a “baffle” they admit some confusion about exactly what that means, (*"Our design project will consist of the design and fabrication of a lightweight, robust, deployable (in a single axis), and reliable optical alignment structure (ie lens mount) featuring a deployable baffle for viewing of space debris"*, Project Description, sharedness=0.554, SP22_65144), (*"We are currently most confused by the baffle, what it is how it works etc [...]"*, Design Consideration, sharedness=0.504, SP22_98921). The team also mentions that their project is a continuation of a previous team’s work, (*"I have been assigned to the Optical Alignment Structures project, which is a continuation project from last semester's senior design"*, Project Description, sharedness=0.370, SP22_03268). Despite this, the team still expresses that they have little information to start with, (*"Unfortunately, the project description we were given is relatively vague and it's difficult to gauge exactly what we will be doing this semester without knowledge of what the previous team has already accomplished"*, Perspective, sharedness=0.129, SP22_03268), (*"We don't know much about our design problem yet"*, Perspective, sharedness=0.080, SP22_98921).

In week six, the team responded to the second iteration of the repeated prompt, P4*. Utterances from these responses had an overall average sharedness decreased to 0.236. Average sharedness of Project Description utterances however increased to 0.577. The team mentions some conflicting perspectives from the Faculty Advisor and Sponsor of the project, (*"In terms of challenge, the most difficult thing currently has been balancing the different expectations from our faculty advisor and corporate sponsor"*, Obstacles, sharedness=0.118, SP22_03268), (*"My biggest uncertainty with the project is that our faculty advisor continues to try changing the scope of our project every time we meet with him"*, Obstacles, sharedness=0.137, SP22_98921). Despite this obstacle, the team's understanding of their problem appears to have narrowed from disparate descriptions of viewing space debris to a more specific structure that meets establish requirements, (*"So far, we have decided that design will consist of three subsystems (lens mount, baffle design, and baffle deployment)"*, Design Consideration, sharedness=0.382, SP22_03268), (*"Our design problem is to design an optical alignment structure with a deployable (in single axis) baffle which can view debris in space at a certain distance of [redacted]"*, Project Description, sharedness=0.636, SP22_65144), (*"Our goals are to create a viewing apparatus with two lenses that will survive the environment of low-earth orbit, be able to extend and retract multiple times, and successfully block out enough stray light to see clear images of space debris"*, Project Description, sharedness=0.507, SP22_27826), (*"The goals of this project are to design and manufacture an optical alignment structure that can view space debris from [redacted] meters away with a positional accuracy of [redacted]"*, Project Description, sharedness=0.634, SP22_04626), (*"The goals of our design problem are just to view space debris"*, Project Description, sharedness=0.532, SP22_98921). Notably, the team seems to have

an increased understanding of the specifics of the project, including the “baffle” which was previously mentioned as a point of confusion, however, new information brings new questions, (*"We are currently struggling with how to angle the baffle with respect to the sun and how sun angle and internal baffle raft angle will affect the amount of light let in – how we can choose and calculate that"*, Design Consideration, sharedness=0.325, SP22_27826).

In the third iteration, P7*, in week fifteen overall average sharedness dipped to 0.177. Average sharedness of Project Description utterances dropped to 0.323. This decrease in sharedness is evident in the differing ways that the team members described their understanding of the project. Some team members described the project summarily at the end of the semester, (*"We have been challenged to design an optical alignment structure that would be mounted on a satellite for the viewing and monitoring of space debris that is moving parallel to the satellite"*, Project Description, sharedness=0.377, SP22_03268), (*"Our goal changed a few times, but it ended up being to just create a device (that can be attached to a black box of controls) which holds lenses to look at space debris"*, Project Description, sharedness=0.307, SP22_27826). Others focused on a specific aspect of their solution, (*"Our goal for the semester was to get a prototype made out of metal instead of just 3D printing, and we successfully did that"*, Project Description, sharedness=0.130, SP22_98921). While there was some variation in project descriptions, some students provided similar explanations of their project goals, (*"Our goals were to design a baffle system that can deploy once in space and hold a pair of lenses a set distance apart from each other"*, Project Description, sharedness=0.382, SP22_18040), (*"The goals of the project are to create an optical alignment structure with two lenses that has a hands*

off, deployable baffle and can view debris at a distance of [redacted] meters away", Project Description, sharedness=0.422, SP22_04626).

The team's average sharedness dips from 0.298 in week two, to 0.236 in week 6, then decreases again to 0.177 in week fifteen. However, the Project Description utterances' average sharedness increases from 0.517 in week two, to 0.577 in week six, then drops significantly to 0.323 in week fifteen. Despite this project being offered as a continuation of work done in the previous semester, Team EK did not have access to the previous team's documentation at the start of the semester. Thus, they started with limited information available from the initial project description. However, unlike Team CP's project, this project was well-defined by the sponsor. Regardless, the team still underwent a process of clarifying project goals as they decided what to use from the previous team's work, as well as reaching a consensus about the project scope with the Faculty Advisor. By week six, the team's Project Description utterances had higher sharedness, but overall sharedness decreased. This suggests that while the team's understanding of the problem improves, as shown in this quote, (*"I definitely have a better understanding of this project than the last time I answered this question"*, Perspective, sharedness=0.037, SP22_27826), they begin to move into increasingly specialized areas with additional unique, and therefore less shared, questions and uncertainties, (*"I'm reasonably comfortable with the project now that I have a defined role and a specific component to focus on"*, Perspective, sharedness=0.182, SP22_65144). This seems to continue to week fifteen, where both Project Description and overall sharedness have decreased. Additional insight can be gained by considering the team's responses to the survey questions and other reflection prompts.

A summary of Team EK's multiple choice survey responses is provided in Figure 4.8.

The team agreed to most questions, with a few exceptions. There was disagreement as to whether the sponsor provided relevant background information. Some context is given by this response from SP22_03268, *"Our contact at [sponsor] was very helpful for getting us as much information as we needed, though there were some specifics that he was unable to provide."* The team was relatively neutral overall on the stability of project goals, the clarity of sponsor expectations, and the clarity of the project starting point. They agreed that the project required clarification and that their understanding has changed, which we can see in these quotes, *"A lot of objectives changed, and our understanding of what deliverables we had to provide to the sponsor changed"*, SP22_27826, and *"We weren't sure in the beginning whether to continue where last semester's group had ended or start the project over. This took a few weeks for us to decide"*, SP22_03268. Despite this, the team agreed that they had a shared understanding of the project, *"Even with the misunderstandings at the beginning, we were all pretty much at the same level of understanding amongst the group members throughout the whole project"*, SP22_27826.

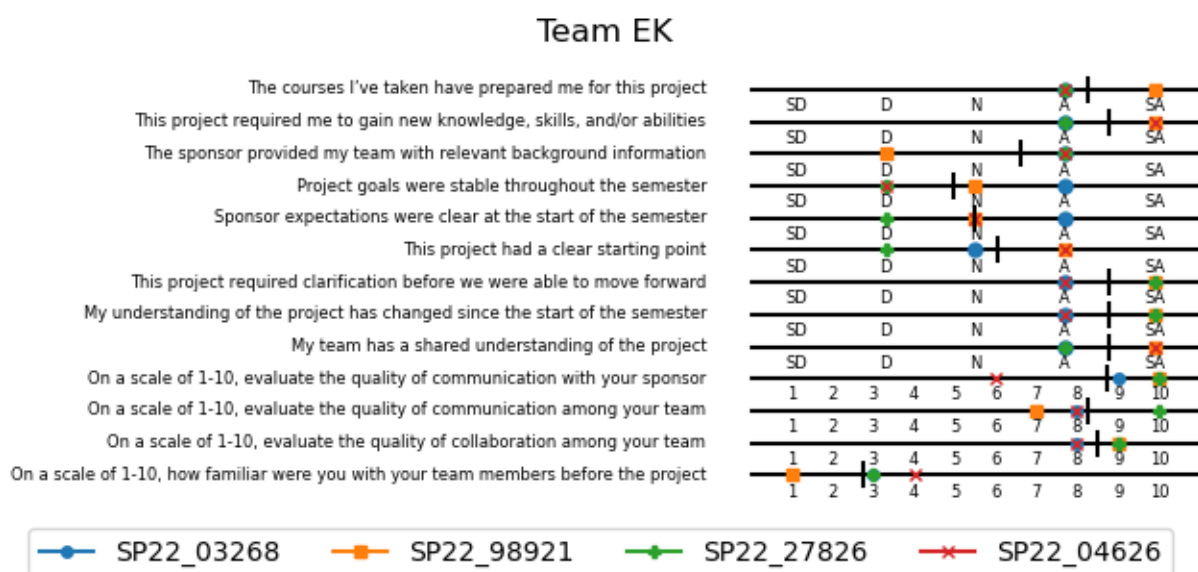


Figure 4.8 - Summary of Team EK's multiple choice survey responses

The team rates communication with their sponsor relatively high, with an average of 8.75. They give some insight into this relationship in responses to P2 and P3. (*"Our sponsor meeting went quite well, our sponsor was friendly and answered the questions we had that he was able to"*, None, sharedness=0.296, SP22_27826), (*"We left the meeting feeling like we had a good idea of his expectations for us and felt confident we could do a good job with the project"*, None, sharedness=0.183, SP22_03268). They also noted the clarity and stability of project objectives as provided by the sponsor, (*"Since this is a national competition within [sponsor], we are working on the same project as many other universities across the country"*, sharedness=0.203, SP22_98921), (*"Because of this, our size, weight, lenses, and other aspects have specific constraints"*, sharedness=0.107, SP22_98921), (*"The scope with [sponsor] has been very clear this entire time"*, sharedness=0.118, SP22_98921).

However, the team faced significant challenges with differing expectations from their different stakeholders, the sponsor and faculty advisor. (*"In terms of challenge, the most difficult thing currently has been balancing the different expectations from our faculty advisor and corporate sponsor"*, Obstacles, sharedness=0.118, SP22_03268), (*"My biggest uncertainty with the project is that our faculty advisor continues to try changing the scope of our project every time we meet with him"*, Obstacles, sharedness=0.137, SP22_98921). Their faculty advisor seems to have suggested a change in the scope of the project as the team explains in week 4 in a response to P3, (*"During the meeting, [Faculty advisor] presented an expansion to the scope of our project, which was an interesting topic but we were a bit worried about pursuing this and losing track of our original scope that our sponsor wanted"*, sharedness=0.463, SP22_03268), (*"Then on Monday we met with our faculty advisor, [redacted], and he had a very different*

scope for us", sharedness=0.339, SP22_98921), ("Our faculty advisor meeting was a little confusing, he seemed to have ideas for our project that were way outside the scope defined by the sponsor", sharedness=0.518, SP22_27826), ("The meeting with my FA resulted in getting conflicting information on the project scope for the optical alignment structure, with him suggesting we try to make the telescope auto-focus and use machine learning and AI to develop image classification software", sharedness=0.398, SP22_04626).

Overall, the team appears to have collaborated well together, rating themselves with an average score of 8.5, with an 8.25 in communication. We can see this dynamic at play in these responses, noting that even in teams with self-reported good communication, some friction should be expected, (*"I think our team has worked quite well together over the course of the project, and we never had any huge disagreements as a group, which made the process of working through difficulties a lot easier."*, Perspective, sharedness=0.281, SP22_27826), (*"I do feel like I do most of the work in the group, as my other group members are just very bad at communicating"*, Perspective, sharedness=0.041, SP22_98921), (*"Our team is confident with one another and have been very communicative and reliable working with each other"*, sharedness=0.104, SP22_18040).

In response to P6, Team EK mentioned how they used a rough prototype to facilitate building a shared understanding, (*"Early after our design decision, I created a low fidelity prototype out of cardboard"*, sharedness=0.217, SP22_98921), (*"This helped us to visualize the product and how the baffles would work"*, sharedness=0.080, SP22_98921), (*"This also helped me show my group that a rack and pinion motor would be a lot more complicated than they previously thought"*, sharedness=0.073, SP22_98921), (*"No one understood until my initial*

prototype that the rack would have to stick out pretty far from the design in order to deploy cylinders out to it", sharedness=0.364, SP22_98921). Their team member explained how this helped improve their understanding, ("One of our members made a rough concept prototype in the beginning of our process to better understand how all the different components of our design would work together", sharedness=0.360, SP22_03268), ("That was really helpful for visualizing what it would look like for all of us", sharedness=0.040, SP22_03268). This ultimately helped facilitate communication, even with other stakeholders ("I think these were useful in helping us communicate our ideas so that everyone was all on the same page, and will be useful for our presentation next week when we are explaining our design", sharedness=0.098, SP22_03268).

CHAPTER 5: DISCUSSION

Applying Team-Task-Design Mental Models Framework

The Team-Task-Design Mental Models framework can be used to explore the sharedness values of different types of utterances. A possible mapping between the types of utterances in the qualitative codebook and the different types of mental models is shown in Table 5.1.

Table 5.1 - Mapping Utterance Types to Mental Models Framework		
Utterance type	Description	Mental Model*
Question	Any question asked. These are mainly clarifying questions about the project. They may be specific, or general	Any
Expectation	Anticipated challenges, expected outcomes, speculations about what might occur	Task-related
Process	Descriptions of what has been done, is currently being done, or what will be done to proceed through the design process	Task-related
Project Description	Overview of the project, may include project goals or objectives	<i>Design-related</i>
Perspective	The individual perspective, how they feel about the project, their team, their advisor, or TA	Any
Obstacle	Obstacles which may impede progress. These are high-level challenges (e.g. advisor and sponsor disagree on scope, scheduling issues, supply chain issues), and should be concrete (potential or anticipated issues may be better coded as Expectations)	Task-related
Design Consideration	A project-specific detail, should be content related (not process related)	<i>Design-related</i>
Other	Utterance does not fit in the above categories or has unidentifiable meaning. Includes sentence fragments due to poor parsing of utterances, unless that fragment can be easily attributed to another code	n/a

Unfortunately team-related mental models were not well addressed by the repeated reflection prompt. However, responses to reflection prompt P2 may provide some additional

insight into team-related considerations (see Table 3.4). For example, this sequence of utterances from Team CP which shows changes in a team-related mental model, (*"I expected to receive more vision and advice regarding the path and deadlines for our project"*, sharedness=0.237, SP22_72514), (*"I think the FA meeting differed from these expectations a bit while the TA meeting was more aligned"*, sharedness=0.248, SP22_72514), (*"Both went well, but the FA meeting presented itself as much more of a technical resource for our project while the TA was the aforementioned guidance for the class structure"*, sharedness=0.255, SP22_72514), (*"Both still went well, but my opinions and expectations are certainly restructured moving forward"*, sharedness=0.098, SP22_72514).

Task-related models would include Process, and Obstacle utterances as these pertain to local task-procedures and likely scenarios (see Table 3.1). Process utterances may also be related to a design-related Process model to the extent that the team is relying on some end-to-end design process. However, process utterances in this study typically described next steps, rather than an overall plan. Expectations would also fall under task-related mental models, though expectations related to one's team, including stakeholders, would be team-related. Team-related expectations were less common than task-related expectations in this study.

