12-2011

Evaluating the Use of Functional Representations for Ideation in Conceptual Design

Benjamin Caldwell
Clemson University, bwcaldw@clemson.edu

EVALUATING THE USE OF FUNCTIONAL REPRESENTATIONS
FOR IDEATION IN CONCEPTUAL DESIGN

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mechanical Engineering

---

by
Benjamin William Caldwell
December 2011

Accepted by:
Gregory M. Mocko, Committee Chair
Joshua D. Summers
Georges M. Fadel
Joel S. Greenstein

ABSTRACT

Functional representations are often used in the conceptual stages of design because they encourage the designer to focus on the intended use and purpose of a system rather than the physical solution. Function models have been proposed by many researchers as a tool to expand the solution search space and guide concept generation, and many design tools have been created to support function-based design. These tools require designers to create function models of new or existing artifacts, but there is limited published research describing what types of functions should be included in a model or the appropriate level of abstraction to model artifacts. Further, there is little experimental evidence that function models are useful for concept generation. Therefore, this research focuses on how artifacts should be modeled to support ideation in conceptual design.

In this research, three functional representations are studied: function models, interaction models, and pruned function models. First, a user study is conducted to test the level of understanding of functional representations by designers. Second, a computational similarity metric is used to identify the appropriate level of abstraction for creating models. Third, a user study is conducted to determine the effects and usefulness of functional representations in concept generation. The three studies show that pruned function models are easier to understand, improve the use of the model by designers, improve the quality of concepts generated, and are more useful for computing functional similarity. Function models contain additional, solution-specific descriptions of

functionality that are not useful in conceptual design for ideation, similarity, or interpretation. The interaction model, which is developed in this research, provides a preliminary representation capable of capturing user actions and interactions in addition to artifact functionality, and shows potential for describing non-functional requirements in a manner that is useful to designers. These outcomes serve as a foundation for guidelines for creating conceptual-level models that support ideation in conceptual design.

To Amanda, Evelyn, and Lucy

ACKNOWLEDGMENTS

I would like to thank my wife, Amanda, for her dedicated love and service to me over the past five years. Without her sacrifice, I would not have been able to complete my dissertation. She is a wonderful mother to Evelyn and Lucy and a great example to those around her. I also want to thank our two girls, Evelyn and Lucy, for their patience and love for their daddy and for the joy they bring me every day.

I want to thank Dr. Mocko for advising me during my master's and doctoral research, helping me grow personally, as a researcher, and as a teacher. He has been a great mentor and I have learned immensely from him over the past five years. He has challenged me in all areas of life and I have truly enjoyed working with him.

I also want to thank Dr. Summers for his valuable feedback on my research and for investing in me during my graduate work at Clemson. Thanks also to Dr. Fadel and Dr. Greenstein for serving on my committee and for their valuable insights in my work.

I want to thank the students of CEDAR, both past and present, for their friendships and collaboration. I have enjoyed working with all of the students in the lab and am grateful the opportunity I had to work with them.

I also want to thank the National Science Foundation, the American Society of Mechanical Engineers, and the Department of Mechanical Engineering for funding during my work as a doctoral student.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1: INTRODUCTION

1.1 Systematic Design Methods

One main goal of engineering design research is to understand how engineers should solve design problems in a manner that is consistent and repeatable. If a general, repeatable process or set of design tools can be developed and taught to engineers, then engineers will be able to address any design problem using the same approach, ensuring success with any project. Many design textbooks have been published describing systematic design processes, most of which follow the same overall approach, shown in Figure 1-1 [1-4].

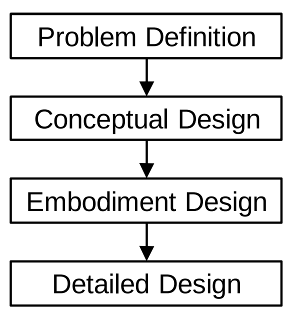


Figure 1-1: General Mechanical Design Process

The problem must first be understood and defined in the problem definition phase. To define the problem, designers must understand the needs of the customer, which can be identified through interviews, focus groups, previous designs, or other methods. Engineers use these customer needs to identify engineering requirements for the design problem, which are more quantitative than customer needs and begin to define the scope of the problem.

After defining the problem, designers move to the conceptual design phase, where they decompose the overall problem into sub-problems and identify means for solving the smaller problems. In conceptual design, engineers may identify high-level ideas of how to solve the problem and begin to sketch out these ideas. Typically, the designer will develop many concepts and will evaluate them, narrowing them down to several that are plausible. Designers will continue development of these plausible concepts in the embodiment design phase.

During embodiment design, designers further develop each concept, ensuring that the ideas can be accomplished. Designers may build prototypes as proof-of-concepts and lay out the preliminary architecture of the final design. Designers begin to identify more-specific means for accomplishing the sub-problems identified in conceptual design, and begin sizing components and subsystems based on the information known at that point in the process. Through embodiment design, designers are able to compare the various concepts that were pursued and typically choose one design—the best solution to the problem—to carry to the detailed design phase.

During detailed design, engineers know the layout of the artifact being designed and can begin to specify existing components (e.g., motors, gears, bolts, screws) or design new components for the final product (e.g., housing). Designers may create Computer Aided Design (CAD) models of the design, build complete prototypes for testing, or analyze the design using Computer Aided Engineering (CAE) tools, such as Finite Element Analysis (FEA) or Computational Fluid Dynamics (CFD).

To this end, design research focuses on the development of tools and methods that support a systematic approach to engineering problems. For example, development of FEA and related software tools has significantly enhanced the ability of engineers to analyze systems without building a physical prototype. Many designers use Failure Modes and Effects Analysis (FMEA), in which they identify failure modes and their likelihood, severity, and detection, prioritizing specific components or systems that are critical. Likewise, tools and methods have been developed for conceptual design to help designers systematically approach concept development. One such tool, a function model, can be used in conjunction with a function-based approach to conceptual design.

1.2 Function-based Conceptual Design

Many design processes prescribe a function-first approach to conceptual design, where designers establish the function of an artifact after identifying engineering requirements [1-4]. There are many differing definitions of the term function [1, 5-8], but all function-based approaches focus on what the designed artifact should do to satisfy the requirements, rather than what the design will look like. For example, if designing an electric drill, a designer may focus on the fact that the drill must create rotational output instead of focusing on using a motor. This allows the designer to explore other ideas besides a motor to accomplish the task of creating rotation. In this manner, a designer may be able to develop ideas such as a pneumatic or gas-powered drill, both of which exist in the consumer market.

The use of function in engineering design has been promoted by many researchers as a means for problem decomposition and concept generation. Although there are many definitions and views of function, it has become an underlying theme behind many design processes, primarily due to its ability to aid in the conceptual design stage where form is not yet critical [1-4, 7, 9-14]. However, many researchers recognize that function-based approaches have limitations and pursue other concepts, such as affordances [5, 15-17], interfaces [18], or usage [19-22]. These approaches have not yet been widely accepted, but they have been introduced more recently than function and are still being developed. In this research, it is postulated that function-based approaches are fundamental to design but do not sufficiently address all aspects of a designed artifact. Therefore, the use of function modeling in conceptual design is studied in addition to complementary and alternative approaches to function in conceptual design.

1.3 Motivation

Many design tools and methods have been developed within the design community to support function-based design. These tools and methods typically rely on previous design knowledge and function models of existing artifacts. These models are created through reverse-engineering and include many details about a device that would not be known at the conceptual phase of design. However, function models are intended to support conceptual design. If a designer creates a model in conceptual design, it will be more abstract than a model of an existing system. When using design tools that are based on previous knowledge, it is important to understand the appropriate level of abstraction to create a model of both the existing system as well as the archived artifacts. There is limited published research describing what types of functions should be included in a function model of a new artifact or the appropriate level of abstraction to model an existing artifact. Further, the modeling methods that have been described have not been validated through user testing. Therefore, this research focuses on how artifacts should be modeled in conceptual design. The development of a modeling method is outside the scope of this research, but the outcomes of this research can be directly used to create a modeling method that should be validated through user experiments. The overall research question pursued is:

> **Overall Research Question:** How should the functionality of mechanical artifacts be modeled to support ideation in conceptual design?

CHAPTER 2: REVIEW OF RELEVANT LITERATURE

2.1 Function-Based Design

2.1.1 Overview

Function models are often used in the conceptual stages of design because they encourage the designer to focus on the intended use and purpose of a system rather than the physical solution. Function models have been proposed by many designer researchers as a tool to expand the solution search space and guide concept generation. For example, Pahl and Beitz [1] suggest that function models provide a means for systematically creating design variants and better exploring the solution space by linking product functions in several ways. Ulrich and Eppinger [2] and Ullman [3] propose problem decomposition, specifically functional decomposition, as a means for addressing a complex design problem, finding solutions for individual functions, and integrating these solutions into the system. Otto and Wood [4] propose the use of function models as a reverse engineering tool to understand the purpose of systems and components of existing products.

Function-based approaches to conceptual design are prescribed by many design texts [1-4], and one focus of recent design research is the area of function modeling. Views and definitions of function vary among researchers [23], but most focus on what an artifact does rather than how it does it. Designers use various representations to describe "what" an artifact must do as opposed to "how" an artifact must complete a task during the conceptual design phase [4]. The definition of function used in this research is a transformative view of function, defined by Pahl and Beitz as "the intended input/output relation of a system whose purpose is to perform a task" [1]. The primary representation pursued in this research is the function structure, which is a graphical representation of the transformation of flows through an artifact. The basic elements of this representation, shown in Figure 2-1, are material flows (bold arrow), energy flows (thin arrow), and information flows (dashed arrow) which are transformed by a function (block).

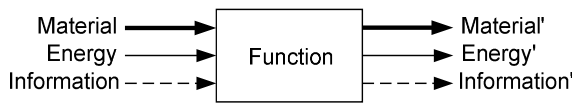


Figure 2-1: Generic Function Block with Flows of Material, Energy, and Information [1]

An artifact can perform many functions, which can be modeled using multiple function blocks and the passage of flows into the artifact's system, among function blocks, and out of the system. An example of a function structure of an electric drill, shown in Figure 2-2, includes four functions performed by the drill: (1) *convert human energy to on/off signal*, which is performed by a switch, (2) *actuate electricity*, which is also performed by a switch, (3) *convert electricity to rotation*, which is performed by a motor, and (4) *increase torque*, which is performed by a gear box. Flows of *electricity* and *human energy* enter the drill, and *rotational energy* is an output. The level of abstraction at which the drill is modeled affects the functions included in the model. For example, the functionality of wires or shafts in the drill could be included in the model.

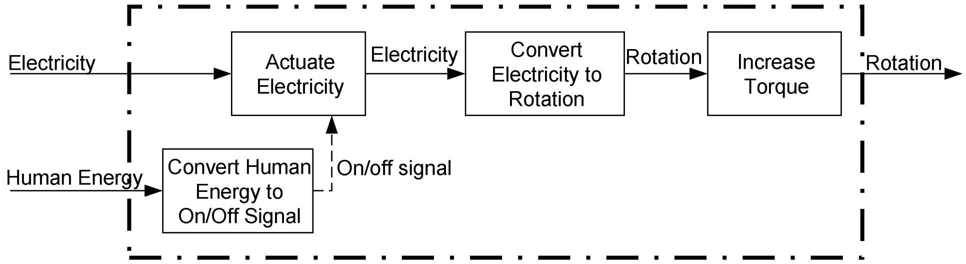


Figure 2-2: Function Structure of a Drill

Function model formalization is important for repeatable and meaningful results [24], and current design research has assisted the formalization of functional modeling, such as the development of a Functional Basis [24], a design repository [25], pruning rules for function structures [26], and development of a physics-based representation of functions [27]. However, much of this research focuses on the reverse engineering and modeling of existing artifacts. Models of existing artifacts can be useful for information archival and a function-based search for solutions to a new design problem. The modeling process can also be useful to the modeler by forcing him or her to understand how the artifact functions and communicate it clearly. The modeling process for forward design may help the designer decompose the problem functionally, understand the problem better, and identify several ways to solve the problem. When creating a function model for a new artifact, the designer must make decisions about the new design as he or she creates the model, resulting in a model or several models that can be used to address the given design problem.

The information gained through modeling a new artifact is different from that gained by modeling an existing artifact. Likewise, the value of the model of a new artifact is different from the value of a model of an existing artifact. Therefore, the purpose of modeling an existing or new artifact must be coupled with the methods used to create the models. The appropriateness of models should reflect their eventual usage by designers, whether for communication, archival, ideation, analysis or other design activities.

Summary

- Function-based design approaches focus on transformations of material, energy, and information through an artifact.
- Function models can be used to describe the functionality of existing or new mechanical artifacts.
- Methods for creating function models should be coupled with the use of the models for design activities.

2.1.2 Functional Basis and Design Repository

Recent efforts in function modeling have focused on the formalization of function models using a controlled vocabulary [12, 13, 24]. The Functional Basis is a functional vocabulary that includes 53 function terms and 45 flow terms and definitions of each. The Functional Basis function and flow sets are each organized in a three-level hierarchy. Primary-level terms, such as *energy*, are more abstract while tertiary-level terms, such as *rotational mechanical energy*, contain more detail. Previous research has shown that the secondary level is the most informative [28, 29] and is used almost exclusively by modelers [30]. Thus, when the Functional Basis vocabulary is used in this research, all functions and flows are modeled using the secondary level.

The Functional Basis has been used to describe the functionality of approximately 130 artifacts, ranging from consumer artifacts to natural systems, in an online design repository [31]. The repository contains functional information about each component of the 130 artifacts in the repository. Each component of an artifact is assigned functions in the form <input flow> <function> <output flow>, where the input and output flows are chosen from the flow vocabulary and the function is chosen from the function vocabulary. Further, a graphical function structure of each artifact can be stored as an image in the design repository describing the functionality of the entire artifact, rather than individual components. Most artifacts in the repository are kitchen appliances, power tools, toys, or electronics, but the repository also includes artifacts from other domains, such as living organisms (e.g., "fly," "lichen," and "heart") and component failure data (e.g., "asm volume 1," "cpsc failure"). The functional information in the repository has been used with many computational design tools, such as automated concept generation [32-39], function-based similarity measures [40, 41], failure and risk analysis [42-47], behavior modeling [48, 49], and biomimicry [50-53]. Since these design tools use the functional information in the repository, it is important that the models stored in the repository capture the appropriate functional information.

The information contained in the repository is used in this research as a source of design knowledge. It is assumed that the information in the repository was systematically obtained using the reverse-engineering methods described by the researchers associated with the design repository [12, 54, 55]. However, these methods allow some freedom for modelers to deviate from the vocabulary (for graphical models) or modeling guidelines. Modelers may use the vocabulary at any hierarchical level or free language as they see fit [30, 56]. The functions in the models are not required to follow laws of conservation of mass or energy, so the resulting models may be logically inconsistent [27, 57]. Furthermore, the traditional transformative view of function has been informally extended in some models to include interactions and assembly relationships, which are explored in more detail in Section 2.1.3.

Summary

- Function models in the design repository do not always adhere to the Functional Basis vocabulary or modeling rules.
- Function models in the design repository have informally extended the traditional definition of function.
- These extensions can be identified and evaluated to determine if they are appropriate for function-based conceptual design.

2.1.3 Current State of Function-Based Design

There are many aspects of artifacts that cannot be described using the traditional definition of function—a transformation of flows. However, recent research has extended this view of function to include assembly relationships, environmental interactions, and human interactions [24]. As an example, the Black and Decker Jigsaw

Attachment (Figure 2-3a) is a consumer power tool that can be attached to a universal driver (Figure 2-3b) to create a typical jigsaw (Figure 2-3c).

Figure 2-3: Black and Decker (a) Jigsaw Attachment, (b) Universal Driver, and (c) Driver-Attachment Assembly (image source: www.blackanddecker.com)

The design repository contains a function structure of the jigsaw attachment [31], reproduced in Figure 2-4. In addition to the function of the artifact, this model contains interactions with the user and other artifacts. First, the chain of functions *import human material, guide human material, export human material* (labeled 1) represents the user physically picking up the artifact and carrying or manipulating it. Second, the function chain *import solid, guide solid, export solid* and the function, *import rotational energy* (labeled 2) represents the physical connection between the universal driver and the jigsaw attachment. Third, the function *import human energy* (labeled 3) represents the user's control of the driver-attachment system by pressing the switch. Fourth, the *import solid,* *secure solid*, *export solid* function chain (labeled 4) represents the physical connection of a saw blade to the jigsaw attachment. These functions take place at the jigsaw attachment's system boundary, not within the system, and are executed by the user. The artifact is designed to allow these interactions to take place, but it does not actively perform these functions.

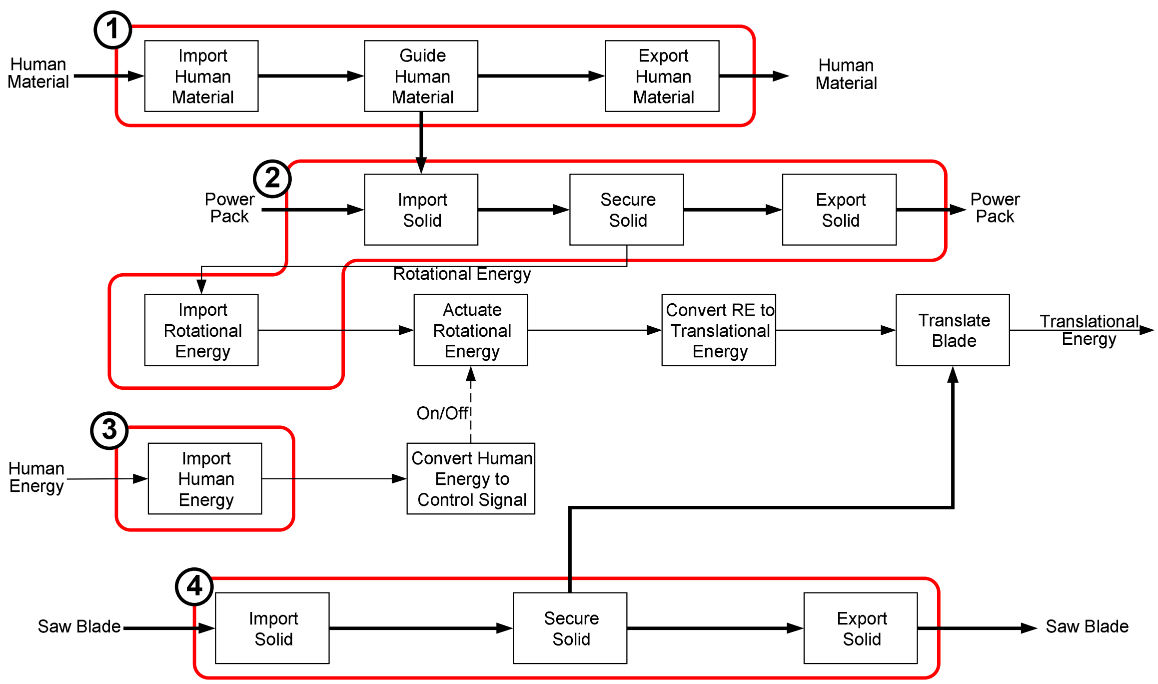


Figure 2-4: Function Structure of a Black and Decker Jigsaw Attachment Showing User and Artifact Interactions (adapted from [31])

Eleven of the functions used to describe the Black and Decker Jigsaw Attachment are not actively performed by the artifact; they are passive functions that do not represent transformative actions. *Passive functions* are defined in this research as functions in which the artifact of interest does not carry the energy used to control the outcome of the function. For example, the Black and Decker Jigsaw Attachment does not provide the energy to import the driver—the user provides the energy. The user also provides the energy to secure the blade to the jigsaw attachment.

> *Passive functions* are functions in which the artifact of interest does not carry the energy used to control the outcome of the function.

A type of passive function appears in previous literature in the form of a supporting function, which is used to describe assembly relationships between components [58]. Supporting functions are modeled separately from a function structure and show physical connections and assemblies. Supporting functions cannot be incorporated into the system-level function structure because components of the system are flows in the supporting function. For example, the supporting function of a screw in a drill assembly may be to *couple the left and right housing*, where the left and right housing are two plastic components that hold the drill assembly together and form the handle of the drill. As shown in Figure 2-5, some components of the system—the left and right housing—are flows in the model, while the screw is represented by a function. The modeling of supporting functions requires the designer to reverse-engineer the artifact because supporting functions describe the functionality of individual components [58]. The goal of the model proposed in this research is to describe interactions at a higher level of abstraction than the component level. For this reason, passive and supporting functions are not included in function structures used in this research.

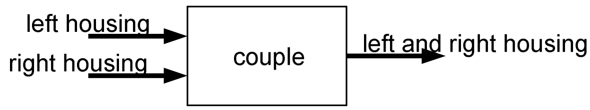


Figure 2-5: Supporting Function of a Screw

Conversely, some of the functions in the previous two examples are transformative functions, which are actively performed by the artifact being modeled. *Active functions* are defined as functions in which the artifact of interest carries an energy flow that is used to control the outcome of the function. For example, the function of *converting electrical energy to rotational energy* is active because the artifact carries the electrical energy that is used to produce the desired outcome of the function—rotational energy.

Active functions are functions in which the artifact of interest carries an energy flow that is used to control the outcome of the function.

Function structures also frequently contain user activities, which are not performed by the artifact, but are performed on the artifact. Because the energy in user activities is provided by the user, they are passive. Kostovich and colleagues have also identified user activities in function structures and have intentionally combined activity models and function structures into an "actionfunction diagram," capturing both user activities and artifact functions [59]. However, in these actionfunction diagrams, both user activities and artifact function are used to describe user activities. The passive functions remain in the model; the functions are simply grouped according to the activity being performed on the artifact when the function is carried out. For example, the authors present an actionfunction diagram of a typical box cutter, shown in Figure 2-6. The diagram describes the purpose of the handle in two different ways: first, with the series of functions, *import hand*, *position hand*, *secure hand*, and second with the activity *grab handle*. The latter is a simpler representation of the same event—the user picking up the artifact. Therefore, one of the two representations is redundant; in the model proposed in this research, user activities are used instead of passive functions to simplify the representation. Additionally, the user activity approach enables the passive functions to be represented actively as user activities, since the energy used to perform these activities is usually carried by the user.

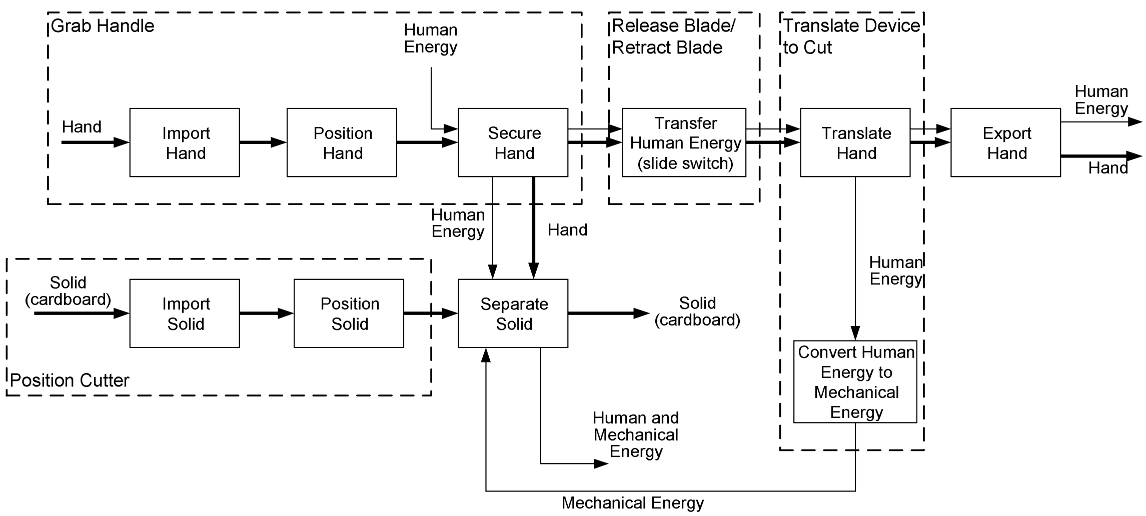


Figure 2-6: Actionfunction Diagram of a Typical Box Cutter [59]

Summary

- Function models have been extended to include passive functions, supporting functions, and interactions.

- Passive functions are performed *on* an artifact by an outside entity, not *by* the artifact being modeled.
- Supporting functions describe assembly relationships among components, do not transform the components, and are not form-independent.
- Function models should include only active functions.

2.1.4 Pruning Rules for Conceptual Modeling

In previous research [26, 60], a set of pruning rules has been developed based on the Functional Basis [24] to increase the level of abstraction of function structures. These rules specify the removal of highly decomposed functions that are less critical at early stages of design [26, 60] as well as passive functions that are not transformative. There are nine function rules (PR 1-9) and six flow rules (PR 10-15). The procedure for applying the rules to a function structure is discussed in detail in [60]. The fifteen pruning rules are:

PR1. Remove all *import* and *export* functions.

PR2. Remove all *channel*, *transfer*, *guide*, *transport*, *transmit*, *translate*, *rotate*, and *allow DOF* functions referring to any type of *energy*, *signals*, or *human material*.

PR3. Remove all *couple*, *join*, and *link* functions referring to any type of *solid*.

PR4. Remove all *support*, *stabilize*, *secure*, and *position* functions.

PR5. Remove all *control magnitude*, *actuate*, *regulate*, *change*, *stop*, *increase*, *decrease*, *increment*, *decrement*, *shape*, *condition*, *prevent*, and *inhibit* functions.

PR6. Remove all *provision*, *store*, *supply*, *contain*, and *collect* functions referring to any type of *energy* or *signal*.

PR7. Remove all *distribute* functions referring to any type of *energy*.

PR8. Remove all *signal*, *sense*, *indicate*, *process*, *detect*, *measure*, *track*, and *display* functions.

PR9. Combine adjacent *convert* functions if the output flows of the first function block are identical to the inputs of the second function block.

PR10. If a flow enters and exits a function block, then the two separate flows should be combined into one flow.

PR11. If a flow enters a function block but does not exit the function block, then the flow should enter the succeeding function block.

PR12. If branch, separate, or distribute is removed, then the flow entering the block should be divided without the use of the function.

PR13. If two convert functions are combined, then the flow between the adjacent functions should be removed.

PR14. If a flow exists without a function, then the flow should be removed.

PR15. If identical flows have the same origin and destination, then the flows should be combined into one flow.

These pruning rules have been demonstrated through their application to eleven function models of consumer, electromechanical artifacts [26, 60], giving confidence that the rules can be used to achieve high levels of abstraction that may be useful in conceptual design. However, there is an opportunity to test the use of these rules by human designers and within computational tools to understand their usefulness in conceptual design.

Summary

- Pruning rules have been developed to achieve a consistent, high-level of abstraction of function structures.
- There is an opportunity to test the pruning rules to understand their usefulness in conceptual design.

2.2 Alternative Approaches to Function-Based Design

Function-based approaches to design, which have been accepted by many researchers, intentionally focus on transformations of material, energy, and information through the artifact. In some cases, proponents of function-based design may purposely ignore non-functional aspects of an artifact early in the design process, viewing this as an advantage of function-based design. In other cases, function-based approaches have been extended to include some of these non-functional aspects, such as assembly relationships or human interactions. Other researchers, however, have identified the value in these non-functional aspects of artifacts and taken non-functional approaches, using concepts such as affordances or interactions.

Affordances, which describe what one artifact provides other artifacts and users, have been extended to mechanical design from the field of perceptual psychology [15]. Artifact-artifact and artifact-user affordances describe the perceived relationships between two artifacts or between an artifact and user, respectively. For example, gears afford mating with other gears, and a lightweight artifact affords being picked up by a user. Affordances are not limited to these relationships; they are used to describe the entire lifecycle of an artifact. Artifacts afford improvement, sustainability, maintenance, manufacturing, and desired purposes, to name a few [16]. Affordances can also be used to describe services, structures, and space. Kim and colleagues have analyzed user activities to determine perceived affordances of a building lobby [61]. In addition, affordances can be an evaluation tool used to identify potential hazards and failure modes in design [5, 62]. The scope of affordances—the complete lifecycle of an artifact, structure, or space—is greater than the scope of this research, which is focused on artifact design. Because of this large scope, affordances are not pursued in this research as a complement to function-based design.

An artifact may interact with other artifacts, a user, or the environment in various ways. Affordances can describe these interactions, but they also cover many other aspects of artifacts. Galvao and Sato describe interactions between an artifact and a user through functional-level and operational-level affordances [17]. It is this subset of affordances—interactions—that is of interest in this research. However, interactions with other artifacts are considered in addition to user interactions discussed by Galvao and Sato [17].

Nagel and colleagues have extended function modeling to support various levels of abstraction, including the environment at the highest level [63]. The environment interacts with the system as a flow of material into the system, requiring all interactions be modeled as flows through a system.

Chandrasekaran and Josephson [6] discuss various views of function, interfaces, and interactions within an ontological model of artifacts. Of particular interest are causal interactions, which are physical interactions that exist between artifacts [6]. The model developed by Chandrasekaran and Josephson is a computational model to support automated reasoning and requires detailed information about an artifact, such as design variables, causal interactions, and structural relations. While these details are not fully known during conceptual design, the ontology and representation may be applicable to this research, so it can be pursued as a potential solution to non-transformative aspects of artifacts. However, this approach alone is not sufficient since it holds a different view of function and does not support graph-based modeling.

Warell [64] discusses three types of functions: operative, structural, and usability functions. Of interest are usability functions, which describe the interactions between an artifact and the user and other systems. Warell demonstrates the use of usability functions through an example of a mobile phone. The usability function of various components, such as the cover or hinges, is described using natural language. The graphical models proposed in this research can extend Warell's research, relating usability functions, or user interactions, to the artifact's technical function in a graphical model.

Summary

- The scope of affordances is inappropriate for application to this research.
- Current integrations of transformative functions with non-transformative functions are limited to directional interactions with material that can be modeled as a flow through the system.
- Alternative approaches have different definitions of function, which do not support a graphical modeling tool.

2.3 Limitations of Function-Based Design

Function-based design approaches intentionally focus on function at early stages of design, so the type of information that can be modeled within function-based approaches is limited. However, customer needs, which are statements about an artifact from a prospective user [2], have a much larger scope in terms of the type of information that they can capture. In this section, customer needs are reviewed to determine how various types of needs can be modeled using a function structure. Any needs that cannot be modeled in a function structure are identified as opportunities for extending function-based design tools.

Customer needs statements describe the desires of eventual customers, are developed before any solution is known, and can be identified through interviews, focus groups, and analysis of existing artifacts [2]. A set of customer need statements for a bicycle suspension is shown in Table 2-1 (bold statements from [2]). Because this artifact exists, the customer needs can be related to the artifact's embodiment and its functionality. In this context, a function is defined as a transformation of material, energy, or information by the artifact of interest. Using this definition, each customer need is viewed from a functional perspective to determine if it can be modeled as a system flow or a transformation of flows. A high-level function structure of a bicycle suspension, shown in Figure 2-7, is used in this analysis.

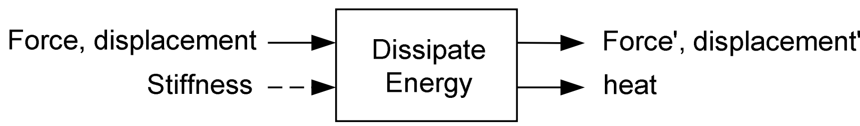


Figure 2-7: High-level Function Structure of a Mountain Bicycle Suspension

The analysis, shown in Table 2-1, relates each customer need statement to the suspension's architecture and identifies any functional element that can be associated with the given need.

Table 2-1: Bicycle Suspension Customer Needs (Bold statements from [2])

	Analysis of Customer Need	Functional Element
1	**The suspension reduces vibration to the hands.** This customer need leads to the functionality of dissipating energy. The input to the artifact is mechanical energy, and the output is another form of energy, such as heat.	energy input
2	**The suspension allows easy traversal of slow, difficult terrain.** This customer need describes the energy input into the artifact, "slow, difficult terrain," implying a low-frequency input with varying amplitude. Therefore, this requirement describes the intended functionality of the suspension.	energy input
3	**The suspension enables high-speed descents on bumpy trails.** Like the second customer need, this need describes the energy input, force and displacement, to the system.	energy input
4	**The suspension allows sensitivity adjustment.** This customer need leads to a user-adjustable input to the system. Therefore, it can be represented as a signal entering the function block.	signal input
5	**The suspension preserves the steering characteristics of the bike.** The bicycle suspension design does not transform material, energy, or information to meet this need—it simply has a similar form as a traditional bicycle fork. Therefore, this customer need leads to a non-functional solution.	none
6	**The suspension remains rigid during hard cornering.** The response of the suspension to various inputs can be shown through various inputs to the function structure.	energy input
7	**The suspension is lightweight.** The weight of the artifact is a property of the system and cannot be represented as a flow or as a transformation of flows.	none
8	**The suspension provides stiff mounting points for the brakes.** This customer need describes the interaction required between the suspension and typical bicycle brakes. This interaction cannot be described as a flow through the suspension or as a transformation by the suspension.	none
9	**The suspension fits a wide variety of bikes, wheels, and tires.** Like the previous customer need, this describes the interaction between the suspension and other bicycle components and cannot be represented in a function structure.	none
10	**The suspension is easy to install.** The ease of installation describes how a user interacts with the system. The installation cannot be shown as a function because it would require that the suspension itself be a material flow.	none

	Analysis of Customer Need	Functional Element
11	**The suspension works with fenders.** Similar to the seventh and eighth customer needs, this customer need requires that the suspension interact with typical fenders. This interaction cannot be shown as a function of the suspension; if the fender enters the function structure, it cannot be transformed by the suspension.	none
12	**The suspension instills pride.** The user's perception of the suspension is subjective and cannot be represented by a transformation of material, energy, or information. Thus, it is non-functional.	none
13	**The suspension is affordable for an amateur enthusiast.** The cost of the suspension cannot be shown in a function structure as a flow or a transformation of flows.	none
14	**The suspension is not contaminated by water.** This need may be represented in a function structure by introducing a new flow of water into the system and showing a transformation of the location of water. However, this need can also lead to the use of corrosion-resistant materials; the designer may choose to represent it using a function structure.	material input
15	**The suspension is not contaminated by grunge.** As in the fourteenth customer need, it is possible to represent this need in a function structure.	material input
16	**The suspension can be easily accessed for maintenance.** This customer need describes the speed that the user can assemble or disassemble the suspension to access components that require maintenance. If this process were shown as a function, it would require that components be flows of material. Since components cannot be flows, this customer need is non-functional.	none
17	**The suspension allows easy replacement of worn parts.** As in the sixteenth customer need, this need cannot be shown in a function structure.	none
18	**The suspension can be maintained with readily available tools.** As in the previous two customer needs, this need describes maintenance, which cannot be shown as a transformation of flows. Therefore, it is non-functional.	none
19	**The suspension lasts a long time.** The life of the product cannot be described using a flow or a transformation of flows.	none
20	**The suspension is safe in a crash.** The safety of suspension is related to its strength, especially in bending. The crash scenario could be shown in a function structure as a different input to the system.	energy input

As shown in Table 2-1, only eight of the twenty customer needs can be modeled using a function structure element, and many could be improved. For example, needs 2, 3, 6, and 20 must all be modeled through different energy inputs into the system. While this is possible, there may be alternative models than can capture these needs in a more meaningful or useful manner. Many systematic design processes prescribe the use of function after identifying customer needs and engineering requirements. Since many of the customer needs cannot be described using a transformation of flows, these needs are not addressed through traditional function structures. Therefore, traditional function-based methods should be supplemented so that designers can address these non-functional needs early in the design process.

Summary

- Function-based design methods support only a subset of customer needs.
- Many customer needs describe interactions, which cannot be modeled using a function structure.
- Some customer needs that can be modeled using function structures can likely be better modeled with other approaches.

2.4 Ideation

Design thinking has been described as a divergent-convergent process, where designers may ask both divergent and convergent questions [65]. Divergent questions lead to many possibilities that can be explored, while convergent questions lead to a deeper understanding of the problem based on engineering knowledge and analysis. The divergent thinking process has been a focus of engineering design research, where the goal is to generate novel solutions to a problem [66-70]. This divergent thinking is important in design since it expands the solution space explored by designers and may lead to innovative ideas. The convergent thinking process is also important for ideation since it may help designers understand concepts that have been developed, evaluate their feasibility, and ultimately converge on a solution to the problem. The convergent thinking process is the focus of this research, and the goal is to support convergence on a high-quality concept rather than a novel or innovative concept. The use of functional representations as a seed for ideation is studied to determine if they yield high-quality functional concepts. The ideation process and ideation techniques are not the focus of this research, but the outcome of the ideation process is used to understand the effects of seed models on ideation.

2.5 User Studies in Design Research

User studies have been conducted in engineering design research to understand the effects of design tools and methods on design activities. For example, Linsey and colleagues studied fixation within design teams by giving design teams of engineering faculty a sample solution to a design problem, intending to induce fixation, along with methods to reduce fixation [71]. Chan and colleagues determined through a user study that far-field, less-common analogies as provocative stimuli improves the novelty of solutions generated by designers [66]. Many other user studies in the field of engineering design have been conducted and use students as participants to evaluate design methods

[67-69, 72-78]. These user studies typically require participants to generate sketches, which are quantified to test the effects of the factors being studied. Metrics of quality, quantity, novelty, and variety [79] or a subset of these metrics are often used to evaluate the sketches. In this research, metrics of quality and quantity are used due to the focus on convergent rather than divergent ideation processes.

Frey and Dym suggest that design research should borrow methods from the medical research field since medical research methods have been used and developed extensively for medical treatments [80]. Frey and Dym state that user studies conducted in a controlled laboratory setting are analogous to in vitro experiments in the medical field, which are part of the overall validation process for medical treatments [80]. Therefore, user studies are conducted in this research with student participants to provide experimental evidence of the effects of functional representations on concept generation in design, providing an experimental layer of validation of the use of functional representations in conceptual design.

CHAPTER 3: RESEARCH APPROACH

3.1 Research Gaps and Opportunities

Based on the review of function-based design, the following research opportunities exist:

Function-Based Design

- Function-based design approaches focus on transformations of material, energy, and information through an artifact.
- Function models can be used to describe the functionality of existing or new mechanical artifacts.
- Methods for creating function models should be coupled with the use of the models for design activities.

Functional Basis and Design Repository

- Function models in the design repository do not always adhere to the Functional Basis vocabulary or modeling rules.
- Function models in the design repository have informally extended the traditional definition of function.
- These extensions can be identified and evaluated to determine if they are appropriate for function-based conceptual design.

Current State of Function-Based Design

- Function models have been extended to include passive functions, supporting functions, and interactions.
- Passive functions are performed *on* an artifact by an outside entity, not *by* the artifact being modeled.
- Supporting functions describe assembly relationships among components, do not transform the components, and are not form-independent.
- Function models should include only active functions.

Pruning Rules for Conceptual Modeling

- Pruning rules have been developed to achieve a consistent, high-level of abstraction of function structures.
- There is an opportunity to test the pruning rules to understand their usefulness in conceptual design.

Alternative Approaches to Function-Based Design

- The scope of affordances is inappropriate for application to this research.
- Current integrations of transformative functions with non-transformative functions are limited to directional interactions with material that can be modeled as a flow through the system.
- Alternative approaches have different definitions of function, which do not support a graphical modeling tool.

Limitations of Function-Based Design

- Function-based design methods support only a subset of customer needs.
- Many customer needs describe interactions, which cannot be modeled using a function structure.
- Some customer needs that can be modeled using function structures can likely be better modeled with other approaches.

Many function-based conceptual design methods in literature use demonstrations to show usefulness of the methods, but few quantitative research studies have been conducted to test the use functional methods and tools by designers. Therefore, this research seeks to both assess and extend function modeling in conceptual design.

3.2 Research Questions

Many function-based design tools have been developed to support ideation in conceptual design, but the models used within these tools may not be useful for conceptual design since they may contain non-transformative descriptions, interactions, component-specific functions, or other extensions of function models. The appropriateness of these extensions and functional descriptions at the conceptual stage of design is the focus of this research. Specifically, the overall research question is:

Overall Research Question: How should the functionality of mechanical artifacts be modeled to support ideation in conceptual design?

Designers may generate or use function models for a variety of tasks, such as (1) understanding and defining a problem by functionally decomposing it, (2) analyzing a functional solution using computational design tools, or (3) generating concepts based on the artifact's desired functionality. In each case, a human must interact with a function model either by creating it and/or using it in the design process. Therefore, it is important to understand how humans interact with a model. The first main research question is:

RQ1: How well do designers understand and use functional representations in conceptual design?

In previous research, a method for creating an abstract description of an artifact from a highly-decomposed description is proposed through function model pruning [26, 60]. The resulting pruned model may be appropriate for use in conceptual design since it is more abstract than the initial, reverse-engineered model, so it is investigated in this research. Further, a new representation—an interaction model—is developed in this research that integrates the pruned representation with a model of a user actions and interactions, addressing many of the limitations of current function-based modeling methods. These two representations, the pruned model (PM) and interaction model (IM), are studied in this research to understand if the way in which artifacts are modeled using each representation is appropriate for conceptual design. Therefore, the second and third research questions pursued are:

RQ2: In what ways do pruned function models support ideation?

RQ3: In what ways do interaction models support ideation?

When generating concepts, Ulrich and Eppinger describe an internal and external search for solutions to the design problem [2]. An external search includes interviewing users, consulting experts, searching literature, or other activities that draw from knowledge outside the design team. Internal searches include brainstorming and other methods that draw from individual and team knowledge [2]. These concept generation classifications are similar to ideation categories intuitive and logical defined by Shah [79]. Intuitive methods draw ideas from designers, while logical methods draw from historical data or use analytical methods to generate ideas [79]. Function models have potential to be used as a stimulus for intuitive methods (internal search) or to drive logical methods based on historical data (external search), so each is explored in this research (the terms *internal* and *external* from Ulrich and Eppinger are used from this point forward). The fourth and fifth research questions are:

RQ4: How well do functional representations support internal search for solutions in conceptual design?

RQ5: How well do functional representations support external search for solutions in conceptual design?

3.3 Research Tasks

The following three research tasks are pursued to address the five research questions: (1) investigate the interpretability of functional representations by humans (interpretability user study), (2) investigate the use of functional representations and abstraction within a similarity metric (similarity study), and (3) investigate the effects of functional representations on concept generation (ideation user study). The relationship between the tasks and research questions is shown in Table 3-1 and discussed in the sections that follow.

Table 3-1: Research Questions and Supporting Research Tasks

	Research Question	Task 1: Interpretability	Task 2: Similarity	Task 3: Ideation
Overall	How should the functionality of mechanical artifacts be modeled to support ideation in conceptual design?	✓	✓	✓
RQ1	How well do designers understand and use functional representations in conceptual design?	✓		✓
RQ2	In what ways do pruned function models support ideation?	✓	✓	✓
RQ3	In what ways do interaction models support ideation?			✓
RQ4	How well do functional representations support internal search for solutions in conceptual design?	✓		✓
RQ5	How well do functional representations support external search for solutions in conceptual design?		✓	

3.3.1 Interpretability User Study

In the interpretability study, participants are provided with function structures and asked to identify an artifact from its function structure alone. Two factors—function language and type—are varied in the function models for this user study. The interpretability study addresses RQ1 since participants' level of understanding of function models is tested by asking them to interpret the model and identify the artifact being modeled. This study addresses RQ2 since the function type factor has two treatments, pruned and reverse-engineered models, assessing the strengths of pruned models for human processing. Finally, RQ4 is addressed by this study because it is important for a human to be able to understand a model if he or she will be using it to generate concepts.

3.3.2 Similarity Study

Design knowledge captured in function models of existing artifacts has been used in previous research to identify artifacts functionally similar to a new design problem, inspiring the development of new concepts. In this study, a published similarity metric is extended and artifacts are compared functionally at three different levels of abstraction to understand the benefits of each level of abstraction in conceptual design. This study addresses RQ2 since the highest level of abstraction used in the study is the pruned model. It addresses RQ5 since the similarity metric can be used to help a designer search externally for solutions to a design problem.

3.3.3 Ideation User Study

Design researchers postulate that function models support creativity in conceptual design because they are abstract models of an artifact, providing freedom for designers to develop many new ideas. However, the focus of this research is on convergent, rather than divergent, thinking. The intent of function models in this study is to help designers converge on a high-quality solution. In this user study, participants generate concepts for a new artifact based on a problem statement, a set of requirements, and a experimental treatment. One of four treatments is provided to each participant: a function model, interaction model, pruned model, or no model. The concepts generated by participants are analyzed for quality of the ideas and conformance (defined as how well the concepts agree with model provided). This study addresses RQ1 through the conformance metric that evaluates whether a designer used the model or deviated from it. Since two of the treatment groups are the pruned model and interaction model, and since the study requires participants to generate ideas based on their own knowledge, this study addresses RQ2, RQ3, and RQ4 as well.

CHAPTER 4: PROPOSED DESIGN APPROACH AND REPRESENTATION

4.1 Integrated Function- and Interaction-Based Design

As demonstrated in Section 2.1.3, there are many non-functional aspects to consider when designing an artifact. Many design texts, however, prescribe a linear, function-based approach to conceptual design, shown in Figure 4-1 [1-4]. Designers begin with customer needs and translate them into engineering requirements. A sub-set of the engineering requirements lead the designer to identify the artifact's function, and a function model is created. Working principles are then identified for each function and, using a morphological chart, working principles are combined into potential concepts. In this approach, the designer intuitively chooses a sub-set of requirements to address through the artifact's functionality.

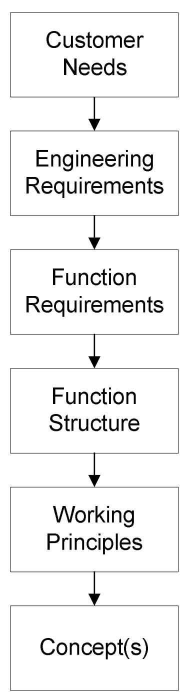


Figure 4-1: Function-based Approach to Conceptual Design

The function-first approach to design does not consider non-functional aspects of the artifact being designed, such as user activities, environment interactions, and artifact interactions. Some researchers use passive functions, such as *import human*, to address these aspects (see Section 2.1.3). However, the approach proposed in this research uses only active functions in a function-based path and includes a complementary, interaction-based path, as shown in Figure 4-2. The function-based approach is included in the left path in the figure, where active functionality of the artifact is addressed. In the right path, interactions are addressed in a similar manner as function:

1. Interaction requirements are identified from the complete list of requirements. Interaction requirements state the context—interactions with users, artifact, and the environment—of the artifact being designed.
2. A solution-independent interaction model is created in conjunction with the function structure. The two models are created together and have an effect on each other, as shown by the arrows in the figure between the two models. A decision made about one model affects the outcome of the other.
3. High-level form principles are identified in conjunction with working principles to embody the interactions in the interaction model. These form principles, like working principles, do not specify an exact geometry; instead, they identify major principles that can be used to satisfy the interaction requirement (e.g., handle, friction-fit, wheels).
4. The working principles and form principles can then be combined using a morphological chart to identify concepts for the artifact being designed.

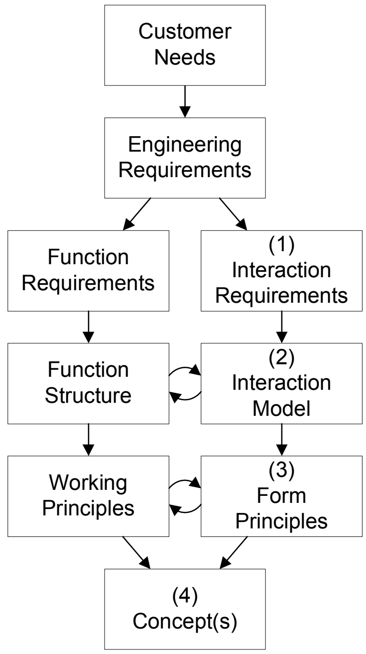


Figure 4-2: Function-based Design Approach with Complementary Interaction-based Approach

Since interactions do not conform to the traditional definition of function, a new model is proposed to capture user and artifact interactions separately from an artifact's function. Artifacts that lie outside the system boundary of the artifact of interest are explicitly modeled and mapped to the artifact of interest through interactions. In Section 4.2, an interaction model is presented that incorporates functions, interactions, and user activities and is demonstrated with the Black and Decker Jigsaw Attachment.

Customer needs may describe an artifact in a manner that either cannot be represented in a function structure or is non-transformative. These descriptions include passive functions, user activities, environment interactions, artifact interactions, and user interactions. Since many extensions of function structures include some or all of these types of descriptions, it is important to clearly define each so that they can be identified and appropriately placed when creating new models.

Function – the transformation of material, energy, and/or information from an input state to an output state [1]

Active Function – a function in which the artifact of interest carries an energy flow that is used to control the outcome of the function

Passive Function – a function in which the artifact of interest does not carry the energy used to control the outcome of the function

User Activity – a change in a property of the artifact or a change in a flow within the artifact in which the user provides the energy to make the change

Artifact of Interest – a clearly defined set of components being studied

Environment – anything that lies outside the artifact of interest; the environment can be decomposed into the natural environment, artifacts, and users

Natural Environment – anything that exists in nature

Artifact – an entity that has been altered from its natural state

User – an entity external to the artifact of interest that initiates interactions with the artifact of interest

Natural Environment Interaction – exists when the artifact of interest changes a property of the environment or when the environment changes a property of the artifact of interest

- *Artifact changes a property of the environment:* a submarine interacts with the environment by changing the water pressure locally near the propeller

- *Environment changing a property of the artifact:* water interacts with a submarine through corrosion of the hull

Artifact Interaction – exists when the artifact of interest physically contacts another artifact or energy or information flows to/from the artifact of interest from/to another artifact

- *Physical contact:* if a flashlight is the artifact of interest, it interacts with a battery because it physical must contact the flashlight to function properly
- *Information flow to the artifact of interest from another artifact:* if a television is the artifact of interest, it interacts with the remote control because it receives a signal from the remote control
- *Information flow from the artifact of interest to another artifact:* if a television remote is the artifact of interest, it interacts with a television because it sends a signal to the television

User Interaction – exists when a user physically contacts the artifact of interest or energy or information flows to/from the artifact of interest from/to the user

- *Physical contact:* A user interaction exists between a drill and a user when the user carries the drill because the user is physically contacting the drill
- *Information flow to the artifact of interest:* A user interaction exists between a user and computer because information flows to the computer from the user.

- *Information flow from the artifact of interest:* A user interaction exists between a user and a vehicle's fuel system because information about the amount of fuel in the tank is displayed to the user.

Context – the set of all interactions between the artifact of interest and the natural environment, artifact, and users

The context of a typical vacuum cleaner includes:

- *Environment Interactions*
  - air, since the vacuum changes a property of the air (pressure)
  - dirt, since the vacuum changes a property of the dirt (location)
- *Artifact Interactions*
  - the wall outlet, since the vacuum physically contacts the wall outlet
  - the floor, since the vacuum sits on the floor
  - floor carpet in a vehicle, if the vacuum is being used to clean the vehicle
- *User Interactions*
  - A user interacts with the vacuum when he carries it around because he is physically contacting it.
  - A user interacts with the vacuum when he turns it on because he is physically contacting it.

The interaction model, presented in the following section, is developed based on these elements and their definitions.

4.2 Interaction Model

The proposed design approach incorporates a function- and an interaction-based approach, which will be supported by a graphical model. The current model being pursued is the interaction model, which integrates a pruned function structure, a user activity model [4], and interactions between these elements. The interaction model entities, shown in Figure 4-3, consist of an system boundary, functions, user activities, and flows of material, energy, signal, and artifact. The system boundaries are shown by a dashed line and used to indicate what is being modeled within an artifact or user. Functions (rectangles) and user activities (hexagons) are included within the artifact and user boundaries, respectively. The functions can have inputs and outputs of material, energy, and signals, while the activities can have these same inputs and outputs in addition to an artifact flow. The artifact flow represents the handling of an artifact by a user, and the artifacts may flow through the user's system. The flows of material, energy, and signal may enter or exit a function or user activity, and they may cross boundaries, passing from the user to an artifact and vice versa.

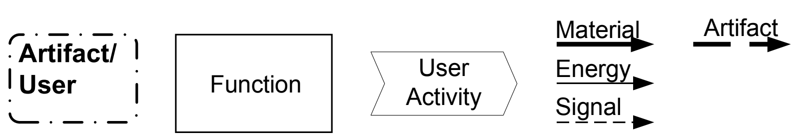


Figure 4-3: Interaction Model Entities

The interaction modeling entities are explained in detail with the example of the Black and Decker Jigsaw Attachment (see Section 2.1.3). The interaction model, shown in Figure 4-4, includes four artifacts: a battery, universal driver, jigsaw attachment, and jigsaw blade. The active function of each artifact is shown and connected to the functionality of other artifacts through flows among them.

- The battery supplies electrical energy to the driver.
- The driver converts human energy to a signal, which actuates the electrical energy. These two functions are accomplished by the switch on the driver.
- The driver then converts the electrical energy to rotational energy through the motor contained inside the driver.
- The driver then changes the rotational energy by reducing the angular velocity through a set of planetary gears.
- The rotational energy flows from the driver to the attachment via a shaft.
- The attachment then converts the rotational energy to translational energy using a cam.
- The translational energy exits the driver's system boundary and passes through the blade.
- The translational energy exits the blade's system boundary and enters the user's boundary, showing that the user is in control of the translational energy output from the system.

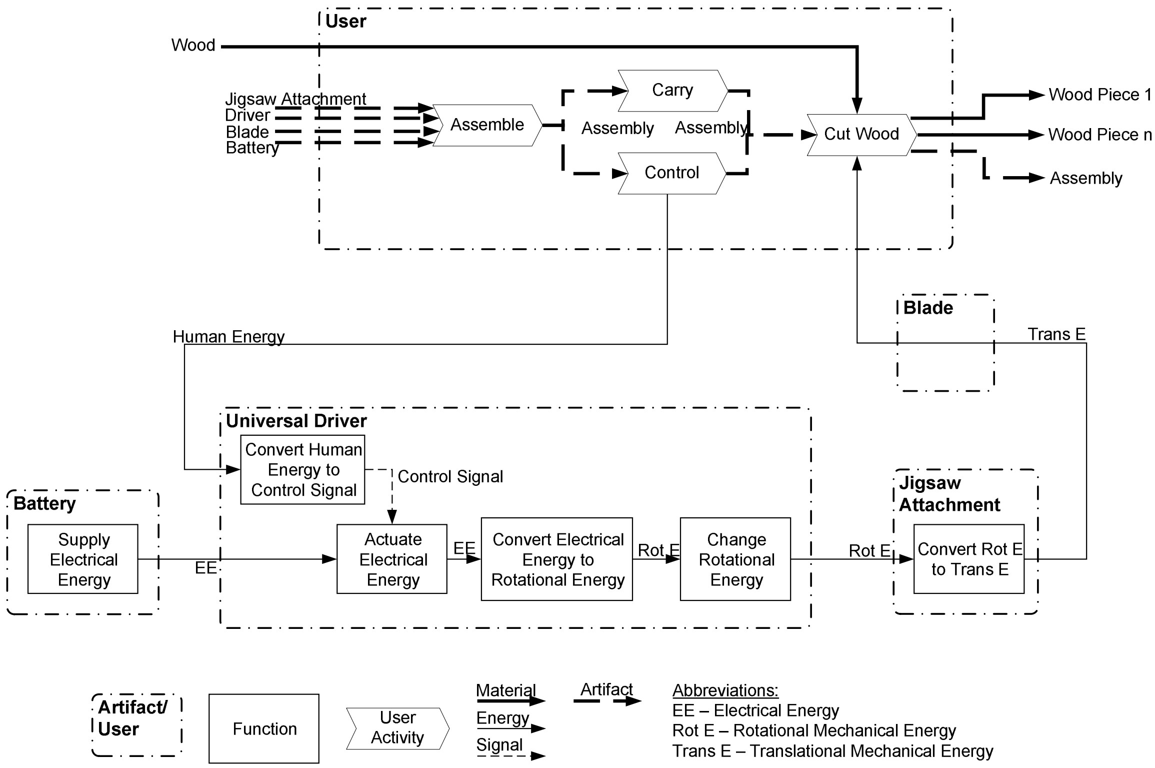


Figure 4-4: Interaction Model of a Black and Decker Jigsaw Attachment

The user and the activities that the user performs are shown separately from the function of the artifact inside the boundary of the user. The hexagons in the model show these activities with flows of artifacts between the activities. The user can carry or control the artifact, cut wood, assemble the artifacts, or perform many other activities. It is important to note that the activity model shown in this example does not include the entire lifecycle of the artifact, as in examples in previous literature [4] (e.g., purchasing, maintaining, recycling). The focus of this model is on routine operation by the end user, so only typical end user activities are shown. Additionally, not all potential user activities are shown as the focus of this research is to capture the relationship between user activities and artifacts.

The flows between artifacts or between the user and the artifact in the model show the interactions among the systems. An interaction must be embodied with a physical form to enable the flow to pass between two artifacts or the artifact and user. The five interactions in the model are explained below.

- The battery and driver interact with each other through a flow of electrical energy from the battery to the driver, represented by the flow between these two artifacts. This flow is embodied through electrical contacts on both the battery and driver as well as other features that enable the battery to be locked in place.
- The driver and jigsaw attachment interact with each other through a flow of rotational energy between the two artifacts, as shown in the model. The driver and attachment both have features that allow them to be secured to each other and two shafts coupled together to allow the passage of rotational energy from one artifact to the other.
- The interaction between the jigsaw attachment and the blade is shown by the flow of translational energy from the jigsaw attachment into the blade. This flow is enabled by a clamping mechanism that secures the blade to the output from the jigsaw attachment.
- The interaction between the blade and a piece of wood is shown by flow from the blade to the activity *cut wood*. The cutting force between the jigsaw's blade and the wood enable this energy passage. The passage of translational energy into the user boundary also shows that the user is in control of the translational energy that is output from the system.

- The user interacts with the system through the flow of human energy from the user to the system. This interaction is embodied by a surface that allows physical contact between the user and the system to take place. The human energy exits the user boundary and enters the driver boundary, indicating that the user is no longer in control of this energy.

4.3 <u>Comparison of Interaction Model and Function Structure</u>

The interaction model of the Black and Decker Jigsaw Attachment contains the same information as the function structure (see Section 2.1.3), but the information is represented differently. In the function structure, the series of functions *import human material, guide human material, export human material* (reproduced in Figure 4-5) describe the human activity of holding the system and manipulating it. These functions are passive because the jigsaw attachment does not provide the energy for these functions to be carried out. Human material does not enter the jigsaw attachment; the two interact with each other. The interaction model describes the relationship between the user and the artifact as a user activity (reproduced in Figure 4-6), capturing the passive functionality described in the original function structure in a more active manner.

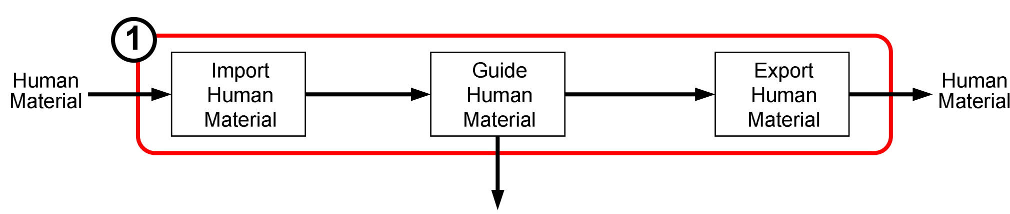


Figure 4-5: User Manipulation of the Black and Decker Jigsaw Attachment Represented Using Passive Functions

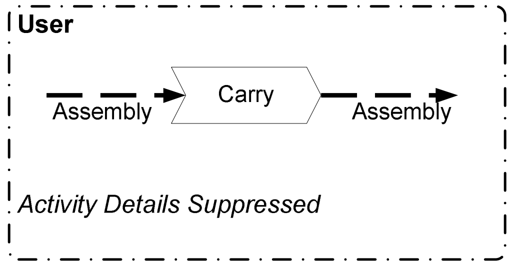


Figure 4-6: User Manipulation of the Black and Decker Jigsaw Attachment Represented Using an Interaction Model

The function structure uses four passive functions to describe the assembly of the jigsaw attachment to the driver (labeled "power pack") and the flow of rotational energy between the two (reproduced in Figure 4-7). The interaction model is simpler, describing the assembly through a user activity and the flow of rotational energy from the driver to the jigsaw attachment as a flow, rather than a function and a flow (reproduced in Figure 4-8).

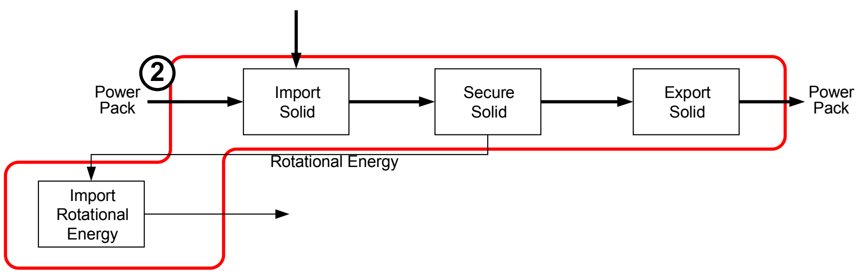


Figure 4-7: Artifact Interaction in the Black and Decker Jigsaw Attachment Represented Using Passive Functions

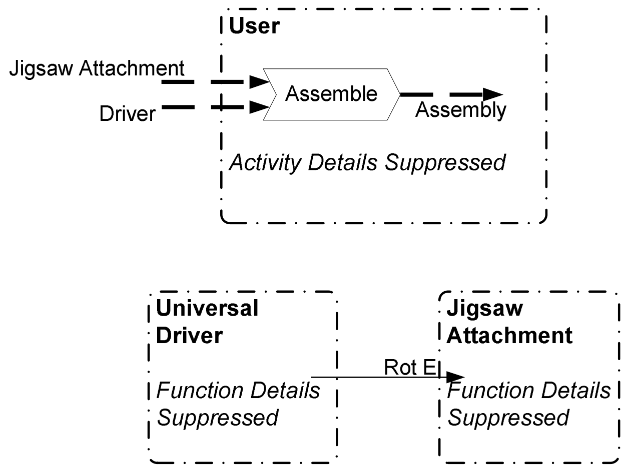


Figure 4-8: Artifact Interaction in the Black and Decker Jigsaw Attachment Represented Using an Interaction Model

The jigsaw attachment function structure describes the user's force that is input to the system to control whether the artifact is on or off with the passive function *import human energy* (reproduced in Figure 4-9). The artifact itself does not forcibly bring human energy into the system; rather, the human energy is provided to the system. The interaction model captures this information more actively by showing that the user controls the assembly and human energy flows from the user to the driver (see Figure 4-10).

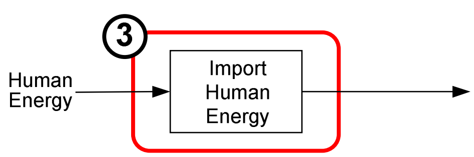


Figure 4-9: User Control of the Black and Decker Jigsaw Attachment Represented Using a Passive Function

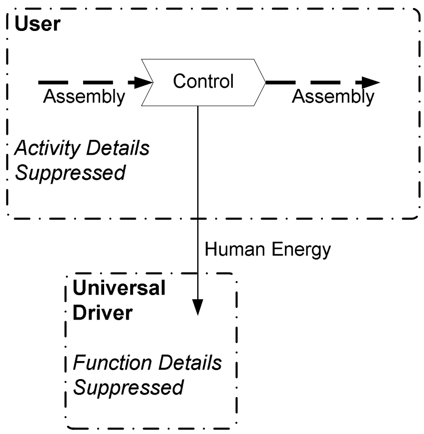


Figure 4-10: User Control of the Black and Decker Jigsaw Attachment Represented Using an Interaction Model

The function structure uses three passive functions to show that the blade can be secured to the jigsaw attachment (reproduced in Figure 4-11). The functions are passive because the energy to perform these functions must be provided something external to the artifacts. The interaction model captures this same information by showing that the user assembles the two components, and that translational energy flows from the jigsaw attachment to the blade (see Figure 4-12).

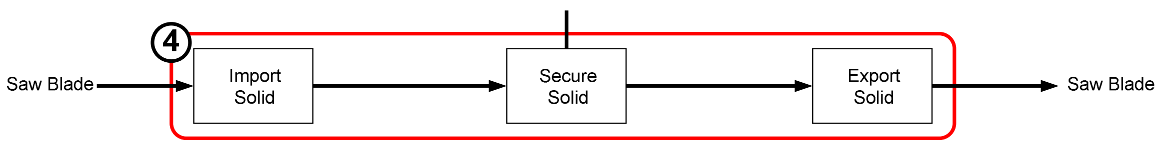


Figure 4-11: Artifact Interaction of the Black and Decker Jigsaw Attachment Represented Using Passive Functions

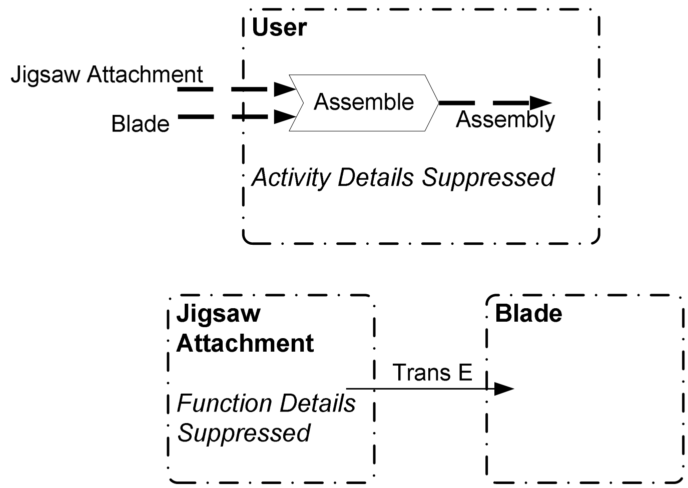


Figure 4-12: Artifact Interaction of the Black and Decker Jigsaw Attachment Using an Interaction Model

Interaction models capture the function of artifacts, interactions among artifacts, interactions between artifacts and users, and user activities. This model describes *what* functions and activities the artifacts and users accomplish, not *how* they accomplish the functions and activities. The functions, activities, and interactions are described at an abstract level to prevent solution-specific models. A separate mapping between interactions and an artifact's form may be used to capture *how* an artifact and user interact for later phases of design or information archival. However, this mapping is outside the scope of this research.

The modeling of functions, activities, and interactions using an interaction model has been demonstrated through the example of the Black and Decker Jigsaw Attachment, capturing all of the information contained in the initial function model. The interaction model, therefore, is able to capture functional requirements and also has potential to address requirements related to user and artifact interactions. In the review of function modeling (see Section 2.3) function structures were shown to be able to address requirements related only to material, energy, and information inputs. In the bicycle suspension example, this covers only eight of the twenty-two customer need statements (see Section 2.3). The interaction model for this example could potentially double the number of customer needs addressed compared to the function structure, including the following customer needs.

The suspension:

- preserves the steering characteristics of the bike
- provides stiff mounting points for the brakes
- fits a wide variety of bikes, wheels, and tires
- is easy to install
- works with fenders
- can be easily accessed for maintenance
- allows easy replacement of worn parts
- can be maintained with readily available tools [2]

The interaction model is not intended to address all types of customer needs, so there will be some customer needs that cannot be addressed using this model. These types of needs include inherent properties of the system, which are based on the system's form. In the bicycle suspension example, these properties include its weight, durability, appearance, and cost (see Section 2.3). Thus, the following customer needs remain unaddressed by both function structures and the interaction model.

The suspension:

- is lightweight
- instills pride
- is affordable for an amateur enthusiast
- lasts a long time [2]

The interaction model presented in this section demonstrates how this representation can be used to model the functionality of several artifacts that interact with each other as well as the user's interaction with the artifact(s). The activity model [4] was selected to model the user's actions because the activity model is a graphical, flow-based representation, similar to function structures. Many alternative representations have potential to be combined with the function structure more effectively than the activity model, but a complete review of user and process representations and their potential for merging with the function structure is outside the scope of this research.

The key elements of this representation are the pruned function model which contains active, conceptual-level artifact functions, a user model that describes the actions a user performs when using the artifact, and the passage of flows between artifacts and between artifacts and the user.

CHAPTER 5: INTERPRETABILITY USER STUDY

This interpretability study was designed, executed, and analyzed in collaboration with Thomas and colleagues [78], and a complete description of this experiment is included in [81]. Thomas analyzed the results of this study using descriptive statistics and qualitative feedback from participants [81]. There is an opportunity to analyze the results statistically, so the data from this study are analyzed and presented in this research using a statistical approach to draw conclusions primarily on the benefits of function structure pruning. Thus, new contributions are included in Sections 5.4 and 5.5.

5.1 Motivation

In either forward design or reverse engineering, it is important for a modeler to be able to communicate his or her ideas clearly using the model. New design problems may not be performed by a single person, so the function models must be understood by an entire design team. Reverse engineered models may be used for information archival and reuse, so the models created for existing artifacts must be understood by anyone using the information. Thus, for any use of a function model, it is important that the ideas in the model are clearly communicated. Multiple models of an artifact may exist, but each model should clearly communicate the functions that the artifact performs. The overall goal of this research is to understand the limitations of current function modeling methods and to improve the usefulness of function models for conceptual design and reverse engineering. As a first step in this overall goal, the level of understanding of reverse engineered function models is assessed by studying the interpretability of models of existing artifacts.

The interpretability of reverse engineered function models will provide insights into the use of these models for communication and archival of functional information. The principles of communication learned through studying the interpretability of reverse engineered function models can then be extended to new design problems, where communication is also essential within design teams.

5.2 Frame of Reference

5.2.1 Interpretability

Research in function structures has focused on consumer, electromechanical artifacts, such as handheld power tools and household appliances. The function structures developed for these artifacts are relatively small and can be created by a single person, so the intent of each element in the model is fully understood by the modeler. However, when a observer unfamiliar with the model uses it, he or she may not understand what the modeler intended. For example, in the hair dryer function structure (see Figure 5-1), the functions *import*, *guide*, and *export human energy* could be interpreted as movement of the whole system or movement of a component of the system, such as a switch. The goal of this research is to understand the interpretability of function structures, or how well designers unfamiliar with a model can understand what is modeled. In this study, interpretability is defined as the ability of a human to correctly identify an artifact by looking only at a model of its function.

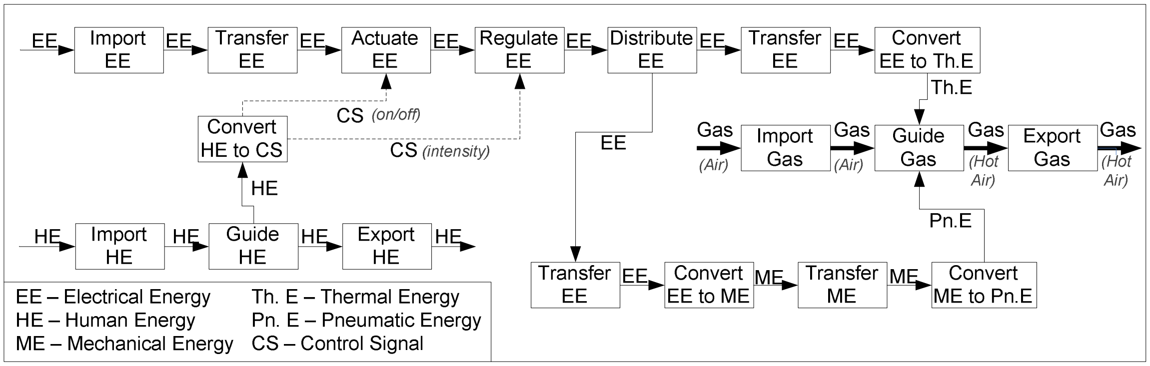


Figure 5-1: Hair Dryer Function Structure (adapted from [31])

5.2.2 Ambiguity

The goal of function modeling in conceptual design is to identify what the artifact should do independent of its final form. Since the final form is not known, a function model supports uncertainty in the design. However, this uncertainty should be clearly identified and communicated by the function model, rather than containing ambiguity that can be misinterpreted by readers [82]. In conceptual design, it is important to explore as much of the available design space as possible, and an abstract model such as a function structure can support this exploration. However, an abstract model should not be ambiguous, but it should clearly outline the design space that is available for exploration. An ambiguous model may seem to be abstract, but it may allow a designer to misinterpret the model and explore areas that are outside the design space. If ambiguity exists in function models of reverse engineered artifacts, then similar models used in forward design may also be ambiguous. This research uses the interpretability of function models to understand if ambiguity exists within function models and, if so, to identify ways to reduce this ambiguity, improving function-based communication and information archival in engineering design

5.3 Research Approach

A user study is conducted to test human interpretability of function structures. A previous study tested the interpretability of three levels of abstraction of function structures [78], leading to the identification of two dimensions of abstraction and further refinement of the experiment [78]. The study was revised and repeated with a larger sample size and an additional level of abstraction that was discovered through the initial study. The primary difference in the refined study is the testing of two independent abstraction factors and the measurement of interpretation speed in addition to accuracy. The two factors tested are the specificity of terms used in the models (Functional Basis or free language) and the type of functions included in the model (reverse-engineered or pruned). In the study, participants are provided with a function model at one of four combinations of abstraction levels and asked to identify the artifact based solely on its function structure.

5.3.1 Function Structure Abstraction Levels

Two levels of abstraction are tested in each of the two dimensions. The function level is tested at the reverse engineered level (RE) and at the pruned level (see Section 2.1.4). The language specificity is tested at the free language level (Free) and using the secondary level of the Functional Basis (FB). Thus, the following four levels of abstraction are obtained: RE-Free, RE-FB, Free-FB, and Free-Pruned.

Four different existing artifacts were selected for this study, and the function model was obtained from the design repository (see Section 2.1.2) [31]. The models in the repository were created independent of this research, and they contain free language as well as Functional Basis terms [30]. Since the artifacts exist, the models have been created at the reverse-engineering abstraction level, containing many specific details about the artifact and many of its individual components. Therefore, the models obtained from the repository are considered to be at the RE-Free level of abstraction. An example of a RE-Free function structure of a rice cooker is shown in Figure 5-2. The key features of this model relative to the FB level of language abstraction are the inclusion of context-specific free language terms, such as *bowl*, *rice*, *water*, *on*, and *off*. In the function dimension, this model contains auxiliary functions and interactions such as *import electrical energy*, *transfer thermal energy*, and *import solid*, which can be identified through reverse engineering but may not be specified in conceptual design.

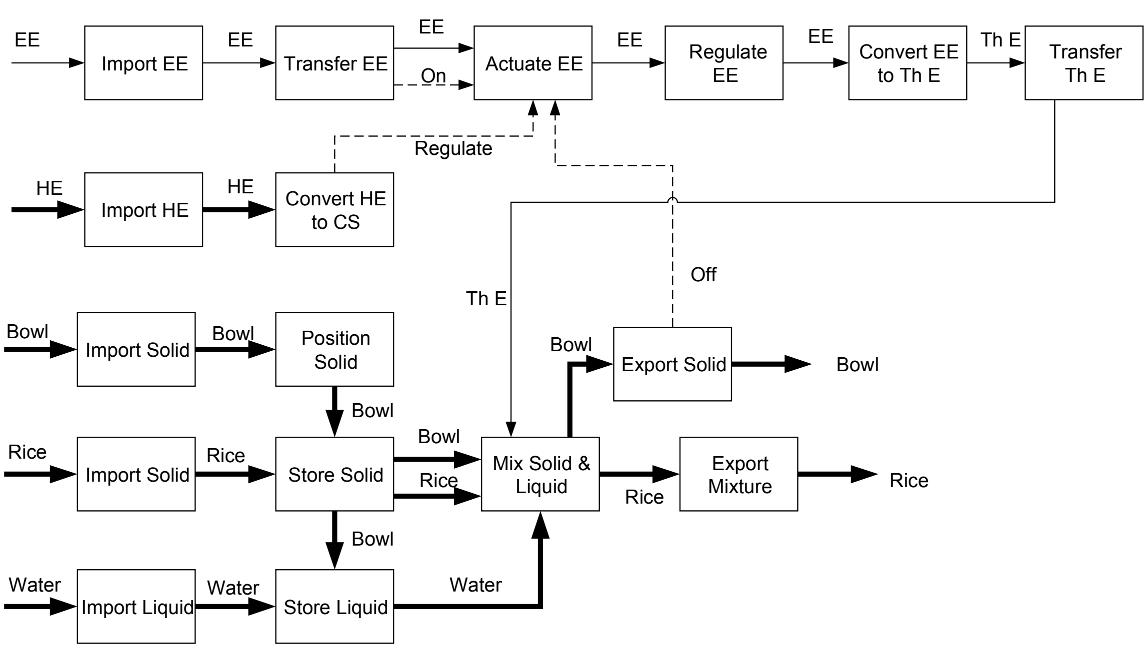


Figure 5-2: Rice Cooker Function Structure at the RE-Free Abstraction Level

The RE-FB level of abstraction is obtained by translating the free language terms in the RE-Free model to Functional Basis terms using guidelines provided with the Functional Basis vocabulary as well as knowledge about the artifact. The number of functions and flows and the relationships among these are identical between the RE-Free and RE-FB levels of abstraction. The RE-FB level of abstraction of the rice cooker is shown in Figure 5-3, where FB terms that required translation are shaded gray. In this model, terms such as *bowl* and *rice* have been translated to *solid*, *on* and *off* to *control signal*, and *water* to *liquid*. The auxiliary functions and interactions remain in the model, as in the RE-Free level of abstraction.

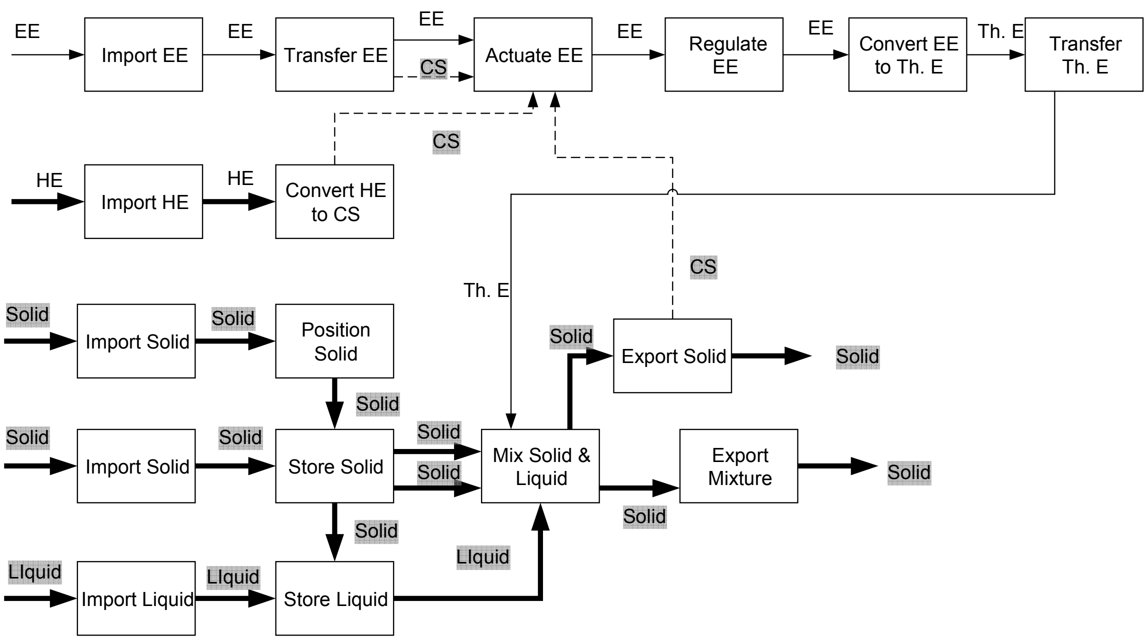


Figure 5-3: Rice Cooker Function Structure at the RE-FB Abstraction Level

The Pruned-FB level of abstraction is obtained by applying pruning rules to the RE-FB model (see Section 2.1.4). The pruning rules remove auxiliary functions and interactions from the models, resulting in a more conceptual-level model compared to the reverse engineered models in the repository (see Section 2.1.4). The pruning process reduces the number of functions and flows in the models but does not change the language. In the Pruned-FB rice cooker model, shown in Figure 5-4, functions such as *import human energy*, *transfer electrical energy*, and *export solid* have been removed.

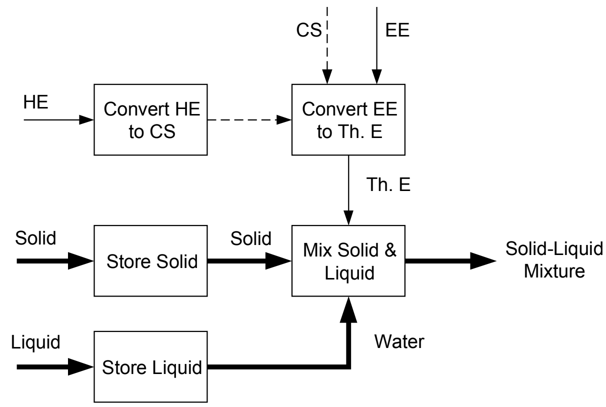


Figure 5-4: Rice Cooker Function Structure at the Pruned-FB Abstraction Level

The final level of abstraction, Pruned-Free, is created by converting the FB terms in the Pruned-FB model back to the free language terms used in the RE-Free level of abstraction, providing additional context that is not included in the Pruned-FB model. As shown in the example of a rice cooker model at the Pruned-Free level of abstraction (see Figure 5-5), the Pruned-Free level of abstraction contains a few, conceptual-level functions with context-specific terms, such as *rice*, *water*, and *on*, rather than *solid*, *liquid*, and *control signal*.

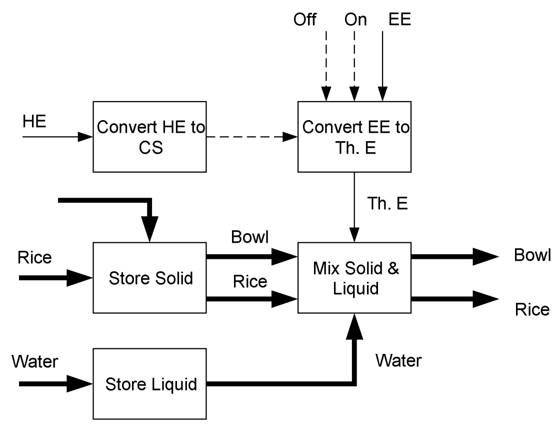


Figure 5-5: Rice Cooker Function Structure at the Pruned-Free Abstraction Level

The following four consumer artifacts were chosen for this study, and each was modeled at the four combinations of abstraction levels, resulting in sixteen unique models:

- Black and Decker Rice Cooker
- DeWalt Sander
- Shopvac Vacuum Cleaner
- Black and Decker Electric Screwdriver

5.3.2 Experimental Design

The experiment was conducted in a graduate-level advanced design course at Clemson University during the Fall 2009 semester. Eighteen students participated in the study during their regularly-scheduled class period. Participants had experience in function modeling through the design course, so they were given a short presentation to remind them of the basics of function structures. Each participant was given the sixteen unique function models (4 artifacts at 4 levels of abstraction each) and asked to identify the artifact that was modeled from a list of 48 artifacts. The models were provided in groups of four to prevent the participants from recognizing repeated artifacts, and participants were not aware that artifacts were used multiple times. The complete details of the experiment, including the models given to participants, artifact selection sheet, and the sequence of models given to participants are discussed in [81].

In the study, participants were presented with an unidentified function structure and asked to identify what artifact was modeled. The participants' responses were classified as exact, non-exact, similar, and dissimilar. Exact responses are those that exactly identify the artifact being modeled, while non-exact responses are the remaining 47 incorrect answer choices. The non-exact responses are further broken down into similar and dissimilar responses. Similar responses are identified as artifacts in the answer packets are functionally similar to the exact answer, while dissimilar artifacts are those that are not functionally similar. Similar artifacts were defined *a priori* by a panel of design researchers based on the high-level purpose of the artifact.

5.3.3 Research Hypotheses

The two factors in this study, language specificity and type of function, are tested to determine if either factor has an effect on the interpretability of function structures and the amount of time required to interpret the function structures. The mean interpretability and time for each factor are compared, with the primary research hypotheses shown in Table 5-1.

Table 5-1: Primary Interpretability Research Hypotheses

Research Hypothesis	
The interpretability of function structures using free language is greater than the interpretability of function structures using the secondary level of the Functional Basis.	$I_{Free} \neq I_{FB}$
The interpretability of pruned function structures is greater than the interpretability of reverse-engineered function structures.	$I_{Pruned} \neq I_{RE}$
The time required to interpret a free-language function structure is less than the time required to interpret a Functional Basis function structure.	$t_{Free} \neq t_{FB}$
The time required to interpret a pruned function structure is less than the time required to interpret a reverse-engineered function structure.	$t_{Pruned} \neq t_{RE}$

Note:
I – Interpretability
t – time
Free – Free Language
FB – Functional Basis
RE – reverse engineered

The secondary research hypotheses test the simple effects of interpretability and time:

- $I_{Pruned\text{-}Free} \neq I_{Pruned\text{-}FB}$
- $I_{RE\text{-}Free} \neq I_{RE\text{-}FB}$
- $I_{Pruned\text{-}Free} \neq I_{RE\text{-}Free}$
- $I_{Pruned\text{-}FB} \neq I_{RE\text{-}FB}$
- $t_{Pruned\text{-}Free} \neq t_{Pruned\text{-}FB}$
- $t_{RE\text{-}Free} \neq t_{RE\text{-}FB}$
- $t_{Pruned\text{-}Free} \neq t_{RE\text{-}Free}$
- $t_{Pruned\text{-}FB} \neq t_{RE\text{-}FB}$

The interpretability hypotheses are tested using two scoring approaches: (1) an exact response is given a score of 1, and a non-exact response is given a score of 0; and

(2) an exact or similar response is given a score of 1, and a dissimilar response is given a score of 0. The definition of exact, non-exact, similar, and dissimilar responses is discussed in Section 5.3.2. The time hypotheses are tested using three approaches: (1) all times are considered, (2) only times of exact responses are considered, and (3) the times of exact and similar responses are considered.

5.4 Statistical Analysis

5.4.1 Interpretability

As discussed in Section 5.3.3, the data collected are used to determine if the function level (Pruned or RE) or language level (Free or DR) has an effect on interpretability of function structures. For the interpretability statistical tests, each of the two scoring approaches discussed in Section 5.3.2 are analyzed assuming a binomial distribution of the responses. Participants and artifacts are both modeled as random effects. The GLIMMIX procedure within SAS/STAT® software is used to analyzed the data and the LSMEANS procedure used to compare the means of interpretability. The interpretability hypotheses and results are shown in Table 5-2, where the values in the table represent the mean interpretability on a scale from 0 to 1. The p-values have not been adjusted for multiple comparisons, as this research is exploratory in nature.

Table 5-2: Results of Statistical Tests of Interpretability

Hypothesis	Exact = 1 Non-Exact = 0 (n = 262)	Exact = 1 Similar = 1 Dissimilar = 0 (n = 262)	Hypothesis Decision
1) $I_{Free} \neq I_{FB}$	$0.68 \neq 0.060$ $p < 0.0001$	$0.91 \neq 0.37$ $p < .0001$	Accept
2) $I_{Pruned} \neq I_{RE}$	$0.33 \neq 0.22$ $p = 0.13$	$0.75 \neq 0.68$ $p = 0.29$	Fail to Accept
3) $I_{Pruned\text{-}Free} \neq I_{Pruned\text{-}FB}$	$0.72 \neq 0.088$ $p < 0.0001$	$0.92 \neq 0.43$ $p < 0.0001$	Accept
4) $I_{RE\text{-}Free} \neq I_{RE\text{-}FB}$	$0.64 \neq 0.041$ $p < 0.0001$	$0.91 \neq 0.31$ $p < 0.0001$	Accept
5) $I_{Pruned\text{-}Free} \neq I_{RE\text{-}Free}$	$0.72 \neq 0.64$ $p = 0.41$	$0.92 \neq 0.91$ $p = 0.65$	Fail to Accept
6) $I_{Pruned\text{-}FB} \neq I_{RE\text{-}FB}$	$0.088 \neq 0.041$ $p = 0.19$	$0.43 \neq 0.31$ $p = 0.28$	Fail to Accept

The interpretability of free language models, using both scoring methods, is significantly better than the interpretability of Functional Basis models ($p < 0.0001$). Using the exact/non-exact scoring, free language models had an average interpretability of 0.68 on a scale from 0 to 1, while Functional Basis models had an average interpretability of 0.066. Using the exact/similar/dissimilar approach, the free language models had an average interpretability of 0.91 while the Functional Basis models had an average interpretability of 0.37. Therefore, the use of free language significantly improves the interpretability of function structures.

The average interpretability of pruned and reverse-engineered function structures using the exact/non-exact scoring method is 0.33 and 0.22, respectively. When using the exact/similar/dissimilar scoring system, the averages are 0.75 and 0.68, respectively. The comparison of these values results in p-values of 0.06 and 0.15, respectively. The hypothesis test was also performed using additional scoring approaches, such as exact responses receiving a score of 2, similar responses a score of 1, and dissimilar responses a score of 0; or non-responses scored as non-exact. In each variation of the analysis, the p-value for this hypothesis test was approximately 0.15. Since the level of significance in this research is 0.05, the second interpretability research hypothesis is not accepted.

The third through sixth hypotheses test for simple effects of the two factors. The results of these hypotheses are consistent with the results of the first two hypotheses, and there are no significant mixed effects.

5.4.2 Time

The time required to interpret each function structure was analyzed using three approaches: (1) all times are considered, (2) only times of exact responses are considered, and (3) only times of exact and similar responses are considered. The procedure GLIMMIX within SAS was also used in the time data analysis. The interpretability times were assumed to be normally distributed, and participants and artifacts were modeled as random effects. The time hypotheses and results are shown in Table 5-3, where the values in the table represent the mean time, in seconds, taken to interpret a function structure.

Table 5-3: Results of Statistical Tests for Time

Research Hypothesis	Time from All Responses (n = 262)	Times from Exact Responses Only (n = 96)	Times from Exact and Similar Responses Only (n = 162)	Research Hypothesis Decision
1) $t_{Free} \neq t_{FB}$	$70.0 \neq 127.6$ $p < 0.0001$	$70.4 \neq 106.7$ $p = 0.014$	$67.7 \neq 103.2$ $p = 0.0002$	Accept
2) $t_{Pruned} \neq t_{RE}$	$79.9 \neq 117.6$ $p < 0.0001$	$81.7 \neq 95.4$ $p = 0.35$	$70.9 \neq 100.0$ $p = 0.0010$	Accept
3) $t_{Pruned\text{-}Free} \neq t_{Pruned\text{-}FB}$	$48.7 \neq 111.1$ $p < 0.0001$	$54.1 \neq 109.3$ $p = 0.0015$	$48.7 \neq 93.1$ $p = 0.0003$	Accept
4) $t_{RE\text{-}Free} \neq t_{RE\text{-}FB}$	$91.3 \neq 144.0$ $p < 0.0001$	$86.7 \neq 104.0$ $p = 0.45$	$86.8 \neq 113.2$ $p = 0.0506$	Accept
5) $t_{Pruned\text{-}Free} \neq t_{RE\text{-}Free}$	$48.7 \neq 91.3$ $p < 0.0001$	$54.1 \neq 86.7$ $p = 0.0024$	$48.7 \neq 86.8$ $p = 0.0001$	Accept
6) $t_{Pruned\text{-}FB} \neq t_{RE\text{-}FB}$	$111.1 \neq 144.0$ $p = 0.003$	$109.3 \neq 104.0$ $p = 0.8402$	$93.1 \neq 113.2$ $p = 0.1640$	Accept

When the times from all responses or exact and similar responses are considered, all of the hypothesis tests are accepted with a significance level of 0.05. Free language models are interpreted significantly faster than Functional Basis models, and pruned models are interpreted significantly faster than reverse-engineered models. Hypotheses 3-6, which test for simple effects, are consistent with the first two hypothesis, so there are no mixed effects. The fastest level of abstraction, therefore, is the Pruned-Free level, which took approximately 49 seconds to interpret.

When the times from only exact responses are considered, the trends in time required to interpret the models are similar but not always significant. The sample size is much smaller in this approach because the times from non-exact responses are not considered. Therefore, the results of the other two approaches are used to accept all of the time research hypotheses.

5.5 Outcomes and Discussion

The interpretability of function structures has been studied to determine how well human users of function structures understand a model. A user study was conducted in which participants were given an function structure and asked to identify what artifact is represented by the model. Function structures varied in terms of language specificity and the level of abstraction of functions to better understand the aspects of a function structure that aid in interpretation. A limitation of the study is that all free language terms in the models were used to describe flows, not functions. Therefore, all conclusions drawn on the Functional Basis are relevant for the flow vocabulary and not necessarily for the function vocabulary. Two major conclusions are drawn from this study:

1) **The use of free language increases the accuracy and speed of interpretability compared to a controlled vocabulary.**

   The statistical analysis shows that free language function structures had a much greater interpretability than Functional Basis function structures. The high specificity of flow terms in free language models provides additional context in the model that helps the user interpret it. In the Functional Basis models, less-specific terms create more ambiguity in the model, and participants are not able to understand the content of the model. One purpose of the Functional Basis is to improve the communication of function models through the use of a controlled vocabulary and specific definitions of terms. This interpretability study, however, shows that Functional Basis terms, specifically flow terms, cause ambiguity in a model rather than clarity. Even though definitions of each term have been provided, the specificity of the terms are not adequate for human communication and interpretability. Thus, either free language should be used in function structures or a more specific flow vocabulary should be developed that enables contextual information to be included in the models.

The speed of interpretation of free language models is significantly higher than Functional Basis models. Participants identified these contextually rich free-language terms and used them to quickly understand the model. In Functional Basis models, the terms were less clear, so they required more time to interpret. The use of free language in communication between human designers, therefore, is enhanced in terms of speed and accuracy when free language is used in the model.

2) **Removing auxiliary functions and interactions from a reverse-engineered function structure increases the speed of interpretation without decreasing interpretability.**

Pruning rules specify the removal of auxiliary functions and interactions in a function model. When this specific set of functions is removed, the average interpretability does not significantly change. Although there is no increase in interpretability, there is also no reduction in interpretability caused by the removal of these functions. Therefore, for human interpretation, auxiliary functions and interactions do not add value to the model. Further, the time required to interpret pruned function structures is significantly lower than that of reverse-engineered functions structures, indicating that the auxiliary functions and interactions divert the interpreter's attention to less important elements in the model. Overall, pruned models are a more efficient representation of function since they are faster to interpret without a sacrifice in accuracy, so pruned models should be used when humans are reading function structures.

The results and conclusions of this study can be used to improve the understanding of artifact functionality in engineering design. The following three applications of this study have been identified:

1) **Model Communication**

When designers use function models to communicate their ideas to other designers, such as in a design report, they should use the Pruned-Free abstraction level. Free language will provide context to those reading the model that will increase the speed and accuracy of their interpretations, reducing the potential for misinterpretation. Further, pruned function structures are more efficient in communication and do not increase the risk of misinterpretation by a reader.

If a designer desires to communicate auxiliary functions or interactions, he or she can include these in a function structure without significantly reducing the ability of the receiver to interpret the model. However, the designer could instead use a separate, complementary model, such as an assembly diagram or a model of interactions, maintaining the efficiency of a pruned function structure while communicating the additional information captured in a reverse-engineered model.

2) **Model Creation**

When creating models in conceptual design, the use of free language and the exclusion of auxiliary functions and interactions from a function structure may support faster identification and increased understanding of critical artifact functionality. Therefore, the pruning rules can be used as guidelines for identifying the types of functions that should be identified first as a problem is decomposed. After a pruned function model is created, auxiliary functions and interactions can be added to the model if desired.

3) **Information Archival**

If functional information is to be captured in a database and retrieved by human users, free language should be used in addition to a controlled vocabulary. The advantage of a controlled vocabulary is increased reasoning on the information, but when this information is returned to a user, it should include free language for easy interpretation. A database should also have the ability to provide pruned models to a human user to further increase the ease of interpretation of models. If free language is captured and pruning rules implemented within a database, all four levels of abstraction investigated in this research will be supported, each of which have different applications. The Pruned-Free level supports quick, accurate communication of functional descriptions between humans, while the RE-Free level supports a more complete but less efficient description of an artifact.

CHAPTER 6: SIMILARITY STUDY

6.1 Motivation

After understanding the function of an artifact and developing a function model, designers can search for potential solutions to each function through benchmarking, patent searches, or catalogs, or they can use their own knowledge to identify solutions [1, 2]. Research in design-by-analogy is currently being conducted by several groups to assist designers in this search for solutions by formally searching for ideas from different domains. Linsey and colleagues have studied the cognitive process that designers use when searching for analogies and have shown that function-based descriptions improve designers' ability to identify potential solutions [83]. Goel and colleagues show that functional and causal design patterns allow designers to identify and apply analogies in design problems [84]. McAdams and colleagues use functional similarity as a basis for analogical comparisons and have demonstrated a method for design-by-analogy through the application of a similarity metric [40] to new design problem [41]. The use of function, therefore, has great potential to help designers make these analogies, aiding in concept generation. However, the level of abstraction at which functional analogies should be made has not been specified in previous research. Therefore, three levels of abstraction are explored for comparing artifacts functionally. The goal of this study is to identify an appropriate level of abstraction for finding existing artifacts that are functionally similar to a new design solution for adaptive design problems [1]. It is assumed that a set of artifacts functionally similar to a new design solution can be used as a seed for the new design, similar to the demonstration in [41]. The level of abstraction is not intended to be used to identify analogies for novel concept generation.

6.2 Frame of Reference

6.2.1 Function-Based Similarity Metric

A quantitative similarity metric has been developed by McAdams and colleagues [40] that uses customer needs and a product function matrix (PFM) to compute similarity. A PFM contains all functions performed by a set of artifacts on the left of the matrix and the list of artifacts across the top of the matrix. The cells in the matrix show the number of times the given artifact performs the given function. For example, the PFM of a coffee maker, vacuum cleaner, and flashlight would include at least the subset of functions shown in Table 6-1. The coffee maker *converts electrical energy to thermal energy* one time. The coffee maker may also *convert electrical energy to electromagnetic energy (light)* to indicate that it is turned on. The flashlight also performs this function, but it is a more important function for the flashlight than for the coffee maker. For this reason, customer needs are used in the similarity metric to give weight to each function for each artifact. The weighted functions are then used in the similarity metric to determine the overall similarity between artifacts.

Table 6-1: Sample Functions in a Product Function Matrix

Functions	Artifacts		
	coffee maker	vacuum cleaner	flashlight
convert electrical energy to thermal energy	1	0	0
convert electrical energy to mechanical energy	0	1	0
convert electrical energy to electromagnetic energy	1	0	1

Functional information for over 130 artifacts is stored in the design repository [31] (see Section 2.1.2) and the product function matrix (PFM) for these artifacts is obtained and used in this research with the similarity metric. Customer needs for each artifact are not included in the design repository, so it is assumed in this research that all functions have equal weighting. Thus, the PFM for all artifacts in the repository can be directly used to compute similarity using the metric developed by McAdams and colleagues [40, 41]. In this metric, each column in the $m \times n$ PFM is treated as a m-dimensional vector and is normalized so that it has a magnitude of one. The similarity of two artifacts is calculated by taking the projection of these vectors [40, 41]. A $n \times n$ artifact similarity matrix can be created that includes these vector projections between each artifact. For example, the similarity of a selection of vacuum cleaners from the repository is shown in Table 6-2. The matrix is symmetric, and the diagonal has values of one since an artifact is exactly similar to itself. The resulting similarity values are used for relative comparisons of similarity between artifacts, not as an absolute measure of similarity [40]. Specific details about this similarity metric can be found in research conducted by McAdams and colleagues [40, 41].

Table 6-2: Similarity of Various Vacuum Cleaners

	A	B	C	D	E	F	G
A Black and Decker Dustbuster	1.00	0.74	0.59	0.64	0.83	0.52	0.75
B Bissell Hand Vac	0.74	1.00	0.72	0.70	0.60	0.87	0.94
C Blowervac	0.59	0.72	1.00	0.49	0.55	0.72	0.73
D Bugvac	0.64	0.70	0.49	1.00	0.34	0.78	0.77
E Dirt Devil Vacuum	0.83	0.60	0.55	0.34	1.00	0.40	0.65
F IRobot Roomba	0.52	0.87	0.72	0.78	0.40	1.00	0.93
G Shopvac	0.75	0.94	0.73	0.77	0.65	0.93	1.00

6.2.2 DSM Clustering

Design Structure Matrices (DSM) can be used to capture relationships between two identical domains. The similarity matrix shown in Table 6-2 is a DSM because it captures artifact-artifact similarity. Algorithms have been developed to help manage the domain of interest by rearranging the rows and columns of the DSM. Thebeau developed a clustering algorithm to improve modularity of components in an elevator system [85]. Since the algorithm identifies and groups closely related items in a DSM, it can be used to identify clusters of similar artifacts in an artifact similarity matrix. The algorithm has several input parameters, such as the maximum cluster size or a penalty for large clusters, that can be changed by the user [85]. In this research, the default values for these parameters are used to ensure an unbiased comparison of abstraction levels.

The clustering algorithm intentionally uses a random starting point for clustering, so each run of the algorithm produces different results. A "likeness" metric is used to compare multiple runs of the algorithm with identical input parameters. The likeness of one cluster to another is twice the intersection of elements in the two clusters divided by the total number of elements in the two clusters. To determine the likeness of one run to another run, the likeness of each cluster in the first run is computed with respect to each cluster in the second run. The closest matching clusters from the two runs are used to determine the likeness of the two runs. A complete discussion and example of the likeness metric can be found in Thebeau's research [85].

6.2.3 Levels of Abstraction

Previous research has identified two dimensions of abstraction in function models—model size and term specificity—through an interpretability user study [78]. In this research, term specificity is held constant through the use of the secondary level of the Functional Basis, while model size is used to vary the level of abstraction of function models. A larger model will tend to describe more details about the artifact than a small model, so the large model is more decomposed, or less abstract, than a small model. It is important to note that model size is used for relative comparisons of abstraction within a single artifact, not for comparisons across artifacts. There are many factors that can affect the size of a model, such as the artifact's complexity, so the size of models for different artifacts are not compared. The three levels of abstraction, from lowest to highest, are:

Level One – Including Supporting Functions

Level Two – Excluding Supporting Functions

Level Three – Pruning Rules Applied

6.2.3.1 Supporting Functions

The functions stored in the design repository are identified as supporting functions if they describe assembly relationships of the artifact [58]. For example, many screws in the repository perform the function *couple solid*, which describes the assembly relationship between two components fastened to each other by the screw. Supporting functions represent a greater level of decomposition than non-supporting, or conceptual level, functions because they describe the physical connections between components [58]. Supporting functions can exist only if the artifact's architecture is already known. When supporting functions are included in the model, the model is at the lowest level of abstraction available in the design repository, defined as Level One in this research.

All functions in the repository are identified as supporting or not, and the PFMs can be obtained from the repository either including or excluding supporting functions. When supporting functions are not included, the size of the model is reduced, increasing the level of abstraction. Further, the functions that remain are conceptual functions, so models that exclude supporting functions are defined as Level Two in this research.

6.2.3.2 Pruning Rules

To further increase the level of abstraction of the function models, additional functions are removed from the models. Therefore, pruning rules (see Section 2.1.4) are used to remove highly decomposed functions. The pruning rules were developed for graphical function models in the repository, so they have been modified for application to PFMs, which relate functions to artifacts by the number of times an artifact accomplishes a particular function. Rules that referred to flows in the function structure are no longer applicable as PFMs are not graph-based. A rule specifying the combination of consecutive convert functions cannot be applied because the order of functions is not captured in PFMs. A rule is also added to remove all *guide solid* functions, which are frequently used to describe assembly relationships in PFMs but did not appear in the function structures that were used when developing the pruning rules. Tertiary terms have been removed from the rules since they are not used in this research. The following nine rules were applied to PFMs to increase their level of abstraction, resulting in conceptual-level models:

- Remove all *import* and *export* functions.
- Remove all *channel*, *transfer*, and *guide* functions referring to any type of *energy*, *signals*, or *human material*.
- Remove all *couple* functions referring to any type of *solid*.
- Remove all *support*, *stabilize*, *secure*, and *position* functions.
- Remove all *control magnitude*, *actuate*, *regulate*, *change*, and *stop* functions.
- Remove all *provision*, *store*, and *supply* functions referring to any type of *energy* or *signal*.
- Remove all *distribute* functions referring to any type of *energy*.
- Remove all *signal*, *sense*, *indicate*, and *process* functions.
- Remove all *guide solid* functions.

An example of the three levels of abstraction used in this research is shown in Table 6-3. The initial PFM, which includes supporting functions, contains 135 functions. When supporting functions are removed, the functions *couple solid*, *guide solid*, *position solid*, and *secure solid* are removed from the PFM, resulting in 49 total functions. Pruning further removes 32 functions, resulting in 17 functions in the pruned model.

Table 6-3: Shopvac PFM at Three Levels of Abstraction

Function	Representation Level		
	Including Supporting Functions (One)	Excluding Supporting Functions (Two)	Pruning Rules Applied (Three)
actuate control to electrical	1	1	
actuate electrical	1	1	
convert electrical to mechanical	1	1	1
convert human energy to control	1	1	1
convert mechanical to pneumatic	1	1	1
couple solid	34		
export electrical	1	1	
export gas	2	2	
export human material	1	1	
export mixture	1	1	
guide gas	5	5	5
guide human material	1	1	
guide mixture	4	4	4
guide solid	16		
import electrical	1	1	
import human energy	1	1	
import human material	3	3	
import mixture	3	3	
position solid	12		
secure solid	24		
separate mixture	1	1	1
separate mixture to gas	1	1	1
stop mixture	1	1	
store control	1	1	
store electrical to acoustic	1	1	
store electrical to mechanical	1	1	
store electrical to pneumatic	1	1	
store human energy to mechanical	1	1	
store human material	1	1	1
store mixture	1	1	1
store mixture to gas	1	1	1
transfer electrical	10	10	
Sum	135	49	17

6.3 Similarity Calculations

It is hypothesized in this research that pruned models are more abstract than unpruned models since they do not contain assembly- and component-specific details about the artifact. To test this general hypothesis, the pruning rules are applied to function models and the similarity of these artifacts is determined using a functional similarity metric. Since this similarity metric has been used in previous research within a conceptual design-by-analogy method [41], the similarity metric can be used to test the usefulness of the pruning rules for this conceptual design activity. The usefulness of the rules for other conceptual design activities is outside the scope of this paper.

6.3.1 Study of Large Artifact Set

The similarity among 128 artifacts was computed using the existing similarity metric and an equal weighting of all functions (see Section 6.2.1) at each of the three levels of abstraction. Due to the size of the results (128×128 matrix), the specific values are not presented, but general trends are discussed. The similarity matrix was then clustered using the DSM clustering algorithm (see Section 6.2.2). These results are summarized due to their length.

6.3.1.1 Results of Similarity Metric

The results of similarity at the each level of abstraction are shown as contour plots in Figure 6-1, Figure 6-2, and Figure 6-3. The values of similarity are not shown, but the trends are depicted by the shading, where darker cells represent a higher level of similarity and lighter cells represent a lower level of similarity between artifacts.