Finally, Project Descriptions and Design Considerations would be part of the design-related mental models. These utterances refer to both the Object and Need models to varying degrees at different points in the design process. At the start of the project, the Object Model is not well developed, so Project Descriptions are more likely to be driven by the Need model. As the Object becomes more refined and additional effort is focused on further developing the object, Project Descriptions and Design considerations may become more associated with an

Object Model. For example, consider these two Project Description utterances from SP22_72514 on Team CP, one from week two of the semester, (*“The goal of the project is to analyze discrepancies between a ground true speed sensor and the vehicle’s speed”*, Project Description, sharedness=0.432, SP22_72514), and the other from week 15 (*“The goal of the project is to design an algorithm that minimizes the effect of slip on a [vehicle] over a variety of different terrains by not allowing the vehicle to go off course by a specific threshold and maintaining a speed within a predetermined range, as well as provide the accompanying sensor suite”* Project Description, sharedness=0.540, SP22_72514). In week two, the focus is on the Need, with little detail about the Object designed to address it. In week fifteen, the focus is on the Object, in this case, an algorithm. However, note that at the end of the semester, the Object and Need are related such that a description of the Object which satisfies the need cannot be separated from a description of the Need itself. Regardless of the weighting of Object versus Need represented in these utterances, Project Description and Design Consideration utterances align more closely to an Object Model or Need Model than to a task-related or team-related mental model.

In general, I found that both Project Descriptions and Design Consideration utterances tended to have higher sharedness than other types of utterances, with Project Descriptions the more shared of the two. Further, the utterances more closely associated with task-related models, Process, Expectations, Obstacles had low or average sharedness. This suggests that design-related models may tend to have higher sharedness than task-related models. This would make sense given the allocation of different tasks to individuals or sub teams, and would help explain the result that task-related mental models are not always similar in successful teams (Bierhals et al., 2007; Smith-Jentsch et al., 2005). Design-related mental models are expected to be related to

multiple sub tasks throughout design and would likely be more central and therefore sharedness would be more important for the team's success. The following sections explore the findings as they relate to each research question and further demonstrate how the Team-Task-Design Mental Models Framework can be used to understand these results.

Research Question 1

RQ1) In what ways can Latent Semantic Analysis be used to understand the shared content of open-ended reflections? What are the limitations of such analysis?

How do reflections differ from other LSA data collection methods?

Reflections are not bound by project duration, unlike transcripts which typically only cover short design sessions. Similarly, reflections are not tied to specific project phases or milestones, unlike project deliverables, so they can be collected at any time. Reflection prompts can also be repeated as needed, so that changes in responses can be measured over time.

Reflections are not primarily intended to communicate design progress to stakeholders, unlike project reports and can therefore be modified to suit instructional or research purposes, distinct from the stakeholder needs. Reflections have the benefit of hindsight. Remembering what happened and reflecting on it may provide a natural synthesis, highlighting specific aspects and presenting them in a cohesive way. In this way, reflections can provide more narrative structure than might otherwise be obtained through project documentation like design notebooks. In general, reflections, project documentation, and project deliverables are complementary resources. They offer different windows into the design teams' mental models and can be used in conjunction with one another. Like good project documentation, engaging in reflection can be beneficial to design teams directly. Reflection has long been seen as an important step for

learning processes, for example in Kolb's Experiential Learning Model (1984) or Schön's reflection-in-action (1992). Reflection prompts can offer useful direction for reflective thinking for students and professionals alike (Turns et al., 2014).

The unit of analysis for LSA is also important, and would differ for reflection responses, transcribed data, or project documentation. Generically, LSA can be applied to any collection of text, but the choice of where the boundaries fall can be important. In the reflection prompts for this study, students were asked to submit 150-word responses to elicit sufficient information, but few responses actually met this length. Nonetheless, each response contained many separate points of discussion that arguably represented separate elements of their mental model. I chose to subdivide reflection responses to better capture the variation within each response as a whole. Sentences would be a natural step down from the roughly paragraph long reflection responses, but not everyone submitted their responses in prose. Thus, I borrowed the term "utterance" from the use of LSA with transcriptions of spoken utterances as it suitably described the sentence fragments and unfinished thoughts that I encountered, even though no one was speaking these utterances aloud. However, due to the necessarily large size of the corpus, parsing each individual utterance was not feasible. Thus, a python script was used to separate responses by sentence ending punctuation, and edge cases were dealt with manually. The size of the "document" is also important for determining the neighborhood within which words are contextualized and therefore, the semantic meaning of those words. An exploration of the sensitivity to this parameter would be beneficial.

How is sharedness defined and calculated in this context?

Sharedness was defined at the utterance level. The sharedness, S_3 , of an utterance is its similarity to a composite representing all utterances from teammates in response to the same prompt. This is different from the “concept centroid” described by Dong as the average of all documents belonging to the team (2005). I have shown that sharedness, S_1 , computed as the cosine similarity between an utterance and this concept centroid is artificially increased due to self-similarity. Since the utterance being compared also contributes to the centroid, there is always a component of the centroid which increases similarity. I suggested two modifications to the composite, S_2 , removing just the utterance being compared, and S_3 , removing all utterances belonging to the speaker. While an improvement over S_1 , S_2 still had flaws related to self-similarity. The best definition of sharedness was S_3 as described above. The calculation for S_3 assuming a five-person team is given in Eq. 4, where each team-member’s utterances are represented by a vector, $\vec{A}, \vec{B}, \vec{C}, \vec{D}, \vec{E}$, taking the form given in Eq. 1.

How should LSA sharedness be interpreted?

For this study, mean sharedness was ~ 0.2 , with low sharedness generally referring to values < 0.2 , and high sharedness generally referring to values > 0.2 . For specific subgroups, the Mann-Whitney U-test was used to determine if the sharedness of the subgroup was significantly higher or lower than the overall group. As sharedness is computed as the cosine similarity between an utterance and a composite vector representing the sum of teammate’s contributions, sharedness increases if there are more members of the team with whom content is similar. Thus, if only two members of a team share some element of their team mental model, the sharedness value will be low even if the degree of sharedness between those team members is high. The

cosine similarity from LSA used to compute sharedness is a measure of the semantic similarity of the documents (*aka* utterances). The semantic meaning is derived from the corpus, so the fidelity with which semantic similarity matches similarity as we would define it outside of the context of LSA is determined in part by the completeness of the corpus. However, even small corpora have been sufficient in some cases (Fu et al., 2010). It is my understanding that a bigger corpus, containing more documents, contributes to more accurate results. Although, if the corpus contains documents from disparate speech communities, where terms are used differently, the semantic meaning of such terms will be spread across their uses, which may complicate interpretation. For example, the corpus used in this study was drawn exclusively from mechanical engineering capstone students, therefore the use of the word “design” has a particular meaning in this context. If the corpus also included documents from students in graphic design or architecture, the meaning of “design” as established through LSA would be more generalized as it would contain contributions from multiple, differing perspectives. Therefore, it is not just the size of the corpus, but also its scope that will determine the semantic meanings of terms and thus the similarity of documents. Additionally, the number of dimensions in the k -reduced semantic space will have some bearing on the results of LSA. While potentially important, these factors have not been optimized or even fully explored in the context of design teams, and it appears that many studies using LSA report good performance even without tuning the parameters of the algorithm. Nonetheless, I believe care should be taken when mapping the similarity or sharedness as assessed by LSA to similarity or sharedness defined elsewhere.

Notably, LSA’s semantic similarity and the measure of sharedness that comes from it, are different than conceptual agreement. Two statements which may be in direct contradiction, can

have high similarity assessed through LSA if they are covering the same topic, as shown in Example A in Table 3.8. The extent to which this is a problem is debatable. In reality, some team members may adopt shared terminology despite disagreeing with the premise or having flaws in their understanding. Sharedness is independent of accuracy. LSA alone doesn't indicate that something is correctly understood, to the extent that a "correct" way exists. Additionally, ideas expressed in reflection responses are not a one-to-one representation of the person's inner mind. Something may be left unstated, or expressed in ways that are contradictory to one's true beliefs. There are even scenarios where such behavior may be advantageous, and might therefore be reasonably expected, such as in the midst of negotiations in an effort to move past contentious, but otherwise insignificant details. Underlying disagreement may not prevent those teammates from being on the same page generally. Ambiguity may sometimes be preserved intentionally, with team members using the same terminology to refer to something generally without having to specify and agree on each detail. Shared language alone is not always an indication of shared understanding, as people may use the same words to mean entirely different things. Fortunately, sharedness determined through LSA is not an indication of the use of the exact same word or phrase in several utterances, but that the utterances include terms which have similar latent semantic meanings. Recall that in LSA, utterances are represented by vectors whose elements roughly correspond to the latent semantic concepts that emerge from the stochastic use of terms in context. This means that LSA is robust in the face of complications of polysemy such as homonyms and homographs, for example being able to distinguish between "charge" as in "the bull charged the matador", as opposed to "the battery's charge was depleted" (Landauer et al., 1998). This should mean the determination of sharedness through LSA is not especially

vulnerable to the misattribution of high sharedness to team members who use the same terms, differently.

Research Question 2

RQ2) *What are the typical elements of a collaborative design team's shared understanding of a design problem?*

What types of utterances were identified?

Responses to the repeated reflection prompt were coded through qualitative data analysis to provide insights into the types of considerations that students included in their responses, and with what frequency. Recall the repeated prompt was, *Describe your current understanding of your design problem. Be sure to include answers to the following questions: What are the goals of the project? What challenges do you anticipate? How comfortable are you with this project? What questions and/or uncertainties do you have? If there is anything else you'd like to mention, please do.* A subset of utterances was used to identify an emergent coding scheme to descriptively label each utterance according to its contents. This coding scheme possessed eight codes: Question, Expectation, Process, Project Description, Perspective, Obstacle, Design Consideration, and Other, defined in Table 3.7.

What was the most common type of utterance?

The most common utterance type was Perspective at 24.5%, followed by Design Considerations at 23.2%. The least common utterance type outside of Other was Questions at 3.5%. Since the reflections are submitted individually and specifically prompt the respondent to describe their current understanding of the design problem, it follows that the most common utterances were related to the individual's perceptions of how things were going (Perspective)

and the specifics of the design problem at hand (Design Considerations). Project Descriptions represented a lower percentage of utterances, at 15.2%, than Design Considerations, at 23.2%. Perhaps this difference is due to their position as high-level descriptors of the project and/or its goals. Project Description utterances serve as the “headline” description of the project, whereas Design Considerations relate to all the details and specifics. Naturally the headlines take up less space than the more nuanced design considerations. Process, Expectations, and Obstacles are all somewhat related to the tasks the team is planning to undertake, as well as the circumstances informing those decisions. Discussing how the team aims to solve the problem rather than discussing the problem or solution directly is a somewhat metacognitive stance. These categories representing a lower percentage of utterances than those discussing the problem or solution directly may suggest that students are spending less time considering their process than applying it. I think anyone who has spent time with beginning designers can recognize this behavior, and it is particularly well described in Crismond and Adams’ The Informed Design Thinking and Learning Matrix (Crismond & Adams, 2012). However, it is unclear what distribution of utterance types might be found in ideal circumstances, and it is possible that expert designers could have similar results to these undergraduate teams for any number of reasons. Finally, the rarest utterance types were Questions and Other. The Other category representing the smallest percentage of utterances is a good sign that the codebook reasonably covered the topics that show up in the sample. However, it is odd that there were so few Question utterances, especially since the prompt also directly asked respondents what questions/uncertainties they may have. It is possible that this is a case of some kind of desirability bias, where respondents are less likely to report things that they believe are undesirable traits. Since reflection responses were submitted

by students through a portal provided by the instructor, there may have been hesitance to be forthcoming with uncertainties or questions as students sought to present themselves as knowledgeable and competent. However, questions and uncertainties are a central aspect of design and thus we'd expect any designer, but especially beginning designers, to have many questions throughout the design process.

How are sharedness values distributed in the sample?

For each utterance a sharedness value was determined through LSA as outlined in Chapter 3. Sharedness has a theoretical range of -1 to 1, though negative values indicate a negative latent semantic relationship and are difficult to interpret practically. A sharedness of 0 represents no shared content, whereas a sharedness of 1 would indicate identical responses from all team members. The average sharedness for coded utterances ($n = 664$), those in response to the repeated prompt, was 0.205, with a standard deviation of 0.157. The utterance with the lowest sharedness was -0.05 and the maximum was 0.713. For reference, the average sharedness for all utterances in response to all prompts throughout the semester ($n = 4655$) was 0.184 with a standard deviation of 0.143. Positive sharedness indicates that teams have at least some baseline-level of shared understanding as represented through semantic similarity. There are at least two contributions to this baseline sharedness, firstly utterances are not random collections of unrelated words, they are semantically coherent responses to a prompt. Because of this, they should be expected to have some semantic similarity to other responses to the same prompt than random utterances solely due to conceptual alignment with the prompt. Secondly, responses from teammates are informed by the same project context. Since members of the team have a common set of shared experiences related to their project, their responses will draw from the same

experience pool and will refer to the same artifacts, players, and circumstances. Since sharedness is computed between an utterance and all other utterances submitted by the speaker's teammates in response to the same iteration of the prompt, these baseline effects should both be present. Thus, a positive average sharedness is expected. The majority (59.5%) of utterances have lower than average sharedness (< 0.2), whereas 40.5% have higher than average sharedness, falling between 0.2 and 0.7129.

How does sharedness vary by utterance type?

The utterance types which have the highest sharedness are expected to be indicative of the commonly shared elements of a design team's shared understanding as posed in RQ2. Two utterance types were found to have significantly higher sharedness than the overall group of all utterances as determined by Mann-Whitney U-Tests. These were Project Description (Average sharedness: 0.370, MWU Test: $U=51123$, $p<0.001$) and Design Consideration utterances (Average sharedness: 0.246, MWU Test: $U=61337$, $p<0.001$), see Table 4.3. This means that Project Description and Design Consideration utterances tended to include language which was the most semantically similar to that used by one's teammates of all utterance types. While unsurprising, this serves as a useful confirmation of what might otherwise be assumed. Project Description utterances contain an overview of the problem, potentially including project goals or objectives. Teams are sometimes characterized by their collaborative effort toward a shared goal, as opposed to a group or other collective which may not have a unified objective. We'd expect undergraduate design teams to be no different in this regard than teams in other contexts, so it would follow that the design teams in this study would also develop some level of shared understanding of their project objective.

Three utterance types were found to have significantly lower sharedness than the overall group as determined by Mann-Whitney U-Tests. These were Process utterances (Average sharedness: 0.161, MWU Test: $U=21128$, $p=0.010$), Perspective utterances (Average sharedness: 0.144, MWU Test: $U=42216$, $p<0.001$), and Obstacles utterances (Average sharedness: 0.115, MWU Test: $U=12961$, $p<0.001$). Utterances coded as Other also tended to have lower sharedness than the group, but these made up only 3.3% of all utterances. Perspective utterances from individual reflections might be expected to be individualistic rather than shared, so it makes sense that the sharedness value for these utterances tended to be low. The low sharedness of Process and Obstacles suggests that some elements of the task-related mental model are not being represented similarly in the reflections. Perhaps this is further evidence that task-related mental model sharedness is less significant in design teams than operational teams, potentially due to the specialization of design team members.

How does sharedness vary by team?

All teams attained a baseline shared understanding, as indicated by the positive overall average sharedness of 0.2. Larger differences in sharedness were found between teams when specific types of utterances were considered separately. This shows that while all teams have a shared understanding, the contents of their shared understandings differ. For the majority of teams however, Project Description and Design Consideration utterances were found to have higher sharedness than the overall group. Seven of eight teams' Project Description utterances were found to be stochastically greater than the overall group of utterances. However, Team LU's Project Description utterances were indistinguishable from the overall group. The sharedness of Project Description utterances varied from team to team, indicating that different

teams expressed different levels of sharedness in their understanding of project objectives, which is almost certainly related to varying project contexts. Design Consideration utterances followed a similar trend, though with lower sharedness on average. Five of eight teams' Design Consideration utterances were found to be stochastically greater than the overall group.