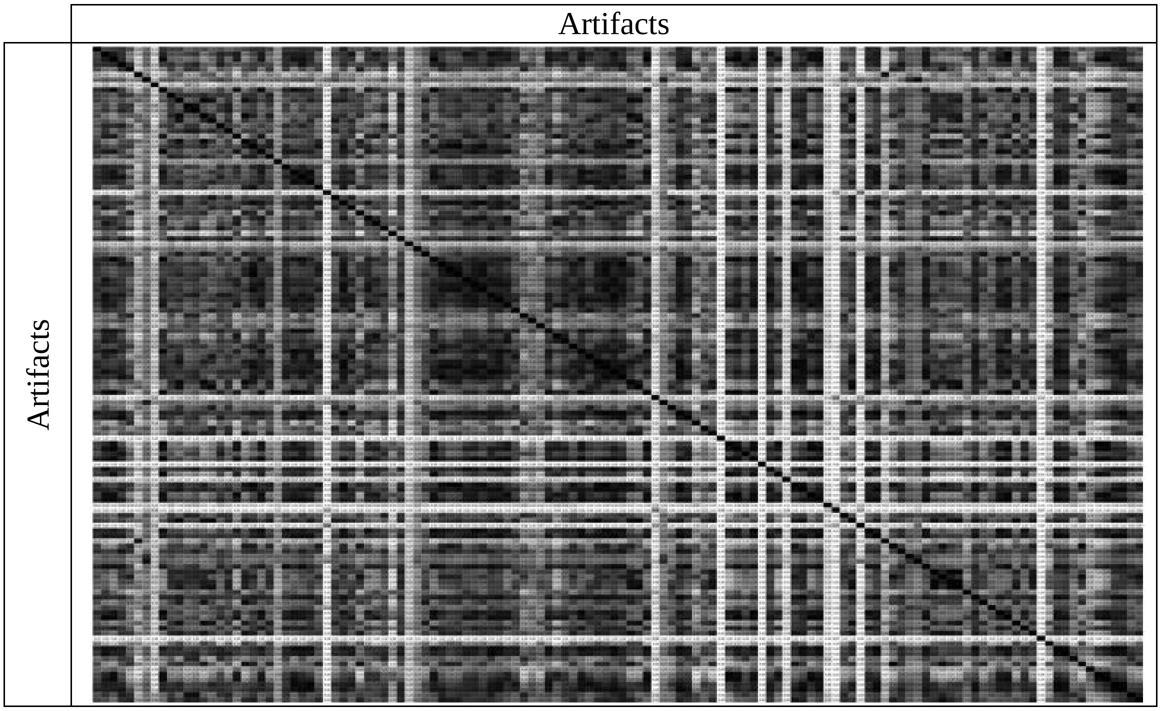


Figure 6-1: Similarity of All Artifacts at Abstraction Level One

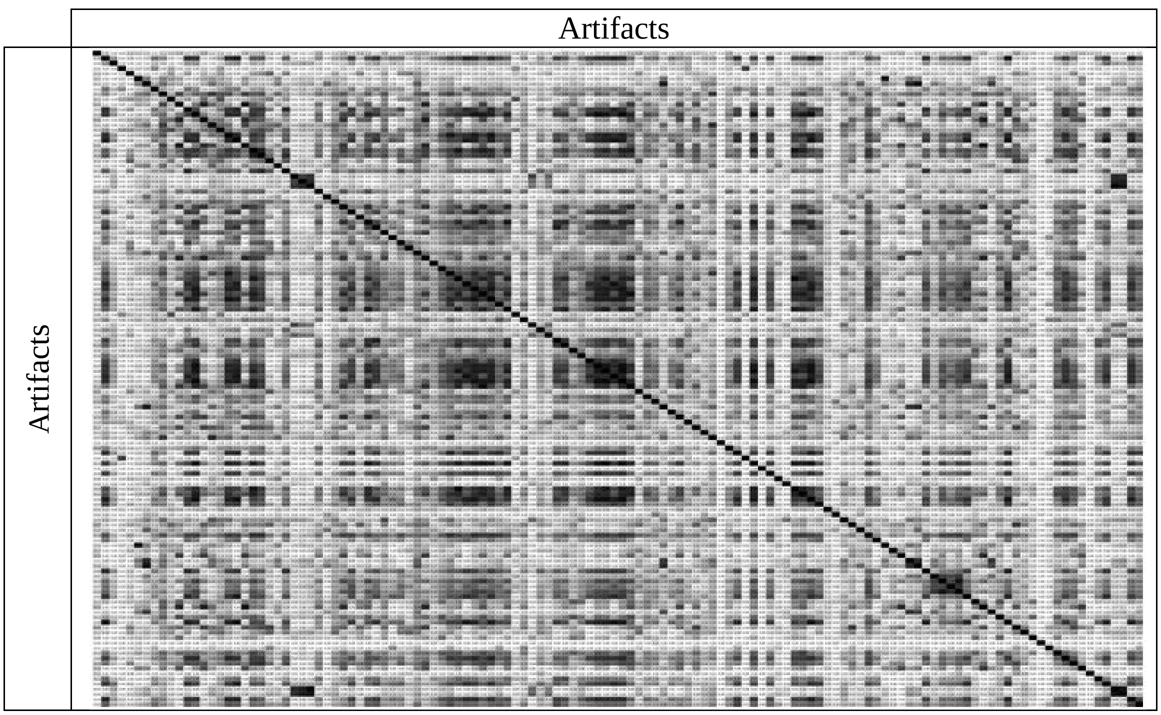


Figure 6-2: Similarity of All Artifacts at Abstraction Level Two

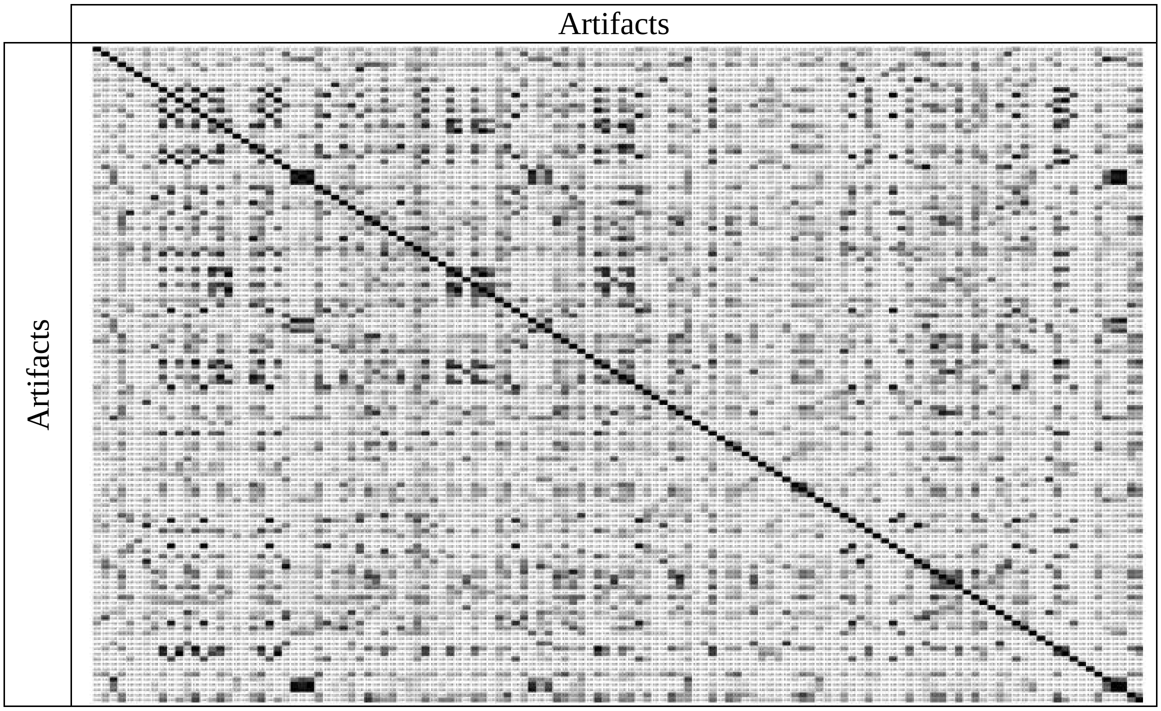


Figure 6-3: Similarity of All Artifacts at Abstraction Level Three

At Level One (Figure 6-1), all artifacts are highly similar to each other, as shown by the darker cells, with the exception of a few artifacts. The light rows represent artifacts that are dissimilar to most other artifacts. Many of these rows correspond to atypical artifacts in the repository: “brake system,” “fly,” “heart,” “jar opener,” “lichen,” “nasa anomaly,” “natural sensing,” “power station,” and “two component regulatory system.” These lighter rows are expected since the artifacts—with the exception of the jar opener—are not the typical power tools, appliances, toys, or electronics in the repository. However, beyond this observation, it is difficult to draw conclusions about the similarity of the remaining artifacts since the values of similarity are all close to each other. A wider distribution of similarity would give greater confidence in the results when comparing pairs of artifacts.

At Level Two (Figure 6-2), pairs of artifacts with a high similarity are easier to identify compared to Level One. The average similarity of all artifacts is smaller, and a greater distinction exists among similarity values, causing closely related artifacts to stand out from the remaining values. This increase in distinction is caused by the exclusion of supporting functions in the models. Supporting functions describe the assembly of components to each other, and are modeled as *position solid*, *guide solid*, *couple solid*, or *secure solid* in the repository 99.7% of the time. Furthermore, there are almost as many supporting functions as non-supporting functions, so at Level One, approximately half of all functions are one of these four supporting functions. Therefore, when including supporting functions, these four functions cause the similarity of all artifacts to be closer together and higher. When the supporting functions are excluded, artifacts are not evaluated on how they are assembled, but on what the artifact does. For this reason, pairs of similar artifacts are more pronounced in Figure 6-2 than in Figure 6-1. It is important to note that the average similarity or measures of the spread of values in the matrix cannot be used to draw conclusions since the desired spread is not known. The average similarity or spread should not necessarily increase or decrease with a higher level of abstraction because it will depend on the artifacts being compared.

Abstraction Level Three—with pruning rules applied—results in an even greater distinction of similarity among artifacts, as shown in Figure 6-3. Pruning rules further increase the level of abstraction of the model by removing functions from the reverse-engineered function structure that would not likely be addressed at the conceptual stage of design, such as *transfer electrical energy*, or *distribute electrical energy*. Like supporting functions, these pruned functions are performed frequently by many artifacts, so they increase the similarity among artifacts and reduce the distinction between values at Level Two compared to Level Three. After removing these functions, closely related pairs of artifacts are more apparent in the figure. The greater distinction in similarity values may also give more confidence when comparing an artifact, A, to two other artifacts, B and C. If the artifacts are compared at Level One, it is likely that the similarity between A and B and the similarity between B and C differs by a small amount. At Level Three, however, these similarity values may differ by a much higher amount, providing a greater confidence that one pair is actually more similar than another pair.

The results of the similarity metric at three levels of abstraction show that higher levels of abstraction provide a greater distinction in similarity values. Thus, when searching for similar artifacts, there will be a smaller set of artifacts that are closely related to the artifact of interest. For example, the similarity of a vacuum cleaner to all other artifacts in the repository is shown for all three levels of abstraction in Figure 6-4. The 128 artifacts are sorted from most similar to least similar on the horizontal axis. At Level One, the sorted list of similar artifacts slowly decreases in similarity for the first sixty artifacts, all of which have a similarity greater than 0.8. At Level Two, there are only a few highly similar artifacts and the remaining artifacts decrease in similarity at a steady rate. At Level Three, the similarity decreases quickly with each artifact, but at a decreasing rate. At this level of abstraction, a few artifacts are highly similar to the vacuum cleaner, while the remainder, which are of less interest, are much less similar.

These trends can also be seen in Table 6-4, where the number of artifacts in various similarity ranges is given for each level of abstraction. At Level One, 59 artifacts have a similarity of greater than 0.80, while only one artifact has this high degree of similarity at Levels Two and Three. Level One has a high percentage of artifacts with a high similarity, while Level Three has a high percentage of artifacts with low values.

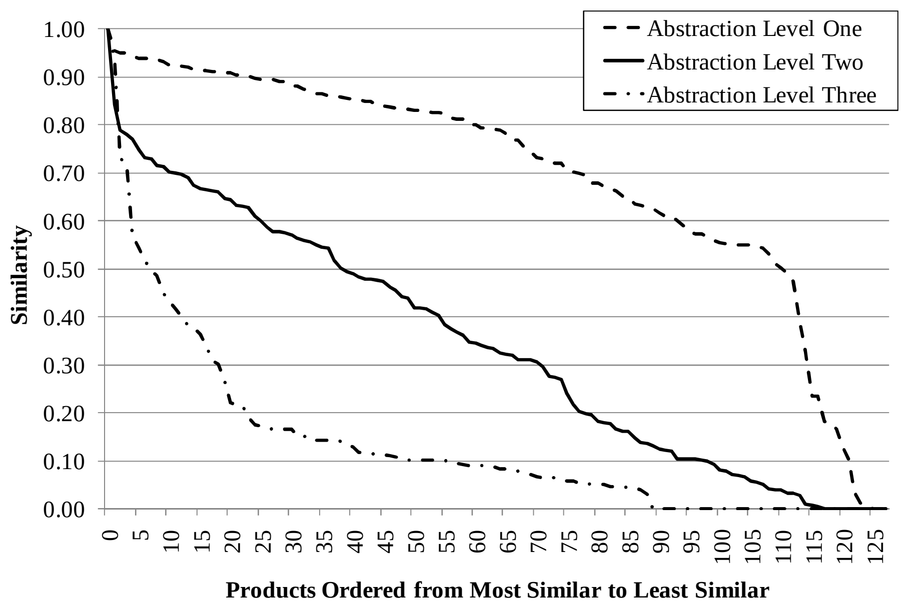


Figure 6-4: Similarity of a Vacuum Cleaner to All Other Artifacts in the Repository at Three Levels of Abstraction

Table 6-4: Degree of Similarity Between a Vacuum Cleaner and All Other Artifacts

	Representation Level		
Similarity Value	One	Two	Three
0.80 - 1.00	59	1	1
0.60 - 0.80	33	24	2
0.40 - 0.60	20	29	8
0.20 - 0.40	4	23	11
0.00 - 0.20	8	39	66

The similarity metric used in these calculations has been used in previous research for a design-by-analogy demonstration by computing the similarity of a new artifact's function model to the functionality of existing artifacts. The artifacts are then ranked according to similarity and an artifact with a high similarity is chosen on which to base the new design [41]. If artifacts are compared at Level Three, the sorted set of results will give more confidence that the first few results in the list are of greater interest than the rest of the artifacts because the similarity decreases quickly.

The high degree of similarity at Level One is caused by the supporting functions in the models. Since most of the artifacts compared contain assembly features, such as screws, then they are found to be highly similar to each other. This assembly-dominated similarity result is not helpful for function-based design-by-analogy. In design-by-analogy, similarity should be used to draw analogies at a functional level, allowing analogies to be made across domains. The similarity calculations at Level One do not provide this type of analogy. At Level Two, artifact similarity is dominated by the means for achieving functions, rather than the functions themselves. Many functions at Level Two represent a one-to-one mapping with system components, such as wires, which are included only to achieve higher-level functions. Since most of the artifacts contain similar means (used to achieve different high-level functions), the similarity metric at Level Two is a means-dominated metric, which will not produce the desired results for design-by-analogy. At Level Three, the assembly- and means-based functions are absent from the model, so the similarity results are based only on the high-level function of the artifact. These high-level functions are best for drawing new analogies across domains because they focus on the transformative purpose of the artifact rather than its embodiment.

6.3.1.2 Results of Clustering

The previous section demonstrates that similarity at higher levels of abstraction results in a smaller set of highly similar artifacts, which can give greater confidence in the results. However, the accuracy of the results has not been evaluated. In order to assess the accuracy of the results from the similarity metric, a DSM clustering algorithm is used to identify groups of similar artifacts in the similarity matrix. These clusters are then evaluated to understand the quality of the values in the matrix. As discussed in Section 6.2.2, the clustering algorithm produces different results each time it is executed. The clustering algorithm was executed many times at each level of abstraction and trends in the clusters were observed. The results of one representative execution of the algorithm at each level of abstraction are presented.

The first five clusters identified by the algorithm at each level of abstraction are shown in Table 6-5. Clusters are labeled A through E in the table for referencing only. There is no relationship between clusters across abstraction levels. The asterisks (*) indicate artifacts that belong to more than one cluster. The first several clusters typically contain five or six artifacts; beyond these first few clusters, the size decreases to two or three artifacts per cluster. The sizes of the resulting clusters are based on input parameters to the clustering algorithm. The default parameters assign a penalty to large clusters, so the largest clusters contained approximately six to seven artifacts. When the penalty was reduced, the clusters increased significantly in size, and it was difficult to determine the similarity between artifacts in a given cluster, as they differed greatly. With smaller clusters, typically there were several artifacts that performed similar functions, so it was assumed that these similar artifacts formed the basis for the cluster. For example, the second cluster at Level Two is assumed to be types of power tools. The number of clusters in each run varied from approximately fifty to sixty clusters.

Table 6-5: First Five Resulting Clusters of Artifacts at Each Level of Abstraction

Cluster	Level One	Level Two	Level Three
A	b and d dustbuster b and d jigsaw dirt devil vacuum *nasa anomaly vibrating razor yoda doll	b and d drill attachment b and d sander attachment razor scooter stapler vise grip	*black 12 cup deluxe coffee dishwasher durabrand iron tractor sprinkler white 4 cup economy coffee
B	b and d power pack dryer hair trimmer skil circular saw skil flashlight	b and d power pack *delta drill delta jigsaw delta sander versapak sander	b and d mini router attachment delta circular saw delta jigsaw firestorm drill giant bicycle
C	b and d can opener b and d sliceright datsun truck *holmes fan irobot roomba	*delta drill *delta nail gun firestorm drill irobot roomba mac cordless dril-driver	ball shooter first shot nerf gun stapler *tippman paintball gun
D	air hawg toy plane brother sewing machine *delta circular saw *delta nail gun firestorm drill	b and d palm sander b and d screwdriver b and d sliceright giant bicycle vibrating razor	*b and d power pack delta drill delta sander slow cooker
E	*b and d drill attachment b and d jigsaw attachment b and d sander attachment tractor sprinkler *ub roller coaster	air purifier coolit drink cooler shopvac supermax hair dryer yoda doll	*black 12 cup deluxe coffee black 12 cup economy coffee black 4 cup regular coffee white 12 cup regular

The results of at least five executions of the clustering algorithm were studied to determine trends at each level of abstraction. At Level One, the clusters typically did not include many similar artifacts. At most, two or three artifacts in the cluster were similar to each other (based on overall functionality). For example, Cluster A contains two vacuum cleaners. It could be argued that the razor and jigsaw are similar because they both remove material, but beyond these two possibilities, these artifacts are not very similar.

Level Two produced better clusters than Level One, as demonstrated by the second column in Table 6-5. Typically, more than half of the artifacts in each cluster were closely related to each other. For example, four of the five artifacts in Cluster C are power tools. However, some clusters, such as Cluster A, did not represent a group of similar artifacts.

The clustering algorithm produced the best clusters at Level Three. Most of the artifacts in each cluster were related by some high-level functionality. For example, all of the artifacts in Cluster A transport water, and four of them heat the water significantly. In Cluster B, four of the five artifacts are power tools, and in Cluster C, three of the four artifacts are toy guns. These results are the most meaningful for function-based similarity because the algorithm results in clusters of functionally-similar artifacts. These types of results would be useful in conceptual design when searching for analogies to a new design problem. If the high-level function of a new artifact is identified, its similarity to known artifacts can be computed and the clustering algorithm will group it with functionally similar artifacts. The artifacts in the same cluster as the new design can then be used to help the designer begin to embody the idea.

6.3.2 Study of Three Groups of Artifacts

To better understand the quality of the abstraction levels for computing similarity, a subset of artifacts is chosen for a more in-depth study. The subset contains three groups of artifacts that are assumed to be similar based on their overall purposes: making coffee, removing dirt from a floor, or producing light. Furthermore, they are similar because they accomplish these high-level purposes with similar working principles. The following three groups of known similar artifacts—coffee makers, vacuum cleaners, and flashlights—were selected for this study:

- Coffee Makers – artifacts that heat water
    - black 12 cup deluxe coffee
    - black 12 cup economy coffee
    - black 4 cup regular coffee
    - white 12 cup regular
    - white 4 cup economy coffee
- Vacuum Cleaners – artifacts that remove dirt from a floor
    - bissell hand vac
    - blowervac
    - bugvac
    - dirt devil vacuum
    - irobot roomba
    - shopvac

- Flashlights – artifacts that produce light
    - delta flashlight
    - firestorm flashlight
    - skil flashlight

In addition to these fourteen artifacts, an artifact similar to each group is chosen from the repository to determine if the similarity metric and clustering algorithm finds them to be similar. An iced tea maker ("mr coffee iced tea maker") is chosen as the artifact most similar to the coffee makers since it shares common functionality with a coffee maker, such as heating water and dripping it over coffee or tea. The tea maker has also been used to validate the results of the similarity metric [40]. The artifact similar to the vacuum cleaner is a hair dryer ("supermax hair dryer") since it, like the vacuum cleaner, creates a flow of air through the system. There are not any artifacts closely related to the flashlights, so a camera is chosen because a secondary purpose of the camera is to produce light. In addition to these three artifacts, an artifact not similar to coffee makers, vacuum cleaners, and flashlights is chosen to determine if it appears in its own cluster. This dissimilar artifact is a computer mouse ("apple usb mouse"), since it does not share overall functionality with these artifacts.

In order to validate the use of pruning rules for similarity, the accuracy and precision of the clusters are computed at each level of abstraction. The accuracy and precision metrics are explained in Sections 6.3.2.3 and 6.3.2.4. Further, an additional random level of abstraction is created to ensure that the specific selection of functions removed from Level Three is responsible for the results, not random chance. To achieve the random level of abstraction, 313 functions—the same number removed through pruning—are randomly removed from the 128-artifact PFM at abstraction Level One. The similarity of these artifacts is computed and the results are used for clustering.

6.3.2.1 Results of Similarity Metric and Clustering

The similarity of the 18 artifacts was computed and the resulting DSM clustered as explained in Section 6.3.2 at the four levels of abstraction. The clustering algorithm was run ten times for each level of abstraction and the trends in the clusters were analyzed. One representative data set from clustering at each level of abstraction is shown in Figure 6-5, Figure 6-6, Figure 6-7, and Figure 6-8. The artifacts are grouped according to the clusters identified by the clustering algorithm, and the similarity values are included in the matrices. The cells are shaded from light to dark based on the lowest and highest values in the given matrix.

			1						2			3			4		5		6	
			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	A	black 12 cup deluxe coffee	1.00	0.96	0.92	0.96	0.92	0.76	0.91	0.93	0.84	0.87	0.81	0.90	0.61	0.88	0.40	0.30	0.38	0.58
	B	black 4 cup regular coffee	0.96	1.00	0.86	0.98	0.96	0.75	0.91	0.89	0.83	0.86	0.81	0.85	0.68	0.89	0.51	0.42	0.53	0.74
	C	irobot roomba	0.92	0.86	1.00	0.87	0.89	0.87	0.91	0.96	0.93	0.78	0.87	0.94	0.72	0.92	0.52	0.40	0.44	0.50
	D	white 12 cup regular	0.96	0.98	0.87	1.00	0.97	0.77	0.93	0.92	0.86	0.89	0.83	0.89	0.64	0.92	0.53	0.48	0.52	0.76
	E	white 4 cup economy coffee	0.92	0.96	0.89	0.97	1.00	0.83	0.96	0.93	0.91	0.82	0.84	0.88	0.74	0.95	0.68	0.60	0.65	0.78
2	F	bissell hand vac	0.76	0.75	0.87	0.77	0.83	1.00	0.85	0.88	0.94	0.70	0.86	0.88	0.72	0.89	0.69	0.60	0.58	0.56
	G	black 12 cup economy coffe	0.91	0.91	0.91	0.93	0.96	0.85	1.00	0.95	0.94	0.77	0.82	0.90	0.68	0.92	0.66	0.61	0.54	0.64
	H	mr coffee iced tea maker	0.93	0.89	0.96	0.92	0.93	0.88	0.95	1.00	0.95	0.84	0.90	0.97	0.67	0.95	0.58	0.51	0.48	0.60
	I	shopvac	0.84	0.83	0.93	0.86	0.91	0.94	0.94	0.95	1.00	0.77	0.89	0.94	0.73	0.95	0.70	0.65	0.58	0.63
3	J	bugvac	0.87	0.86	0.78	0.89	0.82	0.70	0.77	0.84	0.77	1.00	0.85	0.88	0.49	0.87	0.32	0.34	0.34	0.69
	K	delta flashlight	0.81	0.81	0.87	0.83	0.84	0.86	0.82	0.90	0.89	0.85	1.00	0.92	0.70	0.96	0.55	0.49	0.54	0.68
	L	supermax hair dryer	0.90	0.85	0.94	0.89	0.88	0.88	0.90	0.97	0.94	0.88	0.92	1.00	0.60	0.94	0.48	0.46	0.37	0.55
4	M	blowervac	0.61	0.68	0.72	0.64	0.74	0.72	0.68	0.67	0.73	0.49	0.70	0.60	1.00	0.75	0.71	0.55	0.79	0.67
	N	firestorm flashlight	0.88	0.89	0.92	0.92	0.95	0.89	0.92	0.95	0.95	0.87	0.96	0.94	0.75	1.00	0.65	0.58	0.62	0.76
5	O	apple usb mouse	0.40	0.51	0.52	0.53	0.68	0.69	0.66	0.58	0.70	0.32	0.55	0.48	0.71	0.65	1.00	0.83	0.85	0.66
	P	dirt devil vacuum	0.30	0.42	0.40	0.48	0.60	0.60	0.61	0.51	0.65	0.34	0.49	0.46	0.55	0.58	0.83	1.00	0.72	0.63
6	Q	camera	0.38	0.53	0.44	0.52	0.65	0.58	0.54	0.48	0.58	0.34	0.54	0.37	0.79	0.62	0.85	0.72	1.00	0.80
	R	skil flashlight	0.58	0.74	0.50	0.76	0.78	0.56	0.64	0.60	0.63	0.69	0.68	0.55	0.67	0.76	0.66	0.63	0.80	1.00

Figure 6-5 Similarity and Clustering of Subset of Artifacts at Abstraction Level One

			1							2					3			4		
			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	A	bugvac	1.00	0.71	0.77	0.65	0.60	0.64	0.74	0.16	0.15	0.20	0.17	0.13	0.31	0.16	0.23	0.08	0.48	0.69
	B	delta flashlight	0.71	1.00	0.92	0.72	0.71	0.58	0.87	0.03	0.07	0.04	0.05	0.05	0.30	0.23	0.22	0.10	0.56	0.71
	C	firestorm flashlight	0.77	0.92	1.00	0.74	0.73	0.62	0.90	0.04	0.08	0.06	0.06	0.06	0.32	0.22	0.22	0.11	0.59	0.73
	D	irobot roomba	0.65	0.72	0.74	1.00	0.60	0.47	0.76	0.02	0.03	0.02	0.02	0.04	0.51	0.28	0.37	0.01	0.48	0.63
	E	mr coffee iced tea maker	0.60	0.71	0.73	0.60	1.00	0.49	0.74	0.08	0.19	0.16	0.14	0.19	0.27	0.18	0.25	0.20	0.53	0.64
	F	skil flashlight	0.64	0.58	0.62	0.47	0.49	1.00	0.50	0.14	0.19	0.24	0.17	0.16	0.25	0.27	0.24	0.16	0.35	0.44
	G	supermax hair dryer	0.74	0.87	0.90	0.76	0.74	0.50	1.00	0.02	0.04	0.02	0.02	0.03	0.37	0.14	0.35	0.09	0.66	0.84
2	H	black 12 cup deluxe coffee	0.16	0.03	0.04	0.02	0.08	0.14	0.02	1.00	0.70	0.77	0.81	0.78	0.07	0.05	0.07	0.03	0.07	0.14
	I	black 12 cup economy coffe	0.15	0.07	0.08	0.03	0.19	0.19	0.04	0.70	1.00	0.83	0.85	0.86	0.11	0.05	0.15	0.03	0.09	0.12
	J	black 4 cup regular coffee	0.20	0.04	0.06	0.02	0.16	0.24	0.02	0.77	0.83	1.00	0.97	0.91	0.08	0.09	0.08	0.07	0.06	0.08
	K	white 12 cup regular	0.17	0.05	0.06	0.02	0.14	0.17	0.02	0.81	0.85	0.97	1.00	0.92	0.07	0.06	0.08	0.03	0.06	0.10
	L	white 4 cup economy coffee	0.13	0.05	0.06	0.04	0.19	0.16	0.03	0.78	0.86	0.91	0.92	1.00	0.07	0.08	0.15	0.12	0.11	0.10
3	M	blowervac	0.31	0.30	0.32	0.51	0.27	0.25	0.37	0.07	0.11	0.08	0.07	0.07	1.00	0.36	0.47	0.07	0.30	0.31
	N	camera	0.16	0.23	0.22	0.28	0.18	0.27	0.14	0.05	0.05	0.09	0.06	0.08	0.36	1.00	0.23	0.39	0.34	0.15
	O	dirt devil vacuum	0.23	0.22	0.22	0.37	0.25	0.24	0.35	0.07	0.15	0.08	0.08	0.15	0.47	0.23	1.00	0.09	0.40	0.38
4	P	apple usb mouse	0.08	0.10	0.11	0.01	0.20	0.16	0.09	0.03	0.03	0.07	0.03	0.12	0.07	0.39	0.09	1.00	0.40	0.16
	Q	bissell hand vac	0.48	0.56	0.59	0.48	0.53	0.35	0.66	0.07	0.09	0.06	0.06	0.11	0.30	0.34	0.40	0.40	1.00	0.77
	R	shopvac	0.69	0.71	0.73	0.63	0.64	0.44	0.84	0.14	0.12	0.08	0.10	0.10	0.31	0.15	0.38	0.16	0.77	1.00

Figure 6-6: Similarity and Clustering of Subset of Artifacts at Abstraction Level Two

			1						2						3				4	
			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	A	black 12 cup deluxe coffee	1.00	0.84	0.95	0.18	0.95	0.97	0.02	0.00	0.18	0.09	0.09	0.00	0.00	0.00	0.01	0.07	0.00	0.00
	B	black 12 cup economy coffe	0.84	1.00	0.86	0.35	0.86	0.91	0.09	0.10	0.15	0.12	0.14	0.07	0.00	0.09	0.03	0.09	0.00	0.00
	C	black 4 cup regular coffee	0.95	0.86	1.00	0.27	1.00	0.97	0.03	0.00	0.22	0.12	0.11	0.00	0.00	0.00	0.02	0.09	0.00	0.00
	D	mr coffee iced tea maker	0.18	0.35	0.27	1.00	0.27	0.25	0.06	0.20	0.00	0.00	0.14	0.21	0.14	0.35	0.14	0.18	0.00	0.04
	E	white 12 cup regular	0.95	0.86	1.00	0.27	1.00	0.97	0.03	0.00	0.22	0.12	0.11	0.00	0.00	0.00	0.02	0.09	0.00	0.00
	F	white 4 cup economy coffee	0.97	0.91	0.97	0.25	0.97	1.00	0.03	0.00	0.14	0.11	0.10	0.00	0.00	0.00	0.02	0.08	0.00	0.00
2	G	bissell hand vac	0.02	0.09	0.03	0.06	0.03	0.03	1.00	0.22	0.50	0.62	0.94	0.75	0.00	0.07	0.43	0.07	0.00	0.00
	H	blowervac	0.00	0.10	0.00	0.20	0.00	0.00	0.22	1.00	0.38	0.30	0.21	0.35	0.00	0.22	0.13	0.00	0.20	0.31
	I	bugvac	0.18	0.15	0.22	0.00	0.22	0.14	0.50	0.38	1.00	0.46	0.38	0.13	0.00	0.00	0.43	0.17	0.00	0.00
	J	dirt devil vacuum	0.09	0.12	0.12	0.00	0.12	0.11	0.62	0.30	0.46	1.00	0.52	0.64	0.00	0.00	0.13	0.27	0.00	0.00
	K	shopvac	0.09	0.14	0.11	0.14	0.11	0.10	0.94	0.21	0.38	0.52	1.00	0.73	0.00	0.06	0.40	0.00	0.00	0.00
	L	supermax hair dryer	0.00	0.07	0.00	0.21	0.00	0.00	0.75	0.35	0.13	0.64	0.73	1.00	0.00	0.16	0.05	0.00	0.00	0.00
3	M	delta flashlight	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.63	0.49	0.63	0.00	0.14
	N	firestorm flashlight	0.00	0.09	0.00	0.35	0.00	0.00	0.07	0.22	0.00	0.00	0.06	0.16	0.63	1.00	0.63	0.80	0.00	0.18
	O	irobot roomba	0.01	0.03	0.02	0.14	0.02	0.02	0.43	0.13	0.43	0.13	0.40	0.05	0.49	0.63	1.00	0.66	0.00	0.14
	P	skil flashlight	0.07	0.09	0.09	0.18	0.09	0.08	0.07	0.00	0.17	0.27	0.00	0.00	0.63	0.80	0.66	1.00	0.00	0.18
4	Q	apple usb mouse	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.83
	R	camera	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.14	0.18	0.14	0.18	0.83	1.00

Figure 6-7: Similarity and Clustering of Subset of Artifacts at Abstraction Level Three

			1					2				3			4		5		6	
			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	A	black 12 cup deluxe coffee	1.00	0.99	0.91	0.96	0.99	0.43	0.58	0.65	0.59	0.59	0.82	0.80	0.94	0.96	0.94	0.96	0.90	0.92
	B	black 4 cup regular coffee	0.99	1.00	0.87	0.94	0.98	0.50	0.64	0.70	0.65	0.63	0.85	0.84	0.97	0.98	0.94	0.96	0.91	0.93
	C	bugvac	0.91	0.87	1.00	0.93	0.93	0.13	0.29	0.38	0.34	0.40	0.60	0.59	0.74	0.80	0.81	0.85	0.76	0.77
	D	supermax hair dryer	0.96	0.94	0.93	1.00	0.98	0.39	0.54	0.62	0.53	0.62	0.80	0.79	0.88	0.93	0.94	0.96	0.92	0.93
	E	white 12 cup regular	0.99	0.98	0.93	0.98	1.00	0.41	0.56	0.63	0.57	0.58	0.81	0.79	0.92	0.96	0.94	0.97	0.91	0.93
2	F	apple usb mouse	0.43	0.50	0.13	0.39	0.41	1.00	0.86	0.88	0.76	0.76	0.78	0.79	0.64	0.64	0.61	0.57	0.65	0.65
	G	blowervac	0.58	0.64	0.29	0.54	0.56	0.86	1.00	0.94	0.89	0.75	0.92	0.86	0.76	0.75	0.72	0.71	0.74	0.76
	H	camera	0.65	0.70	0.38	0.62	0.63	0.88	0.94	1.00	0.86	0.81	0.93	0.89	0.80	0.81	0.79	0.77	0.84	0.83
	I	dirt devil vacuum	0.59	0.65	0.34	0.53	0.57	0.76	0.89	0.86	1.00	0.73	0.87	0.79	0.76	0.73	0.69	0.69	0.70	0.72
3	J	bissell hand vac	0.59	0.63	0.40	0.62	0.58	0.76	0.75	0.81	0.73	1.00	0.79	0.85	0.71	0.71	0.71	0.67	0.76	0.76
	K	irobot roomba	0.82	0.85	0.60	0.80	0.81	0.78	0.92	0.93	0.87	0.79	1.00	0.96	0.92	0.93	0.92	0.91	0.92	0.94
	L	shopvac	0.80	0.84	0.59	0.79	0.79	0.79	0.86	0.89	0.79	0.85	0.96	1.00	0.91	0.92	0.90	0.89	0.91	0.93
4	M	black 12 cup economy coffee	0.94	0.97	0.74	0.88	0.92	0.64	0.76	0.80	0.76	0.71	0.92	0.91	1.00	0.98	0.94	0.95	0.93	0.95
	N	white 4 cup economy coffee	0.96	0.98	0.80	0.93	0.96	0.64	0.75	0.81	0.73	0.71	0.93	0.92	0.98	1.00	0.97	0.98	0.95	0.98
5	O	mr coffee iced tea maker	0.94	0.94	0.81	0.94	0.94	0.61	0.72	0.79	0.69	0.71	0.92	0.90	0.94	0.97	1.00	0.98	0.96	0.98
	P	skil flashlight	0.96	0.96	0.85	0.96	0.97	0.57	0.71	0.77	0.69	0.67	0.91	0.89	0.95	0.98	0.98	1.00	0.96	0.98
6	Q	delta flashlight	0.90	0.91	0.76	0.92	0.91	0.65	0.74	0.84	0.70	0.76	0.92	0.91	0.93	0.95	0.96	0.96	1.00	0.98
	R	firestorm flashlight	0.92	0.93	0.77	0.93	0.93	0.65	0.76	0.83	0.72	0.76	0.94	0.93	0.95	0.98	0.98	0.98	0.98	1.00

Figure 6-8: Similarity and Clustering of Subset of Artifacts at Random Level of Abstraction

At Level One (Figure 6-5), the clusters are inconsistent. In the ten runs, almost all of the clusters contain artifacts from at least two categories, and many of the clusters contain artifacts from all three categories. For example, the second cluster contains two vacuum cleaners and a coffee maker as well as an added similar artifact, the iced tea maker. The added similar artifacts usually have at least one similar artifact in its cluster, but in some instances these artifacts are clustered with only dissimilar artifacts. Furthermore, the mouse is clustered with all three types of artifacts, and in one run appears in three clusters. The clusters at Level One are difficult to identify from the figure based on the shading of the cells alone. As with the clustering of all 128 artifacts, the clusters are closely related to each other. By inspection of Figure 6-5, it appears that clusters 1, 2, 3, and potentially 4 should be one large cluster.

At Level Two, which excludes supporting functions, the clusters are more consistent between runs of the algorithm. The coffee makers are almost always clustered together (Figure 6-6), but the tea maker does not appear in a cluster with a coffee maker in any of the runs. Other clusters typically have a majority of artifacts that are from one category of artifacts, but almost every non-coffee maker cluster has flashlights and vacuum cleaners as well as some of the additional artifacts. For example, the first cluster in Figure 6-6 contains three flashlights, two vacuum cleaners, the iced tea maker, and the hair dryer. The second cluster contains all of the coffee makers, but it does not include the tea maker, which would be desired. The relationship between clusters is more distinct at this level than the first. For example, the relationship between the first and fourth clusters can be identified by a group of darker shaded cells, which is expected since both clusters contain vacuum cleaners. Furthermore, the coffee maker cluster is not strongly related to any other artifacts, as demonstrated by the lighter cells in the rows containing coffee makers.

At Level Three (Figure 6-7), when pruning rules are applied to the function models, the clusters are much more consistent between runs of the algorithm. Distinct clusters of each type of artifact are apparent in each of the runs at this level of abstraction. The hair dryer is always clustered with vacuum cleaners, and the iced tea maker is clustered with coffee makers in three of the five runs. The camera, however, is not clustered with the flashlights, but instead is paired with the mouse in all five runs. This result, while not anticipated, is acceptable since the camera is also an electronic device. The clustering demonstrates that the camera is more similar to the mouse than the flashlights. The relationship between clusters of artifacts is much lower at Level Three than Levels One and Two, as none of the clusters are strongly related to other clusters. The flashlight cluster is slightly related to the vacuum cleaner cluster because the "irobot roomba" is clustered with the flashlights. This point is discussed further in the next paragraph. Aside from this relationship, all clusters are well defined and make logical sense in terms of similarity.

One interesting result at Level Three is the clustering of the "irobot roomba" vacuum cleaner with flashlights in all five runs. Upon further inspection, the PFM of the this artifact contains eight instances of *converting electrical energy to electromechanical energy*, performed by various sensors, causing it to be more similar to the flashlights than vacuum cleaners. However, this result is not desirable since main functionality of the "irobot roomba" is not to produce light. This discrepancy can potentially be addressed by using customer needs to assign weights to functions in the PFM, as described in the similarity metric used in this research [40]. This would allow the function of *convert* *electrical energy to pneumatic energy* in the vacuum to be weighted much higher than the function of the eight sensors, causing it to be more similar to vacuum cleaners than flashlights. However, this requires knowledge of the customer needs and the mapping between each function and customer needs. A second approach to addressing this problem requires an extension of current functional representations to include flow attributes. If attributes of flows are captured, such as the intensity of the output energy, then the similarity metric could use this information to determine that the sensors on a vacuum cleaner do not fulfill the function of a flashlight. The need for attributes of flows in function models has been identified in related research [30, 86], and is currently being investigated. This approach would also require an additional vocabulary of flow attributes, knowledge of the attributes of all flows, and refinement of the similarity metric to compare the magnitudes of flows.

The results from the random level of abstraction (Figure 6-8), were similar to the results from Level One. Artifacts from all three groups frequently occurred in a single cluster, and it is difficult to distinguish clusters in the figure. The clusters are highly related to each other and are not intuitive. Therefore, the improved results at Level Three are caused by the specific functions removed, not by simply removing any functions at random.

6.3.2.2 Discussion of Similarity Results

The similarity among artifacts within this subset of vacuum cleaners, flashlights, and coffee makers varies greatly depending on the level of abstraction used to compute similarity. In order to understand the similarity metric and results at each level, the “shopvac” is compared to the three groups of artifacts, as shown in Figure 6-9 through Figure 6-11. The abstraction level in which functions are randomly removed is not discussed in this section because it does not represent a true level of abstraction and cannot be placed in the sequence of Levels One, Two, and Three.

The similarity between the “shopvac” and coffee makers is high at Level One, as shown in Figure 6-9, and it is more similar to many of the coffee makers than other vacuum cleaners (compare to Figure 6-11). This high level of similarity is caused by the inclusion of supporting functions, which describe assembly relationships among components. Since both the “shopvac” and coffee makers are assembled together in some manner, they share many common supporting functions, causing them to have this high degree of similarity. At Level Two, the exclusion of supporting functions causes the similarity between the “shopvac” and coffee makers to decrease significantly to approximately 0.1. These values are more desirable than the previous since the “shopvac” and coffee makers do not share the same high-level purpose. At Level Three, the similarity remains approximately the same, indicating that Level Three does not change the level of similarity in this particular case. Thus, the pruning rules used to arrive at Level Three successfully remove the supporting functions from the models that cause a high degree of similarity at Level One.

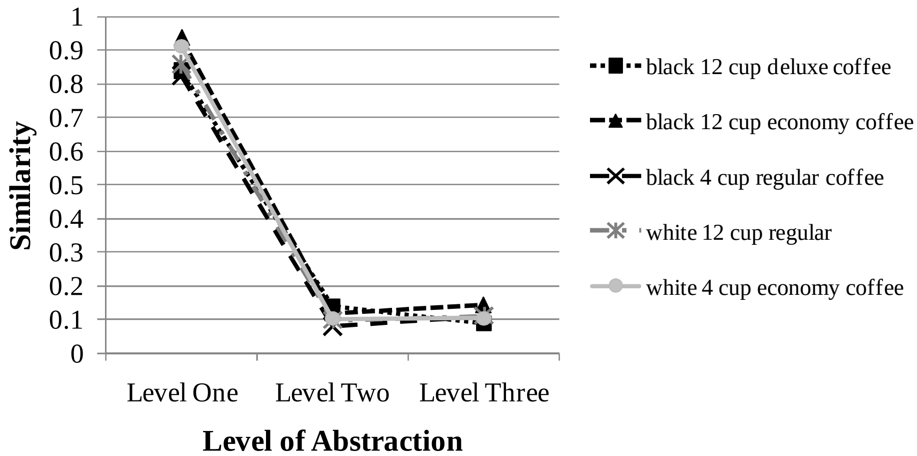


Figure 6-9: Similarity Between Shopvac and Coffee Makers at Three Levels of Abstraction

The similarity between the “shopvac” and the three flashlights (see Figure 6-10) is relatively high at Level One, which is caused by the supporting functions. At Level Two, the similarity between the “shopvac” and flashlights decreases, as does the similarity between the “shopvac” and other vacuums (see Figure 6-11), resulting in two of the flashlights being more similar to the “shopvac” than four of the vacuum cleaners. The removal of supporting functions from flashlight function models, therefore, does not improve the similarity results between the “shopvac” and flashlights, and an additional level of abstraction is required. The pruning rules provide this third level, resulting in a low degree of similarity between the flashlights and the “shopvac”, as shown in Figure 6-10.

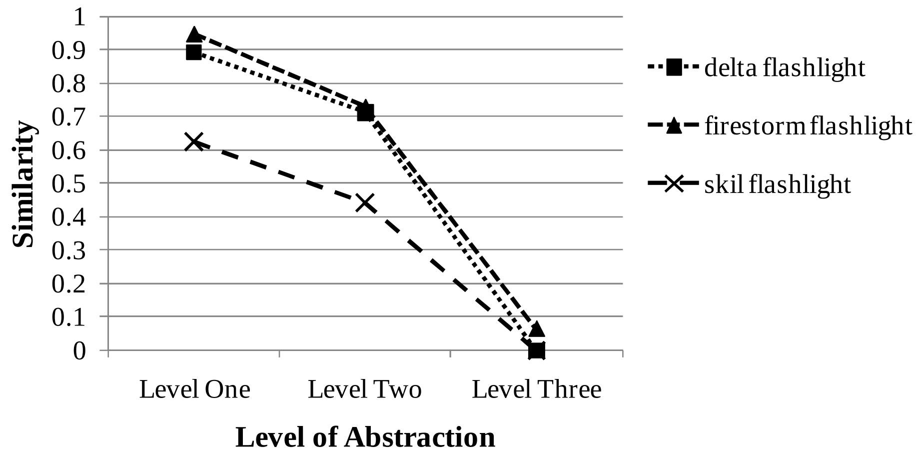


Figure 6-10: Similarity Between Shopvac and Flashlights at Three Levels of Abstraction

The similarity between the "shopvac" and all other vacuum cleaners at Level One is relatively high (see Figure 6-11), and similarity decreases at Level Two. At Level Three, the similarity increases between the "shopvac" and two of the vacuum cleaners, indicating that the pruning rules are improving the results of the similarity metric. Although the similarity of the remaining vacuum cleaners decreases, they do not decrease as much as the flashlights, so the overall results are improved.

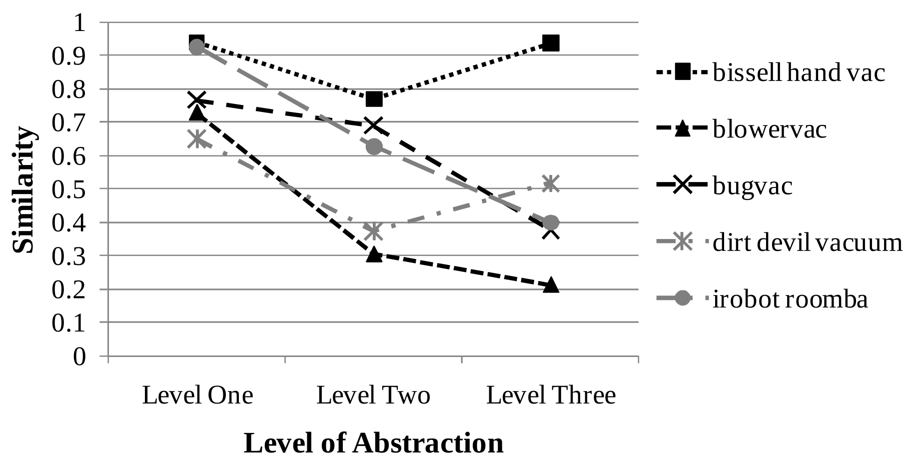


Figure 6-11: Similarity Between Shopvac and Vacuum Cleaners at Three Levels of Abstraction

The average similarity between the "shopvac" and coffee makers, flashlights, and vacuum cleaners is shown in Figure 6-12 and Table 6-6. At Level One, the vacuum cleaners, on average, are least similar to the "shopvac," and coffee makers are most similar. At Level Two, the coffee makers are least similar, but the "shopvac" is still more similar to flashlights than other vacuum cleaners. Only at Level Three is the "shopvac" most similar to vacuum cleaners.

Table 6-6: Average Similarity Between Shopvac and Three Artifact Types at Three Levels of Abstraction

	Representation Level		
Artifact	One	Two	Three
Coffee Makers	0.87	0.11	0.11
Flashlights	0.82	0.63	0.02
Vacuum Cleaners	0.80	0.55	0.49

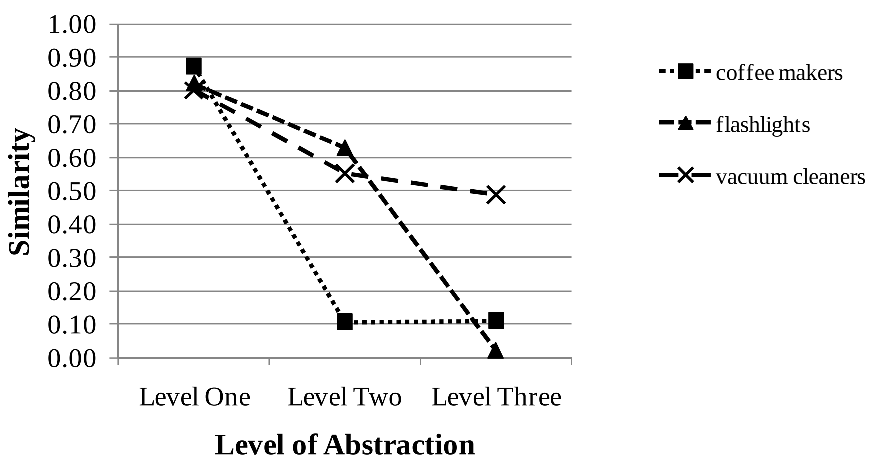


Figure 6-12: Average Similarity Between Shopvac and Three Artifact Types at Three Levels of Abstraction

6.3.2.3 Similarity Precision

The qualitative observations made in Section 6.3.2.1 are further investigated through a quantitative analysis of the precision and accuracy of clustering. The precision is measured by computing the average likeness (see Section 6.2.2) of ten runs to each other at each level of abstraction. The likeness metric, developed by Thebeau [85], quantitatively determines the likeness between each run and all other runs. Each run is given a score between 0 and 1, representing how similar that run is compared to all other runs of the algorithm. The average of these scores is used to compare the consistency, or precision, of the clusters at each level of abstraction.

The results of the likeness calculations for the ten runs at each level of abstraction are shown in Table 6-7. A two-sample t-test is used to compare the means. The hypotheses and resulting t- and p-values are shown in Table 6-8. The pruning rules significantly increase the consistency of the clustering results compared to Level One ($p < 0.0001$), Level Two ($p = 0.074$), and random function removal ($p = 0.0018$). The data also show that clusters computed at abstraction Level Two are more consistent than those computed at Level One ($p < 0.0001$).

Abstraction Level Three is significantly more precise than Levels One and Two, so similarity and clustering at Level Three is the most useful. At Levels One and Two, the higher degree of similarity of the models causes the clusters to be less consistent, resulting in extra noise in the algorithm's output. At Level Three, there is less noise, so there will be fewer artifacts clustered with an artifact of interest, reducing the amount of work required by the designer after the clustering results are obtained.

Table 6-7: Average Likeness Of Each Run to all Other Runs

	Representation Level			
	One	Two	Three	Random
Mean	0.603	0.718	0.759	0.657
Variance	0.0016	0.0011	0.0061	0.0032
Observations	10	10	10	10

Table 6-8: Hypothesis Tests for Clustering Precision

Alternative Hypothesis	Test Statistic, t	p-value
Level Three Precision > Level One Precision	5.63	1.2E-05
Level Three Precision > Level Two Precision	1.51	0.074
Level Three Precision > Random Precision	3.33	0.0018
Level Two Precision > Level One Precision	6.99	7.9E-07

6.3.2.4 Similarity Accuracy

The accuracy of clustering is determined by computing the likeness of an ideal run to the ten runs at each level of abstraction. The ideal run consists of the following four clusters: (1) all vacuum cleaners and the hair dryer, (2) all coffee makers and the iced tea maker, (3) all flashlights and the camera, and (4) the computer mouse. The likeness of this ideal run to all other runs is shown in Table 6-9. A t-test is used to compare the means at each level of abstraction. The hypotheses and resulting t- and p-values are shown in Table 6-10. The data show that the accuracy of the clusters identified by the pruning rules is significantly better than the accuracy of clusters at abstraction Level One ($p < 0.0001$), Level Two ($p = 0.0002$), and the random function removal ($p < 0.0001$). The data do not show that the Level Two accuracy is better than Level One ($p = 0.298$).

Functional analogies for conceptual design of adaptive design problems should be focused on the high-level function of an artifact rather than the means or assembly relationships within the artifacts. Level Three has been shown to focus on these high-level functions, since the ideal clusters were defined in this manner. Level Three, therefore, should be used when making functional comparisons across artifacts and drawing high-level analogies between them.

Table 6-9: Average Likeness of Each Run to Ideal Run

	Representation Level			
	One	Two	Three	Random
Mean	0.609	0.628	0.753	0.566
Variance	0.00333	0.00310	0.00579	0.00278
Observations	10	10	10	10

Table 6-10: Hypothesis Tests for Clustering Accuracy

Alternative Hypothesis	Test Statistic, t	p-value
Level Three Accuracy > Level One Accuracy	4.79	3.97E-06
Level Three Accuracy > Level Two Accuracy	3.64	2.43E-04
Level Three Accuracy > Random Accuracy	6.39	5.47E-09
Level Two Accuracy > Level One Accuracy	0.53	0.298

6.4 Outcomes and Discussion

Two abstraction levels of function models are obtained from existing research, and pruning rules are used to provide a more abstract artifact model for use in conceptual design. The proposed pruning rules are tested using a functional similarity metric to understand their usefulness in conceptual design for design-by-analogy methods. Functional similarity is computed using a metric developed by McAdams [40] and colleagues, and the resulting DSM clustered using the algorithm developed by Thebeau [85]. The similarity of 128 electromechanical artifacts has been evaluated at the following three levels of abstraction:

Level One – Including Supporting Functions

Level Two – Excluding Supporting Functions

Level Three – Pruning Rules Applied.

Similarity computed at Level One compares artifacts based on its function as well as its assembly. Since there are many instances of supporting functions in the models, they have a significant influence on similarity. Therefore, at this level of abstraction, similarity is heavily based on the number of physical connections within an artifact. For this reason, the similarity between many artifacts is high, and the accuracy and precision of clusters at this level is low. Similarity at Level Two reduces the emphasis on component relationships because supporting functions are excluded. Only higher-level functions are used in the models, improving the precision of the results. The accuracy of the results, however, is not significantly better than at Level One (see Table 6-8, Row 4). At Level Three, the application of pruning rules further increases the level of abstraction by removing functions that contain a high level of detail about the artifact. The Level Three comparison reduces the similarity among many artifacts, and only a few artifacts have at a high degree of similarity. This causes an increase in both the accuracy and precision of similarity calculations compared to Levels One and Two (see Table 6-8, Rows 1-2 and Table 6-10, Rows 1-2). These results show that the pruning rules effectively remove decomposed functionality from a model, resulting in a high-level model that is useful for design-by-analogy in the conceptual design phase.

Abstraction Levels One and Two presented in this paper are supported by the design repository containing the function models used in this research. However, Level

Three has been proposed through the pruning of these models following a specific set of rules. The pruning rules used to achieve abstraction Level Three have been shown to significantly improve the accuracy and precision of similarity. Further, it has been shown that this improvement is caused by pruning, not by chance, by showing that pruning is significantly better in terms of accuracy and precision than randomly removing functions from the models. Therefore, the pruning rules have been validated as a means for abstracting a function model when comparing the similarity of consumer electromechanical artifacts. However, the rules have been validated as a complete set, so the effects and validity of each rule individually is not yet known.

Many design researchers suggest the use of function models for understanding existing artifacts through reverse engineering as well as artifact development during conceptual design. However, the amount of detail known about an existing artifact is much greater than that of a new artifact, so the function models of each will be created at different levels of abstraction. If a designer uses a function-based similarity metric to identify artifacts that are similar to a concept being developed, then similarity should be computed at the conceptual level, not a reverse-engineered level. Therefore, the pruning rules proposed in this research should be used to convert reverse-engineered (Level One) models to conceptual (Level Three) models before using a similarity metric in conceptual design. Using the pruned models, the similarity metric will more accurately and consistently identify existing artifacts that can be used as a seed for design-by-analogy.

CHAPTER 7: IDEATION USER STUDY

An initial user study was designed, executed, and analyzed in close collaboration with Ramachandran [87, 88], and a complete description of the initial user study is included in [87]. The outcomes of this initial study have been used to significantly extend the study in the following ways: revise the statistical model, verify statistical assumptions, identify appropriate participants, introduce a new treatment group, introduce a new baseline group, introduce new evaluation metrics, and perform the study with forty-three additional participants. The discussion and outcomes of the initial study and these extensions are new contributions to the initial research and are presented in Sections 7.2.2, 7.2.4, 7.2.5, 7.3, and 7.4.

7.1 Motivation

Recent function modeling research has extended the transformative view of function to include interactions with users, other artifacts, and the environment [89] (see Section 2.1.3). The appropriateness of these extensions for use by humans in conceptual design has not been studied. Rather, these extensions have been studied within the context of computational tools. The usefulness of these extensions within conceptual design, specifically ideation, is the focus of this section.

To understand the usefulness of functional representations for ideation, a user experiment is conducted in which designers are provided different representations of an artifact for a new design problem, a consumer burrito-folding machine. The burrito-folding artifact was selected because participants in the study are familiar with both household appliances and burritos, the artifact requires both mechanical functionality and human interactions, participants can generate ideas for the artifact in a small amount of time, and the artifact has been used in previous research [76, 77]. Four metrics for evaluating sketches commonly used in literature are quality, quantity, novelty, and variety [79]. As mentioned in Section 2.4, the focus of ideation in this research is a convergent rather than divergent process. Since the desired outcome of the ideation process in this research is a high-quality design, novelty or variety of concepts is not studied. These metrics could be studied in the future without affecting the results and conclusions based on quality and quantity.

An overview of the initial experiment procedure is shown in Figure 7-1. Participants were provided with a problem statement, requirements, and a seed model. The participants were then asked to draw from their past experiences to generate concepts that satisfy the problem. The outcome, sketches, were then evaluated using quality and quantity metrics. In the initial study, one group of participants received a function model to aid in concept generation while the other group received an interaction model [89].

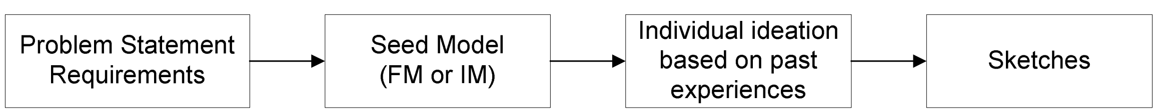


Figure 7-1: Overview of Initial Experiment Procedure

In an extended study, the same design problem and requirement were given to participants, who then received a function model (FM), interaction model (IM), pruned model (PM), or no model (NM). An overview of the extended study is shown in Figure

7-2. The details of the extended experiment and motivation for the additional treatment groups are discussed in Section 7.3.

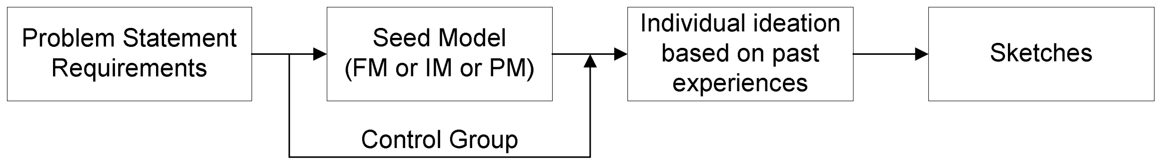


Figure 7-2: Overview of Extended Experiment Procedure

In both studies, the focus is on understanding the effects of using functional representations as a seed for convergent thinking, and participants were instructed to draw from their past experiences to solve the design problem. The study was performed in a setting that was not intended to stimulate ideas, participants were not allowed to work together, and participants were allowed to use both textual and graphical representations to describe their concepts. Thus, while participants were not forced to use certain ideation techniques, they were limited in the techniques that they could use based on the experiment design and setting. The particular ideation techniques used by participants was not evaluated; only the design outcome is assessed in these studies.

To understand if designers are using the models provided, fifteen elements modeled in each representation are analyzed to determine if the designer addresses each element in his or her sketch (referred to as “sketch conformance”). This information includes: seven functions, four user actions, and four artifact-user interactions. The general statistical hypotheses tested are:

Null Hypothesis: The average sketch conformance by participants using each type of representation type is equal.

Research Hypothesis: The average sketch conformance by participants using each type of representation is not equal.

The function, activity, and interaction information in the model is categorically tested to understand whether or not the participants considered the specific information when creating the sketches:

Function Sub-hypothesis: The functional conformance of sketches generated by participants using each type of representation is not equal.

Activity Sub-hypothesis: The activity conformance of sketches generated by participants using each type of representation is not equal.

Artifact-User Interaction Sub-hypothesis: The interaction conformance of sketches generated by participants using each type of representation is not equal.

To understand the effect of the representations on the concepts generated, sketches are evaluated to determine how well the concept addresses the design problem (referred to as "sketch quality"). The quality of a sketch is based on the level of satisfaction of each of nine requirements provided to the participants in the problem statement. The statistical hypotheses to be tested for quality are:

Null Hypothesis: The average quality of sketches generated by participants using each representation is equal.

Research Hypothesis: The average quality of sketches generated by participants using each representation is not equal.

The requirements are also categorized as functional, human activity, or performance requirements. The functional and human activity requirements are compared to understand if the representations affect a subset of the requirements.

To further assess the creativity of the designers using each representation, the quantity of sketches is also measured and compared. The statistical research hypotheses are:

Null Hypothesis: The average number of sketches produced by participants using each representation is equal.

Research Hypothesis: The average number of sketches produced by participants using each representation is not equal.

7.2 Initial Study

The goal of this study is to understand the effects of functional representations on concept generation. Close conformance with a model is desired because it demonstrates that the model is well understood by the designer and it useful to the designer for an adaptive design problem. Designers may deviate from the model if they feel that they have a better idea than that shown in the model. However, the ideas in each representation—FM or IM—were held as closely to each other as possible, so it can be assumed that the variation in conformance due to the designer intentionally ignoring the model is equal for both groups. Therefore, the sketch conformance to the model provides insight into whether or not the designers use the model. The focus of the conformance analysis is on whether or not the designer considered the particular function, activity, or interaction, rather than how well each is satisfied. Sketch quality is also measured to understand the effect of the models on concept quality and ensure that the quality of the ideas is not negatively affected by using a model.

7.2.1 Experiment Design

The user experiment conducted in this research is a single-factor, completely randomized design. The factor, the representation given to the participant, has two levels: function model or interaction model. Forty students—both undergraduate and graduate—participated in the study at Clemson University during the Fall 2010 semester. Participants were assigned to treatment groups in either an alternating or random manner (depending on other conditions of the experiment) to prevent experimental bias. Participants were first trained in the representation before being given the design problem. After training, the participants were given a problem statement, requirements, and the appropriate model for the new design problem, a consumer burrito folder. The participants were then allowed to draw multiple sketches for 30 minutes. The sketches are analyzed to determine how the participants used the model through the conformance metrics discussed in Section 7.2.2. An in-depth discussion of this experiment design and procedure is included in [87, 88], where the quality of the sketches is measured for this experiment. In this research, the sketch conformance metric is developed and measured to understand how the models influence the designers' sketches.

7.2.2 Conformance Scale Development

The function model and interaction model given to each participant in shown in Figure 7-3 and Figure 7-4. The information contained in each of the two models is approximately equal [87], but the information is modeled differently. In the function model, functions and activities are modeled in the same manner, and the designer must infer which functions that the user or the artifact accomplishes. In the interaction model, three of the functions are explicitly shown to be performed by the user and are included in the user boundary in the upper portion of the model.

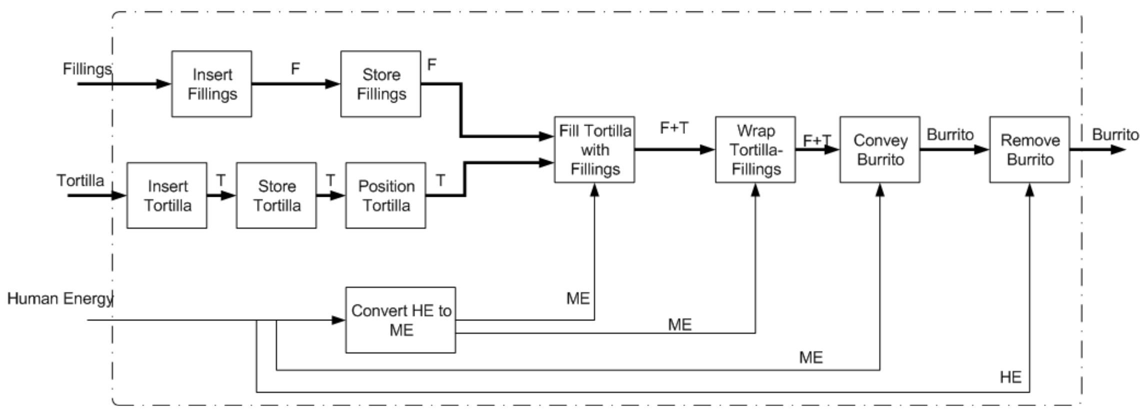


Figure 7-3: Burrito Folder Function Model

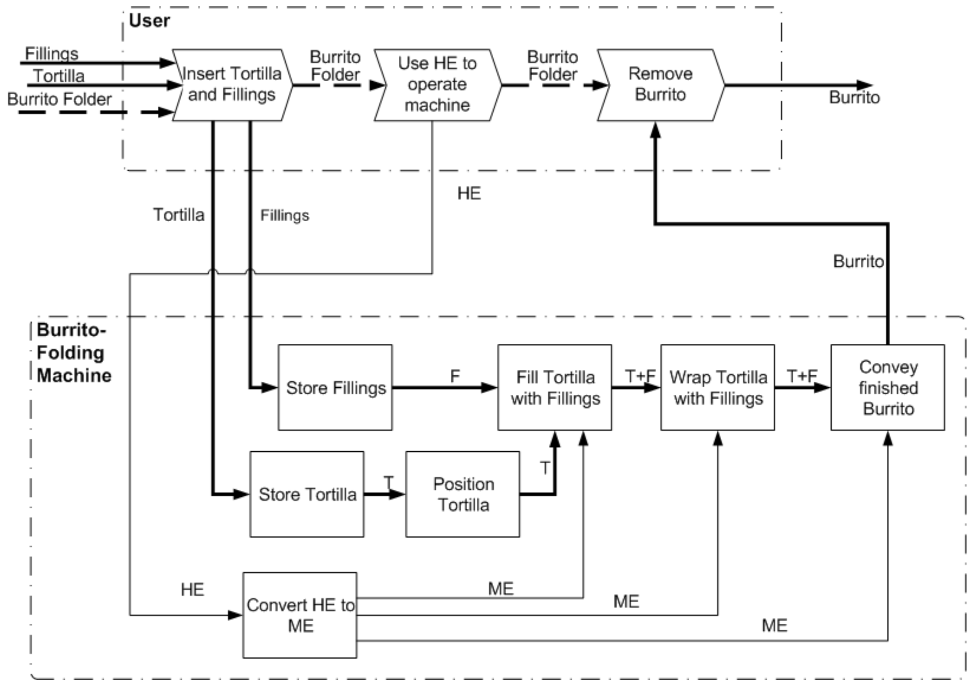


Figure 7-4: Burrito Folder Interaction Model

The two models used in this experiment each contain 15 model elements, categorized as function, activities, and interactions:

Functions:

F1: The artifact stores tortillas

F2: The artifact stores filling

F3: The artifact moves the tortilla into position

F4: The artifact fills the tortilla with fillings

F5: The artifact wraps the tortilla

F6: The artifact conveys the burrito

F7: The artifact converts human energy input into mechanical energy

User Actions:

A1: The user inserts tortillas into the artifact

A2: The user inserts fillings into the artifact

A3: The user operates the artifact

A4: The user removes the burrito

Artifact-User Interactions

I1: The artifact allows the tortilla to enter

I2: The artifact allows the fillings to enter

I3: The artifact allows the human energy to enter

I4: The artifact allows the user to remove a burrito

The interaction model clearly shows who or what is performing the actions in the model, while the function model does not. The goal of the conformance metric is to determine if the designer follows the ideas in the model or deviates from these ideas. For example, the models specify that human energy is the only input to the system. If a designer uses only human energy to accomplish the functions, then the sketch conforms to the model. If, on the other hand, the concept contains an input of electricity, then the sketch does not conform to the model. The intent of the information in the function and interaction models is described in Table 7-1.

Table 7-1: Intent of Information Contained in the Burrito Folder Models

Interaction Model	Function Model	Intent of Information
Store Fillings	Store Fillings	The artifact is able to store the fillings, either individually or together, for some period of time.
Store Tortilla	Store Tortilla	The artifact is able to store multiple tortillas for some period of time.
Position Tortilla	Position Tortilla	The artifact moves a tortilla from the storage location to the location in where it is filled.
Fill Tortilla with Fillings	Fill Tortilla with Fillings	The artifact adds fillings to the open tortilla.
Wrap Tortilla with Fillings	Wrap Tortilla-Fillings	The artifact wraps the tortilla around the fillings.
Convey finished Burrito	Convey Burrito	The artifact moves the folded burrito away from the folding location.
Remove Burrito	Remove Burrito	The user removes the buritto from the artifact.
Insert Tortilla and Fillings	Insert Fillings	The user places fillings in the artifact.
	Insert Tortilla	The user places tortillas in the artifact.
Use HE to operate machine	Convert HE to ME	The user provides energy to the artifact.
Convert HE to ME		The artifact uses human input to perform an action.
Fillings →	Fillings →	Fillings are passed from the user to the artifact.
Tortilla →	Tortilla →	Tortillas are passed from the user to the artifact.
Human Energy →	Human Energy →	Human energy is passed from the user to the artifact.
Burrito →	Burrito →	Burritos are passed from the artifact to the user.

7.2.2.1 Function Conformance Scale

Each of the seven blocks in the burrito folder IM are considered artifact functions, and the sketches are analyzed to determine if the concept addresses each function. A quantitative conformance scale was developed based on the intent of the information in the models. First, a three-category rating scale was developed for the seven artifact functions. The following general scale was used:

Good (1): The function is clearly incorporated in the concept.

Neutral (0): The function is implicitly incorporated in the concept or the function is plausible but not explicitly shown.

Poor (-1): There is a complete absence of the function or there is another function that contradicts the particular function.

Two sketches were fully analyzed and discussed using this scale, and these examples were used to train sketch raters. A random sample of ten sketches was selected from all sketches generated in the study, and the ten sketches were independently rated by two raters for each of the seven functions. The interrater agreement (IRA) of this scale was determined using Cohen's Kappa [90] (see Equation 1), with substantial (0.61 - 0.80) to almost perfect (0.81 - 1.00) agreement desired [91].

$$\kappa = \frac{p_o - p_c}{1 - p_c} \tag{1}$$

where p_o is the proportion of ratings in which the two raters agree, and

p_c is the proportion of ratings in which the two raters are expected to agree by chance.

The actual agreement, chance agreement, and Kappa values for each function (F1-F7) in this first iteration are shown in Table 7-2. As shown in the table, only two functions had substantial or perfect agreement, so the scales should be refined.

Table 7-2: Interrater Agreement for First Iteration of Function Conformance Scale

	F1	F2	F3	F4	F5	F6	F7
Actual Agreement	0.30	0.90	0.70	0.70	0.80	0.60	0.60
Cohen's Chance Agreement	0.29	0.46	0.40	0.40	0.44	0.44	0.30
Cohen's Kappa	0.01	0.81	0.50	0.50	0.64	0.29	0.43

Based on the results of the initial scale, the raters discussed the differences in individual sketch ratings and the scale was refined. A reference sheet with examples of good and bad concepts for each function was developed to assist the raters. The raters individually rated ten additional randomly-selected sketches, and the IRA is shown in Table 7-3.

Table 7-3: Interrater Agreement for Second Iteration of Function Conformance Scale

	F1	F2	F3	F4	F5	F6	F7
Actual Agreement	0.90	1.00	0.60	1.00	0.80	0.70	0.60
Cohen's Chance Agreement	0.34	0.68	0.41	0.68	0.44	0.38	0.38
Cohen's Kappa	0.85	1.00	0.32	1.00	0.64	0.52	0.35

The IRA for many functions improved due to the discussion of differences, clarification of the scale, and the development of the reference sheet. Through discussion of differences in the second iteration, it was determined that the neutral rating (0) in the three-category scale was highly inconsistent. Most of the differences in ratings included a neutral rating by one of the raters. Therefore, in the third iteration of the conformance scale, a binary scale was used. The final functional conformance scale used for this research is:

Good (1): The function is clearly incorporated in the concept.

Poor (0): The function is implicitly incorporated in the concept, the function not explicitly shown, there is a complete absence of the function, or there is another function that contradicts the particular function.

Using the above scale and a revised reference sheet with examples, the two raters achieved substantial agreement on six out of seven function conformance metrics. The seventh metric (F7) had 80% actual agreement, but due to the high chance agreement, IRA is lower than desired. The chance agreement is based on the actual values chosen by the two raters for the ten concepts. Since the ten concepts chosen for this iteration have many poor conformance values (7 of 10), the chance agreement is higher, reducing the IRA. Since the actual agreement of this metric is high, it is consistent with the actual agreement for other metrics, and the IRA still lies in a "moderate" agreement range (0.41-0.60) [91], the value is acceptable and the scale development is complete.

Table 7-4: Interrater Agreement for Third Iteration of Function Conformance Scale

	F1	F2	F3	F4	F5	F6	F7
Actual Agreement	1.00	0.90	0.90	1.00	0.90	0.80	0.80
Cohen's Chance Agreement	0.68	0.74	0.50	0.68	0.50	0.50	0.58
Cohen's Kappa	1.00	0.62	0.80	1.00	0.80	0.60	0.52

It is important to note that added functionality or activities have not been included in this analysis since participants were not instructed to operate under a closed world assumption. For example, if a designer included a heating element to warm the tortillas, the sketch has not been penalized for deviating from the model, which does not include heat flows. However if a designer requires other sources of energy, then the sketch does not conform to activity A3, "The user operates the artifact."

Two examples of functional conformance ratings are discussed to demonstrate the final iteration of the functional conformance scale (see Figure 7-5 and Figure 7-6), and the reference sheet used by the raters is shown in Table 7-5.

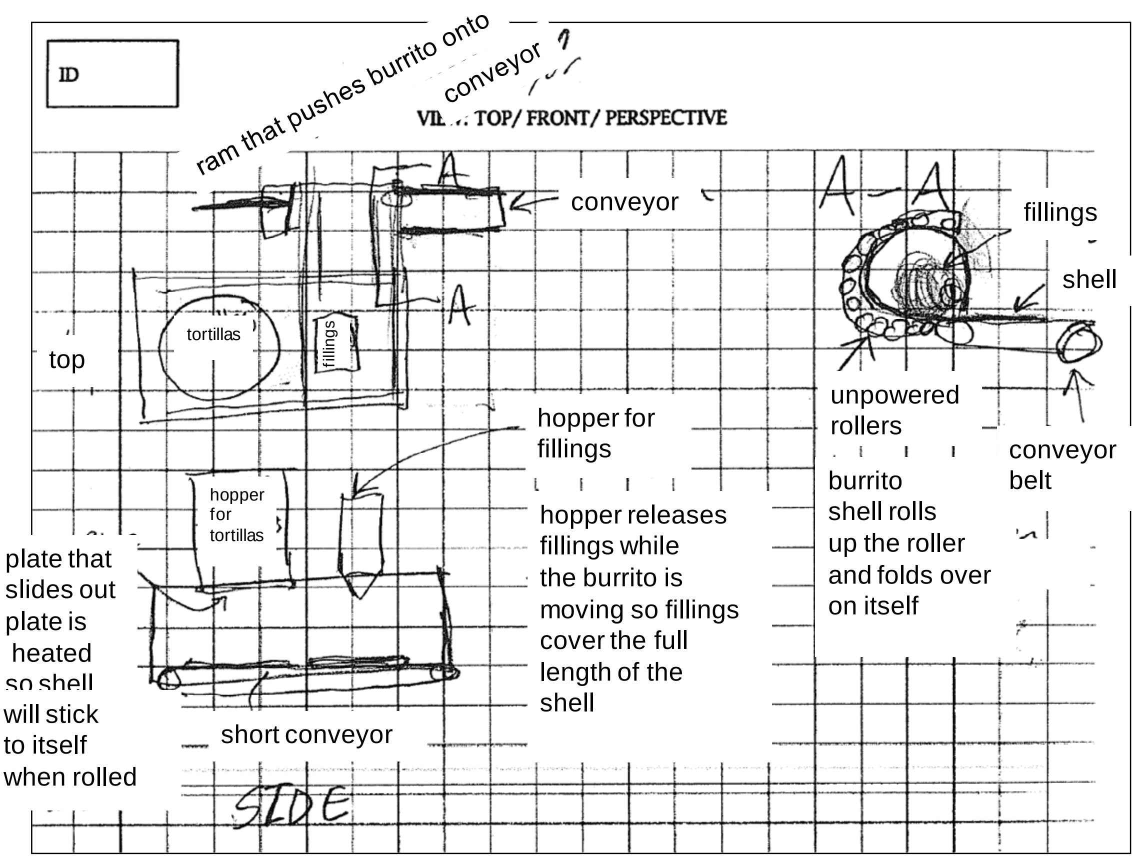


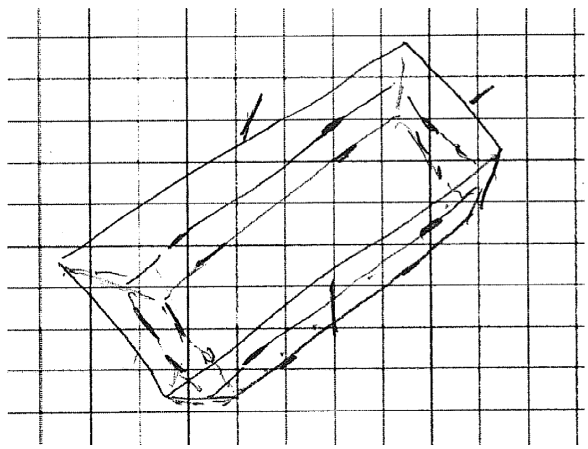
Figure 7-5: Burrito Folder Sketch – Example 1 (text modified to improve readability)

The analysis of each functional conformance metric for the first sample sketch is discussed below (see Figure 7-5):

- **F1: The artifact stores tortillas.** The sketch shows a container for storing tortillas, so the designer considered and explicitly addressed this function. It is given a rating of 1.
- **F2: The artifact stores filling.** The sketch shows a container for storing fillings, so the designer considered and explicitly addressed this function. It is given a rating of 1.
- **F3: The artifact moves the tortilla into position.** The sketch shows a conveyer belt for moving tortillas into position, so the designer considered and explicitly addressed this function. The sketch is given a rating of 1.
- **F4: The artifact fills the tortilla with fillings.** The storage container in the sketch includes a spout showing that the tortilla will be filled by the artifact, so the sketch is given a rating of 1.
- **F5: The artifact wraps the tortilla.** Section A-A in the sketch shows the wrapping functionality of the burrito folder. The quality of the folding process is not evaluated. As long as some form of folding is explicitly shown, the concept is given a rating of 1.
- **F6: The artifact conveys the burrito.** The sketch includes a conveyor system that will move the burrito after being wrapped, so the sketch is given a rating of 1.

F7: The artifact converts human energy input into mechanical energy. The sketch does not show a mechanical user input, such as a crank, so the sketch is given a rating of 0.

The analysis of each functional conformance metric for the second sample sketch is discussed below (see Figure 7-6):



Explanation:

Hinged chamber sides moved by mechanical levers

Figure 7-6: Burrito Folder Sketch – Example 2 (text modified to improve readability)

F1: The artifact stores tortillas. There is no mention of tortilla storage in the sketch. It appears that the tortillas will be folded in the same location as they are placed in the artifact, so the designer likely did not consider the storage function. The sketch is given a rating of 0.

F2: The artifact stores filling. There is no mention of filling storage in the sketch. It is likely that the fillings must be added directly by the user, so the sketch is given a rating of 0.

F3: The artifact moves the tortilla into position. The is no mention of movement of the tortilla in the sketch, and the tortilla is likely folded in the same location where it is placed in the artifact. The sketch is given a rating of 0.

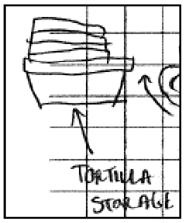
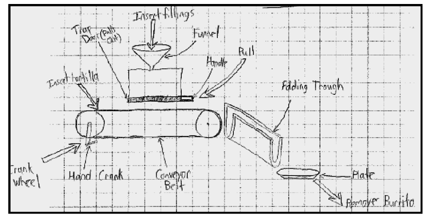
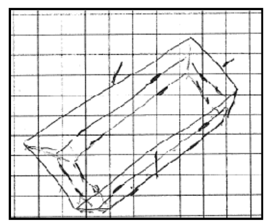
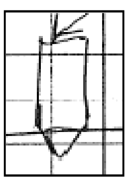
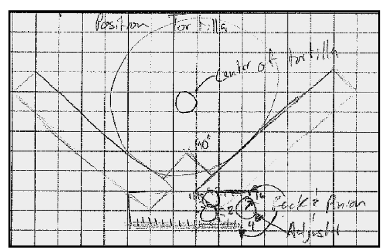
F4: The artifact fills the tortilla with fillings. There is no mention of how fillings are added to the tortilla. It is likely that the user must add them directly because there is no container of fillings incorporated in the sketch. The sketch is given a score of 0.

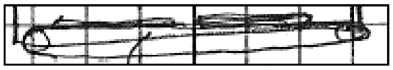
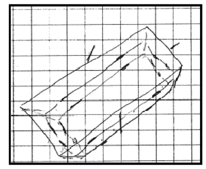
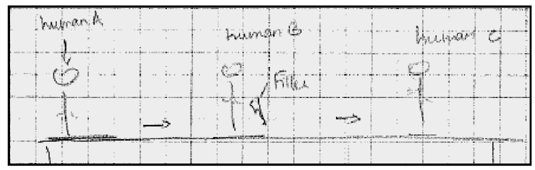
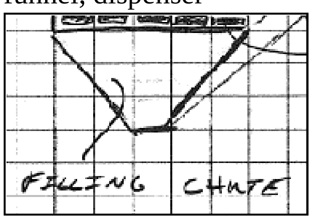
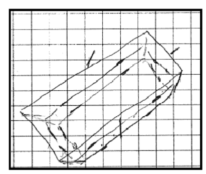
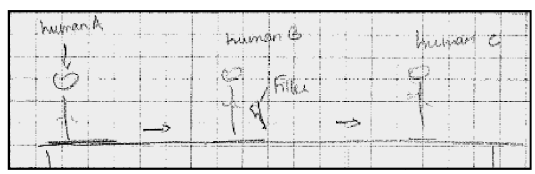
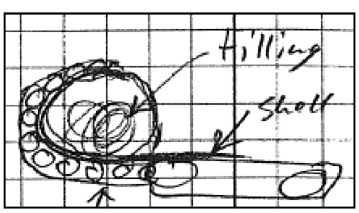
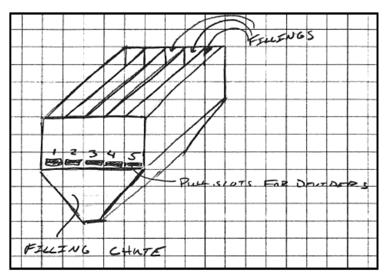
F5: The artifact wraps the tortilla. The description of the artifact states that the sides of the artifact are hinged and have mechanical levers. These features demonstrate that the designer considered how the artifact can wrap a tortilla, so the sketch is given a score of 1.

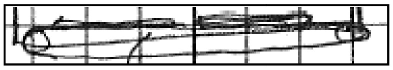
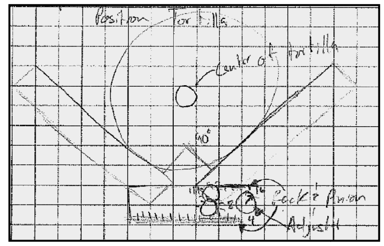
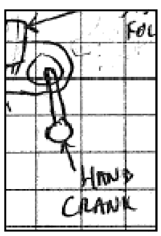
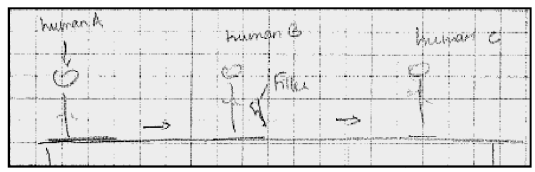
F6: The artifact conveys the burrito. The concept does not move the burrito after being folded, and it is likely that the user must remove it manually. The sketch is given a rating of 0.

F7: The artifact converts human energy input into mechanical energy. The sketch does not show a mechanical user input, such as a crank. It is possible that the user manipulates the mechanical arms, but it is not explicitly stated, so the sketch is given a rating of 0.

Table 7-5: Function Conformance Reference Sheet

	Good (1)	Poor (0)
F1	stack, hopper, bin, ability to feed a stack of tortillas into the artifact 	multiple tortillas in artifact before being filled; may require inserting one at a time  tortilla is filled in the exact location where it is placed in the artifact; only one tortilla is operated on at a time; tortillas are not incorporated in the concept 
F2	hopper, bin, etc. incorporated into the artifact 	fillings must be added directly by the user 

	Good (1)	Poor (0)
F3	conveyer, four-bar mechanism; any movement of the tortilla by the artifact 	the tortilla is filled in the location where it is placed in the artifact; a human moves the tortilla  
F4	spout, funnel, dispenser 	filling is absent; a human fills the tortilla  
F5	folding arms, roller, cup, hinges 	wrapping is absent; a human wraps the burrito 

	Good (1)	Poor (0)
F6	conveyer, four-bar mechanism; any movement of the folded burrito (after being filled) by the artifact 	burrito is removed by the user from the location where it is folded; no burrito exists in the concept 
F7	a human moves a component of the artifact that directly influences the tortilla, fillings, or burrito; crank, arm, handle; at least one conversion of HE to ME (NOT hand-folding the burrito; NOT an electric switch) 	human(s) perform all operations on the tortilla/burrito, rather than on the artifact; a human does not move any system components 

7.2.2.2 Activity Conformance Scale

Each of the three activity blocks in the interaction model (see Figure 7-4) represents actions performed by the user. The activity *insert tortilla and fillings* exists as a single block in the activity model due to the limitations of the model in capturing independent, parallel activities. This activity was separated into the two distinct activities of *insert tortilla* and *insert fillings* for this analysis because the activities can be performed independent of each other.

The activity conformance scale was developed based on the final binary function conformance scale. Based on the lessons learned from the functional conformance scale development, a strict activity conformance scale was used. In the activity conformance scale, if the sketch did not explicitly state that a user performs an activity, it is assumed that the user does not perform that activity. Two examples and a reference sheet were developed to train the raters in the activity conformance scales before two raters individually rated ten randomly selected concepts. The first iteration of the activity conformance scale yielded perfect or substantial levels of agreement, so no further iterations were necessary. The results of this iteration are shown in Table 7-6. Two example sketch ratings are discussed in detail below and the reference sheet is included in Table 7-7.

Table 7-6: Interrater Agreement for Activity Conformance Scale

	A1	A2	A3	A4
Actual Agreement	0.90	1.00	0.90	1.00
Cohen's Chance Agreement	0.50	0.52	0.54	0.52
Cohen's Kappa	0.80	1.00	0.78	1.00

The analysis of each activity conformance metric for the first sample sketch is discussed below (see Figure 7-5):

A1: The user inserts tortillas into the artifact. The tortilla starting location is a hopper. There is no mention of a user placing the tortillas in this hopper, so it is given a score of 0.

A2: The user inserts fillings into the artifact. The filling starting location is a hopper. There is no mention of a user placing the fillings in this hopper, so it is given a score of 0.

A3: The user operates the artifact. The sketch does not mention the use of human power to drive part of the artifact, so it is given a rating of 0.

A4: The user removes the burrito. The sketch indicates that the tortillas are rolled and placed on a conveyor. There is no mention of a user removing the folded burrito, so the sketch is given a rating of 0.

The analysis of each activity conformance metric for the second sample sketch is discussed below (see Figure 7-6):

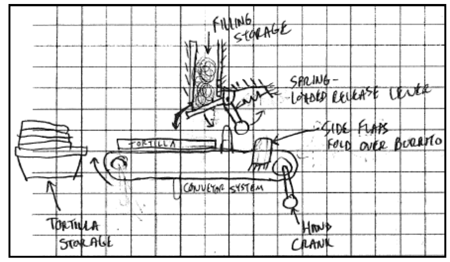
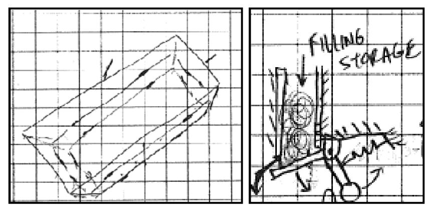
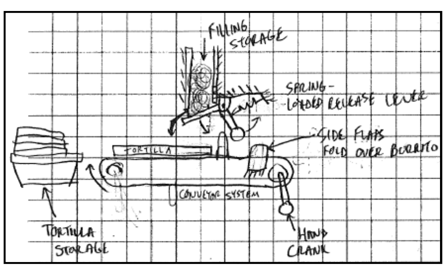
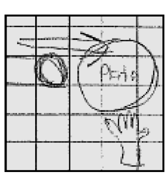
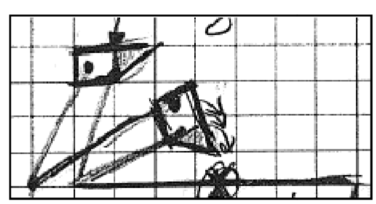
A1: The user inserts tortillas into the artifact. The user is not mentioned in the sketch, so it is given a score of 0.

A2: The user inserts fillings into the artifact. The user is not mentioned in the sketch, so it is given a score of 1.

A3: The user operates the artifact. The sketch does not mention the use of human power to drive part of the artifact, so it is given a rating of 0.

A4: The user removes the burrito. The sketch does not mention a human removing the burrito, so the sketch is given a rating of 0.

Table 7-7: Activity Conformance Reference Sheet

	Good (1)	Poor (0)
A1	The user action is shown with a hand, etc. or described with text. The tortilla & ingredients are placed by hand	User placement of tortillas is not explicitly stated. Tortillas are not present (or mentioned) in the sketch. 
A2	The user action is shown with a hand, etc. or described with text (the fillings may be added directly to the tortilla). Usr selects toppings and insrts them into Filling "bucket"	User placement of fillings is not explicitly stated. Fillings are not present (or mentioned) in the sketch. 
A3	Human energy is explicitly stated as a source of power and it is the only source of power. 	The source(s) of power are unclear. Other energy is (or may be) used to power the artifact. e.g., electricity, motor **Explanation:** With the push of a button, the conveyor belt brings
A4	The user action is shown with a hand, etc. or described with text. the operator can safely remove the completed burrito. 	The removal of burritos from the artifact is not shown (or mentioned). Folded burritos are not present (or mentioned) in the sketch. 

7.2.2.3 Interaction Conformance Scale

Each of the four flows that pass between the user and artifact boundaries in the IM are identified as interactions. Since interactions are flows between systems, rather than actions performed by a system, the interactions are addressed passively by the design. Interactions are closely related to their corresponding function and activity, but they must be analyzed independently of functions and activities. It is possible for an interaction to be addressed without its corresponding activity or function. For example, a sketch may address the interaction *the artifact allows the tortilla to enter* without addressing the function *store tortilla* or the activity *insert tortilla*. However, if the activity *insert tortilla* or the function *store tortilla* is addressed, then the interaction has been addressed. The four interactions in the models are: *the artifact allows the tortilla to enter*, *the artifact allows the fillings to enter*, *the artifact allows the human energy to enter*, and *the artifact allows the user to remove a burrito*.

The interaction conformance scale was developed in the same manner as the activity conformance scale. The same general binary scale was used, and a strict scale was developed to ensure a high interrater agreement. One example was developed describing the rating system and a reference sheet with examples of both good and poor ratings for each of the four interactions was used for training and rating. Ten randomly selected sketches were independently evaluated by two raters, and the IRA was computed for each of the four interactions. In the first iteration of the scale, substantial or perfect agreement was achieved, as shown in Table 7-8. The rating for an example sketch is discussed below and the reference sheet is provided in .

Table 7-8: Interrater Agreement for Interaction Conformance Scale

	I1	I2	I3	I4
Actual Agreement	0.90	0.90	0.80	1.00
Cohen's Chance Agreement	0.62	0.50	0.48	0.52
Cohen's Kappa	0.74	0.80	0.62	1.00

The analysis of each interaction conformance metric for the first sample sketch is discussed below (see Figure 7-5):

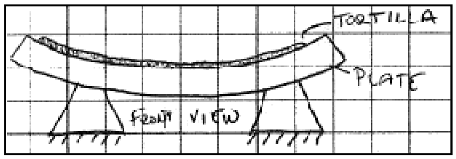
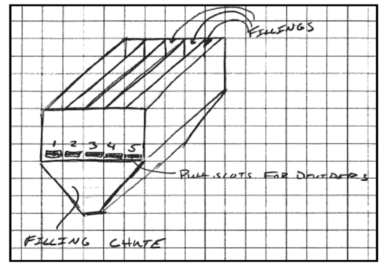
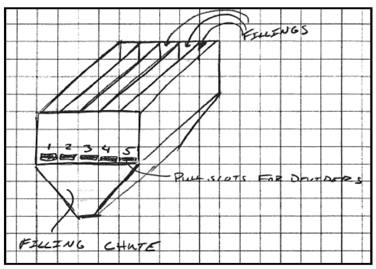
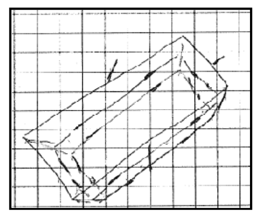
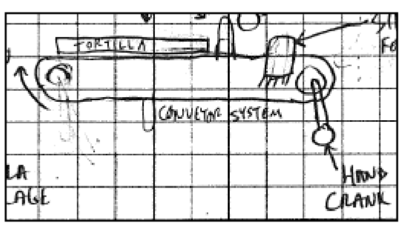
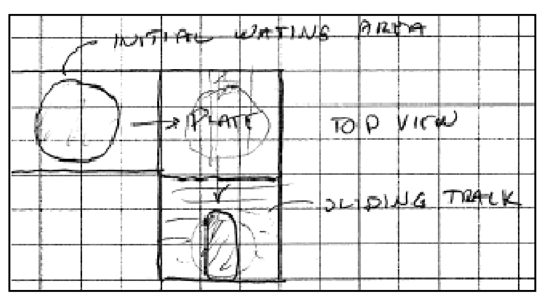
I1: The artifact allows the tortilla to enter. The artifact contacts a tortilla, so it is given a rating of 1.

I2: The artifact allows the fillings to enter. The artifact contacts fillings, so it is given a rating of 1.

I3: The artifact allows the human energy to enter. The sketch does not show if and how a user interacts with the artifact. There are no handles, cranks, etc., so the sketch is given a rating of 0.

I4: The artifact allows the user to remove a burrito. The sketch does not show how a user will remove the folded burritos, so the sketch is given a rating of 0.

Table 7-9: Interaction Conformance Reference Sheet

	1	0
I1	The artifact comes in contact with a tortilla.  	The artifact does not contact a tortilla. 
I2	The artifact comes in contact with fillings.  	The artifact does not contact fillings. The fillings interact only with the tortilla. 
I3	The user comes in contact with the artifact and provides energy to either move, fill, or wrap the tortilla. e.g., handle, crank, pull tab  	The user does not contact the artifact or the user does not provide energy that directly moves, fills, or wraps the tortilla.
I4	A user comes in contact with the burritos. the operator can safely remove the completed burrito.	The user does not contact the burritos. Burritos are not present in the sketch. 

7.2.3 Quality Scale Development

The quality of the ideas generated for this design problem was measured by rating the sketch on a three-level scale for each of the nine requirements provided with the design problem to participants. This scale was developed in close collaboration with Ramachandran and the complete details of the scale discussed in [87]. The same procedure used to achieve high levels of interrater agreement for the conformance scales (see Section 7.2.2) was used to achieve substantial agreement (0.61) using Cohen's Kappa value. Complete details of this scale development are discussed in [87], and the quality scale for each requirement is reproduced in Table 7-10 [87].

Table 7-10: Sketch Quality Scale [87]

Requirement	Low (1)	Medium (3)	High (9)
Position empty tortilla to store fillings	No storage area or conveyor mechanism available	Human has to manually place the tortilla form storage to filling	The tortilla is moved from its stack to the filling zone through any conveyor mechanism
Fill the tortilla after proper positioning	Filling device completely missing	A chamber is present but no other detail is given; Incomplete filling mechanism.	Hopper, funnel, box or any holding device with a provision to fill the empty tortilla.
Wrap burrito over the fillings	Wrapping mechanism is missing	2 sided folding	3 or more sided folding
Deliver completed burritos at rate of at least 4 burritos per minute	When the above three requirements also has low scores. The user does most of the activities.	A chain or gear drive mechanism is used to transfer burritos. The user has to do some actions like position, fill or wrap.	A belt, band or cable drive mechanism is followed. Completely automated.
Easy to use	More than five human activities	Four or five human activities. Either wrapping or inserting is automated.	Three human activities. Fully automated for both wrapping and inserting
The device must fit on a counter top	The size is too big and will not stable if mounted on a table	Either size or stability is not satisfied.	Length= Height and total size is less than 12 “. Satisfies both size and stability criteria
Easy to install	More than 3 independent parts to assemble for the first time.	Has 3 independent parts to assemble	The device looks complete or has two independent parts to assemble.
The device must be easy to clean after use	Disassembly is needed to clean the machine. The user transfers filling and burritos by hand, with more chances of spilling.	Rollers, chains and other surfaces which has crevices. Fill, wrap and delivering completed burritos zones are not continuous.	The device must offer no spillage when moving from one zone to another. After being filled, the transfer mechanism must be uninterrupted.
The device must be safe to use	All parts are completely exposed without a cover. Sharp edges or pinch points which might cause injury during the operation (motor/electrically driven).	Either one (or few) sharp extruding parts or pinch points are present. Hinge (hand driven pinch points). No serious injury will be caused even if some parts are exposed.	No sharp extruding or exposed pinch points

7.2.4 Preliminary Results

The study was conducted in two sessions, one with graduate student participants, and one with undergraduate student participants. The graduate session was conducted during an advanced design methods course at Clemson University. There were 14 participants divided into two groups. Since the participants were taking an advanced design methods course, all participants had already been taught function modeling methods. Therefore, all participants were presented a review of function models and a discussion of interaction models before participating in the experiment. The undergraduate session was conducted during an senior-level design course at Clemson University with 26 participants. Since the participants had not received formal training in function modeling, they were divided into two groups before the representation training began. Participants received training only on the appropriate representation: function or interaction.

7.2.4.1 Selection of Participant Scores

There were 40 participants in this study and a total of 106 sketches created by the participants. Each sketch was evaluated for quality and conformance as discussed in the previous sections. Since the participants were allowed to sketch as few or as many concepts as they desired, there were multiple sketches generated for most participants. The participants, however, were the experimental unit in the study, and the additional sketches can be used only to understand the variation within participants, not between treatment groups. Since the number of sketches generated by each participant varies and some participants produced only one sketch, it is difficult to determine the within-participant quality or conformance variation. Further, this variation is not of interest in this study, so each participant is given a single score based on all of the sketches he or she generated. Several approaches to determining a participant's score were investigated, and two final approaches are used to analyze the data. The approaches are discussed with respect to the conformance metrics, but the same approaches can be used for the quality metrics as well.

The first participant scoring approach considers the participant's best score for each of the fifteen conformance elements, taking into account whether a participant addressed the particular function, activity, or interaction in *any* of his or her sketches. For example, the results of a hypothetical participant's conformance ratings are shown in Table 7-11. The last row in the table shows the participant's score that would result from taking the maximum score for each element, F1-F7, A1-A4, and I1-I4.

Table 7-11: Participant Best Score by Individual Elements

Sketch	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
1	1	0	0	0	1	0	0	0	1	0	1	1	0	0	0
2	0	1	0	0	1	0	0	0	0	0	1	1	1	1	1
3	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1
Score	1	1	0	1	1	0	0	0	1	0	1	1	1	1	1

The second participant scoring approach considers the participants' best sketch within each category (function, activity, or interaction). This approach considers the functional conformance score for all sketches by a participant and uses the values from the sketch with the best functional conformance. The activity and interaction categories are considered independently. For example, if a participant produced sketches with the ratings shown in Table 7-12, the participant's functional score would be based on the third sketch, the participant's activity score would be based on the first sketch, and the participant's interaction score would be based on the second sketch. The participant's final score using this approach is shown in the final row of the table.

Table 7-12: Participant Best Score by Category

Sketch	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
1	1	0	0	0	1	0	0	0	1	0	1	1	0	0	0
2	0	1	0	0	1	0	0	0	0	0	1	1	1	1	1
3	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1
Score	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1

The third scoring approach considers the participants' best overall sketch independent of individual scores or categorical scores. The sum of all conformance values is used to determine the participant's best sketch, and the values from that sketch are used for the final score. For example, if a participant produced sketches with the ratings shown in Table 7-13, the second sketch would be used as the participant's score since it has an overall conformance score of 7, while the first and third sketches have overall scores of 5 and 6, respectively.

Table 7-13: Participant Best Score Overall

Sketch	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
1	1	0	0	0	1	0	0	0	1	0	1	1	0	0	0
2	0	1	0	0	1	0	0	0	0	0	1	1	1	1	1
3	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1
Score	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1

The final scoring approach is to use the participants' average sketch scores, considering the average level of conformance for all sketches. For example, if a participant produced sketches with the rating shown in Table 7-14, the averages for each column would be taken and used for the participant's score.

Table 7-14: Participant Average Score

Sketch	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
1	1	0	0	0	1	0	0	0	1	0	1	1	0	0	0
2	0	1	0	0	1	0	0	0	0	0	1	1	1	1	1
3	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1
Score	0.67	0.67	0	0.33	0.67	0	0	0	0.67	0	0.67	1	0.33	0.33	0.67

There are many other participant scoring approaches that can be used, each of which has advantages and disadvantages. The difficulty of using the best categorical or overall sketches is in the event of a tie. If the sketches both have the same sum but have achieve it through different conformance combinations, then determining which set of scores to use is difficult. For example, if a participant produced three sketches with the functional conformance scores shown in Table 7-15, there would be a tie between the first and third sketches, which conform to different functions in the model. This same problem arises with the best overall sketch scoring approach as well. To address this issue, when a categorical best is used, the average score across the individual ratings is used and the individual ratings themselves are no longer used. In the example below, rather than using the individual conformance scores (F1-F7) using the best categorical approach, the average is used, which is equal for sketches 1 and 3. This same approach is used for the overall best sketches as well.

Table 7-15: Ambiguity Arising in Categorical Best Scoring Approach

Sketch	F1	F2	F3	F4	F5	F6	F7	Average
1	1	0	0	1	1	0	0	0.43
2	0	1	0	0	1	0	0	0.29
3	1	1	0	1	0	0	0	0.43

The selection of appropriate scoring approaches is determined in conjunction with the descriptive statistics from this preliminary study. The preliminary data show that the comparison of individual elements (e.g., F1) between treatment groups would not likely identify significant differences (see 7.2.4.2). Further, a comparison of these individual elements is specific to this design problem, a burrito folding device, and its individual requirements and model elements. For more general findings, and to identify more significant differences, the treatment groups are compared at the categorical and overall levels, rather than at the individual requirement and model element levels.

7.2.4.2 Conformance Descriptive Statistics

The results of this study are first analyzed using basic descriptive statistics to understand relationships and identify statistical tests that should be conducted. The data are analyzed using the four participant scoring approaches discussed previously: best sketch by element, best sketch by category, best sketch overall, and sketch average.

Sketch Scoring Approach: Participant Best by Element

The two types of participants, graduate and undergraduate, are evaluated separately to identify any qualitative differences between the groups. The results of sketch conformance for each function, activity, and interaction are summed for each group and shown in Table 7-16. The numbers in the cells represent the number of participants within the group that conformed to the model in at least one of his or her sketches.

Table 7-16: Conformance Results Using Best Sketch by Element Scoring Approach

Group	Treatment	n	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
Undergrad	IM	13	4	13	10	13	13	7	7	1	2	3	3	13	13	4	3
Undergrad	FM	13	3	12	9	12	12	6	6	5	2	7	4	12	12	6	3
Graduate	IM	7	3	7	4	6	7	3	4	4	1	5	2	7	7	5	2
Graduate	FM	7	3	7	4	7	5	5	3	3	3	3	3	7	7	3	3
Combined	IM	20	7	20	14	19	20	10	11	5	3	8	5	20	20	9	5
Combined	FM	20	6	19	13	19	17	11	9	8	5	10	7	19	19	9	6

There are several key observations and outcomes from these descriptive statistics. First, there does not appear to be a large difference between treatment groups for any particular function, activity, or interaction. Most differences in conformance for an individual element are small, and will likely not be significant using a statistical test. Therefore, comparisons of individual element scores will not be tested formally.

Second, the undergraduate and graduate participant results are inconsistent. For the function conformance, undergraduate participants with the IM treatment consistently conformed to the model better than participants with the FM treatment. While the differences are small for each element, the sum of all functional elements may be significant and will be investigated at the category level. The graduate participants, however, were inconsistent in differences, with three functions being equal (F2, F2, F3), two function conformance sums better within the IM group (F5, F7), and two function conformance sums better within the FM group (F4, F6). For the activity conformance sums, undergraduate participants in the FM group consistently outperformed or equaled participants in the IM group. The graduate participants, however, were inconsistent in differences. Due to the differences in undergraduate and graduate participant results, the participant classification (graduate or undergraduate) will be modeled as a blocking factor in this preliminary study.

Sketch Scoring Approach: Participant Best by Category

The results of the best sketch by category scoring approach are shown in Table 7-17. The numbers in the cells represent the average categorical conformance for participants based on the participant's sketch that best conformed to the model within that individual category (see Section 7.2.4.1).

Table 7-17: Conformance Results Using Best Sketch by Category Scoring Approach

Group	n	Treatment	Function (F1-F7)	Activity (A1-A4)	Interaction (I1-I4)
Undergrad	13	IM	4.62	0.54	2.46
Undergrad	13	FM	3.77	1.38	2.23
Graduate	7	IM	4.71	1.71	3.00
Graduate	7	FM	4.57	1.57	2.86
Combined	20	IM	4.65	0.95	2.65
Combined	20	FM	4.05	1.45	2.45

The outcomes from these results support the outcomes from the previous scoring approach. The graduate and undergraduate participants do not follow the same trends, and the differences in treatment groups within the graduate participants does not appear to be significant for any category. These results further support blocking of the two participant groups.

Sketch Scoring Approach: Participant Best Overall

The results of the best sketch overall scoring approach are shown in Table 7-18. The values represent the average participant conformance based on each participant's sketch that best conformed to the model. The data show that the conformance of the IM treatment group is better than the conformance of the FM treatment group for both undergraduate and graduate participants.

Table 7-18: Conformance Results Using Best Sketch Overall Scoring Approach

Group	n	Treatment	Overall
Undergrad	13	IM	7.62
Undergrad	13	FM	7.38
Graduate	7	IM	9.43
Graduate	7	FM	9.00
Combined	20	IM	8.25
Combined	20	FM	7.95

Sketch Scoring Approach: Participant Average Conformance

The results of sketch conformance for each function, activity, and interaction using the participant average conformance scoring approach are shown in Table 7-19. The conformance is averaged for sketches within a participant and the average across participants is shown in the cells of the table. The trends in conformance using this scoring approach support the previous observations mentioned: the graduate participant outcomes are not consistent with the undergraduate participant outcomes, and the differences between treatments within individual elements is small.

Table 7-19: Conformance Results Using Participant Sketch Average Scoring Approach

Group	n	Trt	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
Undergrad	13	IM	0.21	0.79	0.58	0.79	0.83	0.35	0.34	0.03	0.04	0.15	0.14	0.95	0.79	0.20	0.14
Undergrad	13	FM	0.12	0.56	0.40	0.56	0.64	0.30	0.31	0.19	0.06	0.33	0.19	0.75	0.56	0.31	0.18
Graduate	7	IM	0.19	0.93	0.29	0.86	0.93	0.29	0.57	0.38	0.14	0.57	0.21	0.95	1.00	0.57	0.21
Graduate	7	FM	0.26	0.84	0.45	0.81	0.68	0.39	0.35	0.24	0.23	0.24	0.25	1.00	0.84	0.27	0.25
Combined	20	IM	0.20	0.84	0.48	0.81	0.86	0.33	0.42	0.15	0.08	0.30	0.17	0.95	0.86	0.33	0.17
Combined	20	FM	0.17	0.66	0.42	0.64	0.66	0.33	0.32	0.21	0.12	0.30	0.21	0.84	0.66	0.29	0.20

7.2.4.3 Quality

Concept quality for this study has been evaluated in collaboration with Ramachandran, and a detailed discussion is presented in [87]. Each sketch was evaluated against the following nine requirements provided to participants using the scale discussed in Section 7.2.3:

- R1: Position empty tortilla to store fillings
- R2: Fill the tortilla after proper positioning
- R3: Wrap burrito over the fillings
- R4: Deliver completed burritos at rate of at least 4 burritos per minute
- R5: Be easy to use
- R6: Fit on a counter top
- R7: Be easy to install
- R8: Be easy to clean after use
- R9: Be safe to use

Requirements were categorized as functional (R1-R3), non-functional (R4-R9), and/or human activity (R5, R7, R8) in the analysis, and the treatment groups were compared at the overall level, category level, and individual requirement level. Concept quality was evaluated using all sketches developed by participants rather than using the participant scoring approaches discussed in Section 7.2.4.1, and graduate and undergraduate participants were treated collectively in the analysis. Although a different approach was used, the results are similar to the conformance descriptive statistics (see 7.2.4.2). There were significant differences in the overall average quality of sketches between the two treatment groups, as with categories of requirements. The outcomes of this initial quality study are used to identify the analysis that should be completed in a follow-up study that includes additional treatment groups. Based on the findings through the conformance investigation and this quality study, quality in the new study will be approached in a manner similar to conformance with respect to scoring approaches and participants.