Design Considerations and Project Descriptions are pertinent to the design project as a whole rather than just a specific task. They contain the team's understanding of the design object and the unmet need that they are designing for, and as such relate to the Object and Need Models. As these considerations are important across many phases of design, and to many team members as well as stakeholders, their sharedness is more impactful. Conversely Perspectives and Obstacles were found to have lower sharedness than the overall group in five of eight teams as well. This shows that not all content is shared. Notably, obstacles clearly represent task-related mental model considerations so their low sharedness in this study further suggests that design teams may not always need shared task-related mental models.

How did teams explain the circumstances of their projects?

For the Vehicle Control Team (CP), communication with sponsor was poor with a missing project description at the start and wandering project goals throughout. In week two, the average sharedness of Project Description utterances was 0.359, but overall average sharedness was much lower, 0.166. Since the team was working with a small amount of information at the start of the semester, their descriptions of the project were similar, but other aspects of their understanding differed. Over time their project scope drifted, but they used a block diagram to "keep everyone on the same page" including in meetings with the faculty advisor and TA.

The Mass Customization Team (WZ) had the lowest overall average sharedness, and the lowest average sharedness in each utterance type. This project required a lot of coding, but only a few team members had relevant programming experience. Thus, other team members took on other tasks. The low overall sharedness could be indicative of the specialization of team members.

The Optomechanical Team (EK) had the highest sharedness and worked on a well-defined project that was part of a national program led by their sponsor. However, their faculty advisor disagreed with the project scope and repeatedly tried to suggest different project goals. The team was forced to reconcile different expectations from different stakeholders, but this required them to develop shared language to describe their system and objectives. They also mentioned the importance of a rough prototype to facilitate a shared understanding of how the components of the system interact.

What does this tell us about how shared understanding is developed?

Taken together, these insights suggest several factors which may influence the development of a team's shared understanding. First, multiple teams demonstrated high semantic sharedness in Project Description and Design Consideration utterances early in the project. This suggests that during early phases, when the goal is to understand the objectives and challenges, teams converge to some preliminary shared understanding, even when little concrete information is available. However, Project Descriptions and Design Considerations shift in focus and sharedness over time. Generally, sharedness was not found to increase over the course of a project. It seems that teams with unclear or moving project goals are forced to spend more time

and attention pinning down their objectives, which may result in higher sharedness. Whereas other teams specialize more, potentially yielding lower sharedness.

Secondly, teams noted the importance of prototypes and other representations or models of the system to facilitate communication and shared understanding, for example Team CP's block diagram or Team EK's cardboard prototype. Several authors have also discussed sketches (Goldschmidt, 2007) or prototypes which serve as boundary objects (Bucciarelli, 1994) or transition models (Smulders, 2007) and facilitate communication and understanding across boundaries by providing external representations. Examples of the importance of external representations, especially at boundaries and interfaces, can be found in systems as important and complex as the International Space Station. A team of designers for the ISS built a full-scale cardboard mockup during a meeting with Russian partners to determine details at the interface between US and Russian components, and cited it as a reason behind, "some of the most rapid gains" between the partners (Bacon, Stockman, & Boyle, 2010). It seems one of the best ways to facilitate a shared mental model is to create an external representation upon which all parties can agree and build.

Research Quality

As this research addresses social aspects of engineering design and relies upon the observation and interpretation of constructs situated within a broader context of design research, which is itself defined by socially constructed understandings, special attention must be paid to the considerations of validity and reliability which contribute to overall research quality. That is, what assurances can be provided that this work is grounded in reality? That it is compatible with, and responsive to related, established work and the broader views of the research community?

And that it authentically represents what was observed without undue influence and bias toward a particular perspective? To address these questions, I will use the Qualifying Qualitative Research Quality (Q³) Framework by Walther, Sochacka and Kellam (2015, 2013). The Q³ Framework distinguishes five types of validation, Theoretical Validation, Procedural Validation, Communicative Validation, and Pragmatic Reliability, with the latter four contributing to the overall Theoretical Validation of the work. In the following sections I will reflect on the work presented in light of these five types of validation in both the stages of *making the data* and *handling the data* as described by Walther, Sochacka, and Kellam (2013). A summary of the examples described in the following paragraphs is included in Table 5.2.

Table 5.2 - Summary of application of Q3 Research Quality Framework		
	Data Collection	Data Analysis
Theoretical Validation	• Open-ended reflections capture variation • LSA derived sharedness provides objective methodology • Exploratory and descriptive work seeks to present what was observed	• The Team-Task-Design mental model framework provides a conceptual model which is consistent with past work and lived experience
Procedural Validation	• Capstone design presents an established and realistic undergraduate design experience • Iterative study design	• Maximal variation sampling was used to select descriptive cases to ensure the spectrum of sharedness was represented • Emergent coding
Communicative Validation	• Open-ended reflections allow participants to voice their opinions • Descriptive case analysis further explores individual scenarios	• Iterative, emergent coding grounded researcher interpretations • Research contributions are situated within continuing research efforts
Pragmatic Validation	• Team-Task-Design mental model framework addresses known issues with SMMs in design	• Findings are immediately applicable to student, instructor, researcher, and practitioners of engineering design
Process Reliability	• Open-ended reflections capture variation in project circumstances • LSA provides a systematized numerical measure of sharedness	• Descriptive codebook and interrater reliability • Use of well-documented methods for computational analysis

Theoretical Validation asks, “Do the concepts and relationships of the theory appropriately correspond to the social reality under investigation?” (Walther et al., 2013). The prevailing concept in this work is that of “sharedness” as it manifests in collaborative engineering work, and alludes to the experience of being, “on the same page” as your teammates. To develop this idea of sharedness, I built upon the notion of Team and Shared Mental Models, which have long been used as facilitating tools to discuss the inaccessible inner workings of our minds. The challenge here of course is that we cannot directly measure the contents of mental models and compare them across team members to determine objectively what is shared and to what extent. To combat this, I have developed a metric for sharedness which uses Latent Semantic Analysis to assess semantic similarity, providing a numerical evaluation of the degree to which content expressed in an open-ended reflection response is also expressed by one’s teammates. Further details about this definition of sharedness can be found in Chapter 3. In this way, the measure of sharedness is not subject to variations in the team’s own understanding of what it means to have a shared understanding or to “be on the same page.” Similarly, reflection prompts were created to follow design progression throughout the course of the projects and were intended to elicit relevant responses related to the current phase of design and participants were not prompted to discuss the extent to which their understanding was shared between teammates. However, semantic similarity as determined through LSA and used as the basis for this definition of sharedness may differ from a common understanding of what it means to be shared. Notably, semantic similarity does not imply conceptual agreement. Nonetheless, the complicated team dynamics surrounding shared understanding do not assert that team members must agree on all points to be effective. For example, a breakdown in communication or

collaboration may occur when team members are “talking past each other” or are unaware of a mismatch in understanding of what is being discussed rather than due to a disagreement. The findings of this work are not dependent on a singular interpretation of what it means to have a shared understanding and capture a variety of expressions of sharedness in different team contexts. To further promote the generalizability of these findings, preliminary data was collected from all participating project teams and at all phases of the semester-long project. From this sample, only responses from teams with sufficient response rates to adequately capture team dynamics were considered. Finally, maximal variation sampling was used to select cases for descriptive case analysis in order to ensure that instances of both low sharedness and high sharedness were represented within the analysis.

Procedural Validation asks, “Which features of the research design improve the fit between the reality and the theory generated?” (Walther et al., 2013). The methodology used in this study was designed to capture the experiences of students in collaborative design teams. To this end, a capstone design course was selected as the venue for data collection as the most realistic example of “real-world” design that takes place in most undergraduate design programs. Teams from a variety of project sponsors, representing a wide range of topics were included in the study. As discussed in the previous paragraphs, the primary source of data for this study was open-ended reflections. These reflections were designed to capture the students experience without artificially shifting focus toward phenomena intended to be studied. In this way, instances of sharedness which were observed within the data arose naturally and can be considered to be representative examples of the ways that shared understanding can manifest in undergraduate design teams. Further, the contents of these reflections were coded using an

emergent coding scheme which was developed iteratively to best capture and describe the types of responses that were represented within the data set, and not to express a particular viewpoint of the researcher.

Communicative Validation asks, “Is the knowledge socially constructed within the relevant communication community?” (Walther et al., 2013). While the work presented in this dissertation is of course representative of the viewpoint of its author, my perspective is situated in a broader social context and influenced both passively and actively by my participation and membership of the broader community. Notably, my view of engineering design theory and education is influenced by my involvement in the community of engineering educators and education researchers who focus on design. My connections to the Siebel Center for Design foster an awareness of and connection to Human-Centered Design practices and mindsets which undoubtedly influence my academic perspective. The framing of this work is the result of many revisions and iterations in collaboration with other education and design researchers, including those on my dissertation panel. I have chosen to build upon work in LSA presented by Dr. Andy Dong (2005) and have presented my findings in light of discussions on Shared Mental Models in design largely guided by the special issue of CoDesign on this topic (Badke-Schaub, Lauche, et al., 2007). My contributions to these research efforts are punctuated with language borrowed from popular works from Bucciarelli (e.g. “Object world” becomes “Object model”) and the broader design community (e.g. one’s understanding of the unmet “need”). By positioning my work within the context of the broader research landscape, I seek to leverage the existing knowledge base and contribute to the ongoing thread of research.

Pragmatic Validation asks, “Do the concepts and knowledge claims withstand exposure to the reality investigated?” (Walther et al., 2013). A major contribution of this work is the Team-Task-Design Mental Model Framework which recontextualizes past descriptions of mental models so that they better apply in design contexts. This represents a significant improvement in the pragmatic validity of mental models in design. Notably, the assumption that shared mental models were beneficial for team and project outcomes has been reasonably questioned in design contexts. For example, more sharedness during ideation may not be beneficial in design. By distinguishing design-related mental models such as the Object Model and Need Model from task-related mental models, the extent of sharedness and its implications can be more easily explored on a model-by-model basis. The Team-Task-Design Mental Model Framework thus provides additional tools to explore what content should be shared in design contexts so that teams may be most effective. In this way, this work contributes to the overall pragmatic validation of shared mental models and the concept of sharedness in design contexts.

Process Reliability asks, “How can the research process be made as independent as possible from random influences?” (Walther et al., 2013). The research methodology employed in this study was designed to be independent from the conditions *on the ground*. In this case, this was due to the collaborative relationship between the researcher and the course used for data collection. The researcher had no influence on course design and intended for research practices to be minimally invasive on both the student’s learning experiences as well as the instructors teaching practice. Therefore, this study intended to collect data representing the conditions as they were, without asserting a specific experimental design or condition necessitating the control of environmental factors subject to randomness. The use of open-ended reflections as the

primary source of data for this work facilitates the capture of nuanced contextual details which can then be included in analysis. In essence, rather than controlling random variables, this study sought to capture the state of the randomness so that specific project contexts could be identified and further explored. In the context of design projects, it is unlikely that the exact same project conditions will arise in separate instances, thus generalizing across variations in project context may be a more productive approach than trying to isolate and recreate specific experimental conditions. Elsewhere in the study design, efforts were made to ensure process reliability more explicitly, specifically in qualitative coding. The codebook was developed using an iterative emergent coding process, the output of which was a set of descriptive codes outlining the contents of the utterances which comprised reflection responses and definitions and examples of each code. The repeatability of the application of this coding scheme was assessed through the measurement of interrater reliability (see Chapter 3 for details).

Implications and Future Work

Implications for Design Teams, Researchers, and Educators

While this study did not empirically test the efficacy of any method and thus cannot make recommendations of best practice or suggest the implementation of any tool having proven its effectiveness, I would like to highlight some potential implications. For design teams, a shared understanding is likely important but not all content must be shared or shared at all times. Sharedness is more common when considerations are more central or have more dependencies. Special attention should be given to boundaries between team members, as well as between the team and stakeholders. Boundaries within the system such as interfaces between subassemblies or components are also critical. It is at these boundaries that harmful differences between mental

models are most prevalent. While design teams may not be actively aware of their mental models, a distinction between design-related models of Object and Need may be managed in System Requirements and Stakeholder Requirements documents respectively (INCOSE, 2015). External representations of all forms, including prototypes, sketches and other models are also beneficial for developing a shared Object Model. Likewise, a Gantt Chart could facilitate the development of a shared design-related Process Model, where applicable. Task-related considerations, however, can be less shared as tasks are often allocated to individuals.

Researchers interested in using LSA may be impressed by its general applicability. LSA can be used to process any text-based documents, even those written in other languages. Because of this, it offers a promising way to synthesize large amounts of data. There are also relatively few parameters to adjust, and results appear to be good even without fine tuning. However, care must be taken when generalizing interpretations of LSA results as they may be nonintuitive. Nonetheless, the combination of written reflection responses and Latent Semantic Analysis is a promising means to obtain significant insight without impeding design progress or subjecting participants to exhaustive data collection protocols. This should be taken as a suggestion to incorporate LSA alongside other methods as a relatively lightweight addition.

Educators, especially those in large design courses like capstones, may be interested in using LSA to gain insight into their project teams. There is certainly potential for this method, however I strongly discourage any use of LSA for the purpose of automated assessment. Nonetheless, LSA may be useful in other ways. Most obviously, it can be used to identify plagiarism between responses as similar responses can be quickly found. However, there is also potential for LSA similarity for purposes such as identifying partners for peer review, or flagging

cases of students who are outliers in their groups. Note that all of these applications are contingent on the existence of a suitable corpus for quality LSA output. This means that a course running for multiple years, that has collected many contextually relevant responses for use in its corpus may have better results from LSA than a course without such a mature resource. As mentioned previously, additional work is needed to substantiate these claims.

Directions for Future Work

As mentioned at various points throughout, additional steps are needed to continue the development of the concepts and methods presented herein. These include a sensitivity analysis of LSA, a validation of LSA sharedness against other mental models metrics, as well as a comprehensive ethnographic case study following teams more closely.

While LSA provides useful results without fine tuning, the sensitivity to a variety of parameters is not well understood in the context of design teams. To this end I plan to conduct a sensitivity analysis to systematically explore the effect of different LSA parameters on results relevant to design team research. For example, using a known group of utterances expected to have higher than average sharedness, I would vary the parameters of LSA such as the number of dimensions, k , the range of n-grams, or the stop list, and track the impact of these changes on the sharedness of the known group. Using a grid search, domains could be identified where the output of LSA behaves as expected or aberrantly. In this way a sort of phase-diagram could be constructed. By repeating this analysis with different groups of utterances and for different corpora, a better understanding of target ranges for LSA parameters in design team contexts may be developed.

Additionally, conclusions about mental models drawn from LSA should be compared to conclusions from other mental models methodologies within the context of design teams. This will help resolve the apparent conflict between similarity as described in the mental model literature and semantic similarity in LSA. While LSA shows promise as a method for measuring mental models because it can be used where other methods cannot, it is difficult to locate LSA findings within the context of work done through other means without a direct comparison between LSA and other methods. To most effectively build upon and respond to prior research, LSA should be compared directly with other methods of determining similarity commonly used in mental models research, such as those described by DeChurch & Mesmer-Magnus (2010), especially in design team contexts.

Finally, while this study took a high-level view of teams, synthesizing across many teams over the course of their projects, a closer look would provide additional complementary details that could not be captured in this work. More work is needed to explore the development of shared mental models in the moment, in authentic, long-term design teams not just in short laboratory sessions. Thus, an ethnographic case study approach is needed. Ideally a combination of transcribed design meetings, design documents, reflections, etc., could be used together with interviews and observations to explore mental model development within the Team-Task-Design Mental Models Framework.