7.2.4.4 Quantity

Quantity of ideas was measured by counting the number of sketches produced by each participant. The participants receiving a function model produced significantly more concepts than participants receiving an interaction model [87, 88]. Since there were differences in concept quantity in this study, it will be measured in the same manner in follow-up studies.

7.2.5 Limitations and Outcomes of the Initial Study

The quality and quantity of concepts generated using these two representations have been evaluated statistically and the results are presented in [87, 88]. Further, sketch conformance has been quantified and shown to have a high interrater agreement. The quality and quantity results in [87, 88] and the descriptive statistics of conformance in Section 7.2.4.2 are used to develop a new study to more fully test the use of artifact representations in conceptual design.

7.2.5.1 Scoring Approaches

In the previous study, sketches produced by participants were considered independent observations on the design problem. However, the sketches are dependent on the participant drawing the sketch and multiple sketches produced by a participant provide additional information about the variation within the participant rather than within the treatment group. Since the within-participant variation is not a focus of this study and since participants were allowed to create only a single sketch if he or she desired, a single score will be determined for each participant using several different approaches. Comparisons of sketch conformance will be completed at the category level (functions, activities, or interactions) rather than at the individual level (e.g., F1), due to the small differences between groups at the individual level and to the desire for more general conclusions. Overall conformance will not be assessed since additional treatments are introduced in the new study that do not contain activities and interactions and an overall conformance assessment would not be fair to all treatment groups. Similarly, quality of concepts will be compared at the category level (functional requirements, activity requirements) rather than at the individual level for the same reasons as conformance. In addition, quality will be compare at the overall level to understand the effect of the treatment on the overall quality of the concepts. Both the average and best approaches will be used for conformance and quality to ensure a broader assessment of the participant.

7.2.5.2 Participants

In the initial study, both graduate and undergraduate students participated in the study. However, the graduate participant conformance data are inconsistent with the undergraduate data. One explanation for this difference is due to the way in which the experiment was conducted for graduate participants. Since the graduate participants had an understanding of function modeling prior to the study, they all were also trained in interaction modeling to ensure that each had a similar level of training. Although participants in the FM group did not receive an interaction model, they may have been influenced by the discussion of interaction and human activities immediately before the design problem was given. Further, the background of graduate students is diverse since it includes both domestic and international students, students from different undergraduate institutions and majors, and a wider age range of students compared to undergraduate students. Additionally, after the study was conducted, some international students expressed that they did not know what a burrito was. For these reasons, only senior-level undergraduate students at Clemson University will participate in the new study, and the graduate participant data will not be used.

7.2.5.3 Control Group

The goal of the initial study was to compare the interaction model to the function model, a model well-promoted within the design research community. The results show that the interaction model increases the quality of concepts compared to the function model. However, the effect of the function model on concept quality is not known and has not been rigorously tested through quantitative studies. It is possible that the models have a negative effect on concept quality, and that the function model has a greater negative effect than the interaction model. For this reason, an additional treatment group is added that receives only a problem statement and requirements (no model, NM) with the design problem. This will provide a true baseline to understand how representations affect a designer in the concept generation process. In addition, a pruned function model is tested in the new study since it has been shown previously to be easier to interpret than a function model (see Chapter 5) and to understand the effect of the activity portion of the interaction model. This model is discussed further in Section 7.3.1.

7.3 Extended Study

The initial user experiment was conducted primarily to understand the differences in the effects of two artifact models on concept quality and quantity. These differences and the statistical analysis and conclusions are discussed in detail in [87, 88]. The study has also been used to understand the experiment design and improve upon it for an extended user study based on the initial study. The initial user experiment was used to complete the following tasks:

- create a reliable, quantitative metric of concept quality based on problem requirements [87]
- create a reliable, quantitative metric of sketch conformance

- identify limitations of the statistical model and revise the model for the extended study
- identify scoring approaches to use with the revised statistical model
- identify new treatment groups to further understand the problem

The new treatment groups identified through the previous analysis are the No Model (NM) and Pruned Model (PM) groups. The NM treatment is introduced to serve as a true baseline for the effect of using artifact models to generate concepts. The previous baseline, FM, was used because it is used often in the design community (e.g., see [12, 24, 25, 32, 40-44, 50, 55, 92, 93]). However, after discussing the results of this study and drawing conclusions, the need for a baseline for the FM group was identified. The FM has not been quantitatively shown to support concept generation, so the NM group is introduced to understand the effect of the FM on ideation.

The PM treatment is introduced into this experiment for several reasons. First, the activity portion of the interaction model (see Figure 7-4) has not been researched to the extent of function models. There are many different ways to model human actions and/or processes. The activity model [4], which was chosen to be used with the interaction model, has not been tested for its usefulness in design. The evaluation and selection of an appropriate activity modeling method to be merged with the function modeling method is outside the scope of this research, so the PM treatment is introduced to understand if the selected activity modeling method is advantageous. The PM treatment, shown in Figure 7-7, is the functional subset of the interaction model (compare to Figure 7-4). The PM is identical to the IM with the activity portion removed from the model. The PM has all flows entering the system from the environment, rather than some entering from the user. The PM is also a subset of the FM, where the periphery, interaction-focused functionality (e.g., insert tortilla) has been removed (compare to Figure 7-3). The addition of the pruned model allows for better understanding of specific aspects of the model.

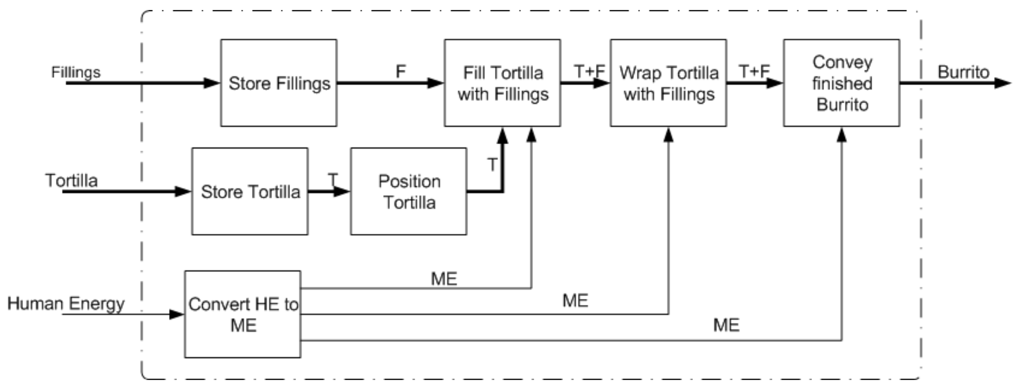


Figure 7-7: Burrito Folder Pruned Model

7.3.1 Experiment Overview

The extended experiment is a single factor, completely randomized design. Participants were solicited from senior design classes at Clemson University to ensure that they had a common educational background and design knowledge. Participants were asked by email and in person to participate, and were offered a small gift for participating. The goal for the study was to obtain approximately 19 participants per treatment groups. The extended user experiment was conducted in a similar manner to the initial experiment, so the undergraduate participant data from the initial study were used with the results of the extended study. Thus, 50 new participants were desired, and 43 eligible participants completed the study over a three-week period during the Fall 2011 semester. Participants were randomly assigned to a treatment group. Due to the reuse of some data, the assignment to a new treatment group was more likely than to a previous treatment group. The potential for bias from this unequal weighting is discussed in Section 7.3.2.

The study took approximately one hour to complete, depending on the treatment group. Participants were scheduled either individually or in groups to participate based on the treatment group to which they were assigned. The general schedule for the user study is shown in Table 7-20. Participants were first read an IRB statement asking for their consent to participate in the study. The general study procedure was then explained to participants. For the FM, IM, and PM treatment groups, a short presentation was given on either function modeling (FM and PM groups) or interaction modeling (IM group) to explain the basics of the representation, how it can be used by a designer, and how the participant should use the model during ideation. Participants in the NM group did not receive any training in the use of a representation. Participants were then explained the expectations for sketching concepts. Participants were instructed that the content of the sketch was being evaluated rather than their artistic abilities, and participants were encouraged to include textual descriptions of the concept to aid the researchers in the evaluation process. Participants were then given the problem statement, requirements, and the appropriate treatment and were allowed to sketch ideas for 30 minutes.

Table 7-20: User Study Schedule

Event	Time Allotted
Read Institutional Review Board Statement and explain general experiment procedure	5 minutes
Train participants in appropriate representation (NM group did not receive any training)	20 minutes
Explain expectations for sketching/concepts	5 minutes
Sketch concepts for design problem	30 minutes

7.3.2 Sources of Variation`

There are several sources of variation that will be accounted for in the experiment design and statistical models used to analyze the data.

Treatment: The representation provided to the participants (FM, IM, PM, or NM) is the source of variation that is being studied. Differences in conformance, quality, and quantity between treatment groups will be studied.

Participant Experience: The participants in this study may vary in terms of design experience, work experience, GPA, and their ability to generate ideas. Since all participants are senior-level undergraduate students from Clemson University, the participants' experience and prior ability to generate ideas was not measured in this study. Participants are randomly assigned to groups, so participant experience is accounted for in the error term of the statistical model.

Participant Environment: The time and location of the study may affect the participants' interest in the training and the design problem. All studies were conducted during the day and in conference rooms within the mechanical engineering building at Clemson University. Participants were familiar with the room setting and participated during hours that classes are typically scheduled. Thus, the testing environment was held constant across all participants.

Conformance and Quality Ratings: The reliability of both conformance and quality ratings was tested using Cohen's Kappa and is discussed in Sections 7.2.2 and 7.2.3. After establishing a reliable rating system, a single rater graded all sketches for conformance and quality in this extended study.

Participant Recruiting: The initial study was performed during a regularly scheduled class, so all students attending class on that particular day participated in the study. In the extended study, participants were asked to volunteer outside of class to attend the study. Participants in the extended study were also provided a thank-you gift of digital calipers, valued at approximately $8.00. There is potential that participants of the initial study were not as interested in the study and may not have put in as much effort since it was conducted during class. However, participants in the extended study may also not have been interested if they attended to receive the gift. Additionally, since a large proportion of the participants the FM and IM groups were from the initial study, there is potential that the FM and IM scores are biased due to the level of interest. The differences within each of these groups between the initial and extended study participants can be tested to determine if there is a significant difference. However, due to the small sample size of participants in the new study for these groups, the tests will not be meaningful. The proportion of participants in the extended study that were drawn from the potential candidate pool is large—approximately one third. There were many participants that signed up as a favor to the researchers rather than a desire to perform a design problem. Due to the large proportion of participants needed and the thank-you gift in the extended study, the participant interest in performing the study was likely not different from the initial study.

Training Presenter: The training presentation conducted during the study to introduce the representation to the participants was conducted primarily by one individual. During one study session in the extended study and during the initial study, an additional presenter was needed to deliver the training in parallel. In each case, the additional presenter was highly involved in the research and familiar with the modeling method that was being presented.

7.3.3 Sample Size Calculations

The data from this initial study are analyzed to approximate the mean squared error (MSE) term and compute the sample size required for the extended study. The sample size is determined using the desired length of confidence intervals, the 90% upper confidence limit for σ^2, and Fisher's LSD comparison procedure. The desired length of the confidence interval is 10% of the difference in the maximum possible score and the minimum possible score, or 0.1 for conformance metrics and 0.8 for quality metrics. The sample sizes required for the desired interval lengths for conformance and quality using each scoring approach is shown in Table 7-21 and Table 7-22. The procedure for calculating sample size is included in Appendix D.

Table 7-21: Conformance Sample Size Calculations

Category	Scoring Approach	MSE	σ^2 90% UCL	interval length	replicates per group	n
Functional	Average	0.1153	0.1767	0.1	137	548
	Best	0.0437	0.0670	0.1	53	212
Activity	Average	0.0055	0.0084	0.1	8	32
	Best	0.0121	0.0185	0.1	15	60
Interaction	Average	0.0307	0.0471	0.1	37	148
	Best	0.0166	0.0254	0.1	21	84

Table 7-22: Quality Sample Size Calculations

Category	Scoring Approach	MSE	σ^2 90% UCL	interval length	replicates per group	n
Overall	Average	1.0714	1.6421	0.8	21	84
	Best	1.1007	1.6870	0.8	21	84
Functional	Average	3.448	5.2847	0.8	65	260
	Best	2.6923	4.1265	0.8	51	204
Activity	Average	1.2269	1.8805	0.8	24	96
	Best	2.0940	3.2095	0.8	40	160

As shown in the tables, the number participants required depends on both the metric and the scoring approach used. The sample size calculations reveal the large amount of error associated with the participants relative to the desired difference in the groups. Due to resource constraints and the availability of participants, the required sample sizes cannot be achieved, so as many participants as possible will be used. The experiment, therefore, will only be able to detect large differences among the means of groups, and small differences will not be identified as statistically significant. This limitation is recognized, but the study will still be conducted to identify large effects among groups. To identify smaller differences, the MSE must be reduced. This can be accomplished by measuring covariates or blocking participants based on some characteristics and incorporating the covariates or blocking factors into the model. However, the appropriate covariates and participant characteristics are not known.

Therefore, the experiment will be conducted with the maximum possible sample size to identify large effects and to understand more about conducting experiments with designers. The goal for this experiment was set at 19 participants per group, or 76 total. This number is based on the number of available participants and the desire to include at least one third new participants with the previous data collected. Previously, two of the treatment groups contained 13 participants, so six additional participants for these two groups would reduce potential bias caused by using data from the previous study.

7.3.4 Quantification of Concept Sketches

Since some of the data from the initial study are used in the extended study, the conformance and quality scales were again tested for intrarater agreement to account for a change in the rater's preferences over time. The previously developed scales were used to rate ten of the initial sketches for conformance. The rater's new scores were checked against the past scores using Cohen's Kappa. All functions, all interactions, and two activities had acceptable levels of IRA. Two of the activities' IRA were low due to the rater being too liberal in rating. This bias was identified, corrected, and checked with ten different sketches from the initial study. The IRA for was acceptable for the second set of concepts, and the final kappa values for all fifteen conformance elements are shown in Table 7-23. Since the rater was consistent across time using this scale, the conformance ratings from the previous study were used and the new sketches were rated by the same rater.

Table 7-23: Conformance Scale IRA Over Time

	Actual Agreement	Cohen's Chance Agreement	Cohen's Kappa
F1	1.00	0.68	1.00
F2	1.00	0.58	1.00
F3	0.90	0.54	0.78
F4	1.00	0.58	1.00
F5	1.00	0.52	1.00
F6	1.00	0.68	1.00
F7	0.90	0.50	0.80
A1	1.00	0.52	1.00
A2	1.00	0.58	1.00
A3	1.00	0.50	1.00
A4	1.00	0.58	1.00
I1	1.00	0.68	1.00
I2	1.00	0.58	1.00
I3	0.80	0.52	0.58
I4	0.90	0.54	0.78

The quality scale was also checked for reliability over time using the same approach as the conformance scale. The quality scale required several iterations for a few of the requirements, but overall maintained a high level of agreement. The final IRA for the quality metrics by a single rater over time is shown in Table 7-24.

Table 7-24: Quality Scale IRA Over Time

	Actual Agreement	Cohen's Chance Agreement	Cohen's Kappa
R1	1.00	0.42	1.00
R2	0.90	0.46	0.81
R3	1.00	0.38	1.00
R4	0.80	0.35	0.69
R5	0.60	0.29	0.44
R6	0.90	0.50	0.80
R7	1.00	0.82	1.00
R8	0.80	0.66	0.41
R9	0.80	0.52	0.58

7.3.5 Quantitative Analysis

The conformance and quality of concepts are measured for all sketches created during the study. The scoring approaches discussed in Section 7.2.5.1 will be used to determine an individual participant’s conformance and quality scores. The quantity of sketches is also measure for each participant. These three metrics—conformance, quality, and quantity—are fit using a linear model:

$$Y_{it} = \mu + \tau_i + \varepsilon_{it} \tag{2}$$

where Y_{it} is the response for the t^{th} participant within the i^{th} treatment

μ is the overall average response

τ_i is the effect of the i^{th} treatment on response

ε_{it} is the error of the t^{th} participant of the i^{th} treatment

A one-way analysis of variance (ANOVA) is conducted to test equality of the means of all treatments on the response:

H_0: $\tau_{FM} = \tau_{IM} = \tau_{PM} = \tau_{NM}$

H_A: at least one τ_i differs for $i = FM, IM, PM, NM$ (3)

Due to the exploratory nature of this research, a significance level, α, of 0.1 is used with the F-test to determine if any of the means are different. If a significant difference is found, all pairwise contrasts are conducted to determine which means are significantly different from the others. The six pairwise contrasts are computed with confidence intervals using the general equation:

$$\bar{y}_{i\cdot} - \bar{y}_{s\cdot} \pm w_{crit} \cdot SE \tag{4}$$

where $i \neq s$

$\bar{y}_{i\cdot}$ is the mean response for the ith treatment

$\bar{y}_{s\cdot}$ is the mean response for the sth treatment

w_{crit} is the critical coefficient for the confidence interval, and

SE is the standard error of the difference.

When multiple contrasts or hypothesis tests are conducted, the overall error rate of the experiment is controlled through multiple comparison procedures. In this research, Tukey's method for all pairwise comparisons is appropriate since all pairwise comparisons are being made if the ANOVA reveals a difference in treatment means. The critical value for Tukey's method in this experiment for a 90% family-wise confidence level is 2.34 ($\nu = 4$, $n - \nu = 65$, $\alpha = 0.1$). However, since this research is exploratory in nature, Fisher's Least Significant Difference (LSD) method may also be used, but the experimental error rate is not controlled. The critical value for Fisher's LSD using a significance level of 0.1 is 1.67 ($n - \nu = 65$, $\alpha/2 = 0.05$). Thus, Tukey's method is a much more conservative comparison than Fisher's LSD since it controls the overall experimental error rate. As a compromise between these two methods, Fisher's LSD will be used with a significance level of 0.05 rather than 0.1, resulting in a critical value of 2.00 ($n - \nu = 65$, $\alpha/2 = 0.025$). This approach is a balance between the two methods, allowing smaller differences to be explored as is appropriate for this type of research. Further, the consequences of a Type I error are minimal, since the outcomes from this study will be used to further study the representations rather than fully rejecting some or all of them. Thus, Fisher's LSD with an individual significance level of 0.05 will be used for all pairwise contrasts, resulting in an uncontrolled experimental error rate.

7.3.6 Experiment Validity

Verification and validation of the results will be completed using the methods presented by Blessing and Chakrabarti [94]:

Statistical conclusion validity: Model assumptions will be checked after all data are collected, and interrater agreement is used to ensure that the response measurements are reliable. The data were collected over several weeks and in slightly different settings, but all factors were controlled as closely as possible to maintain statistical conclusions validity (see discussion in Section 7.3.2).

Internal validity: The causality of the relationship is ensured by randomly assigning participants to treatment groups. Participants are given only one treatment, so there is no potential for bias from learning about the design tool or design problem through practicing within the study. There is potential for the participants to discuss the study with other participants, but the participants were asked not to discuss the design problem with others. The sample size for each treatment was relatively large with at least sixteen participants in each group, reducing the chance that one group is randomly assigned a biased set of participants.

Construct validity: The construct validity for model conformance is ensured by measuring a variety of functional, activity, and interaction information. This coverage tests many different aspects of the model and the participants' level of understanding of these categories in general, rather than the participants' understanding of a specific function for this specific design problem. Construct validity for concept quality is ensured by measuring a variety of requirements that cover functional, activity, and performance characteristics. Conclusions are drawn at the category and overall level rather than at the specific requirement level to test the representations' influence on higher level constructs rather than on specific requirements for this specific design problem. Construct validity is also ensured by not communicating to the participants what is being measured, so the participants did not explicitly consider whether or not the ideas in their sketch matched the information in the model. Participants were also not aware of the quality scale developed to assess the sketches and were not aware that quantity of concepts was being measured.

External validity: The study participants are senior-level undergraduate mechanical engineering students in the first two design courses. Most participants have little work experience, so the results can be generalized to mechanical designers with little formal training in design and little to no work experience. The study was conducted using a single design problem, a burrito-folding device, which is a threat to the generalizability of the findings. This artifact was selected to be representative of a consumer artifact with basic mechanical functionality as well as human interactions, but the findings may be specific to this problem alone. Further testing on a variety of design problems will give confidence in this generalization. The findings will be helpful for understanding the influence of artifact models on novice designers within the conceptual phase of design of consumer mechanical artifacts.

7.3.7 Results

The data analysis is completed using R statistical software [95] and packages `multcomp` [96] and `vcd` [97]. All code is included in Appendix E. Each model is checked to ensure that the following assumptions are met: model fit, outliers, constant variance, and normality. Since participants' names were not associated with the sketches generated, the sequence of participation is not known and independence is assumed to be satisfied. Model fit is checked by plotting standardized residuals against treatments. Data are considered to be potential outliers if they are greater than three standard deviations away from the mean. Constant variance is checked by plotting standardized residuals against fitted values and by comparing the largest treatment variance to the smallest treatment variance. Normality is checked by plotting standardized residuals against their normal scores. The plots and discussion of these assumptions are included in the Appendices. In the case of non-normal data, the Kruskal-Wallis rank sum test is used rather than the linear model and ANOVA.

7.3.7.1 Conformance

The functional, activity, and interaction conformance is compared using the participant average and participant best scoring approaches. First, the functional conformance data are fit with a linear model and a one-way ANOVA performed. The

ANOVA tables for the participant average scoring approach are shown in Table 7-25. All model assumptions are satisfied for participant average scoring approach, but normality is not satisfied for the participant best scoring approach (see Appendix B). Therefore, the linear model and one-way ANOVA cannot be used to compare the groups means using the participant best scoring approach, and its nonparametric analogue, the Kruskal Wallis test [98] is used. This test does not require data to be normally distributed.

The null hypothesis for the Kruskal-Wallis test states that the four treatment distributions (FM, IM, PM, NM) are equal, and the alternative hypothesis states that at least one of the populations yields different results [99]. A test statistic is computed and compared to a critical value of the Chi-square distribution. Using the procedure `kruskal.test` in R, the Kruskal-Wallis rank sum test is performed and the associated p-values computed. The test reveals that there is no significant difference between any of the groups ($p = 0.31$, Chi-square = 3.58).

Table 7-25: Functional Conformance ANOVA Table – Participant Average Scoring Approach

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value	p-value
Treatment	3	0.330	0.1101	2.073	0.112
Error	65	3.451	0.0531		
Total	68	3.781			

The significance level for the hypothesis test that all treatment means are equal is not significant for either scoring approach ($\alpha = 0.1$). However, the p-value for the participant average scoring approach is small, so the pairwise comparisons between treatment groups will be performed. Descriptive statistics are shown in Table 7-26, and the 95% confidence intervals shown in Table 7-27 are used for Fisher's LSD hypothesis tests. If the confidence interval includes 0, then there is no significant difference between the two treatment groups. If the interval does not include 0, then there is a significant difference. See Section 7.3.5 for a discussion on multiple comparison procedures. Therefore, the average functional conformance by participants using a pruned model is greater than that of participants using a function model, using a significance level of 0.05. No other significant differences exist in terms of functional conformance.

Table 7-26: Descriptive Statistics for Functional Conformance – Participant Average Scoring Approach

Treatment	Mean	Standard Deviation	Observations
FM	0.469	0.263	18
IM	0.534	0.191	16
NM	0.515	0.242	18
PM	0.655	0.215	17

Table 7-27: 95% Confidence Interval for All Pairwise Comparisons of Functional Conformance – Participant Average Scoring Approach

Contrast	Estimate	Lower Bound	Upper Bound
IM – FM	0.066	-0.093	0.224
NM – FM	0.046	-0.108	0.199
PM – FM	0.187	0.031	0.342
NM – IM	-0.020	-0.178	0.138
PM – IM	0.121	-0.039	0.282
PM – NM	0.141	-0.015	0.297

The activity conformance data are fit with a linear model and model assumptions are checked. All assumptions are satisfied for each scoring approach, with the exception of normality. As shown in Figure 7-8, the normal probability plots are not linear. Many of the sketches received a score of 0, causing the distribution to be non-normal. Therefore, the Kruskal Wallis test is used.

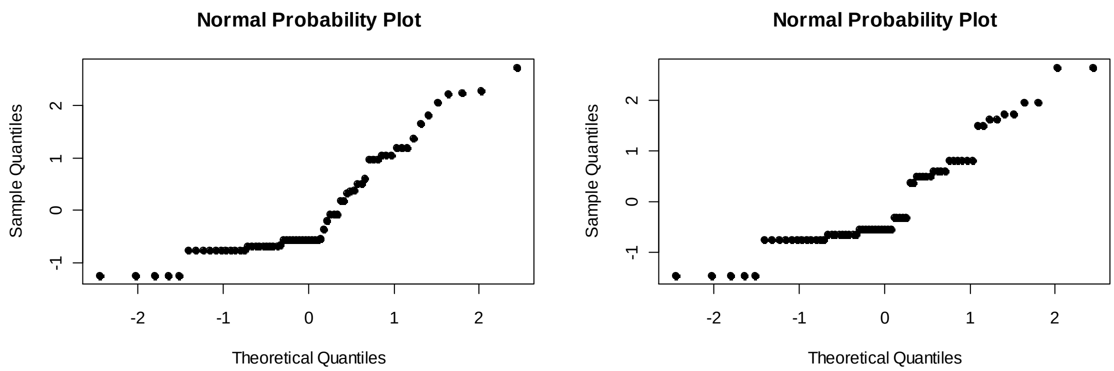


Figure 7-8: Normal Probability Plots for Activity Conformance – Participant Average Scoring Approach (left) and Participant Best Scoring Approach (right)

The resulting test statistics and p-values from the Kruskal-Wallis rank sum test for each scoring approach are shown in Table 7-28. The resulting p-values using this test are close to the p-values for the linear model and ANOVA, which are 0.226 and 0.045 for the participant average and participant best scoring approaches, respectively. This result is expected since the ANOVA F-test is robust against non-normality except for extreme non-normality [99]. Since there is a significant difference between the groups using the participant best scoring approach, pairwise comparisons are made using Mann-Whitney's U test. As with multiple comparison procedures for parametric data (see 7.3.5), these tests can be corrected to control the experiment error rate. However, since this research is exploratory, the overall experimental error rate is not controlled, but a two-sided tests with significance levels of 0.05 are used to compare the groups, similar to the Fisher's LSD procedure with normal distributions. The median values are shown in Table 7-29.

The pairwise Mann-Whitney U tests (see Table 7-30) reveal that the distribution of the FM group is significantly higher than the distribution of the IM group (p = 0.036) and PM group (p = 0.015). The test does not reveal a significant difference between the FM and NM groups (p = 0.056). However, the p-values are approximate due to ties in the data, so a difference between these two groups is possible.

Table 7-28: Kruskal-Wallis Rank Sum Test for Activity Conformance

	Chi-square Test Statistic	p-value
Participant Average Scoring Approach	4.969	0.174
Participant Best Scoring Approach	7.830	0.050

Table 7-29: Descriptive Statistics for Activity Conformance – Participant Best Scoring Approach

Treatment	Mean	Median	Observations
FM	0.319	0.25	18
IM	0.141	0.00	16
NM	0.167	0.00	18
PM	0.118	0.00	17

Table 7-30: All Pairwise Comparisons of Activity Conformance – Participant Best Scoring Approach

Contrast	Mann-Whitney p-value
IM – FM	0.036
NM – FM	0.056
PM – FM	0.015
NM – IM	0.858
PM – IM	0.679
PM – NM	0.860

The interaction conformance data are fit with a linear model and model assumptions are checked. Using the participant average scoring approach, there were two potential outliers in the function model group (FM), one greater than three standard deviations above the mean (3.10) and one close to three standard deviations above the mean (-2.77). These two data points were further investigated, and the low score was considered to be an outlier based on the participant's sketches. It was clear from the sketches that this particular participant was not aware of the sketching expectations. Rather than sketch a design concept, the participant created a new function model and calculated the power required to warm a burrito in the allotted time. This same participant's data were also removed when using the participant best scoring approach, since it was 3.12 standard deviations below the mean. The high scoring participant's data were not removed because further inspection of the sketches revealed that they were good concepts that conformed to the model well. Complete details of the modeling assumptions and outlier removal is included in Appendix B. After removing outliers, the linear model is fit to the data and assumptions are checked. The modeling assumptions for the participant average scoring approach are all satisfied, while the assumption of normality is not satisfied for the participant best scoring approach. Therefore, the Kruskal-Wallis rank sum test is performed for the participant best scoring approach. The ANOVA table for the participant average scoring approach is shown in Table 7-31. The Kruskal-Wallis rank sum test reveals no significant difference between the groups for interaction conformance ($p = 0.13$). Descriptive statistics are shown Table 7-32 and Table 7-33.

Table 7-31: Interaction Conformance ANOVA Table – Participant Average Scoring Approach

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value	p-value
Treatment	3	0.143	0.0477	1.731	0.170
Error	64	1.762	0.0275		
Total	67	1.905			

Table 7-32: Descriptive Statistics for Interaction Conformance – Participant Average Scoring Approach

Treatment	Mean	Standard Deviation	Observations
FM	0.500	0.220	17
IM	0.499	0.143	16
NM	0.442	0.154	18
PM	0.570	0.132	17

Table 7-33: Descriptive Statistics for Interaction Conformance – Participant Best Scoring Approach

Treatment	Mean	Median	Standard Deviation	Observations
FM	0.632	0.50	0.200	17
IM	0.609	0.50	0.182	16
NM	0.500	0.50	0.210	18
PM	0.618	0.50	0.129	17

7.3.7.2 Quality

The overall, functional, and activity quality are compared using the participant average and participant best scoring approaches. The overall quality data are fit with a linear model and a one-way ANOVA is performed. All model assumptions are satisfied for overall quality data using the participant best scoring approach, but normality is not satisfied using the participant average scoring approach (see Appendix C). Therefore, the Kruskal-Wallis rank sum test is used for the participant average scoring approach.

Descriptive statistics for each are shown in Table 7-34 and Table 7-35, and the ANOVA table for the participant best scoring approach is shown in Table 7-36.

Table 7-34: Descriptive Statistics for Overall Quality – Participant Average Scoring Approach

Treatment	Mean	Standard Deviation	Observations
FM	4.933	0.982	18
IM	5.146	0.940	16
NM	5.138	0.971	18
PM	5.502	1.098	17

Table 7-35: Descriptive Statistics for Overall Quality – Participant Best Scoring Approach

Treatment	Mean	Standard Deviation	Observations
FM	5.654	0.983	18
IM	5.583	0.949	16
NM	5.494	0.945	18
PM	5.980	0.988	17

Table 7-36: Overall Quality ANOVA Table – Participant Best Scoring Approach

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value	p-value
Treatment	3	2.318	0.7728	0.827	0.484
Error	65	60.737	0.9344		
Total	68	63.055			

The Kruskal-Wallis rank sum test reveals no significant difference between treatment groups ($p = 0.477$, Chi-square = 2.49) in overall quality using the participant average scoring approach, and the ANOVA reveals that there is no difference between the overall quality group means using the participant best scoring approach ($p = 0.484$). The descriptive statistics show that the pruned treatment group has the highest average quality of concepts, but the differences in group means is not significant in either scoring approach. Therefore, pairwise comparisons between group means are not made.

The functional quality data are fit with a linear model and a one-way ANOVA performed for each of the two scoring approaches. All model assumptions are satisfied for functional quality using the participant average approach (see Appendix C). The ANOVA table for this scoring approach and the descriptive statistics are shown in Table 7-37 and Table 7-38.

Table 7-37: Functional Quality ANOVA Table – Participant Average Scoring Approach

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value	p-value
Treatment	3	26.328	8.7760	2.474	0.069
Error	65	230.584	3.5475		
Total	68	256.912			

Table 7-38: Descriptive Statistics for Functional Quality – Participant Average Scoring Approach

Treatment	Mean	Standard Deviation	Observations
FM	4.483	1.869	18
IM	4.689	1.570	16
NM	5.154	1.990	18
PM	6.092	2.044	17

The functional quality ANOVA table shows that there is a significant difference in at least one of the treatment means, so all pairwise comparisons will be made for the participant average scoring approach. The contrasts and associated 95% confidence intervals are shown in Table 7-39. The pruned model group had an average functional quality of 6.09 compared to 4.69 and 4.48 for the interaction and function model groups, respectively. Participants receiving the pruned model performed significantly better than participants receiving the interaction model or the function model, as demonstrated by the confidence intervals. The pruned model, however, did not result in a higher functional quality score over the no model group using this scoring approach.

Table 7-39: 95% Confidence Interval for All Pairwise Comparisons of Functional Quality – Participant Average Scoring Approach

Contrast	Estimate	Lower Bound	Upper Bound
IM – FM	0.206	-1.086	1.499
NM – FM	0.672	-0.582	1.925
PM – FM	1.609	0.337	2.881
NM – IM	0.465	-0.827	1.758
PM – IM	1.403	0.092	2.713
PM – NM	0.937	-0.335	2.209

When fitting the data for the participant best scoring approach with a linear model, the normality assumption is not satisfied. Therefore, a Kruskal-Wallis rank sum test is performed. This test reveals a significant difference between at least one of the groups ($p = 0.044$, Chi-square = 8.11), so pairwise comparisons are made. Descriptive statistics are shown in Table 7-40, and the resulting p-values from all pairwise comparisons using Mann-Whitney U tests are shown in Table 7-41. The experimental error rate in uncontrolled in these comparisons, so two-sided comparisons and a significance level of 0.05 is used. The results show that the pruned model is significantly better than all three other treatments, and the other three treatment groups did not differ significantly from each other.

Table 7-40: Descriptive Statistics for Functional Quality – Participant Best Scoring Approach

Treatment	Mean	Median	Standard Deviation	Observations
FM	5.630	6.333	1.672	18
IM	5.167	5.000	1.388	16
NM	5.667	5.000	1.786	18
PM	6.961	7.000	1.848	17

Table 7-41: All Pairwise Comparisons of Functional Quality – Participant Best Scoring Approach

Contrast	Mann-Whitney p-value
IM – FM	0.394
NM – FM	0.935
PM – FM	0.041
NM – IM	0.646
PM – IM	0.009
PM – NM	0.045

For both scoring approaches, the average functional quality of concepts developed by participants using a pruned model is greater than that of participants using a function model or an interaction model, using a significance level of 0.05. If considering the best sketch only, the average functional quality produced by participants using a pruned model is greater than that of participants using no model, using a significance level of 0.05.

The activity quality data are fit with a linear model and model assumptions are checked. The plot of treatments against the standardized residuals for the participant average scoring approach revealed a potential lack of fit of the data since many of the data points were below zero. Further, there were three potential outliers identified in the participant average scoring approach, which had standardized residuals of 2.35, 2.65, and

3.13. These points were not removed, however, since there was nothing abnormal about the sketches. The normality assumption is not satisfied since the normal probability plot reveals a heavy-tailed distribution. Using the participant best scoring approach, all assumptions are satisfied except normality, since the normal probability plot is not linear among other problems. The two normal probability plots for activity quality are shown in Figure 7-9. Since normality is not satisfied for either scoring approach, the Kruskal-Wallis rank sum test is used to compare groups.

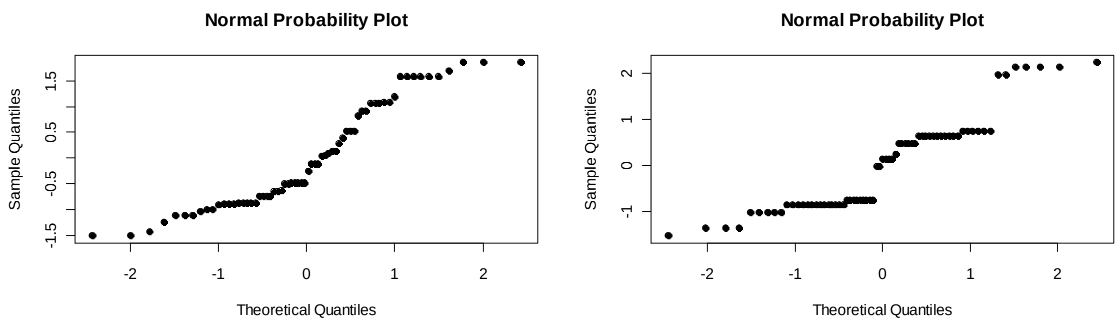


Figure 7-9: Normal Probability Plots for Activity Conformance – Participant Average Scoring Approach (left) and Participant Best Scoring Approach (right)

The Kruskal-Wallis rank sum test is performed for both scoring approaches, and the resulting test statistics and p-values are shown in Table 7-42. Since there are no significant differences between the groups, pairwise comparisons are not performed.

Table 7-42: Kruskal-Wallis Rank Sum Test for Activity Quality

	Chi-square Test Statistic	p-value
Participant Average Scoring Approach	1.705	0.634
Participant Best Scoring Approach	0.481	0.923

Table 7-43: Descriptive Statistics for Activity Quality – Participant Average Scoring Approach

Treatment	Mean	Median	Observations
FM	5.088	5.000	17
IM	5.551	5.778	15
NM	5.630	5.000	18
PM	5.410	5.000	16

Table 7-44: Descriptive Statistics for Activity Quality – Participant Best Scoring Approach

Treatment	Mean	Median	Observations
FM	6.148	6.333	18
IM	6.375	6.666	16
NM	6.148	5.666	18
PM	6.020	5.000	17

7.3.7.3 Quantity

The quantity of concepts produced by each participant is measured by counting the number of sketch sheets that each participant used. Since each participant must produce at least one sketch, quantity is defined by the count of each additional sketch beyond the first. The observed and expected count frequencies assuming a Poisson distribution are shown in Table 7-45. A Chi-square goodness of fit test is conducted to test the hypothesis that the sketch quantity data follow a Poisson distribution. The null hypothesis is that the data are Poisson-distributed and the alternative hypothesis is that they are not. The Chi-square test statistic, computed using the `goodfit` function in R, is 2.49, resulting in a p-value of 0.478. Therefore, the null hypothesis is not rejected, and the data are Poisson-distributed.

Descriptive statistics for the quantity data are shown in Table 7-46. The quantity data are fit with a generalized linear model and all pairwise contrasts are conducted to identify any differences in concept quantity between treatment groups. The group means are compared using Tukey's multiple comparison procedure, using a family-wise confidence level of 0.90. The contrast estimates and confidence intervals are shown in Table 7-47. The quantity of concepts generated by participants receiving the function model is significantly greater than the quantity of concepts generated by participants using the pruned model or no model. There are no other significant differences between groups.

Table 7-45: Expected Counts for Concept Quantity Assuming a Poisson Distribution

Sketch Count (beyond 1)	Observed Counts	Expected Counts
0	16	18.46
1	27	24.34
2	19	16.05
3	4	7.05
4	3	2.33

Table 7-46: Descriptive Statistics for Quantity of Concepts

Treatment	Mean	Standard Deviation	Observations
FM	3.000	1.085	18
IM	2.250	1.000	16
NM	2.000	0.840	18
PM	1.882	0.857	17

Table 7-47: Tukey's 90% Family-Wise Confidence Intervals for All Pairwise Comparisons of Concept Quantity

Contrast	Estimate	Lower Bound	Upper Bound
IM – FM	-0.470	-1.107	0.167
NM – FM	-0.693	-1.353	-0.033
PM – FM	-0.818	-1.521	-0.116
NM – IM	-0.223	-0.966	0.520
PM – IM	-0.348	-1.129	0.432
PM – NM	-0.125	-0.924	0.674

7.3.7.4 Quality Density

Based on the findings above that there are significant differences in the quantity of concepts produced, but no differences in overall quality or activity quality, a new metric is developed, quality density. The quality density is defined as the participant's best quality score (either overall or categorically) divided by the number of sketches produced by the participant, as shown in Equation 5.

$$\text{Quality Density} = \frac{\text{Participant's Best Quality Score}}{\text{Number of Sketches Produced by Participant}} \qquad (5)$$

This quality density is important for concept selection procedures, where a set of good concepts would be selected from a set of all concepts. The quantity of concepts is in the denominator of the metric because a smaller number of concepts would result in a faster concept selection process since the designer would be selecting from fewer concepts.

The quality density data are not expected to be normally distributed, so the Kruskal-Wallis rank sum test is used for these data. Only one scoring approach is used, the participant best scoring approach. The Kruskal-Wallis rank sum test for the overall quality density reveals a significant difference in at least one of the groups (p = 0.027, Chi-square = 9.182), so pairwise comparisons are made. Descriptive statistics for overall quality density are shown in Table 7-48, and the p-values from all pairwise comparisons using Mann-Whitney U tests are shown in Table 7-49.

Table 7-48: Descriptive Statistics for Overall Quality Density

Treatment	Mean	Median	Standard Deviation	Observations
FM	2.258	1.755	1.304	18
IM	3.044	2.389	1.714	16
NM	3.420	2.611	1.904	18
PM	3.889	3.167	1.887	17

Table 7-49: All Pairwise Comparisons of Overall Quality Density

Contrast	Mann-Whitney p-value
IM – FM	0.098
NM – FM	0.041
PM – FM	0.007
NM – IM	0.616
PM – IM	0.134
PM – NM	0.321

Since the overall experimental error is not controlled, the significance level for individual contrasts is 0.05. Therefore, the quality density generated by participants using a pruned model (p = 0.041) or no model (p = 0.007) is significantly greater than that of participants using a function model. No other significant differences exist among the groups (see Table 7-49).

The groups are compared in terms of functional quality density using the Kruskal-Wallis rank sum test, which finds that there is a significant difference in at least one of the groups (p = 0.018, Chi-square = 10.04). All pairwise comparisons are made using

Mann-Whitney U tests, and the resulting p-values from these tests are shown in Table 7-51, and descriptive statistics are included in Table 7-50. The results show that the functional quality density of concepts produced by participants receiving a pruned model is significantly greater than that of participants receiving either a function model ($p = 0.003$) or an interaction model ($p = 0.036$). No other significant difference occur among the groups.

Table 7-50: Descriptive Statistics for Functional Quality Density

Treatment	Mean	Median	Standard Deviation	Observations
FM	2.302	2.111	1.529	18
IM	2.853	2.417	1.809	16
NM	3.698	2.333	2.631	18
PM	4.603	4.500	2.548	17

Table 7-51: All Pairwise Comparisons of Functional Quality Density

Contrast	Mann-Whitney p-value
IM – FM	0.324
NM – FM	0.087
PM – FM	0.003
NM – IM	0.467
PM – IM	0.036
PM – NM	0.191

The Kruskal-Wallis rank sum test is also used to test for differences among groups in terms of activity quality density. The test determines that there is a significant difference in at least one group ($p = 0.028$, Chi-square = 9.06), so pairwise comparisons are made between all groups. Descriptive statistics for these data are shown in Table 7-52, and the p-values resulting from all pairwise Mann-Whitney U tests are shown in Table 7-53. The data show that participants receiving no model ($p = 0.012$) or a pruned model (p = 0.008) produce concepts with a higher activity quality density than participants receiving a function model, using a significance level of 0.05. There are no other significant differences among the groups.

Table 7-52: Descriptive Statistics for Activity Quality Density

Treatment	Mean	Median	Standard Deviation	Observations
FM	2.392	1.667	1.207	18
IM	3.458	3.333	1.966	16
NM	3.802	3.000	2.058	18
PM	3.858	3.500	1.801	17

Table 7-53: All Pairwise Comparisons of Activity Quality Density

Contrast	Mann-Whitney p-value
IM – FM	0.078
NM – FM	0.012
PM – FM	0.008
NM – IM	0.715
PM – IM	0.413
PM – NM	0.714

7.4 Outcomes and Discussion

The results of all significant differences identified in conformance, quality, quantity, and quality density are shown in Table 7-54, and descriptive statistics for each are shown in Table 7-55. As shown in the tables, the pruned model outperforms other models in several areas, including functional conformance, functional quality, and all quality density metrics, and it appears to be the most effective model. The following conclusions are drawn based on significant differences identified in this study (see Table 7-54):

Table 7-54: Summary of Results of All Statistical Tests Comparing Treatment Groups

Metric	Category	Participant Average Scoring Approach	Participant Best Scoring Approach
Conformance	Functional	PM > FM	no significant differences
	Activity	no significant differences	FM > IM FM > PM
	Interaction	no significant differences	no significant differences
Quality	Overall	no significant differences	no significant differences
	Functional	PM > FM PM > IM	PM > FM PM > IM PM > NM
	Activity	no significant differences	no significant differences
Quantity		FM > NM FM > PM	n/a
Quality Density	Overall	n/a	PM > FM NM > FM
	Functional	n/a	PM > FM PM > IM
	Activity	n/a	PM > FM NM > FM

Table 7-55: Descriptive Statistics for All Metrics and Scoring Approaches

Metric	Category	Participant Average Scoring Approach			Participant Best Scoring Approach		
			Mean	Median		Mean	Median
Conformance	Functional	FM	0.469	0.458	FM	0.611	0.643
		IM	0.534	0.548	IM	0.661	0.714
		NM	0.515	0.500	NM	0.603	0.643
		PM	0.655	0.714	PM	0.731	0.714
	Activity		Mean	Median		Mean	Median
		FM	0.177	0.158	FM	0.319	0.25
		IM	0.098	0.000	IM	0.141	0.00
		NM	0.079	0.000	NM	0.167	0.00
		PM	0.109	0.000	PM	0.118	0.00
	Interaction		Mean	Median		Mean	Median
		FM	0.500	0.500	FM	0.632	0.500
		IM	0.500	0.500	IM	0.609	0.500
		NM	0.442	0.500	NM	0.500	0.500
		PM	0.570	0.500	PM	0.618	0.500
Quality	Overall		Mean	Median		Mean	Median
		FM	4.933	4.778	FM	5.654	5.444
		IM	5.146	5.111	IM	5.583	5.333
		NM	5.138	4.944	NM	5.494	5.556
		PM	5.502	5.222	PM	5.980	6.111
	Functional		Mean	Median		Mean	Median
		FM	4.483	4.083	FM	5.630	6.333
		IM	4.689	4.833	IM	5.167	5.000
		NM	5.154	5.000	NM	5.667	5.000
		PM	6.092	5.667	PM	6.961	7.000
	Activity		Mean	Median		Mean	Median
		FM	5.250	5.000	FM	6.148	6.333
		IM	5.767	5.778	IM	6.375	6.667
		NM	5.630	5.000	NM	6.148	5.667
		PM	5.562	5.000	PM	6.020	5.000
Quantity			Mean	Median			
		FM	3.000	3.000			
		IM	2.250	2.000			
		NM	2.000	2.000			
		PM	1.882	2.000			
Quality Density	Overall					Mean	Median
					FM	2.258	1.755
					IM	3.044	2.389
					NM	3.420	2.611
					PM	3.889	3.167
	Functional					Mean	Median
					FM	2.302	2.111
					IM	2.853	2.417
					NM	3.698	2.333
					PM	4.603	4.500
	Activity					Mean	Median
					FM	2.392	1.667
					IM	3.458	3.333
					NM	3.802	3.000
					PM	3.858	3.500

1. **Pruning a function model increases the usage of the model's functions by designers.** The pruned model results in better functional conformance than the function model using the participant average scoring approach. The pruned model contains fewer functions than the function model, and these functions describe only the functionality of the artifact rather than the actions a user performs or the interactions between the artifact and the user. This condensed description of the artifact's function made the model more useful for the designer, helping them address the functions included in the model. When used as a seed for ideation, the model was intended to guide designers toward a particular functional solution. The pruned model did a better job of directing designers toward the desired functions than the function model. It is interesting that there is no significant difference between the control group and the pruned model group, indicating that the pruned model may not actually direct designers toward a particular solution. However, there is a large difference in the means and the medians in these two groups, with the pruned model performing better than no model (see Table 7-55). Due to the high MSE, the power of the test is low, so it is possible that a difference does exist but is not detected by this study.
2. **Pruning a function model reduces the usage of activities in the model.** The function model results in higher activity conformance than the pruned model using the participant best scoring approach. This result seems intuitive, since pruning removes the user activities described in the model. However, it identifies an advantage of including activities in a model due to their ability to direct

designers toward activities that can be used to solve the design problem. The interaction model, like the function model, contains user activities described in an active manner, but the function model outperformed the interaction model as well. This result was not expected and points to difficulties in the activity conformance metric. To ensure high reliability, the activity metric was strict, allowing for little interpretation by the rater. The metric required that the sketch explicitly contain a description of a user performing an activity or a drawing of a user performing an action. Many of the designers probably intended for a user to perform some actions but did not explicitly state it. Since the interaction model explicitly states that a user is performing certain activities, the designers were probably less likely to explicitly include the information in their sketch, since the information would be redundant with the model. This could have caused the interaction model to perform poorly on this metric. Follow-up interviews with participants would have been helpful to clarify this issue, but were not performed in this study.

3. **Pruning a function model increases the functional quality of ideas generated.** The pruned model results in higher functional quality than the function model using both scoring approaches. The pruned model, therefore, is useful to designers, helping them generate ideas that satisfy the functional requirements well. The activities and interactions in a function model are likely taking some of the designer's attention away from the artifact's function. The activity quality, however, is not significantly improved by the function model, so the diverting of the designer's attention away from the function of the artifact is not useful. If the

activities were modeled in a way that the activity quality is improved, then the tradeoffs would need to be explored. However, this is not the case as the activity quality is not improved.

4. **A pruned function model increases the functional quality of ideas compared to no model.** The pruned model resulted in higher functional quality compared to no model using the participant best scoring approach. In this study, the intent of the function model is to help a designer understand a system's functionality and generate ideas based on the desired functionality. Pruned models, therefore, can be used in conceptual design to improve the functional quality of ideas generated.
5. **The inclusion of activities in an interaction model reduces the functional quality of ideas generated.** The pruned model results in higher functional quality than the interaction model using both scoring approaches. As previously mentioned, the activities in the interaction model are likely diverting the designer's attention away from the artifact's function without improving the activity quality. However, the interaction model shows promise that it may help improve the activity quality. The effect may be small and is not detectable with this study, but the interaction model results in the highest mean and median activity quality using both scoring approaches. The modeling of activities, therefore, may be useful to a designer and should be pursued using alternative modeling approaches or a more fully developed activity model.
6. **Function models increase the quantity of ideas generated.** The quantity of concepts generated by designers using the function model is greater than that of

designers with a pruned model. The function model contains descriptions of passive functions and interactions, but it does not specify whether the user or the system will accomplish these activities. This ambiguity gives freedom to the designer to generate concepts in which the user performs the passive functions as well as concepts in which the artifact performs the functions. The pruned model does not include these passive functions, so designers likely do not generate alternative solutions for them, reducing the total number of concepts generated by the designer. Function models also increased the number of concepts generated compared to designers without a model. The reason for this finding is not known, but a possible explanation is that the function model stimulates the concept generation process in designers without restricting them to a particular solution. The inclusion of ambiguous, passive functions stimulates additional ideas. However, this type of ideation is not the focus of this research, which is to direct a designer to a high-quality, functional solution.

7. **Pruned models improve concept generation efficiency.** Pruned models increase the overall, functional, and activity quality density compared to function models, and they increase the functional quality density compared to the interaction model. The pruned model results in fewer concepts than the function model and increases the functional quality and does not reduce the overall or activity quality of the concepts. This results in a significant increase in the quality density of ideas. The pruned model yields greater or equal quality at a lower cost (number of concepts), so the pruned model increases the efficiency of concept

generation (overall, functional, and activity) over the function model. The pruned model also is more efficient functionally than the interaction model, which is expected since the pruned model does not include activities that capture some of the designers' attention and time.

8. **Function models reduce concept generation efficiency.** The overall and activity quality density of the function model is significantly less than that of no model, and the difference in functional quality density is close to significant ($p = 0.086$). Designers using a function model were less efficient in generating quality concepts than those using no model, indicating that the function model hinders the efficient generation of ideas. The function model increases quantity of concept without increasing the quality, so the additional concepts are not very valuable in the concept generation process.

Overall, the pruned model provides functional direction for designers, improving the functional quality of ideas generated compared to the function model. The pruned model also results in a more efficient concept generation process compared to the function model. The interaction model did not perform better than other models, but it shows potential for improvement in activity quality. While the results are not significant, the mean and median activity quality resulting from the interaction is greater than all other groups. The way in which activities are modeled—using an activity model discussed in [4]—has not been studied extensively in this research. The separation of activities and functions shows promise for improvement in functional quality, so a more effective activity model within the interaction model should be identified and tested.

There are only a few instances in which designers without a model were outperformed by designers with a model, suggesting that function-based models in conceptual design may not be useful as ideation seeds. While possible, there are two reasons that models should still be pursued as ideation tools. First, the high amount of variation within groups allows for only large differences among groups to be detected by this study. Therefore, significant differences may exist when they were not found. In order to state with confidence that there is no difference, either the sample size must be increased or a new experiment design must be used. The increase in sample size is not practical due to the number of samples required (see Section 7.3.3), so a new experiment design would be the better approach. The new experiment design should better model the variation among participants using covariates, blocking factors, or other approaches. Second, the quality of ideas resulting from the models are dependent on the content of the models. The models provided to designers represent one functional approach to the problem, and the quality of this idea was not assessed. It is possible that the ideas contained in the models were of low quality, resulting in participants deviating from the ideas in the model or the model reducing the quality of the concepts. The quality of the model likely influences the quality of concepts generated using the model, so alternative models with different working principles should be explored in future studies before eliminating the use of representations as a seed for ideation in conceptual design.

CHAPTER 8: CONCLUSIONS AND RESEARCH OPPORTUNITIES

8.1 Conclusions

The three tasks performed for this research collectively address the overall research question and the five sub-questions. The answers to each of these five questions, discussed in Section 8.1.1, constitute the technical contributions of this research. The outcomes of the individual tasks also provide insight into the research methods used, and the contributions and lessons learned by conducting this research are discussed in Section 8.1.2.

8.1.1 Technical Contributions

The overall question addressed in this research is:

> **Overall Research Question:** How should the functionality of mechanical artifacts be modeled to support ideation in conceptual design?

Three tasks—an interpretability user study, a similarity study, and an ideation user study—were performed to address five sub-questions. The research questions are answered based on the outcomes of the tasks performed. The relationship between the research questions and tasks is shown in Table 8-1.

Table 8-1: Research Questions and Supporting Research Tasks

	Research Question	Task 1: Interpretability	Task 2: Similarity	Task 3: Ideation
Overall	How should the functionality of mechanical artifacts be modeled to support ideation in conceptual design?	✓	✓	✓
RQ1	How well do designers understand and use functional representations in conceptual design?	✓		✓
RQ2	In what ways do pruned function models support ideation?	✓	✓	✓
RQ3	In what ways do interaction models support ideation?			✓
RQ4	How well do functional representations support internal search for solutions in conceptual design?	✓		✓
RQ5	How well do functional representations support external search for solutions in conceptual design?		✓	

8.1.1.1 Designer Understanding and Use of Functional Representations (RQ1)

The first research question, "How well do designers understand and use functional representations in conceptual design?" is first addressed through the interpretability user study (Task 1). The interpretability study tested the effects of the language (Functional Basis or free language) and type of functions (reverse-engineered or pruned) on a user's understanding of models (see Chapter 5). The study shows that interactions and component-specific functions do not improve a user's understanding of the model. When these types of functions are pruned from the model, the level of understanding (i.e., interpretability) is unaffected. Further, the use of free language within a model greatly increases the level of understanding of the model because free language terms contain context that helps the user. Therefore, functional representations used by humans in conceptual design should include context through free language terms as well as high-level conceptual functions (i.e., functions remaining after pruning the model) to ensure a high level of understanding of the model. A high level of understanding will be beneficial for both communication within design teams as well as model creation in conceptual design.

This research question is also addressed through the ideation user study (Task 3), which tests the usage of models in ideation (see Chapter 7). The conformance metric developed to evaluate sketches tests how well the ideas contained in a sketch align with the ideas in the model. The study shows that designers use the functions in a pruned model more than a function model for ideation. Since the pruned model contains fewer, more-active functions than the function model, it is more useful to designers. However, the activities in the function model were used by designers more than the pruned model or interaction model. This may be a result of the strict conformance scale created rather than a true outcome, but it is possible that the function model is more useful for modeling activities than the pruned model or interaction model. No other significant differences in usage of a model were identified through this study, including differences between the baseline group (no model), and other groups. The study is limited to detection of large effect sizes, so there may be medium or small differences between these groups that are not detected by the study. Therefore, the usage of these models by designers should be further investigated.

Based on these two tasks, pruned models are the most useful to designers for conceptual design since they are easy to understand and they improve the usage of functions within the model.

8.1.1.2 Advantages of Pruning for Ideation (RQ2)

The second research question, “In what ways do pruned function models support ideation?” is addressed through all three research tasks. The first task, the interpretability study, shows that the pruned model is a more efficient conceptual representation of an artifact (see Chapter 5). Here, efficiency is defined as a benefit-to-cost ratio. In terms of interpretability, efficiency is the accuracy of interpretation (benefit) compared to the speed of interpretation (cost). The pruned model is much faster to interpret with the same level of accuracy as a function model, so it is more efficient in conceptual design. This more efficient representation reduces the time required to understand a model and generate ideas for a new design problem, allowing faster idea generation and/or more ideas to be generated in the same time frame, leading to higher quality solutions.

The second task, the similarity study, also addresses this research question by testing the appropriateness of the level of abstraction achieved through pruning. This task shows that the pruning rules convert a reverse-engineered function model into a consistent, conceptual-level description that is more precise and accurate for similarity calculations (see Chapter 6). A more precise and more accurate similarity metric will result in better seed examples (accuracy) and fewer poor seed examples (precision) in a design-by-analogy method, saving a designer time sorting through the results of the metric. The pruned representation, therefore, is more efficient for identifying similar artifacts that can be used as seeds for ideation in conceptual design.

The third task, the ideation user study, addresses this research question by comparing the quality of ideas generated from the pruned model to other representations (see Chapter 7). The pruned model significantly increases the functional quality of ideas generated by designers. The pruned model is an efficient representation of functionality and is easily understood by designers, so a designer using the model can quickly generate high quality concepts. The pruned model also increases the quality density of concepts, making it a more efficient representation than other models. Quality density is defined as the quality of the best concept created by a participant divided by the number of concepts generated by that participant, or a benefit (quality of the best concept)-to-cost (number of concepts that must be evaluated) ratio. The pruned model, therefore, is more efficient for concept generation by designers.

Overall, the pruned model is an efficient representation in terms of designer understanding, similarity, and idea generation. Pruned models are created from a reverse-engineered description of an artifact, and pruning rules specify the removal of functions from a model. The pruning rules, therefore, should be inverted to describe what should be modeled at the conceptual design stage for new designs. This topic is discussed in more detail in Section 8.1.1.6.

8.1.1.3 Advantages of Interaction Models for Ideation (RQ3)

The third research question, "In what ways do interaction models support ideation?" is addressed through the ideation user study, which compares the usage and quality resulting from the interaction model to other representations (see Chapter 7). The study does not show a significant improvement in performance by the interaction model over other models, but the study is only able to detect large differences between groups. There are smaller effects that indicate that the interaction model may increase the activity quality of concepts generated, so the interaction model should be further studied.

The interaction model is still in development, and the activity model that was integrated with the pruned model should be further explored. There are other representations for modeling users, processes, or tasks that may be compatible with the function structure and more effective in modeling the actions a user performs. Further exploration of modeling user actions and incorporating them with the interaction model is discussed in Section 8.2.1.

The integrated model of functions and user actions can support ideation within the parallel function- and interaction-based design approach (see Section 4.1), where the interactions and functions of an artifact can be pursued simultaneously in a single model. Initially, the approach anticipated a representation of interactions separate from functions (see Figure 4-2), but the latest iteration of the interaction model incorporates both functions and user actions and the interactions between them. Therefore, a single representation is used to model both paths in the parallel process, and the parallel paths focuses on functions and user actions, as illustrated in Figure 8-1. The interaction model and design approach may be useful for ideation on a function-user continuum, as discussed in Section 8.2.2.1.

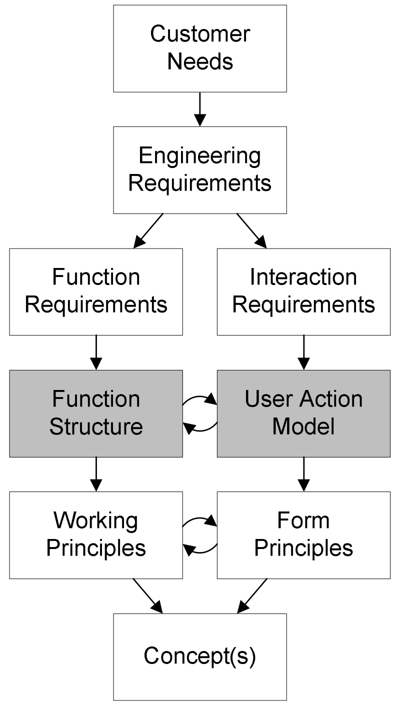


Figure 8-1: Parallel Function- and Interaction-based Design Approach Showing the Location of Application of the Interaction Model

8.1.1.4 Use of Functional Representations for Internal Search (RQ4)

The fourth research question, "How well do functional representations support internal search for solutions in conceptual design?" is addressed by the interpretability and ideation user studies. The interpretability user study (Task 1) shows that pruned models with free language are easier for designers to understand. Therefore, when using function models for internal solution search, a designer should use a the pruned model with free language since it supports a quick understanding of the model. Since the designer understands the model quicker, the concepts developed can more easily be verified to ensure that they meet the functionality described in the model, resulting in more thorough concepts that better address the functionality. The function model is more difficult to interpret, so designers using this representation for ideation will take longer to begin generating ideas and will iterate slower as they verify that their concept addresses the functions.

The ideation user study (Task 3) addresses this research question by comparing the conformance and quality of concepts generated using different models to concepts generated without a model. The study shows that using a pruned model significantly increases the functional quality of concepts, and the pruned model group outperforms the control group in many other categories, but the effect size is too small to detect with significance in this study. The study provides evidence that the pruned model supports ideation in many ways, but further studies must be conducted to support these conclusions statistically. This study also shows that the function model reduces overall and activity quality density, suggesting that the function model hinders concept generation in terms of efficiency. The function model does not reduce the overall quality, but it causes more concepts to be developed without increasing the overall quality. The usefulness of the interaction model for internal solution search is not significant, so this representation may not be useful for ideation in conceptual design. Therefore, the pruned model supports ideation, the function model does not support ideation, and the interaction model may or may not support ideation in conceptual design.

The ideation study tested the use of a single model for each functional representation as a seed for ideation in conceptual design. The use of multiple functional solutions to a design problem may lead to more high-quality concepts. If designers are creating the models and generating ideas from them, the pruned representation may allow faster generation of models and exploration of solutions since they are easier to understand and increase the functional quality and quality density of concepts.

8.1.1.5 Use of Functional Representations for External Search (RQ5)

The fifth research question, “How well do functional representations support external search for solutions in conceptual design?” is addressed through the similarity study (Task 2), where the pruned model is compared to the function model with and without supporting functions. The similarity study shows that the pruned model best supports the similarity metric since it results in more accurate and precise similarity calculations. The similarity metric is useful for design-by-analogy methods, which is one type of external search for solutions. The pruned model, therefore, best supports this external search compared to the function model, either with or without supporting functions. The pruned model is a subset of the interaction model, so the interaction model could also be used to identify functionally similar artifacts as an external search for solutions. The interaction model also has potential to search for similar artifacts based on activities, interactions, or any combination of functions, activities, and interactions. The interaction model supports external solution search functionally and may support external solution search based on activities and interactions, but these metrics have not yet been developed.

8.1.1.6 Functional Representations in Conceptual Design

The five sub-questions support the overall research question, “How should the functionality of mechanical artifacts be modeled to support ideation in conceptual design?". Based on the results of the three tasks, the pruned model is an efficient representation for ideation in conceptual design. The pruned model is easy to understand, supports conceptual-level similarity, improves the functional quality and overall, functional, and activity quality density of ideas generated by designers using the model. Thus, mechanical artifacts should be modeled using the pruning rules as guidelines for creating conceptual-level models. The function model does not represent interactions and user activities in a way that is useful for ideation in conceptual design, and the interaction model provides an alternative manner that does not hinder ideation compared to a pruned model. Therefore, the interaction model should be further developed for application to conceptual design, and the pruning rules should be inverted to provide guidelines for creating conceptual-level models.

The pruning rules were developed for models that use the Functional Basis vocabulary. Since the interpretability study shows that free language should be used to ensure that designers understand the model, the modeling guidelines should be based on another vocabulary or other modeling principles rather than the Functional Basis. These principles have not yet been identified, but a formal physics-based modeling approach developed by Sen [27] shows promise that it will support formal modeling guidelines for conceptual design.

While the goal of this research is to identify formal modeling guidelines, as in the case of the pruning rules, several general modeling guidelines are presented based on the outcomes of this research. These potential guidelines serve only to demonstrate the potential to invert the pruning rules and to use Sen's work to formalize conceptual modeling guidelines. Further, the pruning rules have been tested as a set rather than individually to first develop confidence that the rules are useful. Each of the guidelines developed through inverting the pruning rules should be individually tested to understand its effectiveness and appropriateness for conceptual design activities. It is important to note that function modeling guidelines are discussed in design texts [1-4] but they serve only as general guidelines rather than formal guidelines for creating function models. The guidelines presented below are intended to be formalized in the future based on a formal representation of function, such as the representation presented in [27].

Potential Guidelines for Creating a Conceptual-level Model

- **Model active functions.** Active functions require that the energy used to perform the function be carried by the artifact being modeled. Passive functions, which are performed to or on an artifact, should not be modeled. If a designer wishes to include passive functions, an alternative representation, such as the interaction model, should be explored. This guideline is based on the discussion of active functionality, user actions, and interactions (see Section 2.1.3) and addresses Pruning Rule 1, "Remove all *import* and *export* functions." The functions *import* and *export* are typically passive and describe only interactions of an artifact with its environment. It is possible that these functions describe active functions if they require energy to be performed (e.g., a pump *imports* water into a system), so this modeling guideline better describes the intent of this first pruning rule.
- **Model flows of artifacts only if the function of those artifacts is not in the model.** This modeling guideline is based on Pruning Rule 3, "Remove all *couple*,

join, and *link* functions referring to any type of *solid*" and Pruning Rule 4, "Remove all *support*, *stabilize*, *secure*, and *position* functions," which describe assembly relationships within an artifact. For example, it is appropriate to model the flow of an artifact, such as a battery, in an assembly process since the function of the artifact is not described within the function model of the assembly process. However, the artifact flow of a battery within a function model of a drill is not appropriate if the functionality of the battery is also included in the model. The assembly process should be modeled in a complementary model.

- **Model conduction and radiation of energy as a flow.** Sen decomposes the transfer of all types of energy into conduction, convection, and radiation [27], where conduction of any type of energy does not require net displacement of the material through which the energy flows and radiation does not require any material medium [27]. This guideline describes the intent of Pruning Rule 2, "Remove all *channel*, *transfer*, *guide*, *transport*, *transmit*, *translate*, *rotate*, and *allow DOF* functions referring to any type of *energy*, *signals*, or *human material*." When these functions are used to describe *energy* and *signals*, they typically describe conduction processes (e.g., *transfer electrical energy*). The formalization of conduction and radiation [27] supports a more formal modeling guideline that is based on the physical principles described in the model rather than the vocabulary used. Additionally, conduction and radiation of energy are passive changes in the location of energy that can be represented by a flow rather than a function to improve the level of understanding by designers.

- **Model convection of energy as a function.** Sen defines convection as the transfer of energy through bulk movement of a material flow [27]. This modeling guideline addresses Pruning Rule 2, "Remove all *channel, transfer, guide, transport*, *transmit*, *translate*, *rotate*, and *allow DOF* functions referring to any type of *energy, signals,* or *human material*." in conjunction with the previous modeling guideline. When energy is transferred through convection, a material flow must be transferred within the artifact to carry the energy, and Pruning Rule 2 does not specify the removal of functions describing material transfer (unless it is *human material*, which is outside the scope of this discussion). This bulk material movement is important to consider in conceptual design, so it should be included in a conceptual-level model.

Therefore, mechanical artifacts should be modeled using guidelines developed by inverting pruning rules and using a formal functional representation to describe them. Non-functional or passive aspects of an artifact should be described in a complementary model such as the interaction model. However, this representation is still being developed and has not been proven to be more useful than a pruned representation.

8.1.2 Contribution to Design Research Methods

This research explores the use of functional representations in conceptual design and validates their use using two main approaches analogous to medical research validation, as discussed by Frey and Dym [80]. The first validation approach, used in Task 1 and Task 3, is a laboratory experiment with human subjects, analogous to in vitro experiments in medicine. The two experiments test the use of functional representations for interpretability and ideation using human subjects in a controlled lab setting. There is risk associated with testing in a lab setting since it is different from industry practice, where designers work as part of a team and may be intimately familiar with a domain. The experimental results from students performing design tasks in a controlled setting may not be representative of designers in industry, so generalization to industry may not be appropriate. However, generalization to a student population is appropriate, and it is beneficial to understand the usefulness of design tools for students. If these tools are found to be useful for students, then further tests can be conducted in industry to understand their benefits in industry. Further, these lab experiments can provide insights into proper experimental procedures, which can be applied to the design of controlled experiments for industry. Lab experiments, therefore, are an appropriate step toward validation since controlled studies in industry present many challenges and can be expensive.

The second validation approach, used in Task 2, is a detailed simulation of a design method, analogous to animal models in medicine [80]. The similarity study simulates a designer searching for artifacts that are functionally similar to a model of a new design problem. The "new" design problem in this study is a set of function models with an known level of similarity to a group of artifacts. The similarity of the new design problems to the group of artifacts is determined and the results are compared for different representations. This use of a simulation to validate the representation is appropriate since a computational similarity metric and repository of knowledge exists, and the metric has been shown applicable to design-by-analogy.

The ideation user study (Task 3) is a thorough example and documentation of a completely randomized design applied to ideation in conceptual design. The study includes the development of quantitative metrics to evaluate qualitative data, ensuring the reliability of the ratings through interrater agreement, developing hypotheses, calculating sample size, and checking model assumptions. The desire of these researchers is that this study would help inform other researchers desiring to validate design tools and methods through controlled laboratory experiments. The lessons learned from this experiment can be used to improve future experimental designs, so a few of the practical lessons learned from the ideation user study are discussed:

- A pilot study should be conducted to obtain sample design concepts.
    - A reliable rating scale should be developed from the concepts in the pilot study. A scoring reference sheet with examples should be created and two (or more) raters should independently evaluate concepts. The ratings should be compared using interrater agreement metrics, such as Cohen's Kappa value. If the level of agreement is not acceptable, then the differences in ratings should be discussed among raters and the rating scale should be refined and the process repeated until acceptable levels of agreement are achieved.
    - The mean squared error (MSE) should be estimated from the pilot study and used to predict the sample size.

- MSE using senior-level students at a single university is large in a completely randomized design. This design is only able to detect large differences among groups, and is not powerful for detecting small or medium effect sizes. Other designs should be explored to reduce MSE if large differences among treatment groups are not anticipated.
- Concept scores for conformance and quality were based on the average of several categorical or binary ratings. For example, the functional conformance score was the average of seven binary function ratings. The resulting scores were approximately normal, but were some problems with activities, since many of the participants scored poorly on activities. The rating scales, therefore, should not be too strict or too generous, resulting in many concepts receiving the same score, especially since the goal of the rating system is to separate concepts. Additionally, when only a few individual ratings are averaged into a score, the data tend to be more discrete and there are many ties in the data. A different experiment design may alleviate some of this problems if it includes covariates or another predictor of the response.

8.2 Research Opportunities

8.2.1 Development and Testing of Functional Representations

Three representations of function—the function model, interaction model and pruned model—have been evaluated in this research and compared to a baseline of no model. The studies conducted show that the pruned model is more useful for conceptual design than the function model, but it is not known if the pruned model is more useful than the interaction model or no model. Therefore, the interaction model and pruned model should continue to be developed and tested, since the interaction model may be a useful way to model non-functional requirements, which are not captured in the pruned model and are improperly captured in the function model. The research question that will be explored is:

RQ: How can user actions and interactions be modeled to support ideation in conceptual design?

8.2.1.1 Development of the Interaction Model

The activity model representation [4] was integrated with the pruned model in this research to create the interaction model. The usefulness of the activity model is not known, and did not significantly improve the usefulness of the model for conceptual design. Therefore, this representation and alternative representations of user actions should be investigated to understand their usefulness for conceptual design. Other representations, such as task or process models, may be adapted and integrated with the pruned model to create a useful representation of artifact function, user actions, and interactions between users and artifacts. These models of user actions can be studied independently of function models before integrating them with functional representations. Therefore, the research question pursued is:

RQ: What representation is appropriate to integrate with pruned models to capture user actions?

After identifying an appropriate user modeling method and developing the interaction model, the model should be formalized with a grammar and method for creating the models. The function modeling formalization developed by Sen [27] can likely be used within the functional portion of the interaction model (see Section 8.1.1.6), but user and interaction modeling must also be formalized as each is integrated with the pruned model. Once formalized, interaction modeling methods can be created and tested for repeatability among designers.

8.2.1.2 Testing of the Pruned and Interaction Models

Since the ideation user study conducted in this research is exploratory and since significant differences between the pruned model, interaction model, and no model were not frequently identified, these models should continue to be tested for their usefulness in conceptual design. The research question pursued is an extension to RQ2 and RQ3:

RQ: In what ways to pruned models and interaction models support ideation in conceptual design?

The following areas for further testing have been identified to address this research question:

- **Revise the experiment design:** The ideation user study was effective for identifying large differences among models, but smaller effects cannot be identified due to the high variation among participants. Therefore, a more effective statistical model and experiment design should be identified and used in future studies to reduce sample size and provide better detection of smaller effects.

- **Consider other experimental approaches:** The quantitative experimental approach is effective for identifying differences in the representations, but the reasons for these differences cannot be identified with this approach. Qualitative methods may be more effective, especially during the development of the representations, for understanding how and why designers use the representations for ideation. These qualitative studies may require smaller sample sizes and may provide more insights for the development process so iterations on the representations can be faster and more effective.
- **Test additional factors:** The ideation user study tests a single functional solution to the burrito-folding design problem using three different representations to model the functional solution. There are many different functional solutions that could be used to solve the design problem, ranging from a user-centered solution where a human performs all activities in the burrito folding process to an artifact-centered solution where the process is completely automated. The effect of different models, each containing a different solution to the problem, should be studied using both the pruned model and interaction model, and no model as a baseline. The effect of the representations on ideation may be dependent on the solution described by the model, and different representations may be more appropriate for different types of solutions. For example, an artifact-centered

solution may have few interactions between the artifact and user, so an interaction model may not be appropriate for this type of solution.

- **Test a broader set of participants and design problems:** The ideation study is limited to a single design problem and participants are all senior-level mechanical engineering students at Clemson University. The representations should be tested across broader participants and design problems for broader generalization.

8.2.1.3 Testing of the Modeling Process

In the ideation user study, participants were provided a model that was used as a seed for concept generation. Designers may not typically be provided with a model; they may create the model before using it to generate ideas. The process of creating the model may provide more benefit to the designer than the actual model itself, since the designer must understand the problem and identify a functional approach to the problem as he or she creates the model. The modeling process may help the designer understand and define the problem through decomposition, and the resulting model may be of less benefit to the designer than the insights gained through the exercise.

If the modeling process is more useful than the model itself for ideation, then the representation used to model functionality may not affect the outcome of ideation when the designer generates the model. Further, if a designer creates a model, he or she will know the intent of each element in the model, so the level of understanding of the model by other designers should not affect the outcome of ideation through internal search by the modeler(s). Therefore, the modeling process should be further studied to understand the usefulness of functional representations in conceptual design. The following two research questions can be pursued:

RQ: How does the model development process affect ideation in conceptual design?

RQ: Does the functional representation used to create a model affect the information gained by a modeler?

8.2.2 Integrated Function- and Interaction-based Design Methods

Many systematic design methods prescribe a function-first approach, but in this research a parallel function- and interaction-based approach is proposed (see Section 4.1). The interaction model may be a useful tool within this approach, but methods for using this representation must be developed to support designers.

8.2.2.1 Model Generation Methods on a Function-User Continuum

Since the interaction model contains both artifact functionality and user actions, models can be generated for a design problem on a function-user continuum. For example, if a design problem requires that a user cut curved shapes out of wood, the designer could generate a range of solution ideas (see Figure 8-2).

Figure 8-2: Example of Solutions to Wood-Cutting Problem on a Function-User Continuum (image sources, left to right: sears.com, sears.com, sears.com, rockler.com)

At the user end of the continuum, the user would perform many actions and designed artifact would perform only a few transformative functions (if any). A coping saw would lie at the user end of this continuum. At the functional end of the continuum, the designed artifact would perform many functions and the user would only perform a few actions. In the wood-cutting problem, a CNC wood cutting machine would lie at the functional end of the continuum. Interaction-based design methods could be created around this continuum, encouraging designers to create a broad range of models that can be used for concept generation.

RQ: How does model creation on a function-user continuum affect the quality of ideas generated by designers?

8.2.2.2 Model Evaluation Methods

If a designer uses this interaction-based design approach to develop many models on a function-user continuum, then the designer may generate a large number of concepts from the models. The designer may then evaluate many concepts, reducing them down to a few good concepts to pursue for the final design. If, however, the models rather than the concepts could be evaluated for performance, then the designer could evaluate and eliminate models before generating concepts from each model, reducing the effort required by the designer. Therefore, interaction model evaluation metrics should be developed for comparing models and selecting the models that will lead to the best solutions.

RQ: Can models, rather than sketches or concepts, be effectively evaluated in conceptual design to identify high-quality ideas?

8.2.2.3 Computational Tools to Support Ideation

After creating an interaction model of a new artifact, designers may identify potential solutions to functions, user actions, or interactions through an internal or external search [2]. Internal searches rely on the designers' knowledge and/or experience for idea generation, while external searches require designers to look for existing solutions to the problem or similar problems. New tools may be developed based on the parallel design approach that better support ideation.

RQ: What new computational tools can be created to support external search for solutions in conceptual design?

A formal similarity metric may enable designers to generate an interaction model, compare it to models of existing artifacts, and use the identified similar artifacts as a source for ideas. Function-based similarity metrics exist (see Chapter 6), but user action and interaction similarity metrics must be developed and integrated with function similarity metrics for a complete comparison of all aspects of the artifact captured in an interaction model. This hybrid similarity metric may enable designers to identify artifacts with similar functions, user actions, and/or interactions, supporting concept generation that addresses not only the artifact's function, but also user actions and interactions with users and other artifacts.

RQ: How does a hybrid similarity metric support ideation in conceptual design?

The complete similarity metric will require identifying or developing an interaction model vocabulary, grammar, and method for achieving a consistent level of abstraction.

8.2.3 Design Method Validation

Controlled user studies are gaining popularity in the design research community and are used to evaluate design tools and methods. Most of these user studies are performed in classroom settings with undergraduate students, which may not be representative of designers in industry. The similarities and differences between students and designers in industry should be studied to understand if user studies can be used to validate design methods targeted at industry. These user studies could be conducted with both designers in industry and students as participants and the results compared. If the design outcomes are consistent between industry designers and students, then these types of experiments will be more useful to researchers, allowing for generalization from students to designers in industry. Thus, the following research question is identified:

RQ: Are students appropriate participants for validation of design tools and methods?

User studies are a logical method for design method validation, but they present many challenges due to the nature of design and human subjects. Research methods using human subjects should continue to be explored in other areas and applied to design. Frey and Dym [80] suggest that medical research methods be applied to design research, but there are differences between medical research and design research that will allow for different experimental designs. For example, in design research, participants are given a design problem that is controlled by the researcher, and the participant could be given multiple design problems. In medical research, participants must be found that already have a condition that is being studied. Therefore, repetition within participants is not possible in medical research, but it is possible in design research. Design is also performed by teams in many cases, and medical research methods may not be relevant for testing groups. Therefore, research methods should be explored from many other areas and applied to design. The following two research questions can be pursued to further research in validation techniques:

RQ: Are user studies appropriate for validating design tools and methods?

RQ: What other techniques, within or outside the design community, are appropriate for validating design tools and methods?

APPENDICES

<u>APPENDIX A</u>: EXPERIMENT PACKETS

Each participant in the ideation user study was provided with: (1) a problem statement, (2) either a function model, interaction model, pruned model, or no model, and (3) five sketching sheets. The problem statement and models for each treatment group are shown in Figure A-1, Figure A-2, Figure A-3, and Figure A-4. The sketching sheet provided to all treatment groups is shown in Figure A-5.

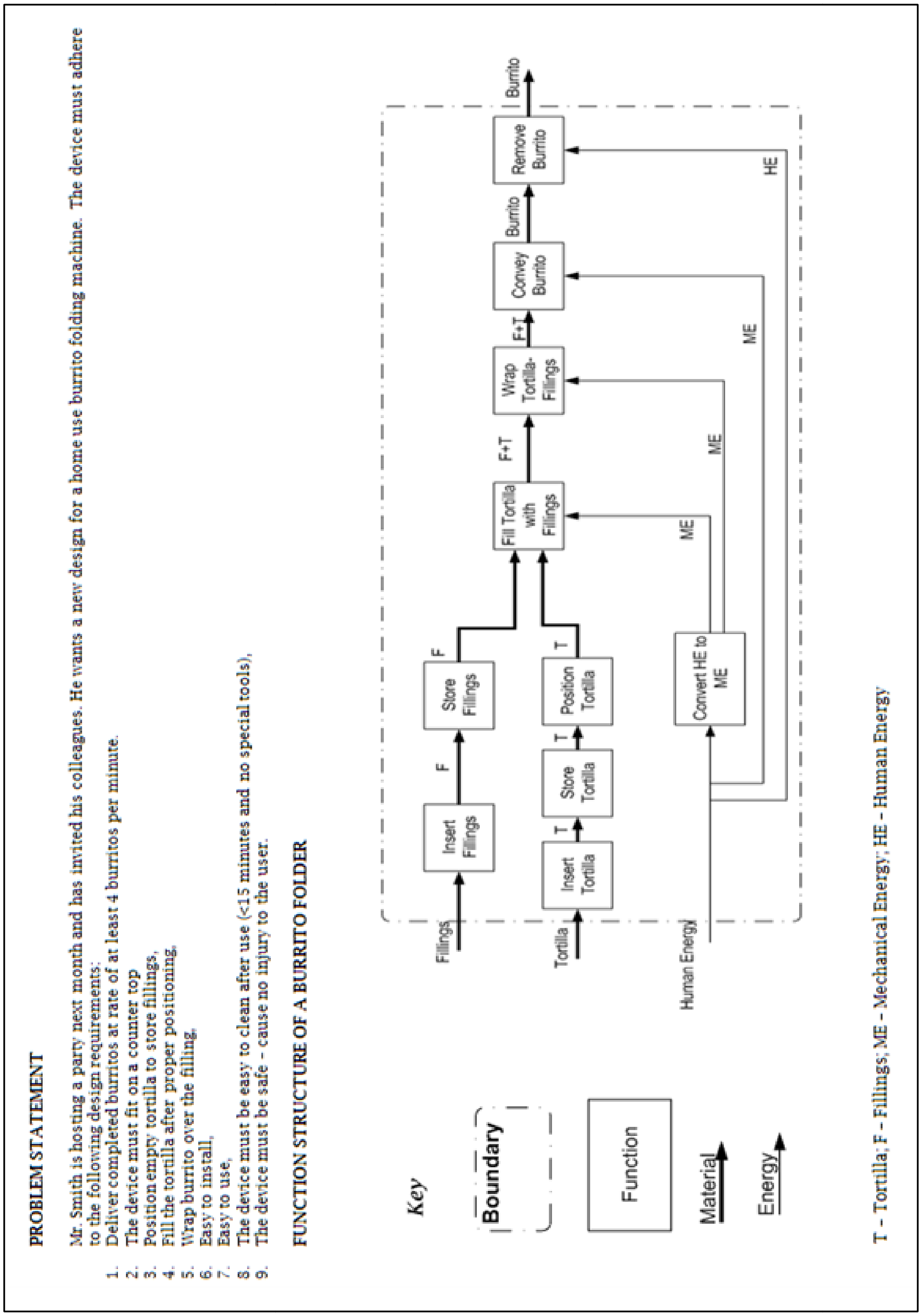


Figure A-1: Burrito Folder Problem Statement and Function Model

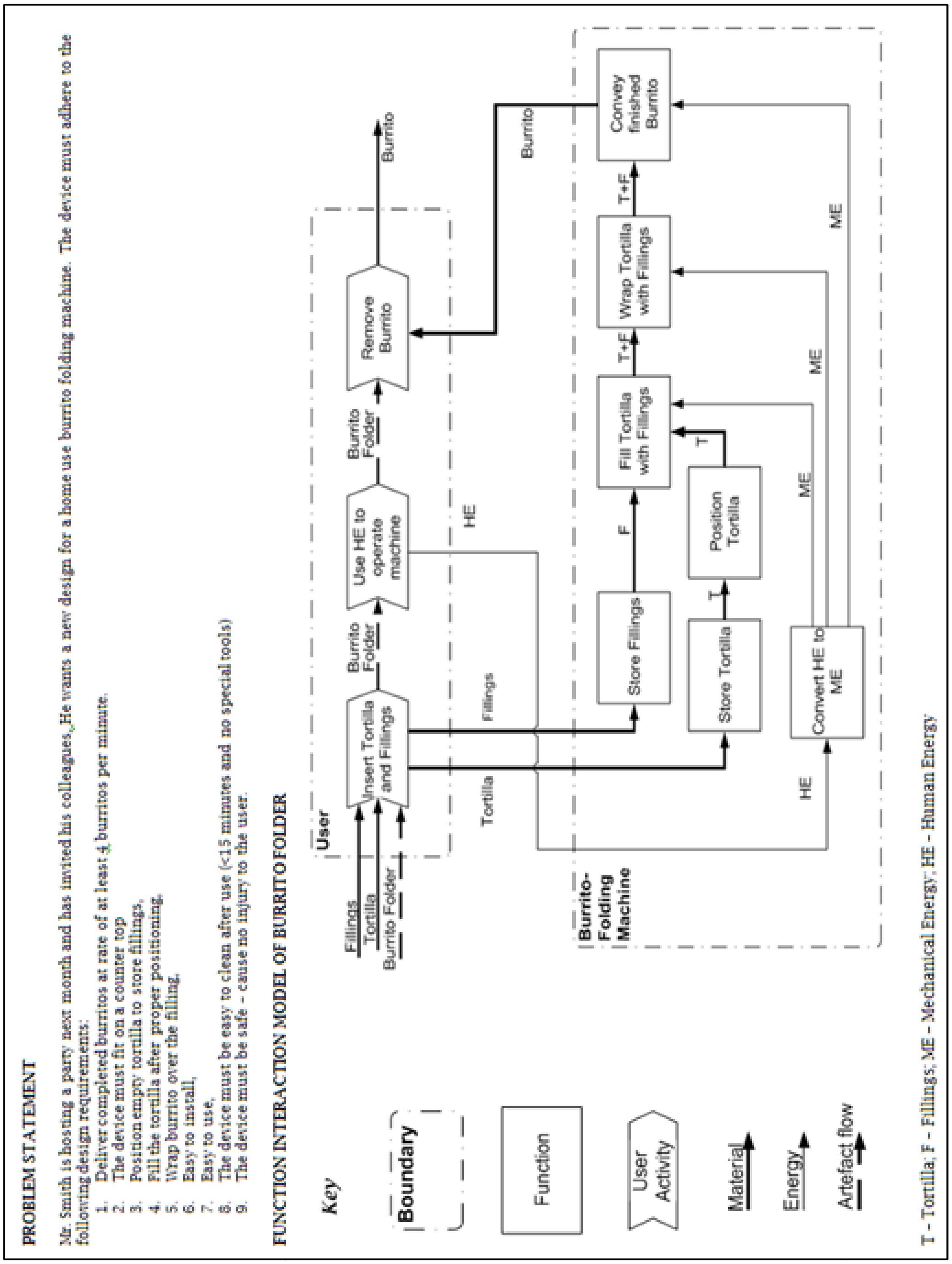


Figure A-2: Burrito Folder Problem Statement and Interaction Model

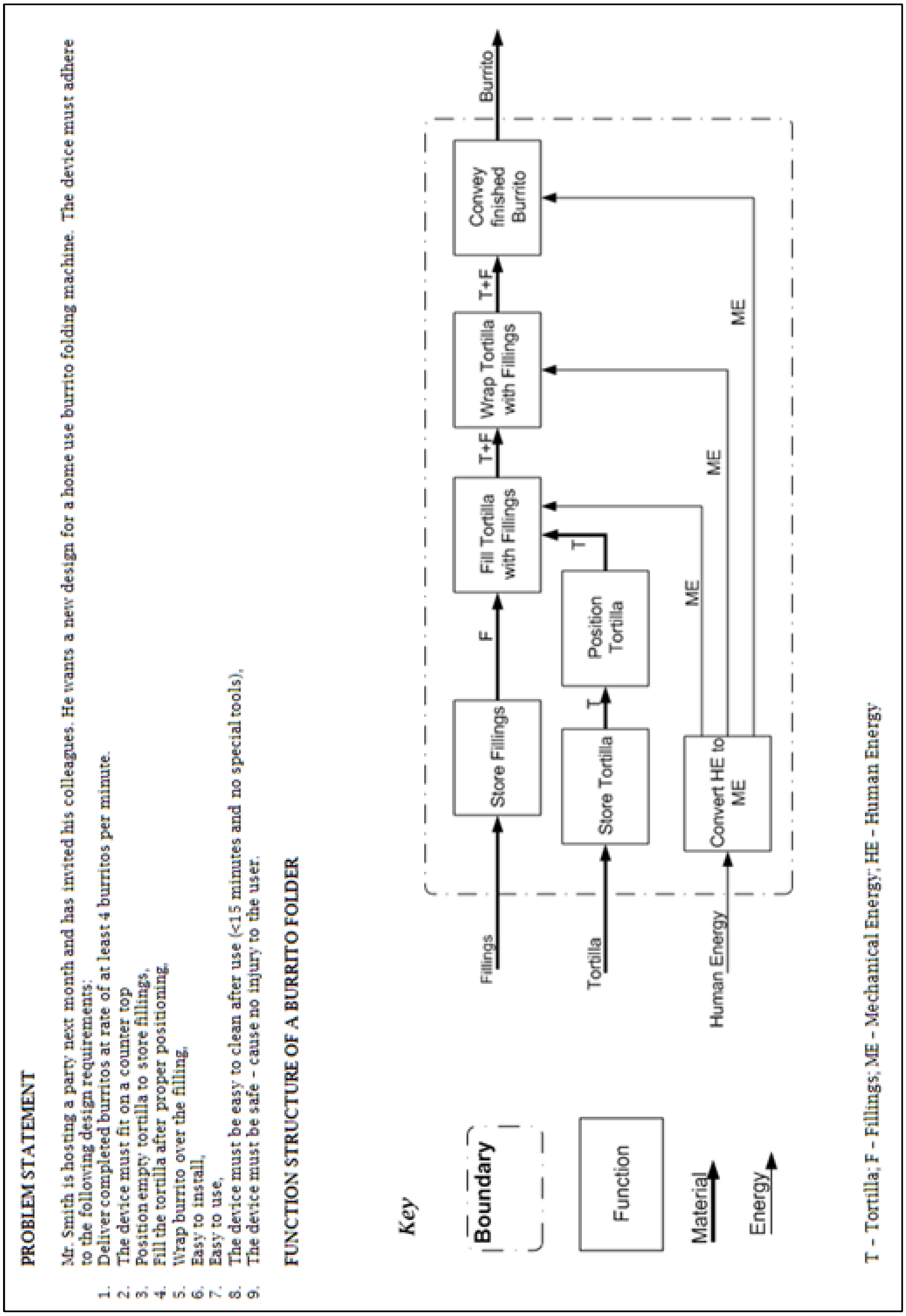


Figure A-3: Burrito Folder Problem Statement and Pruned Model

PROBLEM STATEMENT

Mr. Smith is hosting a party next month and has invited his colleagues. He wants a new design for a home use burrito folding machine. The device must adhere to the following design requirements:

1. Deliver completed burritos at rate of at least 4 burritos per minute.
2. The device must fit on a counter top
3. Position empty tortilla to store fillings,
4. Fill the tortilla after proper positioning,
5. Wrap burrito over the filling,
6. Easy to install,
7. Easy to use,
8. The device must be easy to clean after use (<15 minutes and no special tools),
9. The device must be safe – cause no injury to the user.

Figure A-4: Burrito Folder Problem Statement and No Model

Points to remember:

1. Please specify the view for your concepts. (Top/Front/Perspective)
2. Either a front, top or perspective view is enough. Do not draw the same concept in different view.
3. Please provide properly dimensioned and labeled sketches.
4. Provide explanation for your sketch at the bottom.
5. Only one concept per page.

ID :

VIEW: TOP/ FRONT/ PERSPECTIVE

Explanation:

Figure A-5: Participant Sketch Sheet

<u>APPENDIX B</u>: EXTENDED STUDY MODEL ASSUMPTIONS FOR CONFORMANCE DATA

The data from each conformance metric and scoring approach are first fit with a linear model and all assumptions checked. The assumptions for each model are shown and discussed below.

<u>Functional Conformance – Participant Average Scoring Approach</u>

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure B-1, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

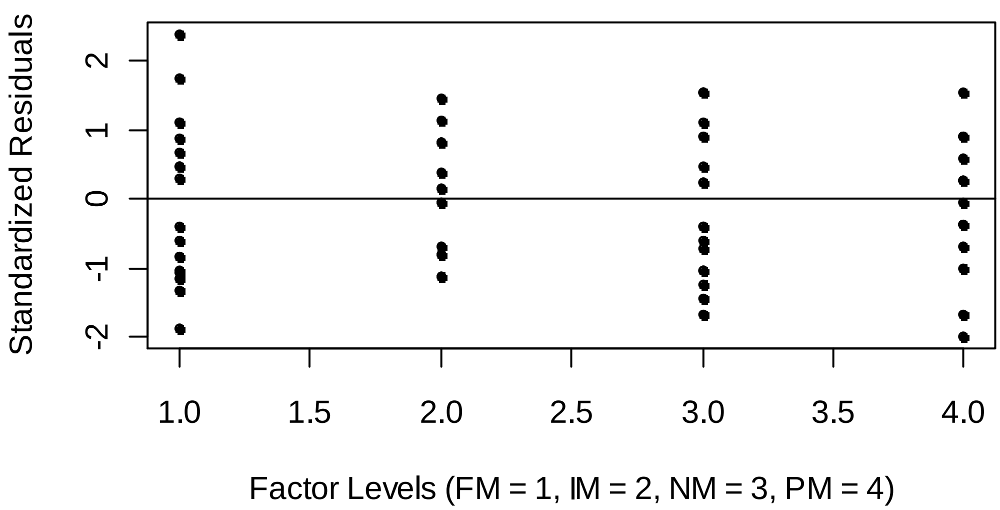


Figure B-1: Linear Model Fit for Functional Conformance – Participant Average Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.37 and -1.97, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure B-2, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 1.9, which is small. Therefore, the constant variance assumption is satisfied.

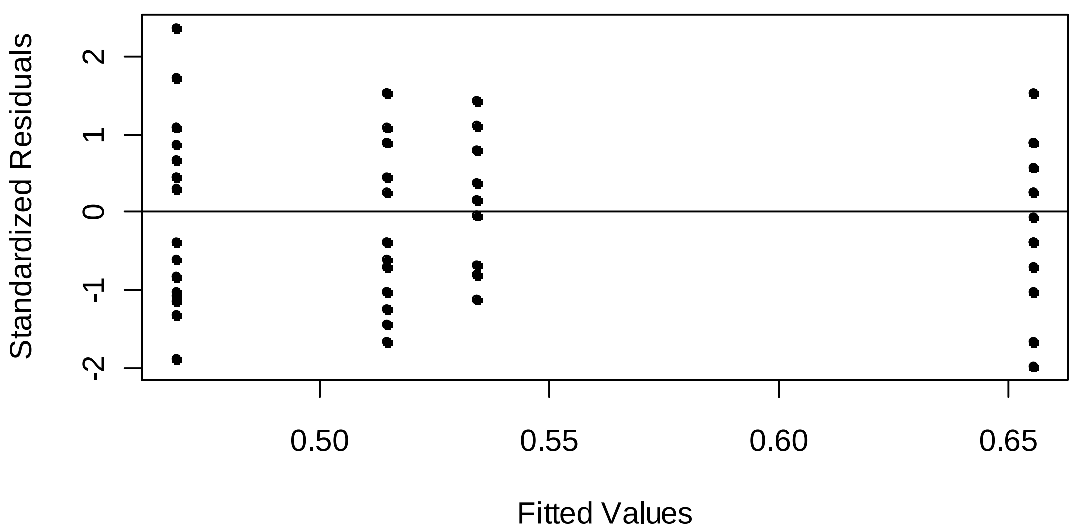


Figure B-2: Plot of Standard Residual versus Fitted Values for Functional Conformance – Participant Average Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure B-3, the normal probability plot shows a linear trend, so the normality assumption is satisfied.

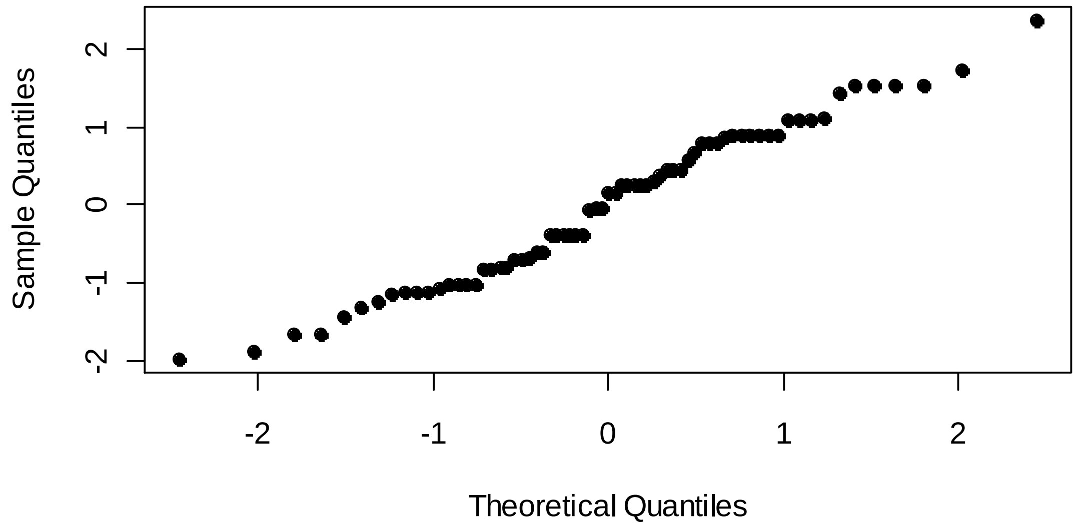


Figure B-3: Normal Probability Plot for Functional Conformance Model – Participant Average Scoring Approach

Functional Conformance – Participant Best Scoring Approach

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure B-4, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

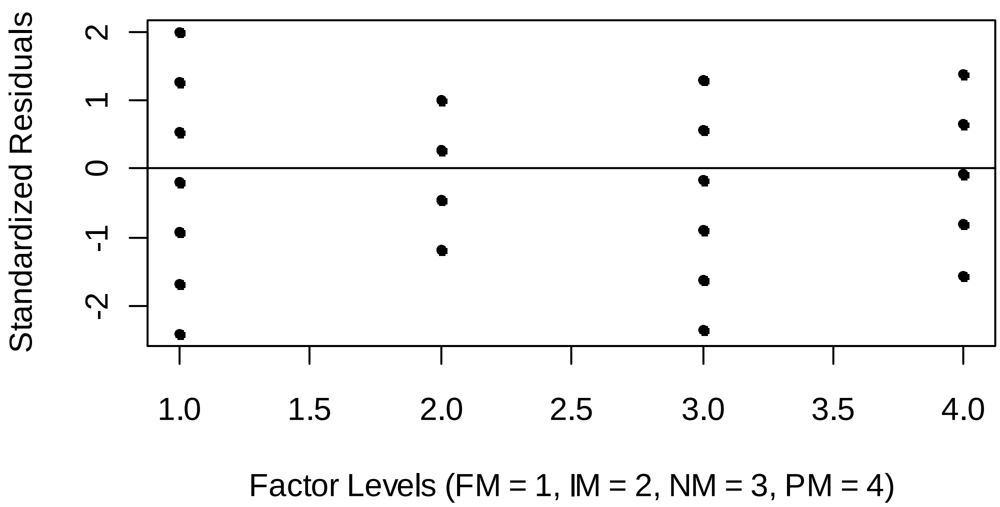


Figure B-4: Linear Model Fit for Functional Conformance – Participant Best Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 1.99 and -2.40, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure B-5, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 2.4, which is small. Therefore, the constant variance assumption is satisfied.

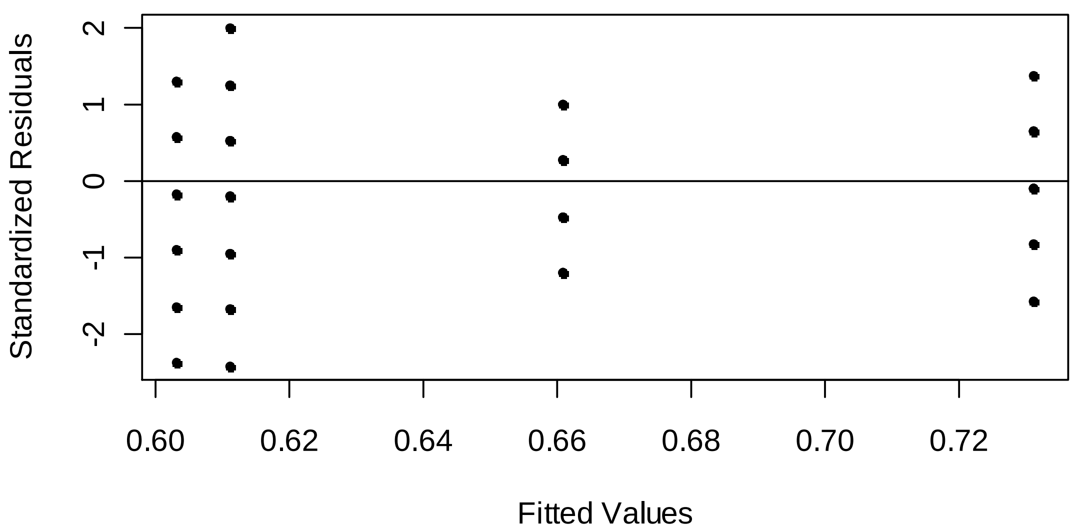


Figure B-5: Plot of Standard Residual versus Fitted Values for Functional Conformance – Participant Average Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure B-6, the normal probability plot is not linear at the ends, so the data may not be normally distributed. A Shapiro-Wilk normality test reveals that the distribution is not normal (p = 0.04). Therefore, nonparametric tests will be performed.

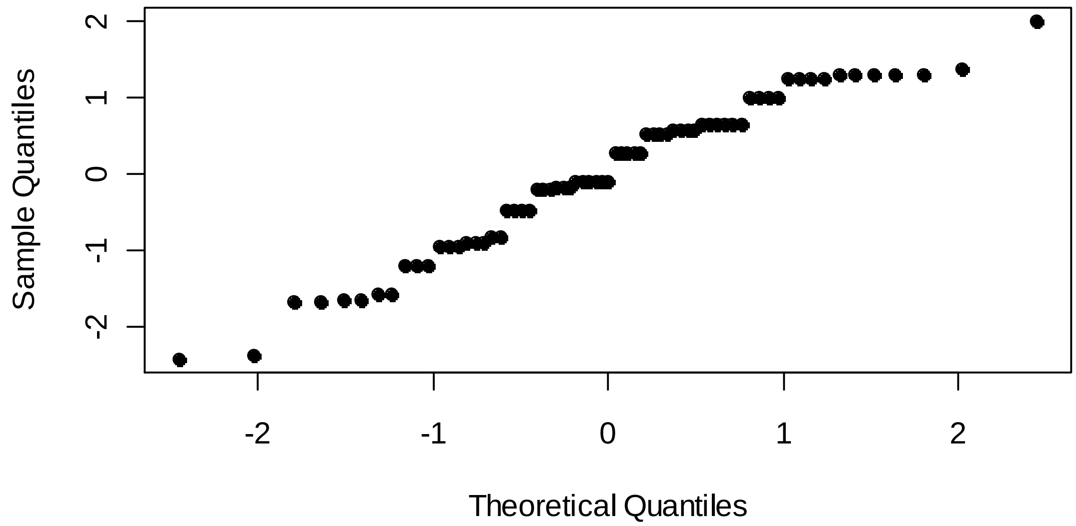


Figure B-6: Normal Probability Plot for Functional Conformance Model – Participant Best Scoring Approach

<u>Activity Conformance – Participant Average Scoring Approach</u>

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure B-7, the residuals appear to exhibit a random pattern, but they are not equally distributed about the mean. In the first treatment group, there are many points below the overall mean, while in the other three groups there are many points above the mean. Thus, the model may not be a good fit for the data.

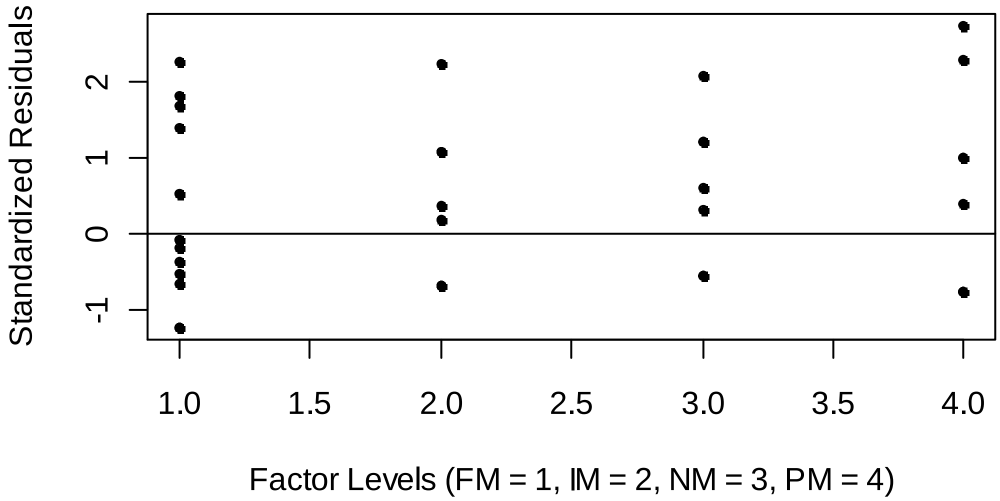


Figure B-7: Linear Model Fit for Activity Conformance – Participant Average Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.72 and -1.23, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure B-8, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 1.9, which is small. Therefore, the constant variance assumption is satisfied.

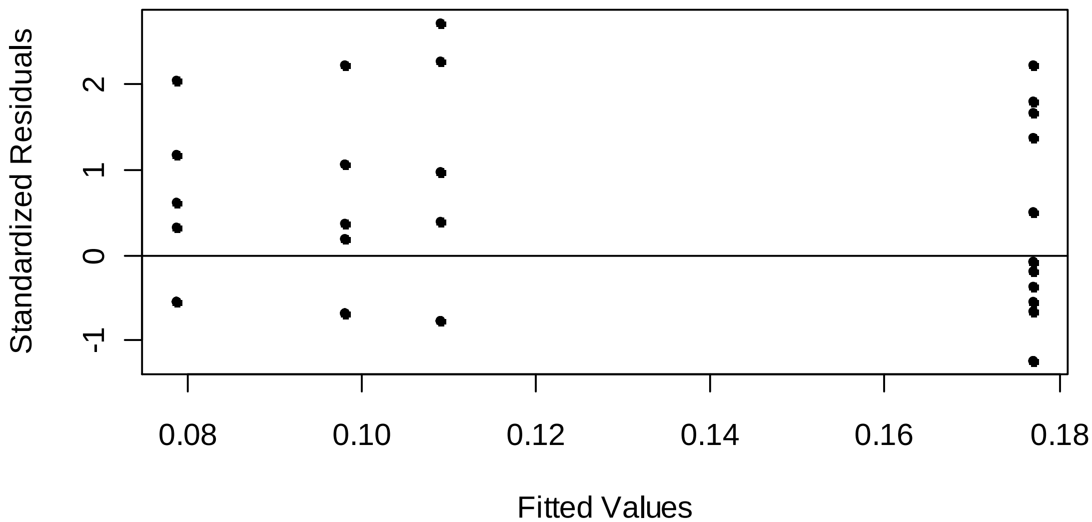


Figure B-8: Plot of Standard Residual versus Fitted Values for Activity Conformance – Participant Average Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure B-9, the normal probability plot is not linear, so the data are not normally distributed. Therefore, nonparametric tests will be performed.