CHAPTER 6: CONCLUSION

Summary

The concept of team mental models, used to describe a team's shared understanding of a collaborative undertaking has been broadly applied to different domains, and has shown utility as an organizing framework to understand and improve teamwork. However, in contexts involving design teams, some findings from past team mental models research cannot be directly applied due to the emergent, and ill-structured nature of design. For example, while similarity in task-related mental models has been linked to team performance for operational teams such as flight crews, design teams can be successful without a shared task model. To address the issues of using mental models in design contexts, I proposed a recontextualization of the existing team-related and task-related mental models paradigm for use in design context, the Team-Task-Design Mental Models Framework. In the Team-Task-Design framework, design-related mental models include the team's understanding of the object they are designing as well as the unmet need they aim to address. Instead of addressing all design considerations, task-related mental models are refocused to cover shorter tasks within the design process, such as creating a CAD model for a motor mount, which may be taken on individually. This helps to explain the finding that design team performance may not be related to task-model similarity (Bierhals et al., 2007) despite seemingly requiring the team to have at least some degree of shared understanding of the object they are designing. An exploratory study was created to establish the groundwork necessary to substantiate these observations.

In this study I explored how LSA can be used to understand the shared content of open-ended reflections (*RQ1*) as a means of identifying the typical elements of a collaborative design

team's shared understanding of a design problem (*RQ2*). Undergraduate student design teams were recruited from a capstone design course in a Mechanical Engineering department at a large midwestern university in Spring 2022. This course had a total enrollment of 175 students at the time of data collection. Teams of 5-6 students were assigned to 32 projects sponsored by industry partners, or university affiliates. Of these, 117 students consented to participate in the study. Eight teams were selected for further study based on response rates - at least three team members responded to each of the repeated reflection prompts. All students were asked to submit reflections throughout the semester, with one prompt repeating three times, see timeline in Figure 3.4. Reflection Prompts listed in Table 3.4. Additionally, a survey was conducted in week 14 to collect demographic information and additional context about the students' perceptions of their project, see Table 3.3.

To identify the commonly shared elements of a design team's shared understanding, an emergent coding scheme was used to label utterances from responses to the repeated reflection prompt according to the type of content contained within each utterance. These utterance types were Question, Expectation, Process, Project Description, Perspective, Obstacle, Design Consideration, and Other, defined in Table 3.7. The most common utterance type was Perspective at 24.5%, followed by Design Considerations at 23.2%. The least common utterance type outside of Other was Questions at 3.5%. Utterances were also processed using Latent Semantic Analysis to approximate each utterance as a vector in the k -reduced semantic space. The sharedness of each utterance was computed as the cosine similarity between the vector representing the utterance and a composite vector representing the sum of utterances from teammates in response to the same reflection prompt. Sharedness was defined this way in

Chapter 3 to avoid the issues related to self-similarity that arise when sharedness is computed as similarity to the concept centroid as defined by Dong *et al.* (Dong, 2005; Dong et al., 2013).

LSA was found to be a promising methodology for the investigation of mental models in design teams, though future work is needed to better understand the impact of various parameters of the algorithm, such as the number of dimensions in the approximated semantic space, k . A notable strength of LSA is the ability to process large amounts of data relatively quickly, offering a means to analyze open-ended responses at a large scale. However, semantic similarity differs from conceptual agreement in that two documents compared by LSA may have high semantic similarity even if they are in direct opposition as long as they relate to similar topics, e.g. “I am very confident” and “I am not very confident.” Because of this, and a dependence on a sufficient corpus for word meanings, care should be taken when interpreting semantic similarity as ‘similarity’ more generally. To this end, LSA results should be directly compared to findings from other methodologies for eliciting and assessing mental models.

Mann-Whitney U-tests (MWU) were used to determine if sharedness varied by utterance type. The groups of sharedness values for each utterance type were compared to the group of sharedness values of all utterances from all teams. This determined if a certain utterance type has stochastically higher or lower sharedness than “normal.” Where “normal” is given by the distribution of all observed sharedness values in the sample. Two utterance types were found to have significantly higher sharedness than the overall group. These were Project Description (Average sharedness: 0.370, MWU Test $U=61337$, $p<0.001$) and Design Consideration utterances (Average sharedness: 0.246, MWU Test: $U=61337$, $p<0.001$). Conversely, Process utterances (Average sharedness: 0.161, MWU Test: $U=21128$, $p=0.010$), Perspective utterances

(Average sharedness: 0.144, MWU Test: $U=42216$, $p<0.001$), and Obstacles utterances (Average sharedness: 0.115, MWU Test: $U=12961$, $p<0.001$) were found to have significantly lower sharedness than the overall group.

Using the Team-Task-Design Mental Models Framework, Project Description and Design Consideration utterances represent aspects of the design-related Object and Need mental models. They largely represent core considerations that are pertinent to all team members and across multiple tasks within the design process, indicative of high *centrality*. These highly central design-related utterances were the most shared of all utterance types. On the other hand, the least shared utterance types were Process, Obstacles and Perspectives. Process and Obstacles utterances represent aspects of the task-related mental model. Design teams face a number of tasks throughout a design project and will often assign tasks to individuals or sub teams with the appropriate skillset. This means that task-related mental models may not be shared in design contexts (Bierhals et al., 2007). Perspective utterances could be representative of team, task, or design-related mental models, but are not very central as they represent an individual's perspective, thus low sharedness is expected for these utterances.

To complement the high-level investigation of sharedness, three descriptive cases were developed more closely following three teams from the sample, The Vehicle Control Team (CP), Mass Customization Team (WZ), and Optomechanical Team (EK). These cases represented the approximate middle, lowest, and highest sharedness averages, respectively. The descriptive cases made clear that high sharedness is not a feature characteristic of well-defined or well-understood projects and, low sharedness was not necessarily indicative of poorly defined or understood ones. Several teams demonstrated high sharedness in Project Description utterances very early on in

the semester, even with limited available information. In general, sharedness was not found to increase over time. It seems that teams with unclear or moving project goals are forced to spend more time and attention pinning down their objectives, which may result in higher sharedness. Whereas other teams specialize more, potentially yielding lower sharedness. These comments are descriptions of cases observed in this sample and are not intended to be applied elsewhere without a verification of their generalizability.

Teams also mentioned the importance of external representations for the purposes of exploration and communication, especially as boundary objects (Bucciarelli, 1994) which bridge the interfaces between individuals with different mental models. External representations, such as Team CP's block diagram or Team EK's cardboard prototype, establish a basis for shared understanding. Shared mental models can then be negotiated and built from this common ground. It seems external representations may be useful to establish a shared understanding at interfaces and boundaries, not just between system components but also between team members and between the team and the stakeholders. These interfaces and boundaries are characterized by dependencies and interconnections (aka high centrality) and thus multiple parts of the system will be impacted by a shared understanding, or the lack of one. There are at least four metrics of team mental models, accuracy, similarity, importance, and centrality (Badke-Schaub, Neumann, et al., 2007). While I did not assess the impact of these on team performance or project outcomes, my observations generally agree with the interactions between these metrics as discussed in prior literature. Similarity is needed where there is also high importance and high centrality. Where importance is low, the consideration is less consequential overall. Where centrality is low, similarity is less needed even if decisions are important.

In conclusion, this dissertation explores the use of LSA to investigate the shared content of open-ended reflections from collaborative design teams. The most common type of utterance found in design team reflections was Design Considerations. Design Considerations ($U=61337$, $p<0.001$) and Project Descriptions ($U=61337$, $p<0.001$) were found to have higher sharedness than the overall group of utterances using the Mann-Whitney U-Test. In the Team-Task-Design framework for mental models, these utterances represent aspects of the design-related Object Model and Need Model. Design-related mental model considerations generally pertain to many tasks and team members and may thus require higher sharedness. Conversely, design tasks are often allocated to individuals with the necessary expertise rather than taken on collaboratively, and therefore task-related mental model considerations may have lower sharedness. Task-related utterance types such as Process ($U=21128$, $p=0.010$) and Obstacles ($U=12961$, $p<0.001$) were found to have lower sharedness than the overall group of utterances using the Mann-Whitney U-Test. The examination of descriptive cases highlighted that the dynamics of shared understanding evolved throughout the project, influenced by factors such as project clarity and team specialization. Importantly, teams underscored the role of external representations as facilitators of shared understanding. This study contributes to the understanding of shared mental models in design teams and their investigation using Latent Semantic Analysis, laying the groundwork for future research to determine the implications of shared understanding on design team performance and project outcomes. In doing so, I hope to contribute to the continuing improvement of engineering design efficacy so that we may be better able to address the many existing problems facing people around the world, as well as the many problems we do not yet realize we have.

Limitations

Generalizability

There are, of course, limitations to the generalizations of these findings. Prominently, the undergraduate population in a capstone course does not represent engineering design practitioners in general. Even within the course, the teams selected for analysis were more female (29% of sample vs 17% of class), less Asian or Pacific Islander (14% of sample vs 22% of class), and more White or Caucasian (43% of sample vs 26% of class). As discussed in Chapter 3, the teams in the sample were determined by response rates. More generally, I believe design should be treated descriptively rather than prescriptively. To that end, the findings reported in this work are intended to describe what took place in the context of this course with these teams. While there are conclusions which may be applied more generally, these are not prescriptions guaranteed to work in any other context.

This brings me to the topic of team performance and design success. It is not my intention to opine on the difference between “good” and “bad” design, nor to draw conclusions about what actions lead to “better” design outcomes. These assessments are mostly subjective and depend on the circumstances defining the design process and outcomes. Because design is such a broad-reaching activity, defining “good” or “bad” design or comparatively “better” or “worse” design choices or outcomes is fraught with traps and obstacles. There are bound to be cases where “bad design” is not so bad, or where “good design” is not so good. Of course, this isn’t to say that designers are free to make decisions without consequences. There are plenty of examples of things that we encounter every day which might be considered poorly designed, or conversely, well-designed. However, these assessments are typically subjective, not universal. It

would follow, then, that any advice to make design *better* should necessarily include a discussion of *better for whom*.

This study makes no assessment of the effect of sharedness on project outcomes or team performance. This is a notable limitation of this study as the ultimate goal of my research is to facilitate better outcomes for design teams. However, since the design teams studied here each worked with a different project sponsor, even if one team were given a better evaluation from their sponsor than another, that would not mean that that the leading team did an objectively better job as they were evaluated on different bases by different stakeholders. Even if all teams worked on the same project, while it would be easier to compare design outcomes and label one approach or process “best”, those findings would apply to that specific project, and may not be generalizable. Design assessment is notoriously tricky (Goldstein, Adams, & Purzer, 2021), especially in educational settings as the interests and needs of instructors differ from those of other stakeholders. Grades on design projects may be unreliable assessments of design quality as they are determined by the instructor, sometimes with input from relevant project stakeholders, but rarely if ever are grades wholly determined by the stakeholders alone. Instructors have a different set of assessment criteria than other project stakeholders. This can easily lead to a scenario where a design team meets course objectives but fails to deliver a successful design resulting in a good grade for poor design outcomes. Other factors, separate from the quality of the design, may also influence grades, such as their bearing on a student’s ability to graduate or their future success. In a capstone course like the one studied here, students may be required to receive a passing grade as a condition for graduation, and the close connection with industry sponsors could mean that poor capstone grades may impact future success and career

opportunities. Students work closely with the sponsor, as well as faculty and teaching assistants to ensure the project stays mostly within course guidelines. Teams that follow poor design practices or struggle for any other reason are more likely to receive some early feedback from their advisors and be given the opportunity to course-correct than to receive a poor or failing grade at the end of the semester. These complications meant that it was not feasible to compare project outcomes or team performance in this study as external factors influenced the assessments that teams received. Thus, I cannot directly draw conclusions about the relationship between sharedness and project success based on, for example, the highest performing team.

Latent Semantic Analysis

Latent Semantic Analysis is a powerful tool for quickly processing a large amount of text-based data, however there are several important considerations that may impact how findings from LSA should be interpreted. The first of these is expressed in the name, *latent* semantic analysis. The meaning of words in LSA is a product of their context of use in the corpus - this has several implications. Primarily, this means that the colloquial usage of terms within the corpus may not perfectly match the definitions of those same terms ascribed by people outside of the represented speech community. Therefore, it is very important that the corpus is ‘sufficiently large’ to be representative. However, there is no universal definition of ‘sufficiently large’, and therefore no guarantee that this standard has been met. The corpus used in this study contains 5162 utterances, or 126524 words. If it were written out in 12pt Times New Roman, single-spaced, with no paragraph breaks, it would be 281.2 pages long. While this is a large amount of data, it represents a small speech community, with all utterances coming from students in the same capstone design course. For context, The Wikipedia Corpus, containing the

full text of Wikipedia, includes roughly 1.9 billion words. It would be very interesting to repeat this analysis with a much larger corpus representing a broader variety of speaker perspectives.

Second is the critical distinction between semantic similarity or sharedness, and agreement. As I started working on semantic similarity and began presenting these concepts to others, I quickly realized that there was a tendency for myself and others to default to agreement when considering the semantic similarity of utterances. For example, one might assume that “the sky is bright and sunny” would have low semantic similarity to “the sky is dark and stormy” because they are conceptually opposed to one another, they disagree. However, depending on the corpus, the terms represented in these utterances are likely to end up in similar contexts and are therefore contextually related. Thus, they will likely have relatively high semantic similarity. One way to distinguish between semantic similarity and conceptual agreement is that utterances which are semantically similar tend to discuss the same topic or subject but may take opposing stances.

Additionally, common words or phrases may impact the semantic similarity of utterances determined by LSA. These words or phrases can be removed from the corpus so that they are ignored completely. These ignored words are referred to as, “stop-words.” In this study, I used an existing list of English stop-words included in the python *scikit-learn* implementation of the LSA algorithm. This list contains words such as, “the”, “it”, “her”, “and”, “as”, etc. The choice of stop-words will have an impact on the similarity results, though the scope of this impact varies significantly. There may be words or phrases that were not included in the stop-words that artificially raised or lowered the sharedness values reported in this study. One possible example of this is the phrase, “the goal of this project is to...”. This phrase commonly appears in Project

Description utterances and may contribute to their relatively high sharedness. However, the extent to which this is a problem is debatable. One could argue that the use of this specific phrase indicates a specific mindset, and the sharedness of that mindset is important. Conversely, it could be said that the details that follow the phrase, “the goal of this project is to...” are more significant than the phrase itself. A systematic review of the impact of the choice of stop-word lists would provide useful insight into the significance of this choice.

Finally, in LSA the term-document matrix is approximated using singular value decomposition, with the k -most significant singular values remaining. This means that each utterance is approximated by a vector of length k , where k is the number of dimensions defining the concept space. For this study, $k = 200$ was used as was suggested in other applications of LSA on English language corpora. However, the resulting sharedness values may be impacted by this choice of k . It is possible that for sufficiently complex corpora, more conceptual dimensions are needed. A sensitivity analysis, for example showing how the sharedness of a particular utterance varies as a function of k would provide useful insight into the significance of this choice.

Sampling Errors

In a related area, the selection of the unit of analysis also has an impact on the findings. For this work, the unit of analysis for LSA was the utterance, meaning that each utterance was represented by its own vector, and thus each utterance was assigned its own sharedness value. Early in the analysis, the unit of analysis was the entire reflection response, such that each response had a single sharedness value. I decided that that was not sufficient to capture the variation within each response as a single response might contain some offerings that were more

shared and others that were more individual. Thus, reflection responses were broken down into “utterances”, with that term borrowed from other applications of LSA, often with transcribed data divided into spoken utterances. In the context of written responses, I chose to divide my utterances using sentence ending punctuation, such that each utterance was roughly one sentence long. The caveat to this is that not all responses were written out in prose. For those responses which used a different formatting method, for example a bulleted list, utterances were manually divided bullet by bullet. However, this approach also meant that there are utterances which contain multiple semantic concepts. An utterance can even contain opposing or semantically unrelated concepts. Taking utterances as the unit of analysis and parsing them to be each roughly a sentence long was a compromise between ease of processing at scale and the potential loss of accuracy. The question remains as to what size unit of analysis is best.

The frequency of data collection could also potentially influence the results. Reflection responses were collected weekly in this study, but the contents of these reflections might change if they were collected more or less frequently. For example, in studies which use transcribed audio of students engaging in a project using a think-aloud protocol, the moment-to-moment interactions between students are represented in the transcript. However, these exchanges are not represented in a weekly written reflection. It is possible that the team had additional areas of shared understanding that were not captured by the data collection methods used here. One possible example of this could be a shared task procedure that might be discussed and established in an in-person team meeting but was not included in reflection responses. In a similar sense, the questions asked in reflection prompts guided students to provide specific

details. This could lead to blind spots in the study. For example, the repeated reflection prompt did not provide much insight into the team-related mental models of the participants.

Study Design

Finally, there are several notable limitations as a consequence of the study design. As with many real-world efforts, this study was subjected to multiple constraints during its conception and throughout data collection, and compromises were made. A large amount of data was collected from reflection responses, but often only one or two members of a team submitted responses. This limited the number of teams that could be considered as responses from multiple team members were needed to provide insight into what was shared amongst the team.

Additionally, many of the projects were protected by non-disclosure agreements, further limiting the data which could be collected and reported. Ideally, a more detailed look into one or more teams using ethnographic research methods would have been available. All teams were asked to participate in an interview, but unfortunately none responded. This was likely due to the timing, at the end of their senior year. However, the study would have benefited greatly from additional sources of data that could help with triangulation and explanation of what happened within each team on a more detailed level.

REFERENCES

- Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design Studies*, 24(3), 275–294. [https://doi.org/10.1016/S0142-694X\(02\)00056-X](https://doi.org/10.1016/S0142-694X(02)00056-X)
- Akgün, A. E., Byrne, J., Keskin, H., Lynn, G. S., & Imamoglu, S. Z. (2005). Knowledge networks in new product development projects: A transactive memory perspective. *Information and Management*, 42(8), 1105–1120. <https://doi.org/10.1016/j.im.2005.01.001>
- Bacon, J., Stockman, B., & Boyle, J. (2010). International Space Station Systems Engineering Case Study.
- Badke-Schaub, P., Lauche, K., & Neumann, A. (2007). Team mental models in design. *CoDesign*, 3(1), 1–3. <https://doi.org/10.1080/15710880601170743>
- Badke-Schaub, P., Neumann, A., Lauche, K., & Mohammed, S. (2007). Mental models in design teams: a valid approach to performance in design collaboration? *CoDesign*, 3(1), 5–20. <https://doi.org/10.1080/15710880601170768>
- Bierhals, R., Schuster, I., Kohler, P., & Badke-Schaub, P. (2007). Shared mental models—linking team cognition and performance. *CoDesign*, 3(1), 75–94. <https://doi.org/10.1080/15710880601170891>
- Boos, M. (2007). Optimal sharedness of mental models for effective group performance. *CoDesign*, 3(1), 21–28. <https://doi.org/10.1080/15710880601170776>
- Borrego, M., Karlin, J., McNair, L. D., & Beddoes, K. (2013). Team effectiveness theory from

industrial and organizational psychology applied to engineering student project teams: A research review. *Journal of Engineering Education*, 102(4), 472–512.

<https://doi.org/10.1002/jee.20023>

Bucciarelli, L. L. (1988). An ethnographic perspective on engineering design. *Design Studies*, 9(3), 159–168. [https://doi.org/10.1016/0142-694X\(88\)90045-2](https://doi.org/10.1016/0142-694X(88)90045-2)

Bucciarelli, L. L. (1994). *Designing Engineers*. MIT Press. Retrieved from <https://books.google.com/books?id=ECSKxrTrHPsC>

Bucciarelli, L. L. (2002). Between thought and object in engineering design. *Design Studies*, 23(3), 219–231. [https://doi.org/10.1016/S0142-694X\(01\)00035-7](https://doi.org/10.1016/S0142-694X(01)00035-7)

Carley, K. M. (1997). Extracting team mental models through textual analysis. *Journal of Organizational Behavior*, 18(S1), 533–558.
[https://doi.org/https://doi.org/10.1002/\(SICI\)1099-1379\(199711\)18:1+<533::AID-JOB906>3.0.CO;2-3](https://doi.org/https://doi.org/10.1002/(SICI)1099-1379(199711)18:1+<533::AID-JOB906>3.0.CO;2-3)

Casakin, H., & Badke-Schaub, P. (2017). Sharedness of team mental models in the course of design-related interaction between architects and clients. *Design Science*, 3, 1–21.
<https://doi.org/10.1017/dsj.2017.15>

Chan, L. K., & Wu, M. L. (2002). *Quality function deployment: A literature review*. *European Journal of Operational Research* (Vol. 143). [https://doi.org/10.1016/S0377-2217\(02\)00178-](https://doi.org/10.1016/S0377-2217(02)00178-9)

9

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring Team

- Knowledge. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(1), 151–173. <https://doi.org/10.1518/001872000779656561>
- Craik, K. J. W. (1943). *The nature of explanation*. Cambridge: University Press.
- Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education*, 101(4), 738–797. <https://doi.org/10.1002/j.2168-9830.2012.tb01127.x>
- Cross, N., Christiaans, H., & Dorst, K. (1996). *Analysing design activity*. Chichester ; Wiley.
- Davison, G., & Blackman, D. (2005). The role of mental models in innovative teams. *European Journal of Innovation Management*, 8(4), 409–423. <https://doi.org/10.1108/14601060510627795>
- Dechurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics*, 14(1), 1–14. <https://doi.org/10.1037/a0017455>
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407. [https://doi.org/https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASI1>3.0.CO;2-9](https://doi.org/https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9)
- Descombes, V., & Schwartz, S. A. (2001). *The Mind's Provisions*. Princeton University Press. <https://doi.org/10.2307/j.ctv1nxcv3h>
- Dong, A. (2005). The latent semantic approach to studying design team communication. *Design Studies*, 26(5), 445–461. <https://doi.org/10.1016/j.destud.2004.10.003>

- Dong, A., Hill, A. W., & Agogino, A. M. (2004). A document analysis method for characterizing design team performance. *Journal of Mechanical Design*, 126(3), 378–385.
<https://doi.org/10.1115/1.1711818>
- Dong, A., & Kleinsmann, M. (2016). Methods for Studying Collaborative Design Thinking. In P. Cash, T. Stanković, & M. Štorga (Eds.), *Experimental Design Research* (pp. 83–96). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-33781-4_5
- Dong, A., Kleinsmann, M. S., & Deken, F. (2013). Investigating design cognition in the construction and enactment of team mental models. *Design Studies*, 34(1), 1–33.
<https://doi.org/10.1016/j.destud.2012.05.003>
- Dong, A., & Sarkar, S. (2014). Generalized Design Knowledge and the Higher-Order Singular Value Decomposition. *Design Computing and Cognition '12*, 415–432.
https://doi.org/10.1007/978-94-017-9112-0_23
- Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1), 188–230. <https://doi.org/https://doi.org/10.1002/aris.1440380105>
- Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering Design Thinking, Teaching, and Learning. *Journal of Engineering Education*, 94(1), 103–120.
<https://doi.org/10.1002/j.2168-9830.2005.tb00832.x>
- Eccles, D. W., & Tenenbaum, G. (2004). Why an expert team is more than a team of experts: A social-cognitive conceptualization of team coordination and communication in sport. *Journal of Sport and Exercise Psychology*, 26(4), 542–560.
<https://doi.org/10.1123/jsep.26.4.542>

- Elena, M. V., & Summers, J. D. (2019). Requirement generation: Lecture intervention impact on variety and novelty. *Proceedings of the ASME Design Engineering Technical Conference*, 3, 1–10. <https://doi.org/10.1115/DETC2019-97528>
- Ensley, M. D., & Pearce, C. L. (2001). Shared cognition in top management teams: Implications for new venture performance. *Journal of Organizational Behavior*, 22(2), 145–160. <https://doi.org/10.1002/job.83>
- Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective. *Performance Improvement Quarterly*, 26(2), 43–71. <https://doi.org/10.1002/piq.21143>
- Espinosa, J. A., Lerch, F. J., & Kraut, R. E. (2004). Explicit versus implicit coordination mechanisms and task dependencies: One size does not fit all. In *Team cognition: Understanding the factors that drive process and performance*. (Vol. 7, pp. 107–129). Washington: American Psychological Association. <https://doi.org/10.1037/10690-006>
- Faulkner, W. (2015). ‘Nuts and Bolts and People’ Gender Troubled Engineering Identities. In S. H. Christensen, C. Didier, A. Jamison, M. Meganck, C. Mitcham, & B. Newberry (Eds.), *Engineering Identities, Epistemologies and Values. Philosophy of Engineering and Technology* (pp. 23–40). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-16172-3_2
- Forrester, J. W. (1962). Industrial dynamics. [Cambridge, Mass.]: M.I.T. Press. Retrieved from <file://catalog.hathitrust.org/Record/001084975>
- Fu, K., Cagan, J., & Kotovsky, K. (2010). Design team convergence: The influence of example

solution quality. *Journal of Mechanical Design*, 132(11), 1–11.

<https://doi.org/10.1115/1.4002202>

Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. University of Chicago Press. Retrieved from <https://books.google.com/books?id=6Gcu92U8CwYC>

Gero, J. S., & Kannengiesser, U. (2004). The situated function-behaviour-structure framework. *Design Studies*, 25(4), 373–391. <https://doi.org/10.1016/j.destud.2003.10.010>

Goldschmidt, G. (2007). To see eye to eye: the role of visual representations in building shared mental models in design teams. *CoDesign*, 3(1), 43–50.
<https://doi.org/10.1080/15710880601170826>

Goldstein, M. H., Adams, R. S., & Purzer, S. (2021). Understanding Informed Design through Trade-Off Decisions With an Empirically-Based Protocol for Students and Design Educators. *Journal of Pre-College Engineering Education Research*, 11(2), 25–41.
<https://doi.org/10.7771/2157-9288.1279>

Gottschling, V. (2009). Mental Models. In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of Neuroscience* (pp. 2333–2337). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_3434

Goucher-Lambert, K., Gyory, J. T., Kotovsky, K., & Cagan, J. (2020). Adaptive inspirational design stimuli: Using design output to computationally search for stimuli that impact concept generation. *Journal of Mechanical Design*, 142(9), 1–10.
<https://doi.org/10.1115/1.4046077>

Hill, A. W., Dong, A., & Agogino, A. M. (2002). Towards Computational Tools for Supporting the Reflective Team. *Artificial Intelligence in Design '02*, 305–325.

https://doi.org/10.1007/978-94-017-0795-4_15

INCOSE. (2015). *INCOSE Systems Engineering Handbook : A Guide for System Life Cycle Processes and Activities* (Vol. Fourth edi). Hoboken, New Jersey: Wiley.

Johnson-Laird, P. N. (Philip N. (1983). *Mental models : towards a cognitive science of language, inference, and consciousness* . Cambridge, Mass: Harvard University Press.

Johnson-Laird, P. N. (1980). Mental Models in Cognitive Science. *Cognitive Science*, 4(1), 71–115. https://doi.org/10.1207/s15516709cog0401_4

Johnson, T. E., Lee, Y., Lee, M., O'Connor, D. L., Khalil, M. K., & Huang, X. (2007). Measuring sharedness of team-related knowledge: Design and validation of a shared mental model instrument. *Human Resource Development International*, 10(4), 437–454.
<https://doi.org/10.1080/13678860701723802>

Kameda, T., Ohtsubo, Y., & Takezawa, M. (1997). Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context. *Journal of Personality and Social Psychology*, 73(2), 296.

Kim, E. (2019). A systematized literature review of the characteristics of team mental models in engineering design contexts. *ASEE Annual Conference and Exposition, Conference Proceedings*. <https://doi.org/10.18260/1-2--32009>

Kleinsmann, M., Valkenburg, R., & Buijs, J. (2007). Why do(n't) actors in collaborative design

- understand each other? An empirical study towards a better understanding of collaborative design. *CoDesign*, 3(1), 59–73. <https://doi.org/10.1080/15710880601170875>
- Klimoski, R., & Mohammed, S. (1994). Team mental model: construct or metaphor? *Journal of Management*, 20(2), 403–437. [https://doi.org/10.1016/0149-2063\(94\)90021-3](https://doi.org/10.1016/0149-2063(94)90021-3)
- Kolb, D. A. (1984). Experiential Learning: Experience as The Source of Learning and Development. *Prentice Hall, Inc.*, (1984), 20–38. <https://doi.org/10.1016/B978-0-7506-7223-8.50017-4>
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*. US: American Psychological Association. <https://doi.org/10.1037/0033-295X.104.2.211>
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), 259–284. <https://doi.org/10.1080/01638539809545028>
- Landauer, T. K., & Psotka, J. (2000). Simulating text understanding for educational applications with Latent Semantic Analysis: Introduction to LSA. *Interactive Learning Environments*, 8(2), 73–86.
- Langan-Fox, J., Anglim, J., & Wilson, J. R. (2004). Mental models, team mental models, and performance: Process, development, and future directions. *Human Factors and Ergonomics In Manufacturing*, 14(4), 331–352. <https://doi.org/10.1002/hfm.20004>
- Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of

- the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, 27(4), 403–418. <https://doi.org/10.1002/job.387>
- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *The Journal of Applied Psychology*, 87(1), 3–13. <https://doi.org/10.1037/0021-9010.87.1.3>
- Marshall, N. (2007). Team mental models in action: a practice-based perspective. *CoDesign*, 3(1), 29–36. <https://doi.org/10.1080/15710880601170784>
- Mathieu, J. E., Goodwin, G. F., Heffner, T. S., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, 85(2), 273–283. <https://doi.org/10.1037/0021-9010.85.2.273>
- Mathieu, J. E., Rapp, T. L., Maynard, M. T., & Mangos, P. M. (2009). Interactive effects of team and task shared mental models as related to air traffic controllers' collective efficacy and effectiveness. *Human Performance*, 23(1), 22–40. <https://doi.org/10.1080/08959280903400150>
- Minneman, S. (1991). *The Social Construction of a Technical Reality*. Stanford University.
- Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, 36(4), 876–910.

<https://doi.org/10.1177/0149206309356804>

Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The Measurement of Team Mental Models: We Have No Shared Schema. *Organizational Research Methods*, 3(2), 123–165.

<https://doi.org/10.1177/109442810032001>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. Retrieved from <http://jmlr.org/papers/v12/pedregosa11a.html>

Peterson, E., Mitchell, T. R., Thompson, L., & Burr, R. (2000). Collective Efficacy and Aspects of Shared Mental Models as Predictors of Performance Over Time in Work Groups. *Group Processes & Intergroup Relations*, 3(3), 296–316.

<https://doi.org/10.1177/1368430200033005>

Plato, & Jowett, B. (1941). *Plato's the republic*. New York: The Modern library.

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. *Computer Supported Collaborative Learning*, 69–97.

https://doi.org/10.1007/978-3-642-85098-1_5

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6), 1296–1308. <https://doi.org/10.1109/21.199457>

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349.