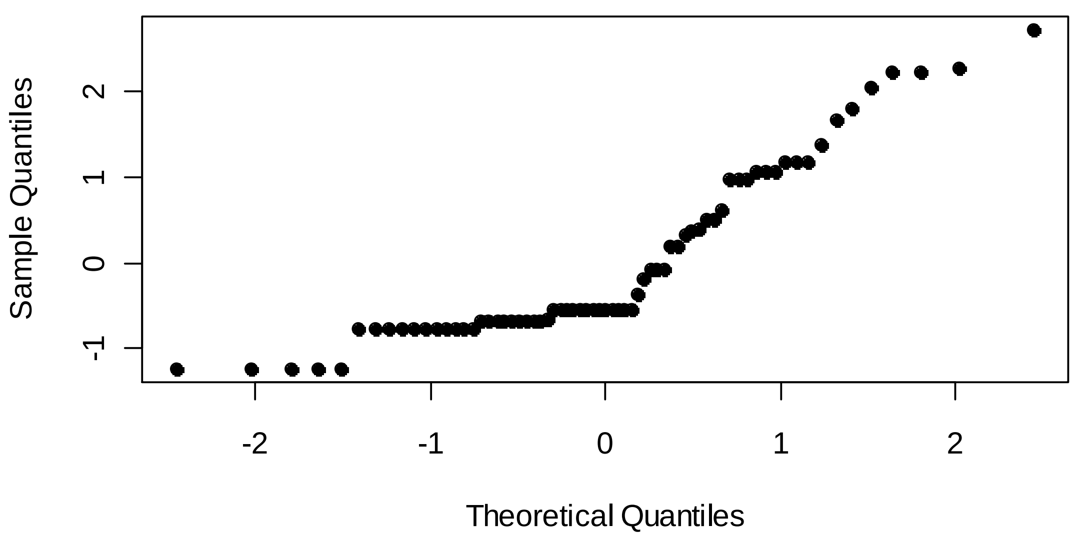


Figure B-9: Normal Probability Plot for Activity Conformance Model – Participant Average Scoring Approach

Activity Conformance – Participant Best Scoring Approach

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure B-10, the residuals appear to exhibit a random pattern around the mean, so the linear model is appropriate. The residuals are evenly spaced since the participant best scoring approach results in relatively discrete data.

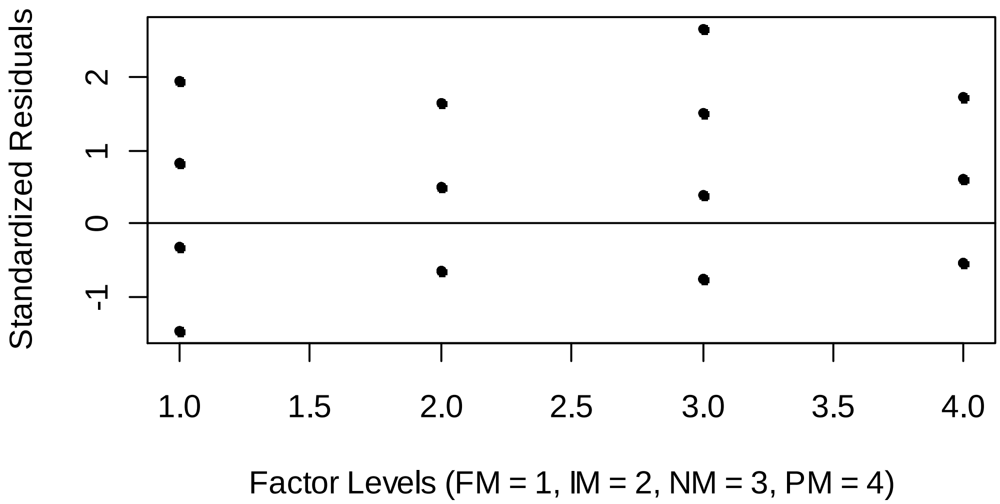


Figure B-10: Linear Model Fit for Activity Conformance – Participant Best Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.64 and -1.44, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure B-11, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate.

The ratio for this scoring approach is 2.3, which is small. Therefore, the constant variance assumption is satisfied.

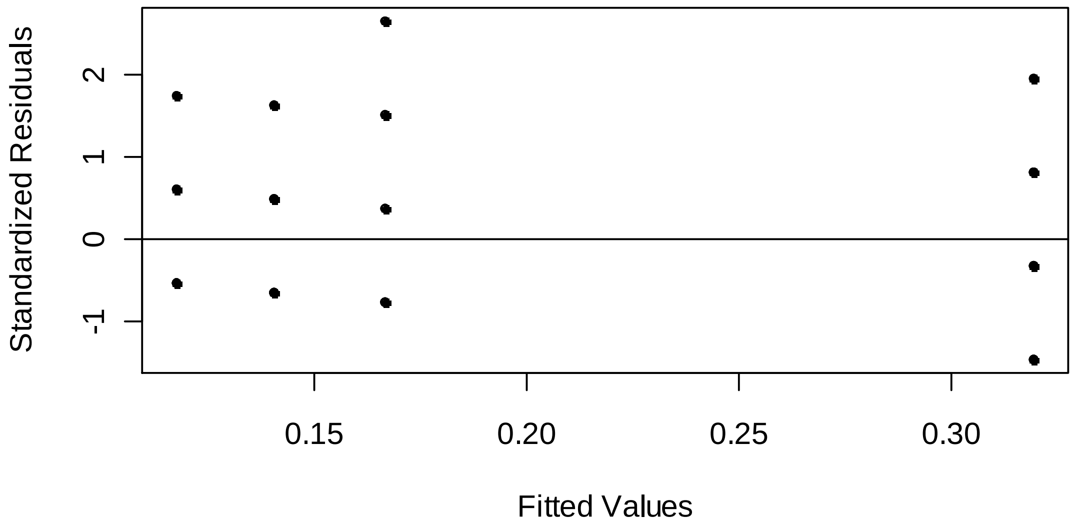


Figure B-11: Plot of Standard Residual versus Fitted Values for Activity Conformance – Participant Best Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure B-12, the normal probability plot is not linear, so the data are not normally distributed. Therefore, nonparametric tests will be performed.

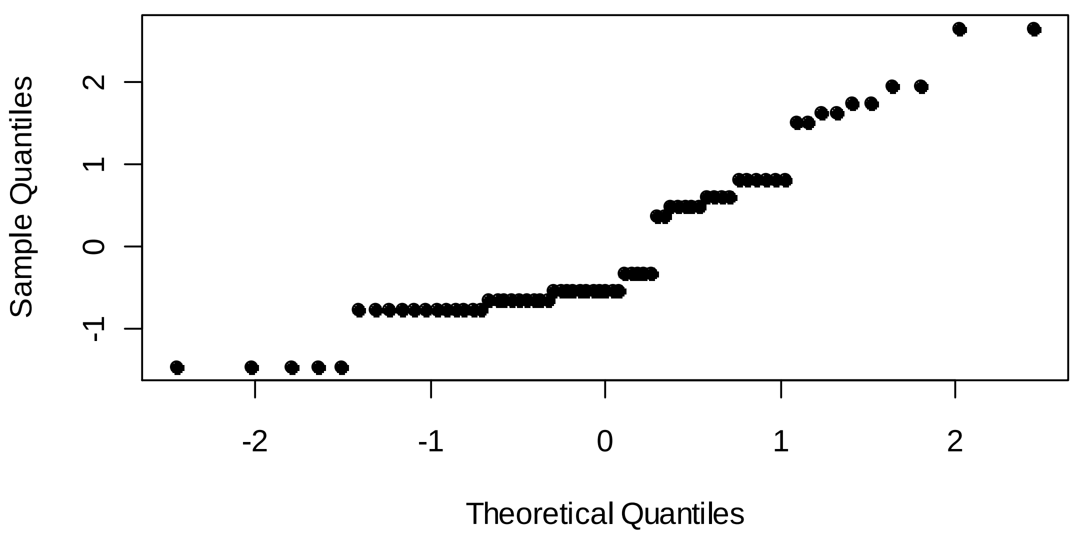


Figure B-12: Normal Probability Plot for Activity Conformance Model – Participant Best Scoring Approach

Interaction Conformance – Participant Average Scoring Approach

The initial check of assumptions for this model revealed one outlier that was removed from the data because the participant clearly did not understand what was expected in the sketching exercise (see Section 7.3.7.1). The assumptions after removal of this outlier are checked and described below.

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure B-13, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

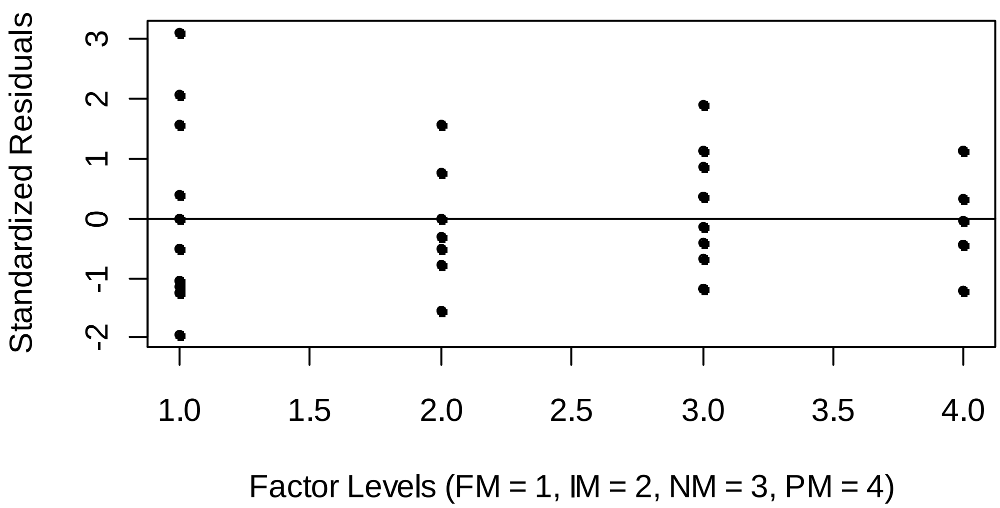


Figure B-13: Linear Model Fit for Interaction Conformance – Participant Average Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 3.11 and -1.94, respectively. The sketches and data associated with the high score are reviewed and it is determined that this participant created good sketches, so the data point is not removed from the model.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure B-14, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 2.8, which is small. Therefore, the constant variance assumption is satisfied.

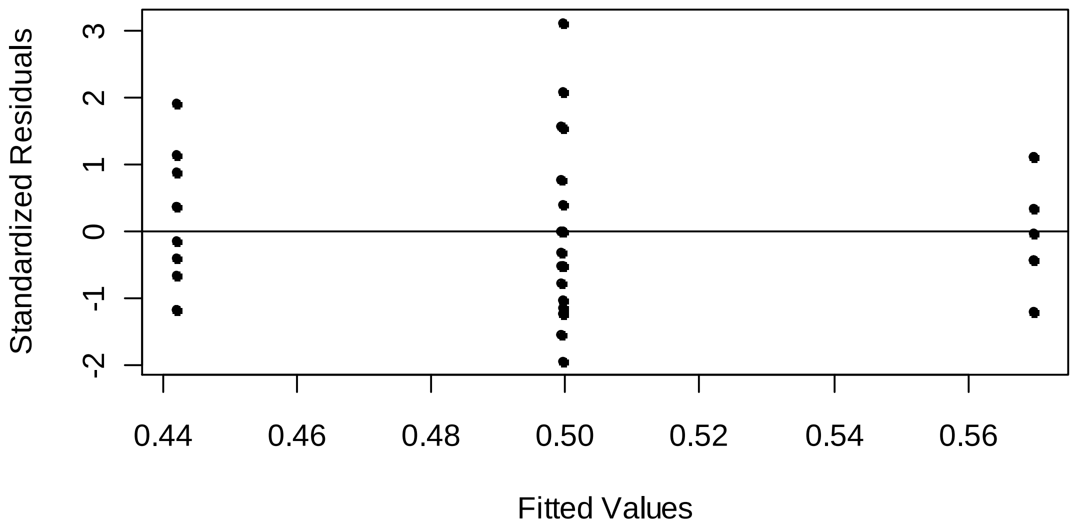


Figure B-14: Plot of Standard Residual versus Fitted Values for Interaction Conformance – Participant Average Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure B-15, the normal probability plot is approximately linear, but the plot contains steps in the data. A Shapiro-Wilk normality test shows that the data are normally distributed ($p = 0.28$), so this assumption is satisfied.

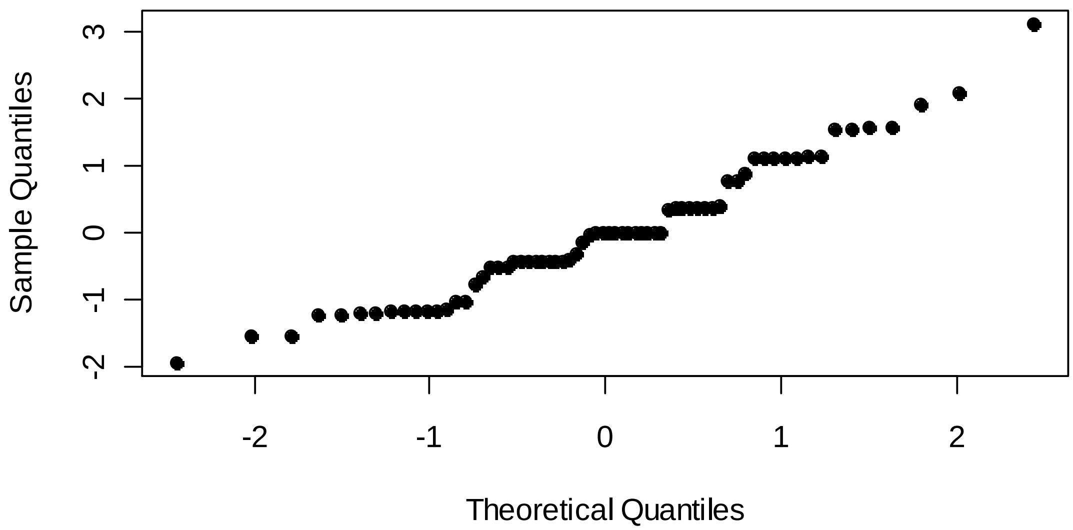


Figure B-15: Normal Probability Plot for Interaction Conformance Model – Participant Average Scoring Approach

Interaction Conformance – Participant Best Scoring Approach

The initial check of assumptions for this model revealed one outlier that was removed from the data because the participant clearly did not understand what was expected in the sketching exercise (see Section 7.3.7.1). The assumptions after removal of this outlier are checked and described below.

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure B-16, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

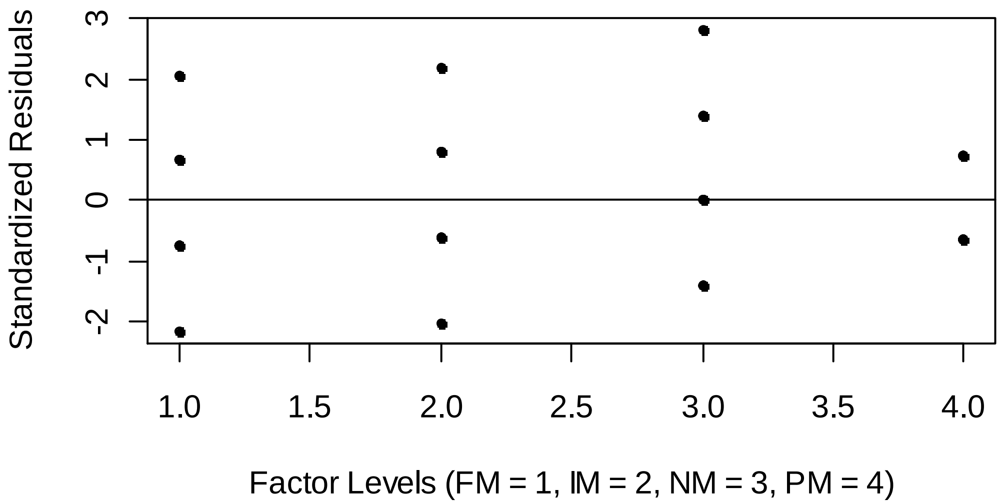


Figure B-16: Linear Model Fit for Interaction Conformance – Participant Best Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.81 and -2.15, respectively. These values are not considered outliers, since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure B-17, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 2.7, which is small. Therefore, the constant variance assumption is satisfied.

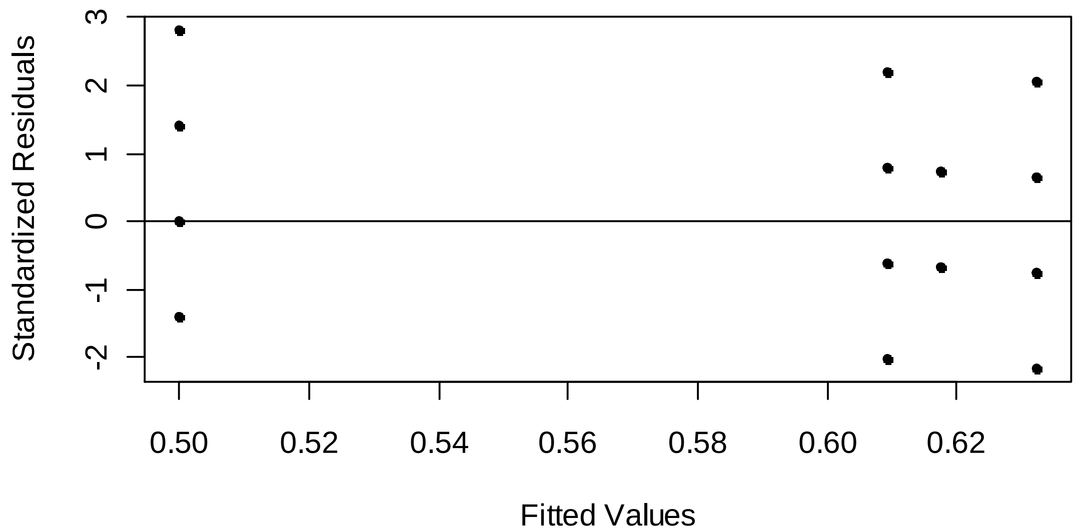


Figure B-17: Plot of Standard Residual versus Fitted Values for Interaction Conformance – Participant Best Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure B-18, the normal probability plot is approximately linear, but there are distinct steps in the data due to repeated scores. A Shapiro-Wilk normality test confirms that the data are not normally distributed ($p = 0.002$), so nonparametric tests will be used.

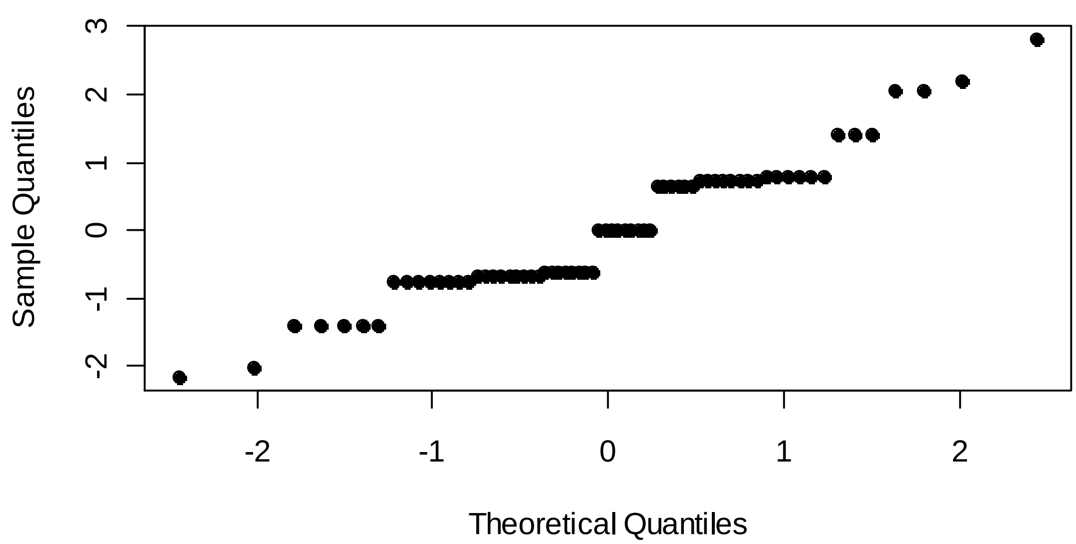


Figure B-18: Normal Probability Plot for Interaction Conformance Model – Participant Best Scoring Approach

APPENDIX C: EXTENDED STUDY MODEL ASSUMPTIONS FOR QUALITY DATA

The data from each quality metric and scoring approach are first fit with a linear model and all assumptions checked. The assumptions for each model are shown and discussed below.

Overall Quality – Participant Average Scoring Approach

To check the fit of the model, the standardized residuals are plotted against the factor levels. As shown in Figure C-1, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

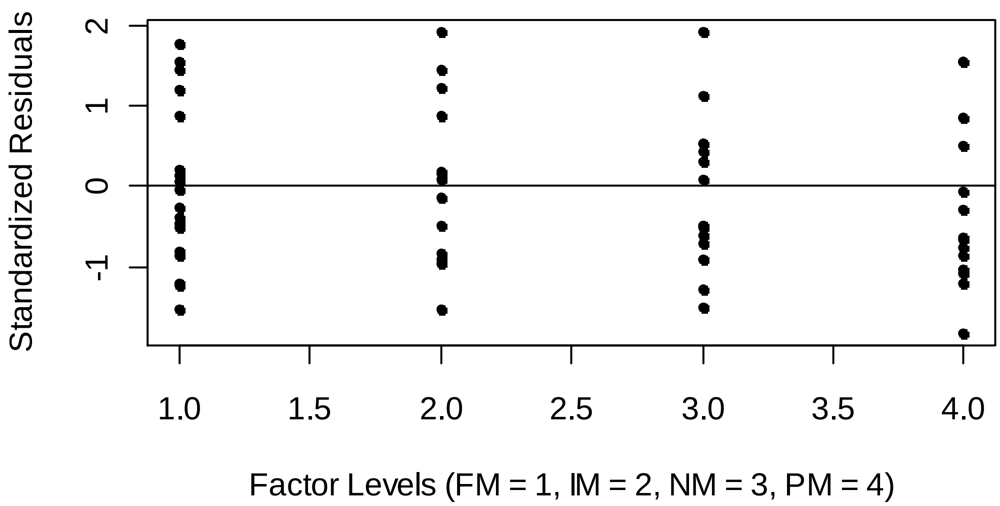


Figure C-1: Linear Model Fit for Overall Quality – Participant Average Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 1.92 and -1.82, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure C-2, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 1.4, which is small. Therefore, the constant variance assumption is satisfied.

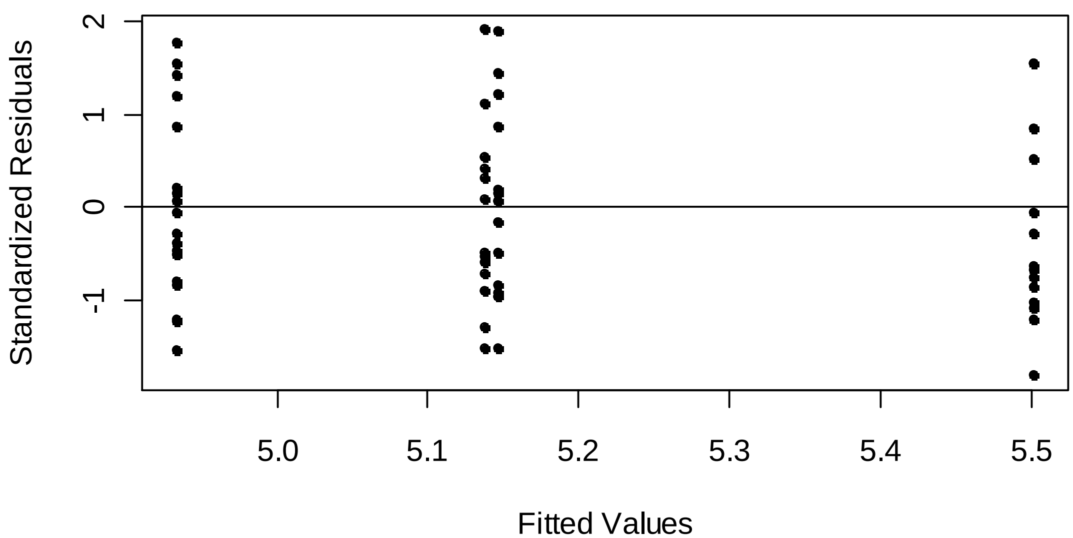


Figure C-2: Plot of Standard Residual versus Fitted Values for Overall Quality – Participant Average Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure C-3, the normal probability plot is not linear at the ends of the plot, so the data are not normally distributed. Therefore, nonparametric tests will be performed.

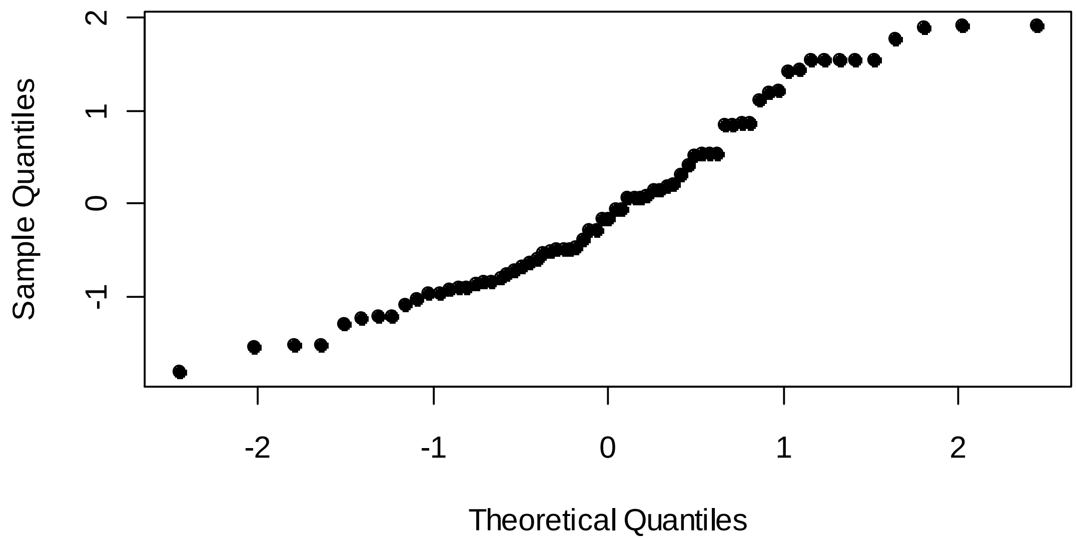


Figure C-3: Normal Probability Plot for Overall Quality Model – Participant Average Scoring Approach

Overall Quality – Participant Best Scoring Approach

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure C-4, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

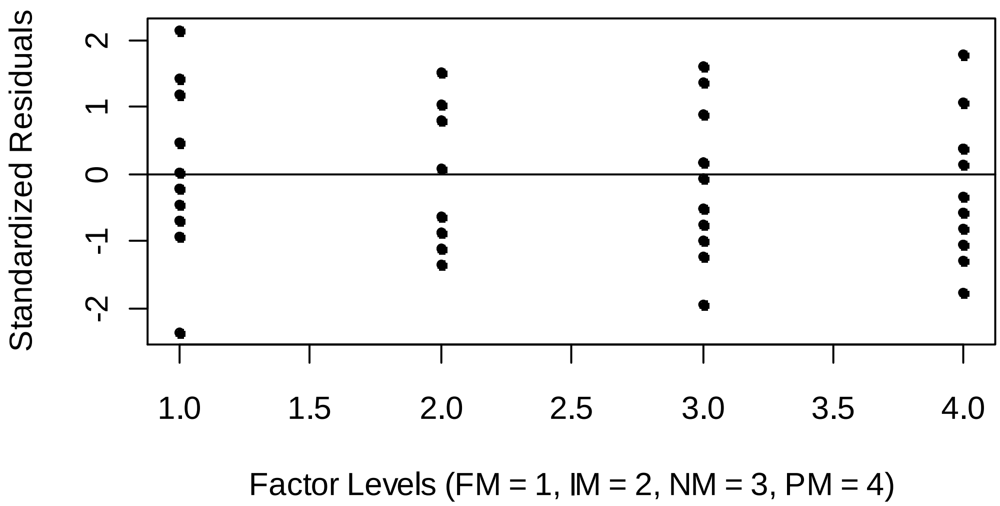


Figure C-4 Linear Model Fit for Overall Quality – Participant Best Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.15 and -2.35, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure C-5, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 1.1, which is small. Therefore, the constant variance assumption is satisfied.

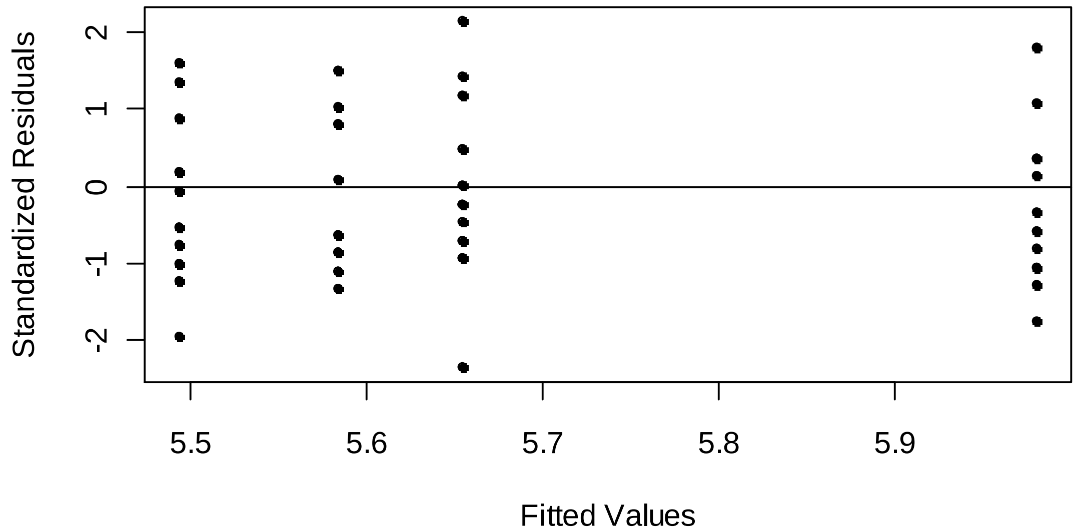


Figure C-5: Plot of Standard Residual versus Fitted Values for Overall Quality – Participant Best Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure C-6, the normal probability plot shows a linear trend, so the normality assumption is satisfied.

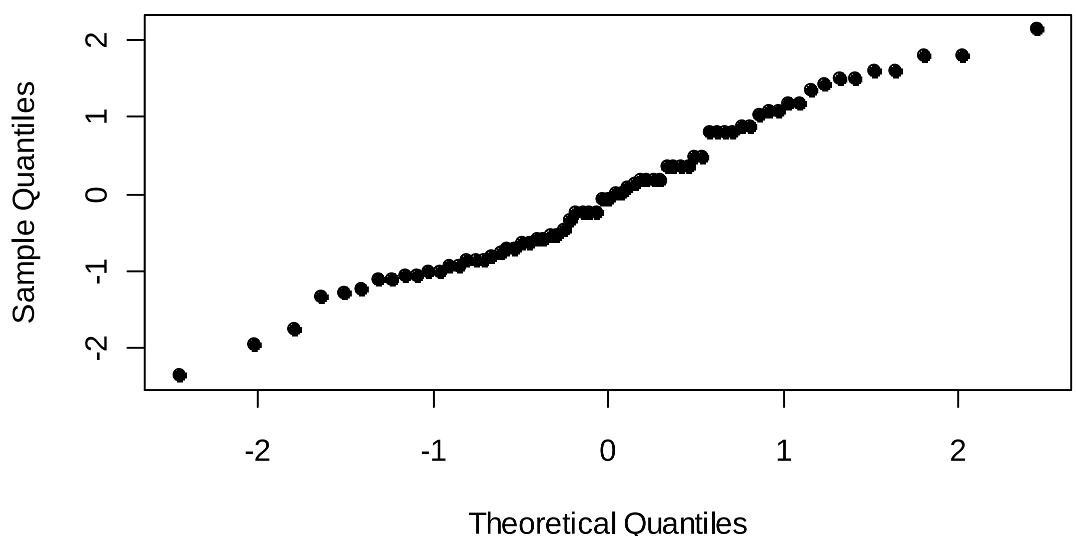


Figure C-6: Normal Probability Plot for Overall Quality Model – Participant Best Scoring Approach

Functional Quality – Participant Average Scoring Approach

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure C-7, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

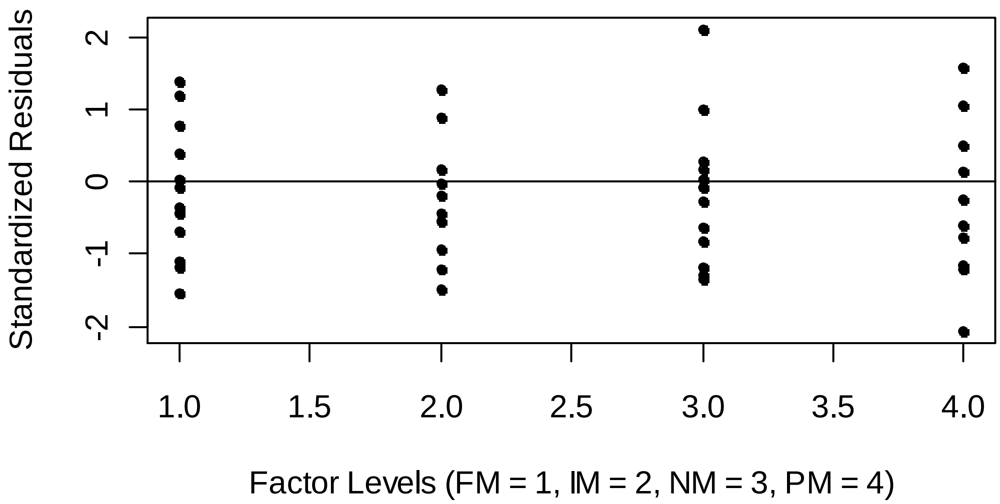


Figure C-7 Linear Model Fit for Functional Quality – Participant Average Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.10 and -2.06, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure C-8, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 1.7, which is small. Therefore, the constant variance assumption is satisfied.

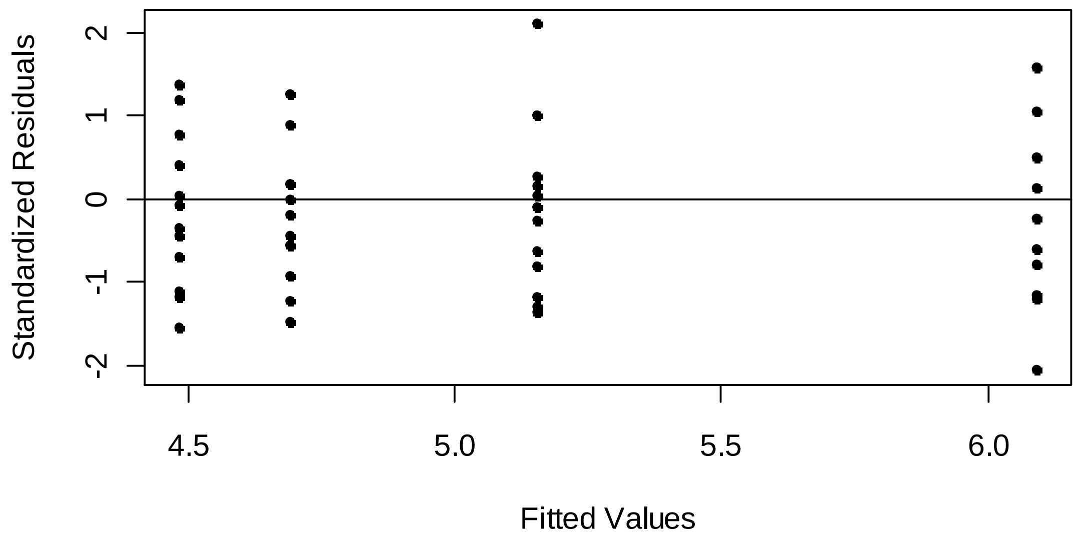


Figure C-8: Plot of Standard Residual versus Fitted Values for Functional Quality – Participant Average Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure C-9, the normal probability plot shows a linear trend, so the normality assumption is satisfied.

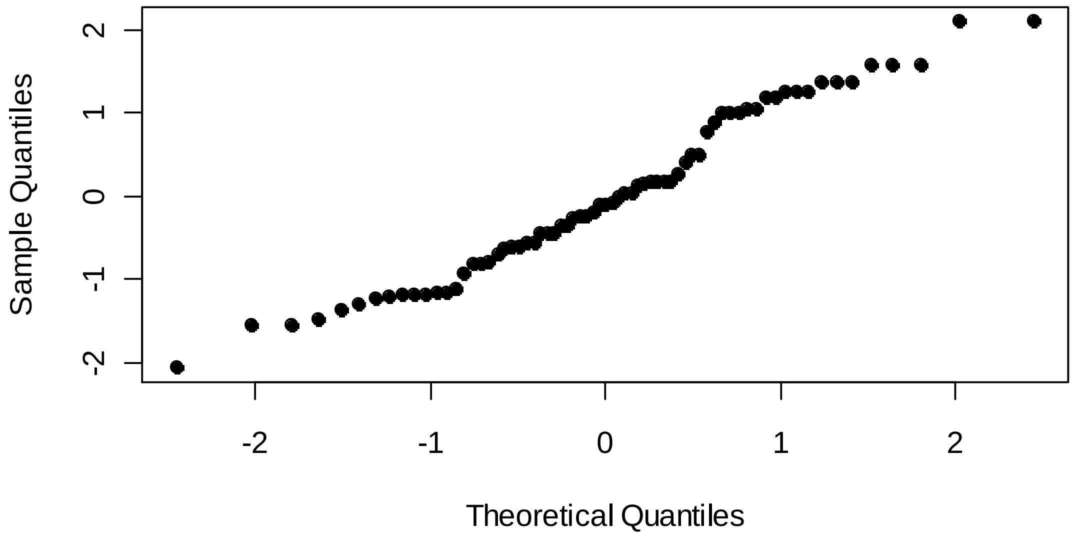


Figure C-9: Normal Probability Plot for Functional Quality Model – Participant Average Scoring Approach

<u>Functional Quality – Participant Best Scoring Approach</u>

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure C-10, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

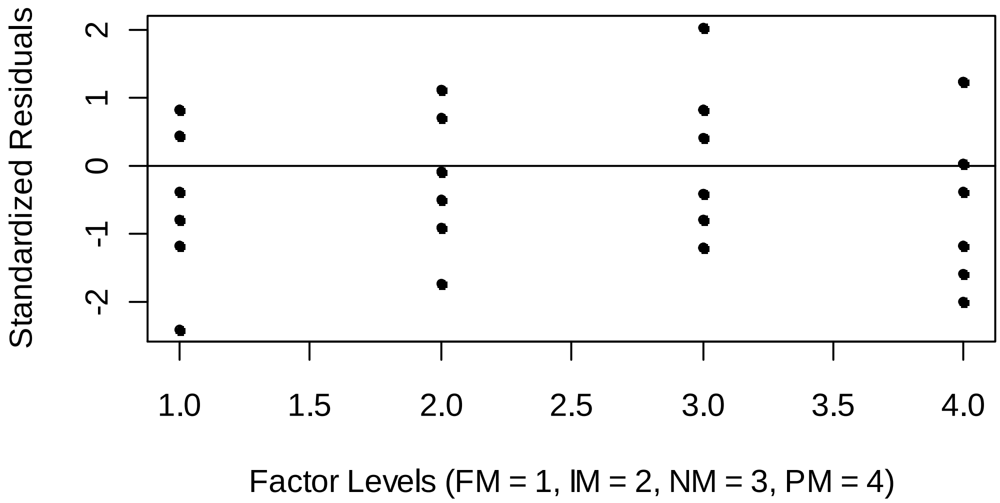


Figure C-10 Linear Model Fit for Functional Quality – Participant Best Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.03 and -2.42, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure C-11, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 1.8, which is small. Therefore, the constant variance assumption is satisfied.

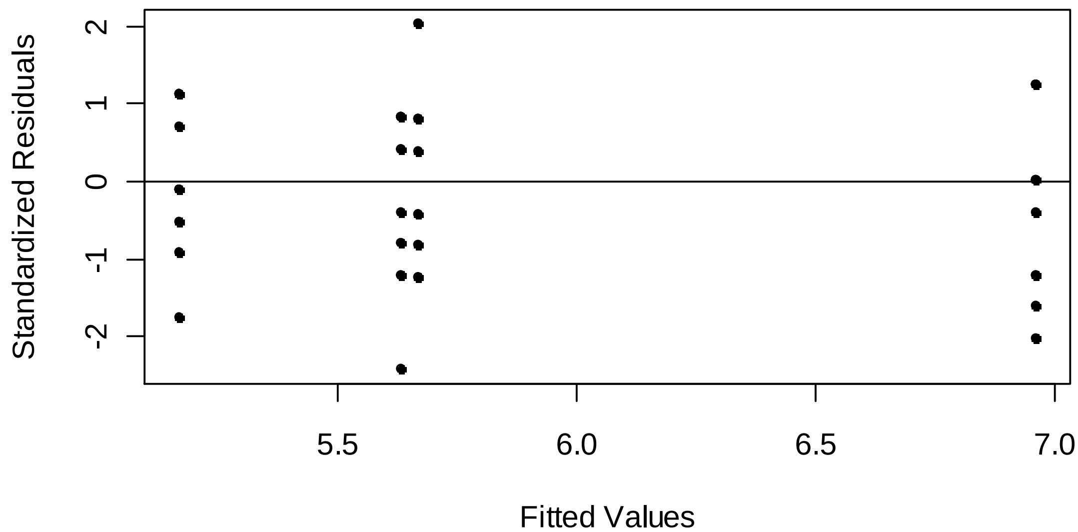


Figure C-11: Plot of Standard Residual versus Fitted Values for Functional Quality – Participant Best Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure C-12, the normal probability plot contains steps and deviates from linearity, so there may be problems with the normality assumption. A Shapiro-Wilk normality test shows that the distribution is not normal ($p = 0.03$), so nonparametric tests will be used.

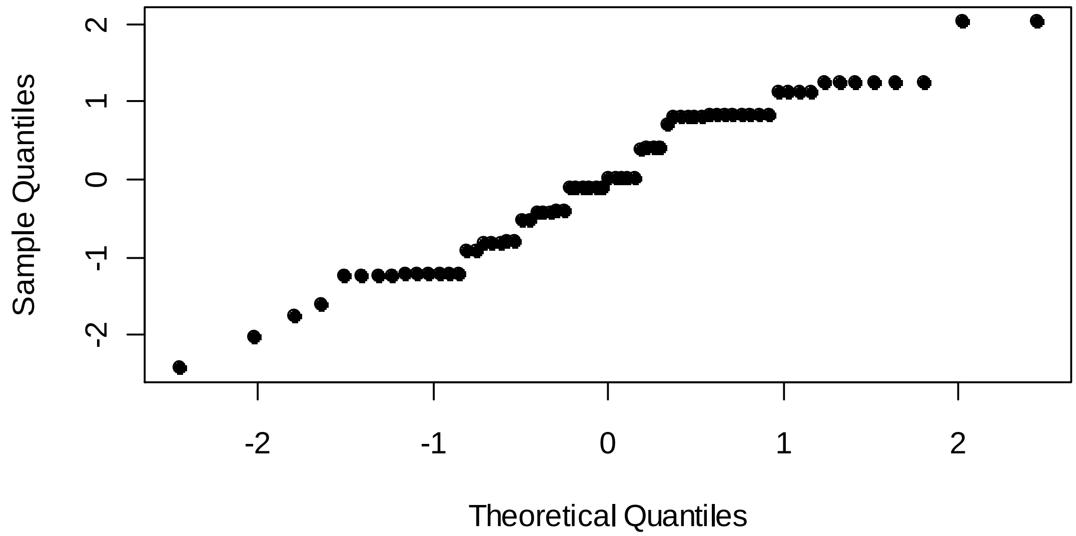


Figure C-12: Normal Probability Plot for Functional Quality Model – Participant Best Scoring Approach

Activity Quality – Participant Average Scoring Approach

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure C-13, the residuals appear to exhibit a random pattern, but they are not equally distributed about the mean. There are many points below the overall mean, with a few potential outliers. Thus, the model may not be a good fit for the data.

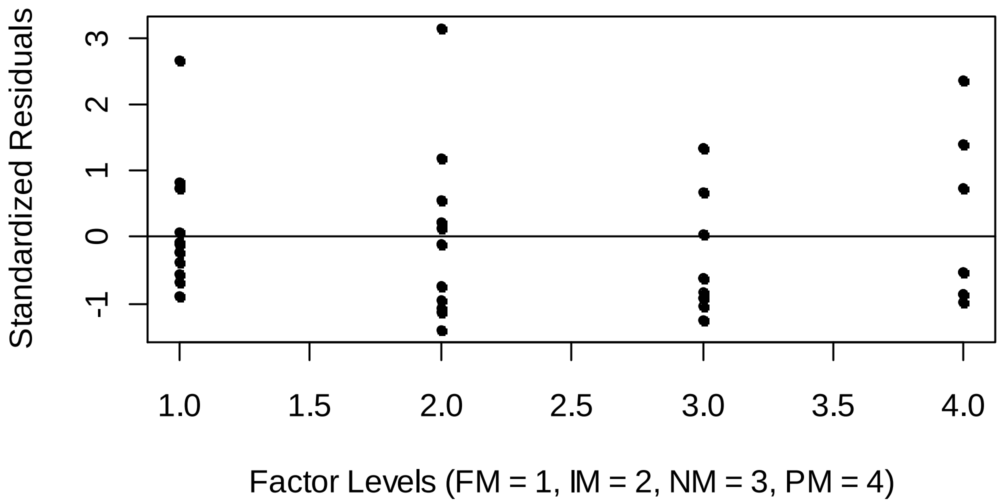


Figure C-13: Linear Model Fit for Activity Quality – Participant Average Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 3.13 and -1.39, respectively. There were three data points with high standardized residuals relative to the rest of the data, 2.35, 2.65, and 3.13. The sketches did not reveal any problems and the three points are each in different treatment groups, so the data are not considered outliers.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure C-14, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate. The ratio for this scoring approach is 1.6, which is small. Therefore, the constant variance assumption is satisfied.

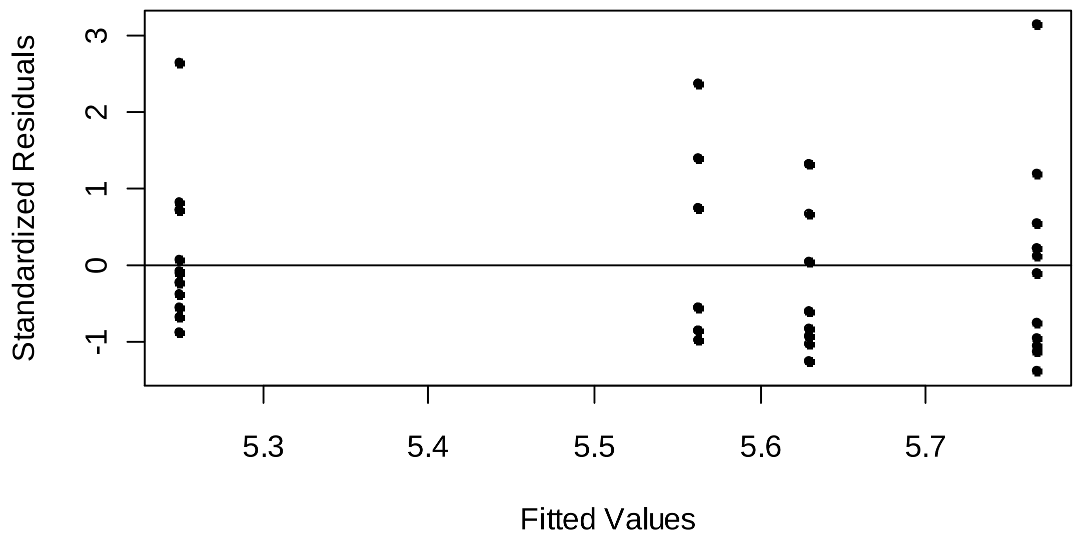


Figure C-14: Plot of Standard Residual versus Fitted Values for Activity Quality – Participant Average Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure C-15, the normal probability plot is not linear, so the data are not normally distributed. Therefore, nonparametric tests will be performed.

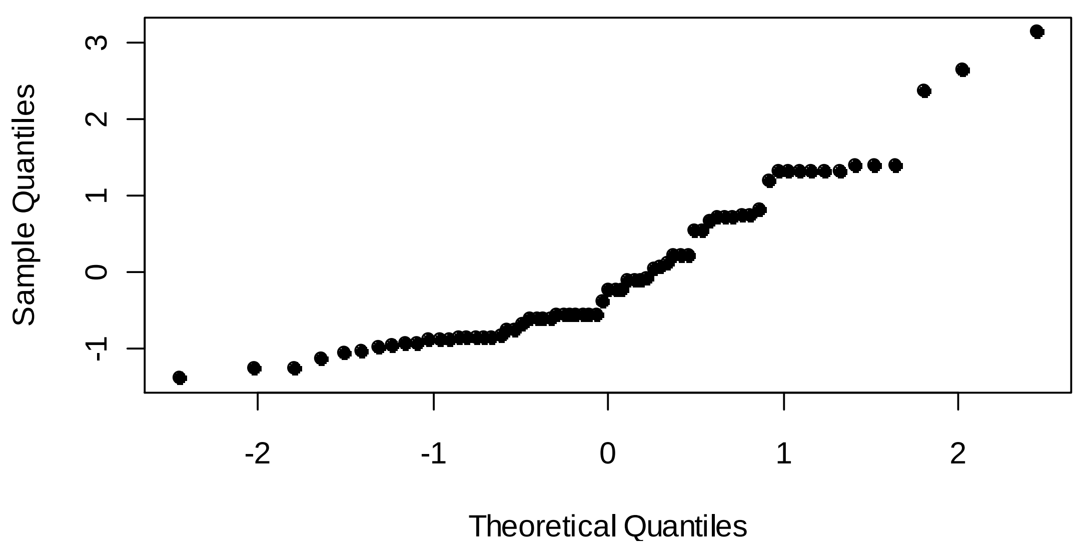


Figure C-15: Normal Probability Plot for Activity Quality Model – Participant Average Scoring Approach

Activity Quality – Participant Best Scoring Approach

To check the fit of the model, the standardized residuals are plotted against the factor levels . As shown in Figure C-16, the residuals appear to exhibit a random pattern, so the linear model is appropriate.

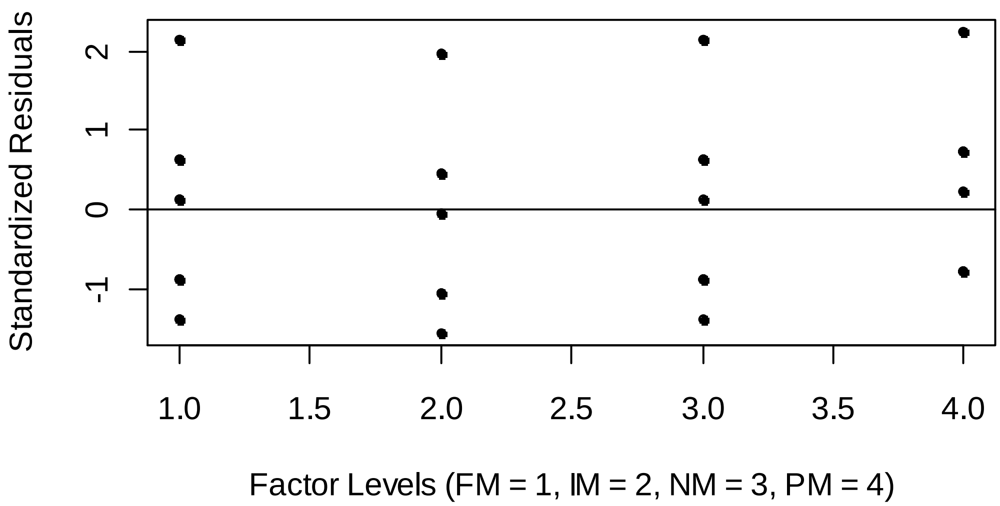


Figure C-16: Linear Model Fit for Activity Quality – Participant Best Scoring Approach

To check for outliers, the standardized residuals are sorted from smallest to largest. The maximum and minimum standardized residuals are 2.23 and -1.53, respectively. These values are not considered outliers since they are within three standard deviations of the mean.

To check for constant variance, the standardized residuals are first plotted against the fitted values. As shown in Figure C-17, there is no trend in variance so the data appear to satisfy the constant variance assumption. The second check for constant variance is to compare the largest variance estimate with the smallest variance estimate.

The ratio for this scoring approach is 1.3, which is small. Therefore, the constant variance assumption is satisfied.

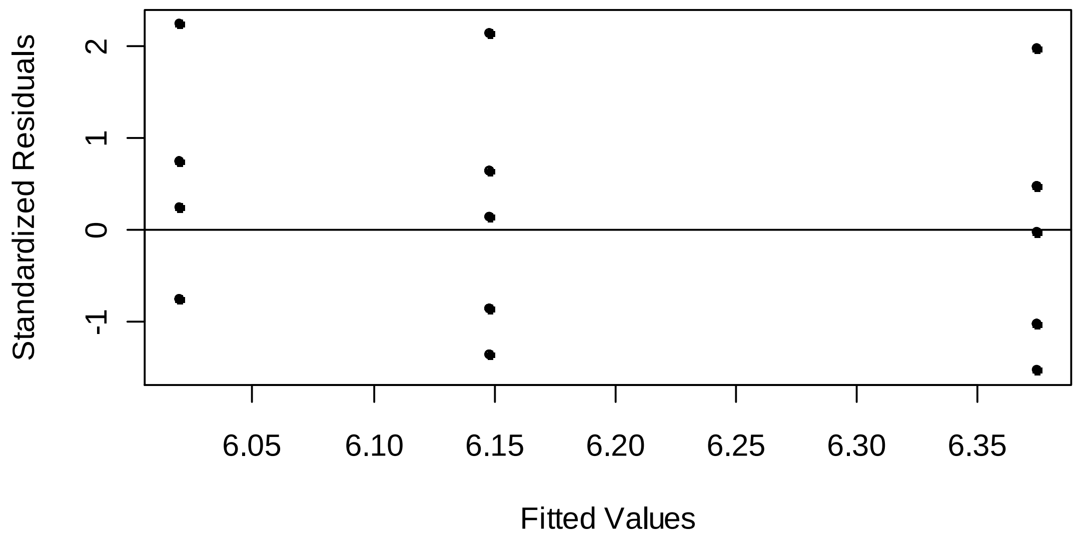


Figure C-17: Plot of Standard Residual versus Fitted Values for Activity Quality – Participant Best Scoring Approach

To check for normality, the standardized residuals are plotted against their normal scores. As shown in Figure C-18, the normal probability plot is not linear, so the data are not normally distributed. Therefore, nonparametric tests will be performed.