- Schaub, H. (2007). The importance of the characteristics of the task to understand team mental models. *CoDesign*, 3(1), 37–42. <https://doi.org/10.1080/15710880601170800>
- Schön, D. A. (1992). *The Reflective Practitioner: How Professionals Think in Action (1st ed.)*. Routledge. <https://doi.org/10.4324/9781315237473>
- Smith-Jentsch, K. A., Mathieu, J. E., & Kraiger, K. (2005). Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology*, 90(3), 523–535. <https://doi.org/10.1037/0021-9010.90.3.523>
- Smulders, F. E. (2007). Team mental models in innovation: means and ends. *CoDesign*, 3(1), 51–58. <https://doi.org/10.1080/15710880601170834>
- Song, S., Dong, A., & Agogino, A. M. (2003). Time variation of design “story telling” in engineering design teams. *Proceedings of the International Conference on Engineering Design, ICED, DS 31*.
- Stempfle, J., & Badke-Schaub, P. (2002). Thinking in design teams - An analysis of team communication. *Design Studies*, 23(5), 473–496. [https://doi.org/10.1016/S0142-694X\(02\)00004-2](https://doi.org/10.1016/S0142-694X(02)00004-2)
- Stevens, A. L., & Gentner, D. (1983). *Mental Models*. Hillsdale, N.J.: Psychology Press.
- Stevens, R., Johri, A., & O’connor, K. (2015). Professional engineering work. *Cambridge Handbook of Engineering Education Research*, 119–138. <https://doi.org/10.1017/CBO9781139013451.010>
- Suh, N. P. (1995). Axiomatic design of mechanical systems. *Journal of Mechanical Design*,

Transactions of the ASME, 117(B), 2–10. <https://doi.org/10.1115/1.2836467>

Teasley, S. D., Fischer, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., & Chi, M. (2008). Cognitive convergence in collaborative learning. *Computer-Supported Collaborative Learning Conference, CSCL*, (PART 3), 360–367.

Turns, J. A., Sattler, B., Yasuhara, K., Borgford-Parnell, J. L., & Atman, C. J. (2014). Integrating reflection into engineering education. *ASEE Annual Conference and Exposition, Conference Proceedings*. <https://doi.org/10.18260/1-2--20668>

Waller, M. J., Gupta, N., & Giambatista, R. C. (2004). Effects of adaptive behaviors and shared mental models on control crew performance. *Management Science*, 50(11), 1534–1544. <https://doi.org/10.1287/mnsc.1040.0210>

Walther, J., & Sochacka, N. (2015). Qualifying qualitative research quality (The Q3 project): An interactive discourse around research quality in interpretive approaches to engineering education research. *Proceedings - Frontiers in Education Conference, FIE*, 2015-Febru(February), 5–8. <https://doi.org/10.1109/FIE.2014.7043988>

Walther, J., Sochacka, N. W., & Kellam, N. N. (2013). Quality in interpretive engineering education research: Reflections on an example study. *Journal of Engineering Education*, 102(4), 626–659. <https://doi.org/10.1002/jee.20029>

Webber, S. S., Chen, G., Payne, S. C., Marsh, S. M., & Zaccaro, S. J. (2000). Enhancing Team Mental Model Measurement with Performance Appraisal Practices. *Organizational Research Methods*, 3(4), 307–322. <https://doi.org/10.1177/109442810034001>

- Wegner, D. M. (1987). Transactive Memory: A Contemporary Analysis of the Group Mind. *Theories of Group Behavior*, 185–208. https://doi.org/10.1007/978-1-4612-4634-3_9
- Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61(6), 923–929. <https://doi.org/10.1037//0022-3514.61.6.923>
- Weick, K. E., & Roberts, K. H. (1993). Collective Mind in Organizations: Heedful Interrelating on Flight Decks. *Administrative Science Quarterly*, 38(3), 357. <https://doi.org/10.2307/2393372>
- Wood, M., Chen, P., Fu, K., Cagan, J., & Kotovsky, K. (2014). The Role of Design Team Interaction Structure on Individual and Shared Mental Models. *Design Computing and Cognition '12*, 209–226. https://doi.org/10.1007/978-94-017-9112-0_12
- Xiao, Y., Parker, S. H., & Manser, T. (2013). Teamwork and Collaboration. *Reviews of Human Factors and Ergonomics*, 8(1), 55–102. <https://doi.org/10.1177/1557234X13495181>
- Yi, S., Damen, N. B., & Toh, C. A. (2020). Back and Forth: Using Conversation Analysis to Explore Dialogues of Sharedness of Mental Models. In *Volume 8: 32nd International Conference on Design Theory and Methodology (DTM)* (pp. 1–11). American Society of Mechanical Engineers. <https://doi.org/10.1115/DETC2020-22515>
- Yin, R. K. (2017). *Case Study Research and Applications: Design and Methods*. SAGE Publications. Retrieved from <https://books.google.com/books?id=uX1ZDwAAQBAJ>

APPENDIX A: IRB LETTER



OFFICE OF THE VICE CHANCELLOR FOR RESEARCH & INNOVATION

Office for the Protection of Research Subjects
805 W. Pennsylvania Ave., MC-095
Urbana, IL 61801-4822

Notice of Exempt Determination

March 1, 2022

Principal Investigator	Blake Johnson
CC	Molly Goldstein, Alexander Pagano
Protocol Title	<i>Continuous improvement of ME470</i>
Protocol Number	22770
Funding Source	Unfunded
Review Category	Exempt 1
Determination Date	March 1, 2022
Closure Date	February 28, 2027

This letter authorizes the use of human subjects in the above protocol. The University of Illinois at Urbana-Champaign Office for the Protection of Research Subjects (OPRS) has reviewed your application and determined the criteria for exemption have been met.

The Principal Investigator of this study is responsible for:

- Conducting research in a manner consistent with the requirements of the University and federal regulations found at 45 CFR 46.
- Requesting approval from the IRB prior to implementing major modifications.
- Notifying OPRS of any problems involving human subjects, including unanticipated events, participant complaints, or protocol deviations.
- Notifying OPRS of the completion of the study.

Changes to an **exempt** protocol are only required if substantive modifications are requested and/or the changes requested may affect the exempt status.

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

IORG0000014 • FWA #00008584
217.333.2670 • irb@illinois.edu • oprs.research.illinois.edu

APPENDIX B: SHAREDNESS DATA BY TEAM

Team BK

Table B.1 - Mann Whitney U-Test by Utterance Type – Team BK		
Utterance Type	Less than	Greater than
Questions	U=47.5, p=0.737	U=47.5, p=0.279
Expectations	U=207.0, p=0.481	U=207.0, p=0.527
Process	U=315.0, p=0.085	U=315.0, p=0.917
Project Description	U=626.0, p=0.949	U=626.0, p=0.052
Perspective	U=3716.0, p=0.100	U=3716.0, p=0.901
Obstacles	U=27.5, p=0.366	U=27.5, p=0.652
Design Consideration	U=774.0, p=0.933	U=774.0, p=0.069
Other	U=97.0, p=0.155	U=97.0, p=0.851

Table B.2 - Comparing Sharedness Descriptive Statistics by Utterance type - Team BK				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	1 (1.4%)	0.269	0.269	
Expectations	6 (8.6%)	0.233	0.204	0.106
Process	12 (17.1%)	0.135	0.164	0.148
Project Description	14 (20.0%)	0.282	0.295	0.147
Perspective	14 (20.0%)	0.109	0.159	0.142
Obstacles	1 (1.4%)	0.162	0.162	
Design Consideration	18 (25.7%)	0.299	0.285	0.150
Other	4 (5.7%)	0.151	0.126	0.082
All Combined	70 (100.0%)	0.191	0.223	0.153

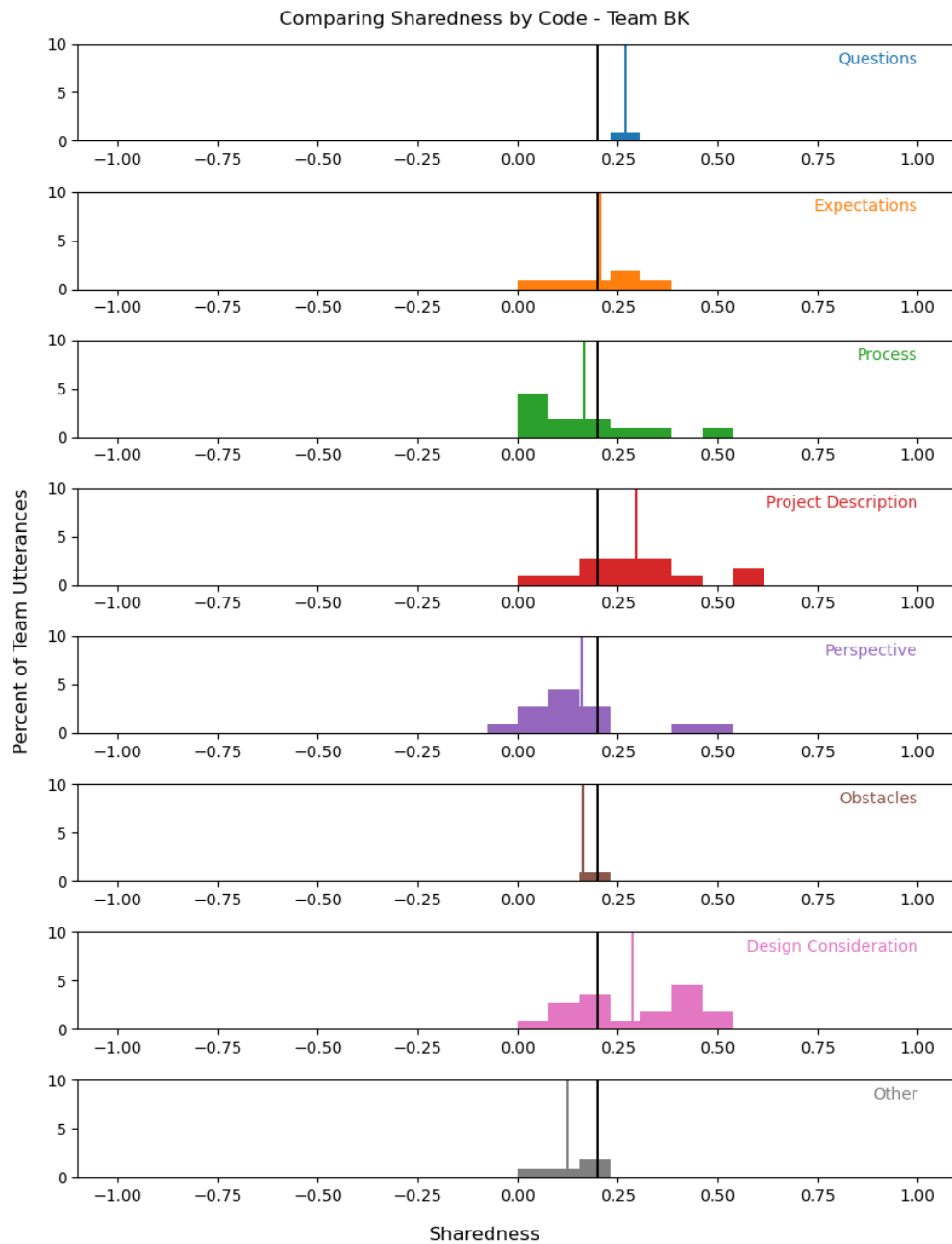


Figure B.1 – Histograms showing the distributions of sharedness values for each code (Team BK).

Team CP

Table B.3 - Mann Whitney U-Test by Utterance Type – Team CP		
Utterance Type	Less than	Greater than
Questions	U=74.5 p=0.899	U=74.5 p=0.109
Expectations	U=480.5 p=0.536	U=480.5 p=0.468
Process	U=168.5 p=0.027	U=168.5 p=0.974
Project Description	U=878.0 p=0.997	U=878.0 p=0.003
Perspective	U=3682.0 p=0.018	U=3682.0 p=0.982
Obstacles	U=140.0 p=0.032	U=140.0 p=0.970
Design Consideration	U=1479.5 p=0.933	U=1479.5 p=0.068
Other	U=48.0 p=0.147	U=48.0 p=0.859

Table B.4 - Comparing Sharedness Descriptive Statistics by Utterance type - Team CP				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	1 (1.2%)	0.414	0.414	
Expectations	11 (12.8%)	0.180	0.214	0.120
Process	7 (8.1%)	0.096	0.100	0.084
Project Description	14 (16.3%)	0.399	0.401	0.218
Perspective	16 (18.6%)	0.103	0.121	0.095
Obstacles	6 (7.0%)	0.089	0.100	0.072
Design Consideration	29 (33.7%)	0.238	0.256	0.122
Other	2 (2.3%)	0.099	0.099	0.095
All Combined	86 (100.0%)	0.185	0.224	0.167

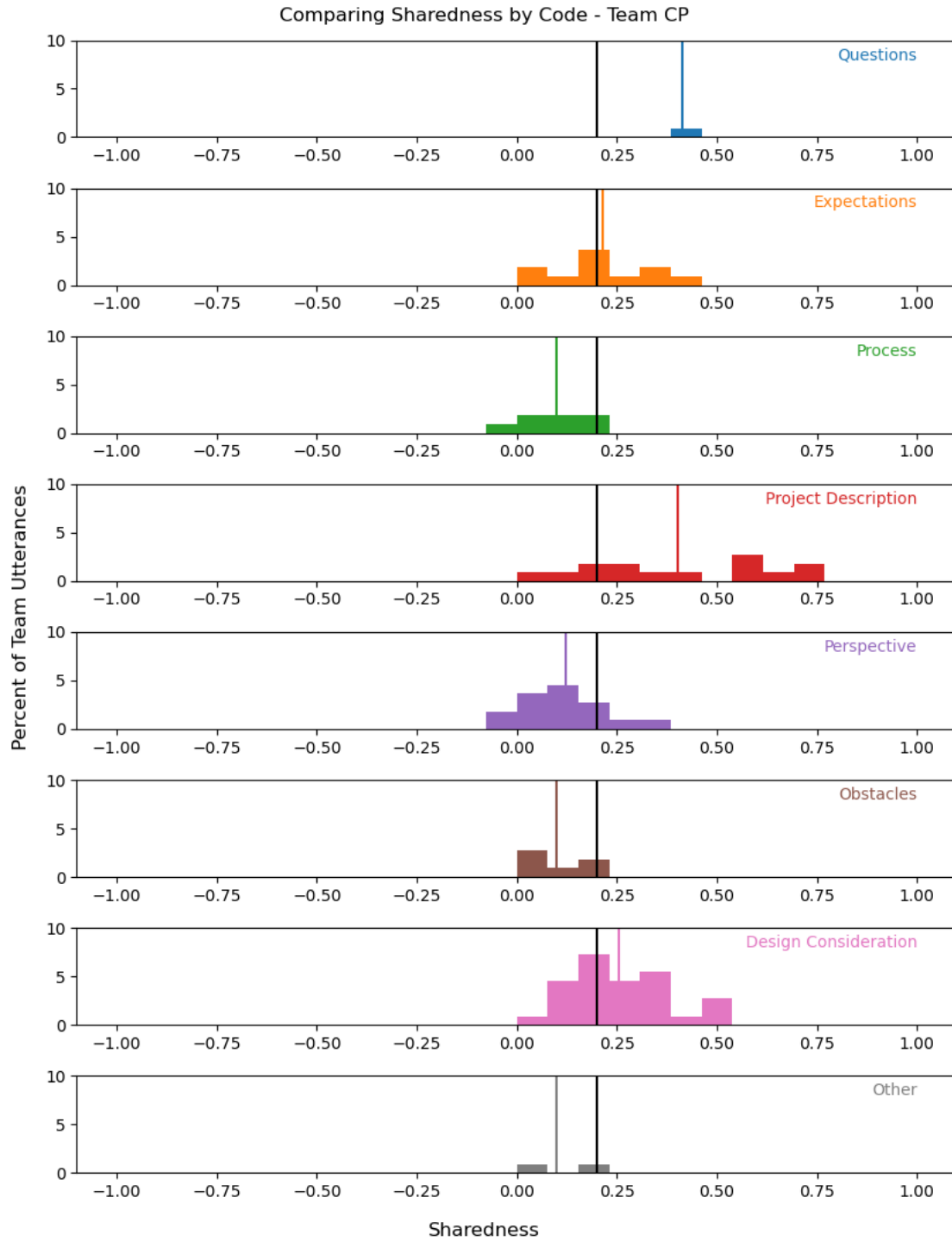


Figure B.2 – Histograms showing the distributions of sharedness values for each code (Team CP).

Team EK

Table B.5 - Mann Whitney U-Test by Utterance Type – Team EK		
Utterance Type	Less than	Greater than
Questions	U=544.5 p=0.692	U=544.5 p=0.311
Expectations	U=392.0 p=0.067	U=392.0 p=0.934
Process	U=481.5 p=0.134	U=481.5 p=0.868
Project Description	U=1287.0 p=1.000	U=1287.0 p<0.001
Perspective	U=8513.5 p=0.146	U=8513.5 p=0.855
Obstacles	U=251.0 p=0.022	U=251.0 p=0.979
Design Consideration	U=1681.0 p=0.918	U=1681.0 p=0.083
Other	U=145.5 p=0.367	U=145.5 p=0.640

Table B.6 - Comparing Sharedness Descriptive Statistics by Utterance type - Team EK				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	9 (8.2%)	0.308	0.278	0.200
Expectations	10 (9.1%)	0.136	0.143	0.092
Process	11 (10.0%)	0.111	0.182	0.151
Project Description	14 (12.7%)	0.474	0.465	0.141
Perspective	29 (26.4%)	0.142	0.162	0.103
Obstacles	8 (7.3%)	0.128	0.109	0.059
Design Consideration	26 (23.6%)	0.260	0.278	0.167
Other	3 (2.7%)	0.141	0.178	0.090
All Combined	110 (100.0%)	0.180	0.234	0.173

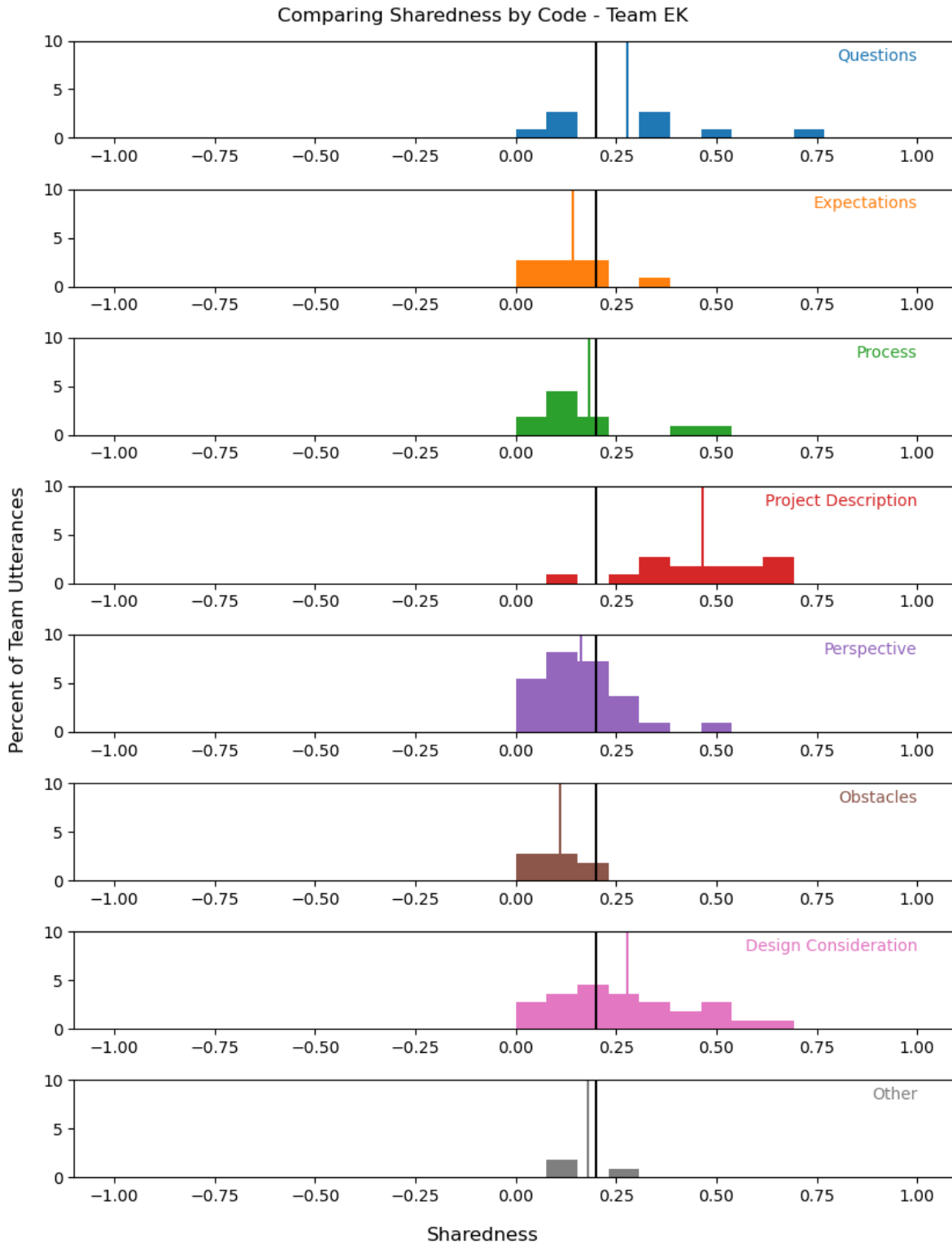


Figure B.3 – Histograms showing the distributions of sharedness values for each code (Team EK).

Team GR

Table B.7 - Mann Whitney U-Test by Utterance Type – Team GR		
Utterance Type	Less than	Greater than
Questions	U=50.5 p=0.535	U=50.5 p=0.479
Expectations	U=358.5 p=0.038	U=358.5 p=0.963
Process	U=352.5 p=0.172	U=352.5 p=0.831
Project Description	U=1538.5 p=1.000	U=1538.5 p<0.001
Perspective	U=8504.0 p=0.088	U=8504.0 p=0.912
Obstacles	U=386.0 p=0.029	U=386.0 p=0.971
Design Consideration	U=902.0 p=0.976	U=902.0 p=0.024
Other	U=6.5 p=0.071	U=6.5 p=0.933

Table B.8 - Comparing Sharedness Descriptive Statistics by Utterance type - Team GR				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	1 (1.0%)	0.200	0.200	
Expectations	11 (11.3%)	0.125	0.130	0.088
Process	9 (9.3%)	0.127	0.172	0.138
Project Description	19 (19.6%)	0.448	0.440	0.127
Perspective	30 (30.9%)	0.131	0.154	0.093
Obstacles	12 (12.4%)	0.110	0.126	0.092
Design Consideration	14 (14.4%)	0.284	0.306	0.143
Other	1 (1.0%)	0.025	0.025	
All Combined	97 (100.0%)	0.183	0.226	0.164

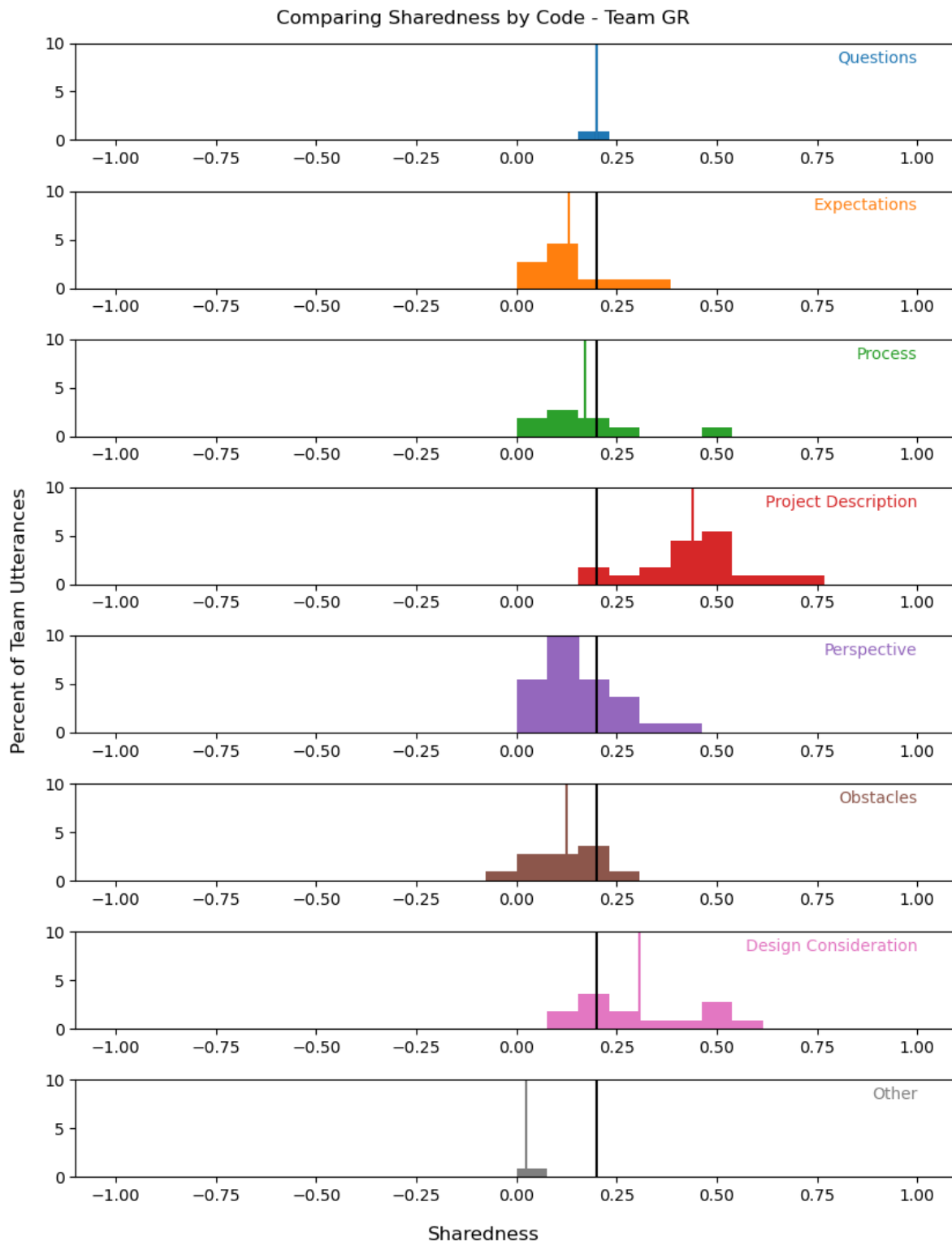


Figure B.4 – Histograms showing the distributions of sharedness values for each code (Team GR).

Team KL

Table B.9 - Mann Whitney U-Test by Utterance Type – Team KL		
Utterance Type	Less than	Greater than
Questions	U=122.5 p=0.514	U=122.5 p=0.495
Expectations	U=401.0 p=0.867	U=401.0 p=0.136
Process	U=641.0 p=0.475	U=641.0 p=0.529
Project Description	U=656.5 p=0.995	U=656.5 p=0.006
Perspective	U=3842.0 p=0.029	U=3842.0 p=0.971
Obstacles	U=241.0 p=0.019	U=241.0 p=0.982
Design Consideration	U=633.5 p=0.605	U=633.5 p=0.398
Other	U=48.0 p=0.167	U=48.0 p=0.840

Table B.10 - Comparing Sharedness Descriptive Statistics by Utterance type - Team KL				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	3 (3.7%)	0.175	0.162	0.067
Expectations	8 (9.9%)	0.196	0.203	0.083
Process	16 (19.8%)	0.183	0.145	0.087
Project Description	11 (13.6%)	0.303	0.286	0.142
Perspective	16 (19.8%)	0.117	0.124	0.076
Obstacles	10 (12.3%)	0.088	0.089	0.077
Design Consideration	15 (18.5%)	0.156	0.173	0.097
Other	2 (2.5%)	0.094	0.094	0.024
All Combined	81 (100.0%)	0.129	0.163	0.110

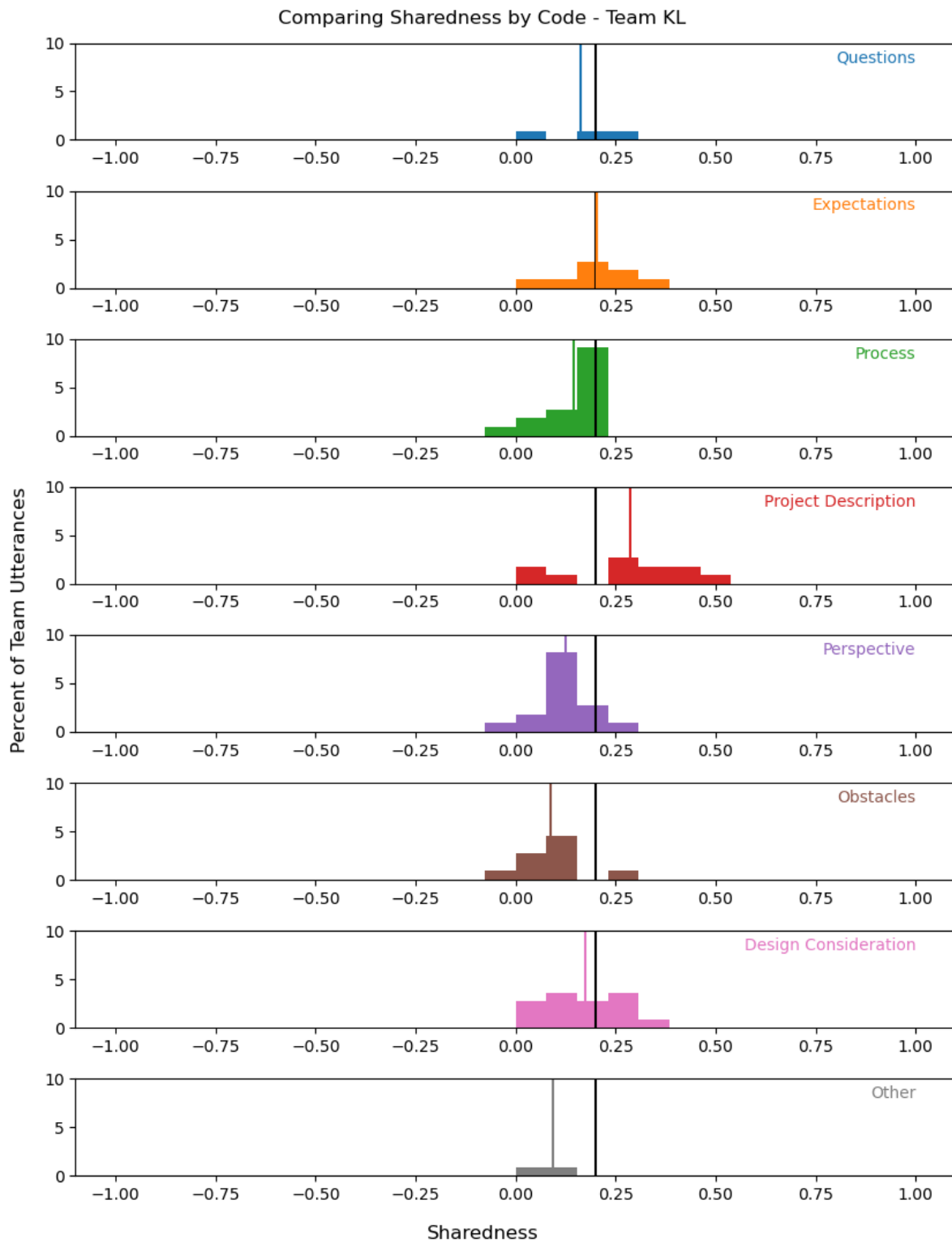


Figure B.5 – Histograms showing the distributions of sharedness values for each code (Team KL).

Team LU

Table B.11 - Mann Whitney U-Test by Utterance Type – Team LU		
Utterance Type	Less than	Greater than
Questions	U=56.0 p=0.172	U=56.0 p=0.835
Expectations	U=277.0 p=0.491	U=277.0 p=0.515
Process	U=481.5 p=0.773	U=481.5 p=0.230
Project Description	U=509.0 p=0.958	U=509.0 p=0.043
Perspective	U=4403.0 p=0.001	U=4403.0 p=0.999
Obstacles	U=310.5 p=0.102	U=310.5 p=0.900
Design Consideration	U=1933.0 p=0.999	U=1933.0 p=0.001
Other	U=84.5 p=0.001	U=84.5 p=0.999

Table B.12 - Comparing Sharedness Descriptive Statistics by Utterance type - Team LU				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	2 (2.2%)	0.099	0.099	0.002
Expectations	6 (6.5%)	0.141	0.181	0.121
Process	9 (9.7%)	0.143	0.231	0.159
Project Description	8 (8.6%)	0.246	0.273	0.154
Perspective	22 (23.7%)	0.089	0.103	0.073
Obstacles	9 (9.7%)	0.125	0.120	0.089
Design Consideration	30 (32.3%)	0.226	0.251	0.101
Other	7 (7.5%)	0.004	0.032	0.055
All Combined	93 (100.0%)	0.156	0.179	0.131

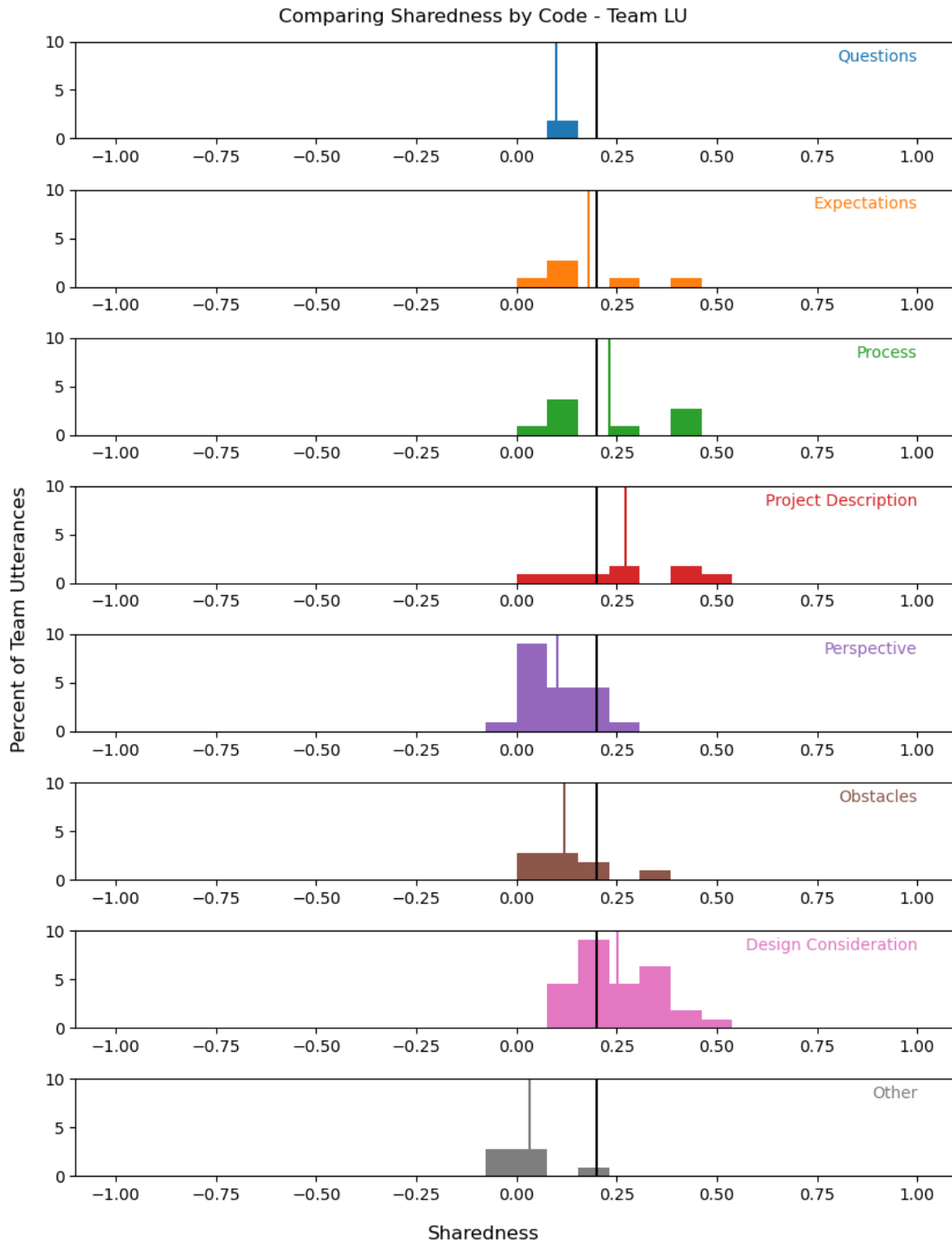


Figure B.6 – Histograms showing the distributions of sharedness values for each code (Team LU).

Team WZ

Table B.13 - Mann Whitney U-Test by Utterance Type – Team WZ		
Utterance Type	Less than	Greater than
Questions	U=197.0 p=0.278	U=197.0 p=0.728
Expectations	U=281.5 p=0.181	U=281.5 p=0.822
Process	U=149.0 p=0.076	U=149.0 p=0.927
Project Description	U=625.5 p=0.995	U=625.5 p=0.006
Perspective	U=5800.5 p=0.275	U=5800.5 p=0.725
Obstacles	U=100.0 p=0.122	U=100.0 p=0.883
Design Consideration	U=771.5 p=0.645	U=771.5 p=0.358
Other	U=77.5 p=0.171	U=77.5 p=0.835

Table B.14 - Comparing Sharedness Descriptive Statistics by Utterance type - Team WZ				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	6 (7.8%)	0.085	0.119	0.086
Expectations	9 (11.7%)	0.096	0.104	0.050
Process	6 (7.8%)	0.056	0.087	0.097
Project Description	11 (14.3%)	0.222	0.257	0.116
Perspective	19 (24.7%)	0.175	0.194	0.181
Obstacles	4 (5.2%)	0.064	0.079	0.040
Design Consideration	19 (24.7%)	0.189	0.156	0.111
Other	3 (3.9%)	0.098	0.078	0.055
All Combined	77 (100.0%)	0.132	0.158	0.134

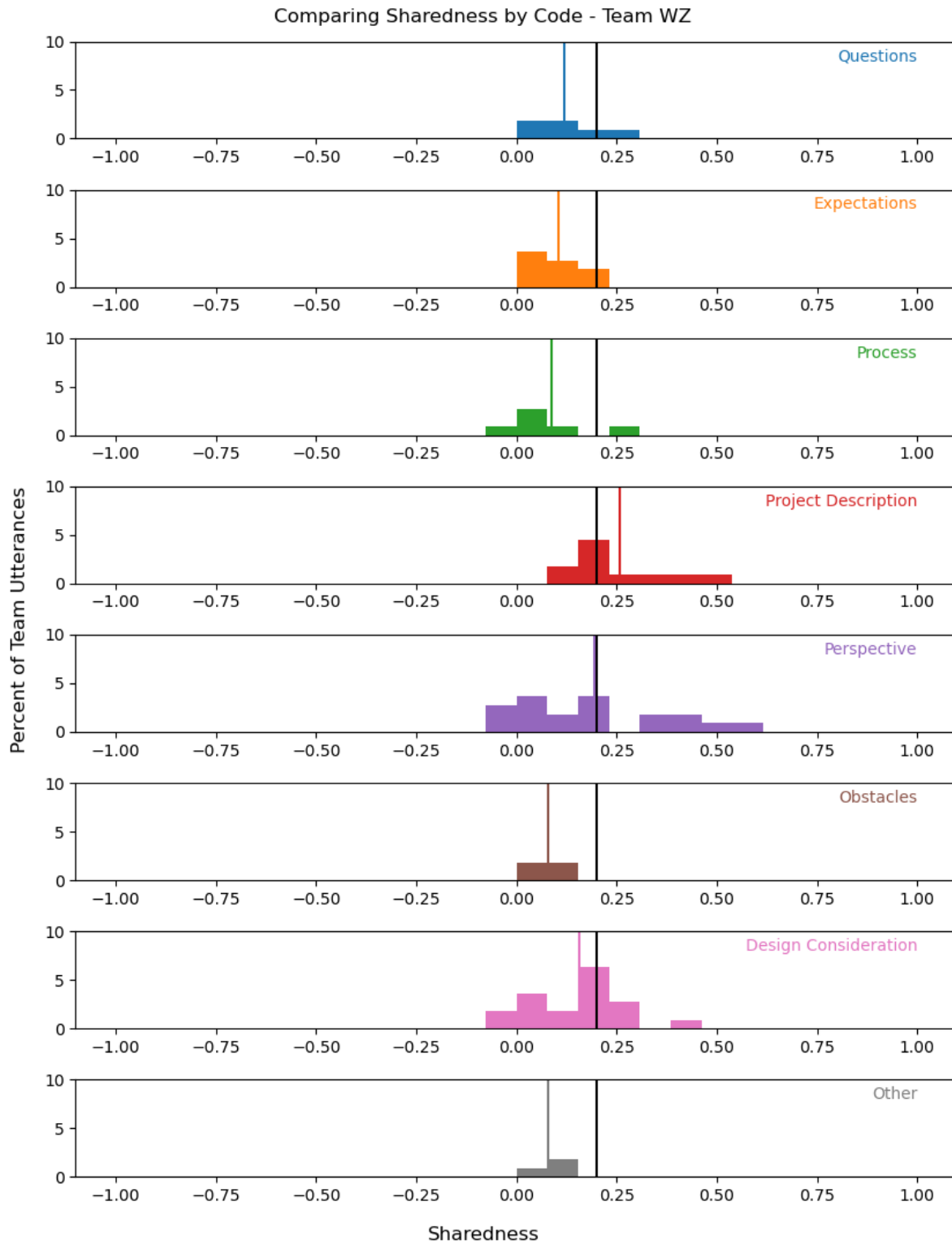


Figure B.7 – Histograms showing the distributions of sharedness values for each code (Team WZ).

Team ZH

Table B.15 - Mann Whitney U-Test by Utterance Type – Team ZH		
Utterance Type	Less than	Greater than
Questions		
Expectations	U=153.5 p=0.802	U=153.5 p=0.206
Process	U=124.0 p=0.250	U=124.0 p=0.759
Project Description	U=409.0 p=0.999	U=409.0 p=0.001
Perspective	U=3755.5 p=0.009	U=3755.5 p=0.991
Obstacles	U=187.5 p=0.218	U=187.5 p=0.788
Design Consideration	U=73.5 p=0.485	U=73.5 p=0.531
Other		

Table B.16 – Comparing Sharedness Descriptive Statistics by Utterance type - Team ZH				
Utterance Type	Count	Sharedness Median	Sharedness Mean	Sharedness Std. Dev.
Questions	0 (0.0%)			
Expectations	5 (10.0%)	0.202	0.293	0.188
Process	6 (12.0%)	0.104	0.179	0.180
Project Description	10 (20.0%)	0.522	0.458	0.159
Perspective	17 (34.0%)	0.072	0.126	0.133
Obstacles	9 (18.0%)	0.093	0.152	0.132
Design Consideration	3 (6.0%)	0.155	0.219	0.200
Other	0 (0.0%)			
All Combined	50 (100.0%)	0.153	0.226	0.200

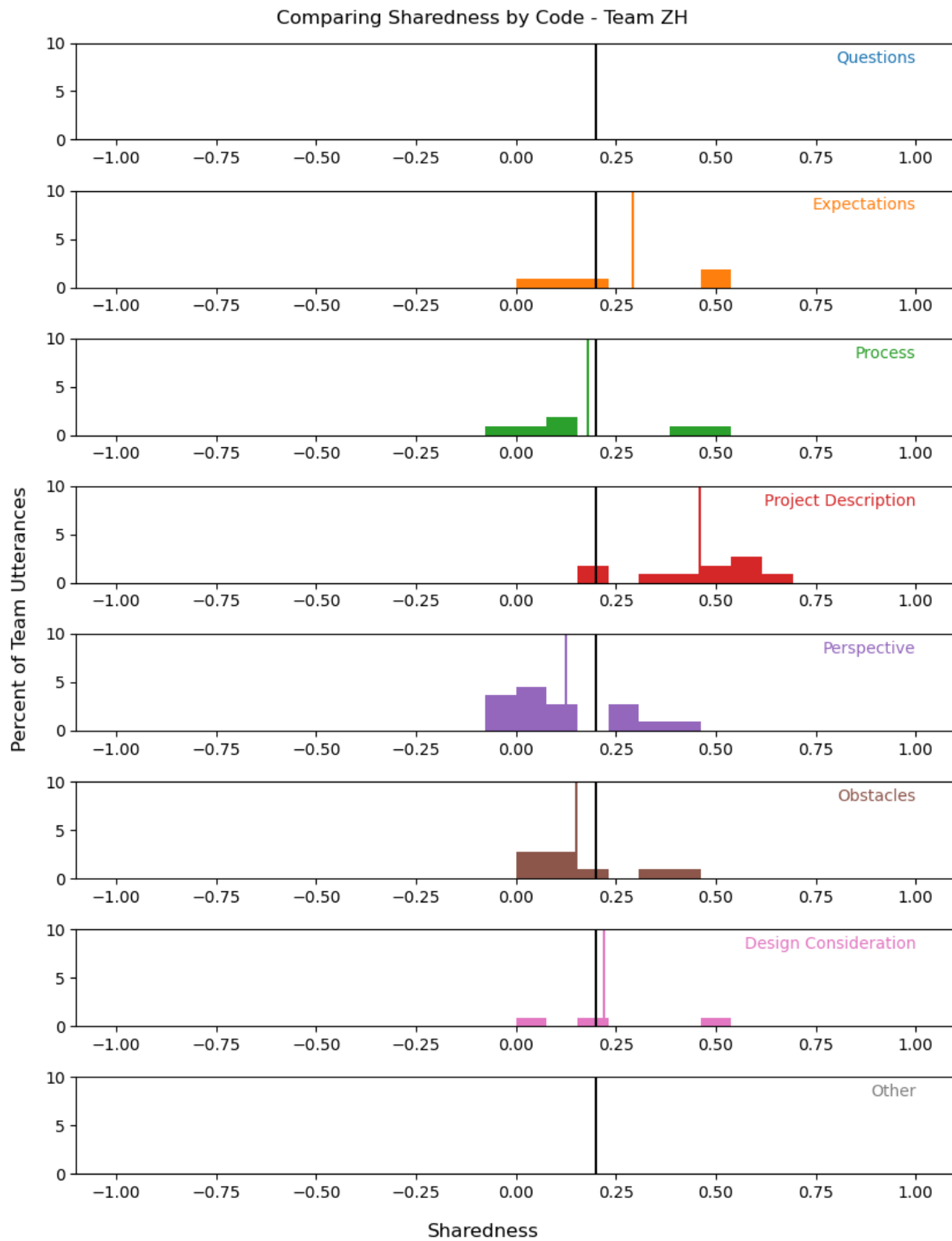


Figure B.8 – Histograms showing the distributions of sharedness values for each code (Team ZH).