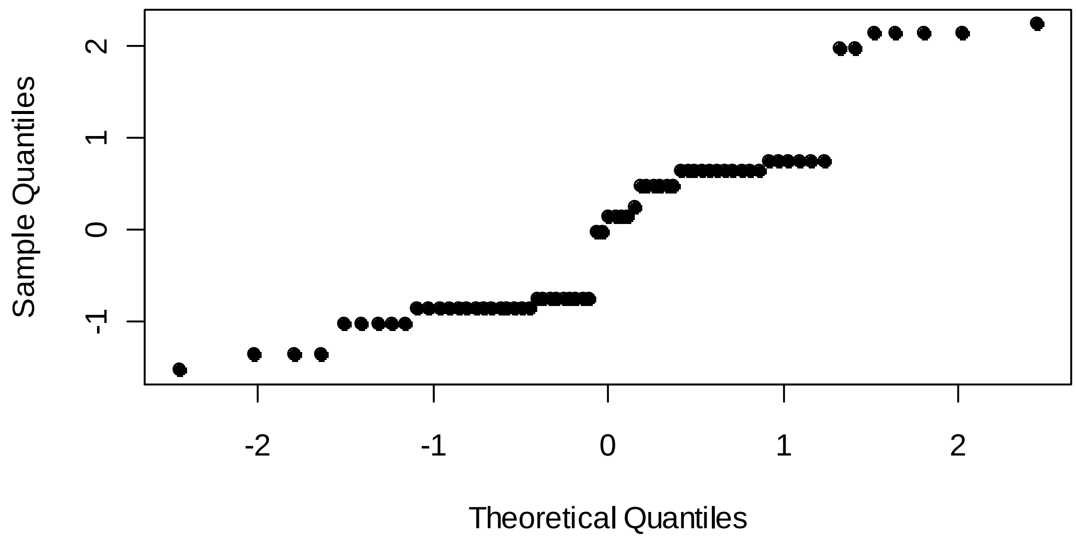


Figure C-18: Normal Probability Plot for Activity Quality Model – Participant Average Scoring Approach

APPENDIX D: SAMPLE SIZE CALCULATIONS

Sample size for the extended user study is calculated from the contrasts that will be performed using Fisher's Least Significant Difference comparison procedure. The contrast equation is:

$$\Delta \pm t_{dfe,\alpha/2}\sqrt{MSE\left(2/r\right)} \qquad \text{(D-1)}$$

where Δ is the difference in means of the treatment groups,

MSE is the estimated mean squared error, and

r is the number of replicates per treatment group.

The experiment is a completely randomized design, so $dfe = n - v = vr - v = v(r-1)$. In order to identify a difference as significant the difference in means, Δ, must be greater than the margin of error.

$$\Delta \geq t_{dfe,\alpha/2}\sqrt{MSE\left(2/r\right)} \qquad \text{(D-2)}$$

Equation D-2 can be rearranged as follows

$$\left(t_{dfe,\alpha/2}\right)^2 \leq \frac{r\Delta^2}{2MSE} \qquad \text{(D-3)}$$

The experiment variables used to calculate sample size for overall quality are:

Δ = 10% of the response range = 10%(max - min) = 10%(9-1) = 0.8

ν = 4 (FM, IM, PM, NM)

α = 0.05

To estimate MSE, a 90% upper confidence limit on MSE from the initial study (n=26, ν=2) is calculated as follows:

$$\sigma^2 \le \frac{SSE}{\chi^2_{n-\nu,1-\alpha}} \qquad \text{(D-4)}$$

$$\sigma^2 \le \frac{(n-\nu)MSE}{\chi^2_{n-\nu,1-\alpha}}$$

$$\sigma^2 \le \frac{(26-2)1.07}{\chi^2_{26-2,1-0.1}}$$

$$\sigma^2 \le \frac{(26-2)1.07}{15.659}$$

$$\sigma^2 \le \frac{(26-2)1.07}{15.659}$$

$$\sigma^2 \le 1.642$$

Therefore, 1.642 is used as MSE to calculate sample size for the extended study. The values for this experiment are substituted into Equation D-3, resulting in Equation D-5.

$$\left(t_{4(r-1),0.025}\right)^2 \le \frac{r(0.8)^2}{2(1.642)}$$

$$\left(t_{4(r-1),0.025}\right)^2 \le 0.195r \qquad \text{(D-5)}$$

A table of values for r and the two terms in Equation D-5 is created and values for r are iterated until the inequality holds for the smallest integer value of r.

Table D-1: Iterations for Calculating Sample Size

r	4(r-1)	$\left(t_{4(r-1),0.025}\right)^2$	$0.195r$	Action
10	36	4.113	1.949	increase r
15	56	4.013	2.923	increase r
20	76	3.967	3.897	increase r
25	96	3.940	4.872	decrease r
23	88	3.949	4.482	decrease r
22	84	3.955	4.287	decrease r
21	80	3.960	4.092	r = 21

As shown in Table E-1, the sample size required to detect a difference of 0.8 in overall quality using the participant average scoring approach is 21 participants per group, or 84 total participants. This procedure is repeated for each metric and scoring approach, and the results are shown in Table 7-21 and Table 7-22.

APPENDIX E: USER STUDY DATA

The quality and conformance ratings for all participants' sketches is shown in Table E-1. The sketch ID is coded as <treatment group><participant number within treatment> - <sketch number within participant>. For example, the ID P8-2 refers to the second sketch created by Participant 8 in the pruned model group.

Table E-1: Participant Quality and Conformance Scores

ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
F1-1	3	9	9	9	3	9	9	3	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0	1
F1-2	1	9	9	9	9	3	9	3	9	0	1	1	1	1	1	0	1	0	0	1	1	1	0	1
F2-1	3	9	9	9	9	3	9	9	9	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0
F2-2	3	9	9	9	3	3	3	3	3	0	1	1	1	1	1	0	1	0	0	0	1	1	0	0
F3-1	1	1	1	1	1	3	9	3	3	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0
F3-2	1	9	1	1	1	9	9	9	9	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0
F3-3	1	1	3	3	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
F3-4	1	1	1	1	1	9	9	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F3-5	1	1	1	1	1	9	9	3	9	0	0	0	0	0	1	1	0	0	1	1	0	0	1	0
F4-1	3	1	1	1	1	9	9	3	3	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
F4-2	3	9	1	3	3	3	9	3	9	0	1	0	1	0	0	1	0	0	1	0	1	1	1	0
F4-3	1	1	1	1	1	3	9	1	9	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0
F5-1	1	1	3	1	1	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
F5-2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F5-3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
F6-1	9	9	3	9	9	3	9	9	3	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
F6-2	9	9	3	9	9	3	9	3	3	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
F7-1	9	9	3	9	9	3	9	1	9	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
F7-2	1	1	9	1	1	9	9	3	3	0	0	0	0	1	0	1	1	1	1	0	1	1	1	0
F7-3	3	3	3	3	9	3	9	1	9	0	1	1	1	1	1	0	1	0	0	0	1	1	0	0
F7-4	1	1	3	1	1	9	9	1	9	0	0	0	0	1	0	1	1	1	1	0	1	0	1	0
F8-1	1	9	1	1	1	9	9	9	9	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
F8-2	1	1	9	1	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1
F8-3	1	9	9	3	9	3	9	3	3	0	1	1	1	1	0	0	0	0	0	0	1	1	0	0
F9-1	9	9	1	3	3	3	9	3	9	0	1	1	1	0	0	0	0	0	0	0	1	1	0	0
F9-2	1	1	9	3	3	9	9	3	9	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
F9-3	9	1	1	1	1	9	9	3	9	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
F9-4	1	3	1	1	1	3	9	3	3	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
F10-1	1	9	1	1	3	3	9	3	9	0	1	0	1	0	0	1	0	0	1	0	0	1	1	0
F10-2	3	1	1	1	1	3	9	9	9	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
F10-3	1	3	3	1	3	9	9	3	9	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
F10-4	1	1	9	3	3	9	9	3	9	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
F10-5	1	1	1	1	1	3	9	3	3	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0
F11-1	1	9	3	3	3	9	9	3	9	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0
F11-2	1	9	3	3	3	9	9	9	9	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0
F12-1	1	3	9	3	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
F12-2	3	9	3	9	9	3	3	3	3	0	1	1	1	1	0	0	1	1	0	0	1	1	0	0
F12-3	1	1	3	1	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0

ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
F13-1	1	9	1	1	1	9	9	1	3	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0
F13-2	1	1	9	1	3	9	9	9	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
F13-3	1	9	1	1	1	3	9	3	9	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
F13-4	1	1	9	1	1	9	9	9	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
F14-1	9	9	1	3	3	3	9	3	3	1	1	1	1	0	0	1	0	0	0	0	1	1	0	0
F14-2	1	1	3	1	3	9	9	3	3	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0
F14-3	1	1	3	3	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
F15-1	1	9	3	3	3	9	9	3	3	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0
F15-2	9	9	3	9	3	3	9	3	3	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0
F15-3	1	9	9	3	1	9	9	1	3	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1
F16-1	9	9	3	3	9	3	9	3	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	0
F16-2	9	1	9	3	1	3	3	3	3	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0
F17-1	9	9	3	9	9	3	3	3	3	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
F17-2	1	1	9	3	3	9	9	3	1	0	0	0	0	1	0	1	1	1	0	0	1	0	0	0
F17-3	3	9	3	3	3	9	9	3	3	0	0	1	1	1	1	1	0	0	1	0	1	1	1	0
F19-1	9	9	3	9	9	3	3	1	9	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
I1-1	1	1	3	3	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
I1-2	1	3	3	3	9	9	9	1	3	0	1	0	1	1	0	0	0	0	0	1	1	1	1	1
I2-1	3	9	3	9	9	3	9	3	3	0	1	1	1	1	1	0	0	0	0	1	1	1	0	1
I2-2	3	9	3	9	9	3	3	3	3	0	1	1	1	1	1	0	0	0	0	1	1	1	0	1
I3-1	9	9	3	9	9	3	9	3	3	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
I3-2	9	9	3	9	9	3	9	3	3	1	1	0	1	1	0	0	0	0	0	0	1	1	0	0
I4-1	3	9	9	9	9	3	9	3	3	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0
I4-2	9	9	3	9	9	3	3	3	3	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0
I5-1	1	1	9	3	3	9	9	3	3	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0
I5-2	1	9	1	1	1	9	9	3	9	0	1	0	1	0	0	1	0	0	0	0	0	1	0	0
I6-1	9	1	1	1	1	9	9	3	3	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
I6-2	1	9	3	3	3	9	9	1	3	0	1	0	1	1	0	1	0	0	0	0	1	1	0	0
I6-3	9	9	3	9	9	3	9	9	3	1	1	1	1	0	1	0	0	0	0	0	1	1	0	0
I7-1	1	1	9	3	3	9	9	1	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
I7-2	1	9	1	1	3	3	9	1	9	0	1	0	1	0	0	1	0	0	0	0	1	1	0	0
I8-1	3	9	9	9	9	3	9	9	3	0	1	1	1	1	1	1	0	0	0	0	1	1	0	0
I9-1	3	9	3	9	9	3	9	3	3	0	1	1	1	1	1	0	0	0	0	1	1	1	0	1
I9-2	1	1	3	3	3	9	9	3	3	0	0	0	0	1	0	0	1	1	0	0	1	0	0	0
I9-3	3	9	3	9	9	3	9	3	9	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0
I10-1	3	1	1	1	1	9	9	3	3	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0
I10-2	1	9	1	1	1	3	9	3	3	0	1	0	1	0	0	0	0	1	0	0	1	1	1	0
I10-3	1	1	1	1	1	9	9	3	1	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0
I10-4	1	1	3	1	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0

ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
I10-5	1	9	3	3	3	9	9	3	3	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0
I11-1	3	9	3	9	3	3	9	9	9	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0
I11-2	1	9	3	3	3	3	9	3	3	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0
I12-1	1	9	9	9	9	9	9	1	3	0	1	1	1	1	0	0	0	0	0	0	1	1	0	0
I13-1	3	9	3	9	3	3	9	3	3	0	1	1	1	1	0	0	0	0	0	0	1	1	0	0
I14-1	1	9	3	3	1	9	9	3	3	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0
I14-2	1	9	3	3	3	3	9	3	3	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0
I14-3	1	9	3	3	3	3	9	3	3	0	1	0	1	1	1	0	0	0	0	0	1	1	0	0
I15-1	3	9	3	3	3	9	9	3	3	0	1	0	1	1	0	1	1	0	1	0	1	1	1	0
I15-2	3	9	3	3	3	9	9	3	3	0	1	1	1	1	0	0	0	1	1	0	0	1	0	0
I15-3	3	9	3	3	3	9	9	3	3	0	1	1	1	1	0	1	0	0	1	0	0	1	1	0
I16-1	3	3	9	3	9	1	9	3	3	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0
I16-2	1	1	1	1	1	9	9	3	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N1-1	9	9	9	9	9	3	9	3	3	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0
N2-1	3	9	3	9	9	3	9	3	1	0	1	1	1	1	0	0	0	0	0	0	1	1	0	0
N2-2	9	9	3	3	9	3	9	3	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
N2-3	1	9	3	3	3	9	9	3	9	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0
N3-1	1	1	9	3	9	9	9	3	1	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0
N3-2	1	1	9	3	1	9	9	3	3	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0
N4-1	9	9	3	9	9	3	3	3	1	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0
N4-2	9	9	3	3	9	3	1	3	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0
N4-3	9	9	3	9	9	3	3	3	3	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
N5-1	1	9	9	3	3	9	9	3	9	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0
N5-2	9	3	3	3	9	3	9	9	9	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0
N6-1	1	1	3	1	3	9	9	3	9	0	0	0	0	1	0	1	1	1	1	0	1	0	1	0
N6-2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N6-3	1	9	1	1	1	3	9	3	3	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0
N7-1	3	9	3	3	9	3	9	9	9	0	1	1	1	1	0	1	0	0	0	0	1	1	0	0
N7-2	1	9	3	3	3	3	9	3	9	0	1	0	1	0	0	1	1	0	1	0	1	1	1	0
N8-1	1	1	3	1	3	9	9	3	3	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0
N8-2	1	9	3	3	3	9	9	3	9	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0
N9-1	3	9	9	3	3	9	9	3	3	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
N10-1	9	9	9	9	9	3	9	3	3	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
N11-1	9	9	3	9	9	9	9	1	3	1	1	0	1	1	0	0	0	0	0	0	0	1	0	0
N11-2	1	9	3	3	3	3	1	3	1	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
N11-3	1	9	3	3	3	3	9	3	3	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0
N12-1	1	1	9	3	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
N12-2	1	1	3	1	3	9	9	3	3	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0
N12-3	1	1	9	3	3	9	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0

ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
N13-1	1	1	9	3	3	9	9	3	3	0	0	0	0	1	0	0	1	1	1	0	1	0	1	0
N13-2	1	9	3	3	3	9	9	3	3	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0
N14-1	1	1	9	3	3	9	9	1	9	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
N14-2	1	1	9	3	3	3	9	3	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
N15-1	3	9	9	3	9	3	9	3	3	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0
N16-1	3	1	3	3	3	9	9	3	3	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0
N16-2	9	1	3	3	3	9	3	3	3	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0
N16-3	1	1	3	3	3	9	9	3	9	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
N17-1	3	9	3	3	9	1	1	3	1	0	1	1	1	1	1	1	0	0	0	0	1	1	1	0
N18-1	9	3	3	3	9	9	9	3	3	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0
P1-1	9	9	9	9	3	3	9	3	3	1	1	1	1	1	0	1	0	0	0	0	1	1	0	1
P2-1	3	9	9	3	3	9	9	3	3	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0
P2-2	1	1	1	1	1	9	9	3	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P3-1	1	1	9	3	1	9	9	3	3	0	0	0	0	1	0	1	0	1	1	0	1	0	1	0
P3-2	1	9	9	3	1	9	9	3	3	0	1	0	1	1	0	1	0	1	1	0	1	1	1	0
P3-3	1	1	3	1	3	9	9	3	3	0	0	0	0	1	0	1	1	1	0	0	1	0	0	0
P3-4	1	9	3	3	3	9	9	3	3	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0
P4-1	9	9	3	9	9	3	9	3	3	0	1	1	1	1	1	1	0	0	1	0	1	1	1	0
P4-2	3	9	1	3	1	1	3	3	3	0	1	1	1	0	1	1	0	0	1	0	1	1	1	0
P5-1	3	9	3	3	3	3	9	3	3	0	1	1	1	1	1	1	0	0	1	0	1	1	1	0
P6-1	9	9	3	9	9	9	9	3	9	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
P6-2	9	9	3	9	9	3	3	9	3	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
P7-1	1	9	3	3	3	9	9	3	3	0	0	0	1	1	0	1	0	0	1	0	1	1	1	0
P7-2	3	9	3	3	9	3	3	3	9	0	1	1	1	1	1	1	0	0	1	0	0	1	1	0
P8-1	1	9	3	3	3	3	9	3	3	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0
P8-2	9	9	3	9	9	3	3	3	1	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0
P9-1	9	9	9	9	9	3	9	9	3	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0
P9-2	9	9	3	9	9	3	9	3	3	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0
P10-1	3	9	9	9	9	3	9	3	3	0	1	1	1	1	0	0	1	0	0	1	1	1	0	1
P11-1	9	9	9	3	9	3	9	3	3	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0
P11-3	1	9	1	1	1	9	3	3	9	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
P12-1	9	9	9	9	9	3	9	3	3	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0
P13-1	9	9	9	9	9	3	3	3	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
P13-2	9	9	3	9	9	9	1	3	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
P14-1	3	3	3	9	9	3	9	3	3	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0
P14-2	1	3	9	3	3	9	9	3	3	0	1	0	1	1	1	0	0	0	0	0	1	1	0	0
P14-3	1	9	3	3	9	3	9	3	3	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0
P15-1	9	9	9	9	9	9	3	3	3	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0
P16-1	3	9	3	3	9	9	9	1	3	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0

ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	F1	F2	F3	F4	F5	F6	F7	A1	A2	A3	A4	I1	I2	I3	I4
P17-1	1	9	1	1	1	9	9	3	9	0	1	0	1	0	0	1	0	0	1	0	1	1	1	0
P17-2	3	1	1	1	1	3	9	3	3	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
P17-3	1	1	3	1	3	9	9	3	3	0	0	0	0	1	0	1	0	0	1	0	1	0	1	0

<u>APPENDIX F</u>: EXTENDED USER STUDY ANALYSIS CODE

The following R code is used to complete the data analysis.

```
# Load Required Libraries
library(multcomp, pos=4)
library(vcd)

# Load all ratings for all sketches
# File contains all sketches with column headings
#       R1-R9: nine requirement quality ratings
#       F1-F7: seven functional conformance ratings
#       A1-A4: four activity conformance ratings
#       I1-I4: four interaction conformance ratings
#       ORAvg: sketch overall requirement score (average of R1:R9)
#       FRAvg: sketch functional requirement score (average of R1,R2,R3)
#       ARAvg: sketch activity requirement score (average of R5,R7,R8)
#       FCAvg: sketch functional conformance score (average of F1:F7)
#       ACAvg: sketch activity conformance score (average of A1:A4)
#       ICAvg: sketch interaction conformance score (average of I1:I4)

Sketch_Scores_All <- read.csv("C:/Documents and Settings/bwcaldw/My
       Documents/Research/Situatedness/User Study Ideation 2/Sketch
       Ratings/Sketch_Scores_All.csv", header=TRUE)
Sketch_Scores_All$Stud_Unique <- as.factor(Sketch_Scores_All$Stud_Unique)

# Create list of participants
Participants <- subset(Sketch_Scores_All, subset=Sketch==1,
       select=c(Stud_Unique,TRT))
Participants <- Participants[order(Participants$Stud_Unique),]

# ----- Scoring Approach: Participant Average Sketch -----

# Compute the average quality score for each participant

Sketch_Scores_All_Quality <- subset(Sketch_Scores_All,
       select=c(TRT,Stud,Stud_Unique,Sketch,ID,BlindID,R1,R2,R3,R4,R5,R6,R7,R8,R
       9,ORAvg,FRAvg,ARAvg))
Sketch_Scores_All_Quality$Stud_Unique <-
       as.factor(Sketch_Scores_All_Quality$Stud_Unique)

Sketch_Scores_All_Quality_no_NA <-
       Sketch_Scores_All_Quality[!is.na(Sketch_Scores_All_Quality$ORAvg),]
Sketch_Scores_Avg_Quality <-
       aggregate(Sketch_Scores_All_Quality_no_NA[,c("R1","R2","R3","R4","R5","R6
       ","R7","R8","R9","ORAvg","FRAvg","ARAvg"), drop=FALSE],
  by=list(Stud_Unique=Sketch_Scores_All_Quality_no_NA$Stud_Unique), FUN=mean)
Sketch_Scores_Avg_Quality <- cbind(Participants$TRT, Sketch_Scores_Avg_Quality)
names(Sketch_Scores_Avg_Quality)[c(1)] <- c("TRT")


# --- Overall Quality, Participant Average Scoring Approach ---
# Model
AnovaModel.1 <- aov(ORAvg ~ TRT, data=Sketch_Scores_Avg_Quality)
summary(AnovaModel.1)
```

```
# Descriptive Statistics
numSummary(Sketch_Scores_Avg_Quality$ORAvg ,
       groups=Sketch_Scores_Avg_Quality$TRT, statistics=c("mean", "sd",
       "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Avg_Quality$TRT), rstandard(AnovaModel.1), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.1))
min(rstandard(AnovaModel.1))
sort(rstandard(AnovaModel.1))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.1), rstandard(AnovaModel.1), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Avg_Quality$ORAvg ,
       groups=Sketch_Scores_Avg_Quality$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(ORAvg_LM), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.1))

# Perform nonparametric test
tapply(Sketch_Scores_Avg_Quality$ORAvg, Sketch_Scores_Avg_Quality$TRT, median,
       na.rm=TRUE)
kruskal.test(ORAvg ~ TRT, data=Sketch_Scores_Avg_Quality)


# --- Functional Quality, Participant Average Scoring Approach ---
# Model
AnovaModel.2 <- aov(FRAvg ~ TRT, data=Sketch_Scores_Avg_Quality)
summary(AnovaModel.2)

# Descriptive Statistics
numSummary(Sketch_Scores_Avg_Quality$FRAvg ,
       groups=Sketch_Scores_Avg_Quality$TRT, statistics=c("mean", "sd",
       "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Avg_Quality$TRT), rstandard(AnovaModel.2), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.2))
min(rstandard(AnovaModel.2))
sort(rstandard(AnovaModel.2))

# Check Model Assumptions - Constant Variance
```

```
plot(fitted(AnovaModel.2), rstandard(AnovaModel.2), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)
temp <- numSummary(Sketch_Scores_Avg_Quality$FRAvg ,
       groups=Sketch_Scores_Avg_Quality$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.2), main = "Normal Probability Plot", pch = 19)

# Compute All Pairwise Contrasts
.Pairs <- glht(AnovaModel.2, linfct = mcp(TRT = "Tukey"))
summary(.Pairs) # pairwise tests
confint(.Pairs, calpha = qt(0.975, 65)) # CI Fisher LSD, alpha=0.05, dfe=65
remove(.Pairs)

# --- Activity Quality, Participant Average Scoring Approach ---
# Model
AnovaModel.3 <- aov(ARAvg ~ TRT, data=Sketch_Scores_Avg_Quality)
summary(AnovaModel.3)

# Descriptive Statistics
numSummary(Sketch_Scores_Avg_Quality$ARAvg ,
       groups=Sketch_Scores_Avg_Quality$TRT, statistics=c("mean", "sd",
       "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Avg_Quality$TRT), rstandard(AnovaModel.3), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.3))
min(rstandard(AnovaModel.3))
sort(rstandard(AnovaModel.3))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.3), rstandard(AnovaModel.3), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Avg_Quality$ARAvg ,
       groups=Sketch_Scores_Avg_Quality$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.3), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.3))

# Perform nonparametric test
tapply(Sketch_Scores_Avg_Quality$ARAvg, Sketch_Scores_Avg_Quality$TRT, median,
       na.rm=TRUE)
kruskal.test(ARAvg ~ TRT, data=Sketch_Scores_Avg_Quality)


# Compute the average conformance score for each participant
```

```
Sketch_Scores_All_Conf <- subset(Sketch_Scores_All,
       select=c(TRT,Stud,Stud_Unique,Sketch,ID,BlindID,F1,F2,F3,F4,F5,F6,F7,FCAv
       g,A1,A2,A3,A4,ACAvg,I1,I2,I3,I4,ICAvg))
Sketch_Scores_All_Conf$Stud_Unique <-
       as.factor(Sketch_Scores_All_Conf$Stud_Unique)
Sketch_Scores_Avg_Conf <-
       aggregate(Sketch_Scores_All_Conf[,c("F1","F2","F3","F4","F5","F6","F7","F
       CAvg","A1","A2","A3","A4","ACAvg","I1","I2","I3","I4","ICAvg"),
       drop=FALSE],
  by=list(Stud_Unique=Sketch_Scores_All_Conf$Stud_Unique), FUN=mean)
Sketch_Scores_Avg_Conf <- cbind(Participants$TRT, Sketch_Scores_Avg_Conf)
names(Sketch_Scores_Avg_Conf)[c(1)] <- c("TRT")
Sketch_Scores_Avg_Conf$TRT <- as.factor(Sketch_Scores_Avg_Conf$TRT)

# --- Functional Conformance, Participant Average Scoring Approach ---
# Model
AnovaModel.4 <- aov(FCAvg ~ TRT, data=Sketch_Scores_Avg_Conf)
summary(AnovaModel.4)

# Descriptive Statistics
numSummary(Sketch_Scores_Avg_Conf$FCAvg , groups=Sketch_Scores_Avg_Conf$TRT,
       statistics=c("mean", "sd", "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Avg_Conf$TRT), rstandard(AnovaModel.4), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.4))
min(rstandard(AnovaModel.4))
sort(rstandard(AnovaModel.4))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.4), rstandard(AnovaModel.4), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Avg_Conf$FCAvg ,
       groups=Sketch_Scores_Avg_Conf$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.4), main = "Normal Probability Plot", pch = 19)

# Compute All Pairwise Contrasts
.Pairs <- glht(AnovaModel.4, linfct = mcp(TRT = "Tukey"))
summary(.Pairs) # pairwise tests
confint(.Pairs, level=0.9) # confidence intervals (TUKEY)
confint(.Pairs, calpha = qt(0.975, 65)) # CI Fisher LSD, alpha=0.05, dfe=65
remove(.Pairs)

# --- Activity Conformance, Participant Average Scoring Approach ---
# Model
AnovaModel.5 <- aov(ACAvg ~ TRT, data=Sketch_Scores_Avg_Conf)
summary(AnovaModel.5)

# Descriptive Statistics
```

```
numSummary(Sketch_Scores_Avg_Conf$ACAvg , groups=Sketch_Scores_Avg_Conf$TRT,
       statistics=c("mean", "sd", "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Avg_Conf$TRT), rstandard(AnovaModel.5), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.5))
min(rstandard(AnovaModel.5))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.5), rstandard(AnovaModel.5), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Avg_Conf$ACAvg ,
       groups=Sketch_Scores_Avg_Conf$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.5), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.5))

# Perform nonparametric test
tapply(Sketch_Scores_Avg_Conf$ACAvg, Sketch_Scores_Avg_Conf$TRT, median,
       na.rm=TRUE)
kruskal.test(ACAvg ~ TRT, data=Sketch_Scores_Avg_Conf)

# --- Interaction Conformance, Participant Average Scoring Approach ---
# Model
AnovaModel.6 <- aov(ICAvg ~ TRT, data=Sketch_Scores_Avg_Conf)
summary(AnovaModel.6)

# Descriptive Statistics
numSummary(Sketch_Scores_Avg_Conf$ICAvg , groups=Sketch_Scores_Avg_Conf$TRT,
       statistics=c("mean", "sd"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Avg_Conf$TRT), rstandard(AnovaModel.6), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.6))
min(rstandard(AnovaModel.6))
sort(rstandard(AnovaModel.6))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.6), rstandard(AnovaModel.6), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Avg_Conf$ICAvg ,
       groups=Sketch_Scores_Avg_Conf$TRT, statistics=c("mean", "sd"))
```

```
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.6), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.6))

# Remove Outliers
Sketch_Scores_Avg_Conf_no_outlier <- Sketch_Scores_Avg_Conf[c(1:4,6:69),]
AnovaModel.6b <- aov(ICAvg ~ TRT, data=Sketch_Scores_Avg_Conf_no_outlier)
summary(AnovaModel.6b)

# Descriptive Statistics
numSummary(Sketch_Scores_Avg_Conf_no_outlier$ICAvg ,
       groups=Sketch_Scores_Avg_Conf_no_outlier$TRT, statistics=c("mean", "sd",
       "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Avg_Conf_no_outlier$TRT),
       rstandard(AnovaModel.6b), xlab = "Factor Levels (FM = 1, IM = 2, NM = 3,
       PM = 4)", ylab = "Standardized Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.6b))
min(rstandard(AnovaModel.6b))
sort(rstandard(AnovaModel.6b))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.6b), rstandard(AnovaModel.6b), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Avg_Conf_no_outlier$ICAvg ,
       groups=Sketch_Scores_Avg_Conf_no_outlier$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.6b), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.6b))

# Perform nonparametric test
tapply(Sketch_Scores_Avg_Conf_no_outlier$ICAvg ,
       Sketch_Scores_Avg_Conf_no_outlier$TRT, median, na.rm=TRUE)
kruskal.test(ICAvg ~ TRT, data=Sketch_Scores_Avg_Conf_no_outlier)



# ----- Scoring Approach: Participant Best Sketch -----

# Compute the best quality score for each participant

Sketch_Scores_Best_Quality <-
       aggregate(Sketch_Scores_All_Quality_no_NA[,c("ORAvg","FRAvg","ARAvg"),
       drop=FALSE],
  by=list(Stud_Unique=Sketch_Scores_All_Quality_no_NA$Stud_Unique), FUN=max)
Sketch_Scores_Best_Quality <- cbind(Participants$TRT,
       Sketch_Scores_Best_Quality)
names(Sketch_Scores_Best_Quality)[c(1)] <- c("TRT")
Sketch_Scores_Best_Quality$TRT <- as.factor(Sketch_Scores_Best_Quality$TRT)
```

```
# --- Overall Quality, Participant Best Scoring Approach ---
# Model
AnovaModel.7 <- aov(ORAvg ~ TRT, data=Sketch_Scores_Best_Quality)
summary(AnovaModel.7)

# Descriptive Statistics
numSummary(Sketch_Scores_Best_Quality$ORAvg ,
       groups=Sketch_Scores_Best_Quality$TRT, statistics=c("mean", "sd",
       "quantiles"))
# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Best_Quality$TRT), rstandard(AnovaModel.7), xlab
       = "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.7))
min(rstandard(AnovaModel.7))
sort(rstandard(AnovaModel.7))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.7), rstandard(AnovaModel.7), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Best_Quality$ORAvg ,
       groups=Sketch_Scores_Best_Quality$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.7), main = "Normal Probability Plot", pch = 19)


# --- Functional Quality, Participant Best Scoring Approach ---
# Model
AnovaModel.8 <- aov(FRAvg ~ TRT, data=Sketch_Scores_Best_Quality)
summary(AnovaModel.8)

# Descriptive Statistics
numSummary(Sketch_Scores_Best_Quality$FRAvg ,
       groups=Sketch_Scores_Best_Quality$TRT, statistics=c("mean", "sd",
       "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Best_Quality$TRT), rstandard(AnovaModel.8), xlab
       = "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.8))
min(rstandard(AnovaModel.8))
sort(rstandard(AnovaModel.8))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.8), rstandard(AnovaModel.8), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
```

```
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Best_Quality$FRAvg ,
       groups=Sketch_Scores_Best_Quality$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.8), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.8))

# Compute All Pairwise Contrasts
.Pairs <- glht(AnovaModel.8, linfct = mcp(TRT = "Tukey"))
summary(.Pairs) # pairwise tests
confint(.Pairs, calpha = qt(0.975, 63)) # CI Fisher LSD, alpha=0.05, dfe=65
remove(.Pairs)

# Perform nonparametric test
tapply(Sketch_Scores_Best_Quality$FRAvg, Sketch_Scores_Best_Quality$TRT,
       median, na.rm=TRUE)
kruskal.test(FRAvg ~ TRT, data=Sketch_Scores_Best_Quality)
pairwise.wilcox.test(Sketch_Scores_Best_Quality$FRAvg,
       Sketch_Scores_Best_Quality$TRT, p.adjust.method = "none", paired=FALSE)


# --- Activity Quality, Participant Best Scoring Approach ---
# Model
AnovaModel.9 <- aov(ARAvg ~ TRT, data=Sketch_Scores_Best_Quality)
summary(AnovaModel.9)

# Descriptive Statistics
numSummary(Sketch_Scores_Best_Quality$ARAvg ,
       groups=Sketch_Scores_Best_Quality$TRT, statistics=c("mean", "sd",
       "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Best_Quality$TRT), rstandard(AnovaModel.9), xlab
       = "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.9))
min(rstandard(AnovaModel.9))
sort(rstandard(AnovaModel.9))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.9), rstandard(AnovaModel.9), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Best_Quality$ARAvg ,
       groups=Sketch_Scores_Best_Quality$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.9), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.9))
```

```
# Perform nonparametric test
tapply(Sketch_Scores_Best_Quality$ARAvg, Sketch_Scores_Best_Quality$TRT,
       median, na.rm=TRUE)
kruskal.test(ARAvg ~ TRT, data=Sketch_Scores_Best_Quality)


#---- Conformance Best ----

# Compute the best conformance score for each participant

Sketch_Scores_Best_Conf <-
       aggregate(Sketch_Scores_All_Conf[,c("FCAvg","ACAvg","ICAvg"),
       drop=FALSE],
  by=list(Stud_Unique=Sketch_Scores_All_Conf$Stud_Unique), FUN=max)
Sketch_Scores_Best_Conf <- cbind(Participants$TRT, Sketch_Scores_Best_Conf)
names(Sketch_Scores_Best_Conf)[c(1)] <- c("TRT")
Sketch_Scores_Best_Conf$TRT <- as.factor(Sketch_Scores_Best_Conf$TRT)

# --- Functional Conformance, Participant Best Scoring Approach ---
# Model
AnovaModel.10 <- aov(FCAvg ~ TRT, data=Sketch_Scores_Best_Conf)
summary(AnovaModel.10)

# Descriptive Statistics
numSummary(Sketch_Scores_Best_Conf$FCAvg , groups=Sketch_Scores_Best_Conf$TRT,
       statistics=c("mean", "sd", "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Best_Conf$TRT), rstandard(AnovaModel.10), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.10))
min(rstandard(AnovaModel.10))
sort(rstandard(AnovaModel.10))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.11), rstandard(AnovaModel.10), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Best_Conf$FCAvg ,
       groups=Sketch_Scores_Best_Conf$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.10), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.10))

kruskal.test(FCAvg ~ TRT, data=Sketch_Scores_Best_Conf)

# --- Activity Conformance, Participant Best Scoring Approach ---
# Model
AnovaModel.11 <- aov(ACAvg ~ TRT, data=Sketch_Scores_Best_Conf)
summary(AnovaModel.11)

# Descriptive Statistics
```

```
numSummary(Sketch_Scores_Best_Conf$ACAvg , groups=Sketch_Scores_Best_Conf$TRT,
       statistics=c("mean", "sd", "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Best_Conf$TRT), rstandard(AnovaModel.11), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.11))
min(rstandard(AnovaModel.11))
sort(rstandard(AnovaModel.11))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.11), rstandard(AnovaModel.11), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Best_Conf$ACAvg ,
       groups=Sketch_Scores_Best_Conf$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.11), main = "Normal Probability Plot", pch = 19)

# Perform nonparametric test
tapply(Sketch_Scores_Best_Conf$ACAvg, Sketch_Scores_Best_Conf$TRT, median,
       na.rm=TRUE)
kruskal.test(ACAvg ~ TRT, data=Sketch_Scores_Best_Conf)
pairwise.wilcox.test(Sketch_Scores_Best_Conf$ACAvg,
       Sketch_Scores_Best_Conf$TRT, p.adjust.method = "none")

# Compute All Pairwise Contrasts
.Pairs <- glht(AnovaModel.11, linfct = mcp(TRT = "Tukey"))
summary(.Pairs) # pairwise tests
confint(.Pairs, level=0.9) # confidence intervals (TUKEY)
confint(.Pairs, calpha = qt(0.975, 65)) # CI Fisher LSD, alpha=0.05, dfe=65
remove(.Pairs)

# --- Interaction Conformance, Participant Best Scoring Approach ---
# Model
AnovaModel.12 <- aov(ICAvg ~ TRT, data=Sketch_Scores_Best_Conf)
summary(AnovaModel.12)

# Descriptive Statistics
numSummary(Sketch_Scores_Best_Conf$ICAvg , groups=Sketch_Scores_Best_Conf$TRT,
       statistics=c("mean", "sd"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Best_Conf$TRT), rstandard(AnovaModel.12), xlab =
       "Factor Levels (FM = 1, IM = 2, NM = 3, PM = 4)", ylab = "Standardized
       Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.12))
min(rstandard(AnovaModel.12))
sort(rstandard(AnovaModel.12))
```

```
# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.12), rstandard(AnovaModel.12), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Best_Conf$ICAvg ,
       groups=Sketch_Scores_Best_Conf$TRT, statistics=c("mean", "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.12), main = "Normal Probability Plot", pch = 19)

# Remove Outliers
Sketch_Scores_Best_Conf_no_outlier <- Sketch_Scores_Best_Conf[c(1:4,6:69),]

# Model
AnovaModel.12b <- aov(ICAvg ~ TRT, data=Sketch_Scores_Best_Conf_no_outlier)
summary(AnovaModel.12b)

# Descriptive Statistics
numSummary(Sketch_Scores_Best_Conf_no_outlier$ICAvg ,
       groups=Sketch_Scores_Best_Conf_no_outlier$TRT, statistics=c("mean", "sd",
       "quantiles"))

# Check Model Assumptions - Model Fit
plot(as.numeric(Sketch_Scores_Best_Conf_no_outlier$TRT),
       rstandard(AnovaModel.12b), xlab = "Factor Levels (FM = 1, IM = 2, NM = 3,
       PM = 4)", ylab = "Standardized Residuals", pch = 20)
abline(0,0)

# Check Model Assumptions - Outliers
max(rstandard(AnovaModel.12b))
min(rstandard(AnovaModel.12b))
sort(rstandard(AnovaModel.12b))

# Check Model Assumptions - Constant Variance
plot(fitted(AnovaModel.12b), rstandard(AnovaModel.12b), xlab = "Fitted Values",
       ylab = "Standardized Residuals",
 main = "Standardized Residuals vs. Fitted Values", pch = 20)
abline(0, 0)

temp <- numSummary(Sketch_Scores_Best_Conf_no_outlier$ICAvg ,
       groups=Sketch_Scores_Best_Conf_no_outlier$TRT, statistics=c("mean",
       "sd"))
max(temp$table[,2]^2)/min(temp$table[,2]^2)

# Check Model Assumptions - Normality
qqnorm(rstandard(AnovaModel.12b), main = "Normal Probability Plot", pch = 19)
shapiro.test(rstandard(AnovaModel.12b))

# Perform nonparametric test
tapply(Sketch_Scores_Best_Conf_no_outlier$ICAvg,
       Sketch_Scores_Best_Conf_no_outlier$TRT, median, na.rm=TRUE)
kruskal.test(ICAvg ~ TRT, data=Sketch_Scores_Best_Conf_no_outlier)


# ---- Sketch Quantity ----
```

```
# Compute the number of sketches created by each participant

Sketch_Scores_Quantity <- subset(Sketch_Scores_All,
       select=c(TRT,Stud,Stud_Unique,Sketch,ID,BlindID))
Sketch_Scores_Quantity$Stud_Unique <-
       as.factor(Sketch_Scores_Quantity$Stud_Unique)
Sketch_Scores_Quantity <- aggregate(Sketch_Scores_Quantity[,c("Sketch"),
       drop=FALSE],
  by=list(Stud_Unique=Sketch_Scores_Quantity$Stud_Unique), FUN=max)

# correct for participant who skipped sketch no. 2 in packet
Sketch_Scores_Quantity[45,2]=2

Sketch_Scores_Quantity <- cbind(Participants$TRT, Sketch_Scores_Quantity)
names(Sketch_Scores_Quantity)[c(1)] <- c("TRT")
Sketch_Scores_Quantity$TRT <- as.factor(Sketch_Scores_Quantity$TRT)

# Check Distribution of Data
summary(goodfit(Sketch_Scores_Quantity$Sketch-1,type= "poisson",method=
       "MinChisq"))
summary(goodfit(Sketch_Scores_Quantity$Sketch-1,type= "poisson",method= "ML"))

# Model
GLM.1 <- glm(Sketch-1 ~ TRT, family="poisson", data=Sketch_Scores_Quantity)
summary(GLM.1)
aov(GLM.1)

# Compute All Pairwise Contrasts
.Pairs <- glht(GLM.1, linfct = mcp(TRT = "Tukey"))
summary(.Pairs) # pairwise tests
confint(.Pairs, level = 0.9) # confidence intervals
confint(.Pairs, calpha = qt(0.975, 65)) # CI Fisher LSD, alpha=0.05, dfe=65
remove(.Pairs)


#---- Quality Density (Best/Avg) ----

# Compute Quality Density

Sketch_Scores_Quality_Density <- cbind(Sketch_Scores_Best_Quality,
       Sketch_Scores_Quantity$Sketch)
names(Sketch_Scores_Quality_Density)[c(6)] <- c("Quantity")

#---- Quality Density Overall ----
# Descriptive Statistics
numSummary(Sketch_Scores_Quality_Density$ORAvg/(Sketch_Scores_Quality_Density$Q
       uantity) , groups=Sketch_Scores_Quality_Density$TRT, statistics=c("mean",
       "sd", "quantiles"))

# Perform nonparametric test
tapply(Sketch_Scores_Quality_Density$ORAvg/(Sketch_Scores_Quality_Density$Quant
       ity), Sketch_Scores_Quality_Density$TRT, median, na.rm=TRUE)
kruskal.test(ORAvg/(Quantity) ~ TRT, data=Sketch_Scores_Quality_Density)
pairwise.wilcox.test(Sketch_Scores_Quality_Density$ORAvg/(Sketch_Scores_Quality
       _Density$Quantity), Sketch_Scores_Quality_Density$TRT, p.adjust.method =
       "none")

#---- Quality Density Function ----
# Descriptive Statistics
```

```
numSummary(Sketch_Scores_Quality_Density$FRAvg/(Sketch_Scores_Quality_Density$Q
       uantity) , groups=Sketch_Scores_Quality_Density$TRT, statistics=c("mean",
       "sd", "quantiles"))

# Perform nonparametric test
tapply(Sketch_Scores_Quality_Density$FRAvg/(Sketch_Scores_Quality_Density$Quant
       ity), Sketch_Scores_Quality_Density$TRT, median, na.rm=TRUE)
kruskal.test(FRAvg/(Quantity) ~ TRT, data=Sketch_Scores_Quality_Density)
pairwise.wilcox.test(Sketch_Scores_Quality_Density$FRAvg/(Sketch_Scores_Quality
       _Density$Quantity), Sketch_Scores_Quality_Density$TRT, p.adjust.method =
       "none")

#---- Quality Density Activity ----
# Descriptive Statistics
numSummary(Sketch_Scores_Quality_Density$ARAvg/(Sketch_Scores_Quality_Density$Q
       uantity) , groups=Sketch_Scores_Quality_Density$TRT, statistics=c("mean",
       "sd", "quantiles"))

# Perform nonparametric test
tapply(Sketch_Scores_Quality_Density$ARAvg/(Sketch_Scores_Quality_Density$Quant
       ity), Sketch_Scores_Quality_Density$TRT, median, na.rm=TRUE)
kruskal.test(ARAvg/(Quantity) ~ TRT, data=Sketch_Scores_Quality_Density)
pairwise.wilcox.test(Sketch_Scores_Quality_Density$ARAvg/(Sketch_Scores_Quality
       _Density$Quantity), Sketch_Scores_Quality_Density$TRT, p.adjust.method =
       "none")
```

REFERENCES

[1] Pahl, G., Beitz, W., Feldhusen, J., and Grote, K. H., 2007, *Engineering Design: A Systematic Approach*, 3rd ed. Springer-Verlag. London.

[2] Ulrich, K. T. and Eppinger, S. D., 2008, *Product design and development*, 4th ed. McGraw-Hill. New York.

[3] Ullman, D. G., 2010, *The mechanical design process*, 4th ed. McGraw-Hill. New York.

[4] Otto, K. N. and Wood, K. L., 2001, *Product design : techniques in reverse engineering and new product development* Prentice Hall. Upper Saddle River, NJ.

[5] Brown, D. C. and Blessing, L., 2005, "The relationship between function and affordance," *17th International Conference on Design Theory and Methodology*, Long Beach, CA, United States, 24-28 Sep 2005.

[6] Chandrasekaran, B. and Josephson, J. R., 2000, "Function in Device Representation," *Engineering with computers.*, **16**(3) p. 162.

[7] Hubka, V. and Eder, W. E., 2001, "Functions Revisited," *13th International Conference on Engineering Design*, Glasgow, 21-23 August 2001.

[8] Vermaas, P. E., 2007, "The Functional Modelling Account of Stone and Wood: Some Critical Remarks," *16th International Conference on Engineering Design*, Paris, France, 28-31 August 2007.

[9] Gero, J. S., 1990, "Design prototypes: a knowledge representation schema for design," *AI Magazine*, **11**(4) pp. 26-36.

[10] Hundal, M. S., 1990, "Systematic method for developing function structures, solutions and concept variants," *Mechanism & Machine Theory*, **25**(3) pp. 243-256.

[11] Kirschman, C. F. and Fadel, G. M., 1998, "Classifying functions for mechanical design," *Journal of Mechanical Design*, **120**(3) pp. 475-482.

[12] Stone, R. B. and Wood, K. L., 2000, "Development of a Functional Basis for Design," *Journal of Mechanical Design*, **122**(4) pp. 359-370.

[13] Szykman, S., Racz, J. W., and Sriram, R. D., 1999, "The Representation of Function in Computer-Based Design," *11th International Conference on Design Theory and Methodology*, Las Vegas, Nevada, 12-15 Sept 1999.

[14] Erden, M. S., Komoto, H., van Beek, T. J., D'Amelio, V., Echavarria, E., and Tomiyama, T., 2008, "A review of function modeling: Approaches and applications," *AI EDAM*, **22**(02) pp. 147-169.

[15] Maier, J. and Fadel, G., 2009, "Affordance based design: a relational theory for design," *Research in Engineering Design*, **20**(1) pp. 13-27.

[16] Maier, J. R. A. and Fadel, G. M., 2003, "Affordance-based methods for design," *15th International Conference on Design Theory and Methodology*, Chicago, IL, United States.

[17] Galvao, A. B. and Sato, K., 2005, "Affordances in Product Architecture: Linking Technical Functions and Users' Tasks," *17th International Conference on Design Theory and Methodology*, Long Beach, California, USA, Sep 24-28, 2005.

[18] Götz, A. and Maier, T., 2007, "An Adaptive Product Development Process for Engineers and Industrial Design Engineers," *16th International Conference on Engineering Design*, Paris, France, 28-31 August 2007.

[19] van der Vegte, W. F. and Horváth, I., 2002, "Consideration and Modeling of Use Processes in Computer-Aided Conceptual Design: A State of the Art Review," *Journal of Integrated Design and Process Science*, **6**(2) pp. 25-59.

[20] van der Vegte, W. F., Vergeest, J. S. M., and Horvath, I., 2001, "Towards A Unified Description Of Product Related Processes," *J. Integr. Des. Process Sci.*, **5**(2) pp. 53-63.

[21] Green, M., Linsey, J., Seepersad, C. C., and Wood, K. L., 2006, "Frontier Design: A Product Usage Context Method," *18th International Conference on Design Theory and Methodology*, Philadelphia, Pennsylvania, USA.

[22] Green, M., Rajan, P., and Wood, K. L., 2004, "Product Usage Context: Improving Customer Needs Gathering and Design Target Setting," *16th International Conference on Design Theory and Methodology*, Salt Lake City, Utah, USA.

[23] Vermaas, P. E., 2010, "Technical Functions: Towards Accepting Different Engineering Meanings with One Overall Account," *International Symposium on Tools and Methods of Competitive Engineering (TMCE)*, Ancona, Italy, April 12-16, 2010.

[24] Hirtz, J., Stone, R. B., McAdams, D. A., Szykman, S., and Wood, K. L., 2002, "A functional basis for engineering design: Reconciling and evolving previous efforts," *Research in Engineering Design*, **13**(2) pp. 65-82.

[25] Bohm, M. R., Stone, R. B., and Szykman, S., 2005, "Enhancing virtual product representations for advanced design repository systems," *Journal of Computing and Information Science in Engineering*, **5**(4) pp. 360-72.

[26] Caldwell, B. W. and Mocko, G. M., 2008, "Towards Rules for Functional Composition," *34th Design Automation Conference*, Brooklyn, New York, August 3-6, 2008.

[27] Sen, C., 2011, *A Formal Representation of Mechanical Functions to Support Physics-Based Computational Reasoning in Early Mechanical Design*, Department of Mechanical Engineering. Clemson University, Clemson, SC.

[28] Sen, C., Caldwell, B. W., Summers, J. D., and Mocko, G. M., 2010, "Evaluation of the Functional Basis using an Information Theoretic Approach," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, **24**(1)

[29] Sen, C., 2009, *A study in the information content, consistency, and expressive power of function structures in mechanical design*, Department of Mechanical Engineering. Clemson University, Clemson, SC.

[30] Caldwell, B. W., Sen, C., Mocko, G. M., and Summers, J. D., 2010, "An Empirical Study of the Expressiveness of the Functional Basis," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM in press*,

[31] Oregon State University, Design Engineering Lab Repository. <http://repository.designengineeringlab.com>. Accessed on 10 Sep 2009.

[32] Bohm, M. R., Vucovich, J. P., and Stone, R. B., 2008, "Using a design repository to drive concept generation," *Journal of Computing and Information Science in Engineering*, **8**(1) pp. 014502-1.

[33] Vucovich, J., Bhardwaj, N., Ho, H., Ramakrishna, M., Thakur, M., and Stone, R., 2006, "Concept generation algorithms for repository-based early design," *26th Computers and Information in Engineering Conference*, Philadelphia, PA, United States, 10-13 Sep 2006.

[34] Strawbridge, Z., McAdams, D. A., and Stone, R. B., 2002, "A Computational Approach To Conceptual Design," *14th International Conference on Design Theory and Methodology*, Montreal, Canada.

[35] Bryant Arnold, C. R., Stone, R. B., and McAdams, D. A., "Memic: An interactive morphological matrix tool for automated concept generation," in *Proceedings of the IIE Annual Conference and Exposition* Vancouver, BC, Canada: Institute of Industrial Engineers, 2008.

[36] Bryant, C. R., McAdams, D. A., Stone, R. B., Kurtoglu, T., and Campbell, M. I., 2006, "A validation study of an automated concept generator design tool," *18th International Conference on Design Theory and Methodology*, Philadelphia, PA, United states.

[37] Kurtoglu, T. and Campbell, M. I., 2009, "An evaluation scheme for assessing the worth of automatically generated design alternatives," *Research in Engineering Design*, **20**(1) pp. 59-76.

[38] Kurtoglu, T., Campbell, M. I., and Linsey, J. S., 2009, "An experimental study on the effects of a computational design tool on concept generation," *Design Studies*, **30**(6) pp. 676-703.

[39] Kurtoglu, T., Swantner, A., and Campbell, M. I., "Automating the Conceptual Design Process: From Black-box to Component Selection," in *Design Computing and Cognition '08*, 2008, pp. 553-572.

[40] McAdams, D. A., Stone, R. B., and Wood, K. L., 1999, "Functional Interdependence and Product Similarity Based on Customer Needs," *Research in Engineering Design*, **11**(1) pp. 1-19.

[41] McAdams, D. A. and Wood, K. L., 2002, "A Quantitative Similarity Metric for Design-by-Analogy," *Journal of Mechanical Design*, **124**(2) pp. 173-182.

[42] Grantham Lough, K. A., Stone, R. B., and Tumer, I. Y., 2008, "Failure prevention in design through effective catalogue utilization of historical failure events," *Journal of Failure Analysis and Prevention*, **8**(5) pp. 469-481.

[43] Stone, R. B., Turner, I. Y., and Stock, M. E., 2005, "Linking product functionality to historic failures to improve failure analysis in design," *Research in Engineering Design*, **16**(1-2) pp. 96-108.

[44] Tumer, I. and Stone, R., 2003, "Mapping function to failure mode during component development," *Research in Engineering Design*, **14**(1) pp. 25-33.

[45] Vucovich, J. P., Stone, R. B., Liu, X., and Tumer, I. Y., "Risk assessment in early software design based on the software function-failure design method." vol. 1 Beijing, China: Inst. of Elec. and Elec. Eng. Computer Society, 2007, pp. 405-412.

[46] Mitchell, B. A., McAdams, D. A., Stone, R. B., and Tumer, I. Y., "Computational methods to predict and avoid design failure," 2 ed. vol. 118B Orlando, FL, United states: American Society of Mechanical Engineers, 2005, pp. 721-732.

[47] Stone, R. B., Tumer, I. Y., and Van Wie, M., 2005, "The function-failure design method," *Journal of Mechanical Design, Transactions of the ASME,* **127**(3) pp. 397-407.

[48] Hutcheson, R. S., McAdams, D. A., Stone, R. B., and Tumer, I. Y., 2008, "Effect of Model Element Fidelity Within a Complex Function-Based Behavioral Model," *28th Computers and Information in Engineering Conference*.

[49] Hutcheson, R. S., McAdams, D. A., Stone, R. B., and Tumer, I. Y., 2007, "Function-Based Behavioral Modeling," *19th International Conference on Design Theory and Methodology*.

[50] Nagel, R. L., Midha, P. A., Tinsley, A., Stone, R. B., McAdams, D. A., and Shu, L. H., 2008, "Exploring the use of functional models in biomimetic conceptual design," *Journal of Mechanical Design,* **130**(12)

[51] Stroble, J. K., Watkins, S. E., Stone, R. B., McAdams, D. A., and Shu, L. H., "Modeling the cellular level of natural sensing with the functional basis for the design of biomimetic sensor technology," Piscataway, NJ, USA: IEEE, 2008, pp. 27-32.

[52] Tinsley, A., Midha, P. A., Nagel, R. L., McAdams, D. A., Stone, R. B., and Shu, L. H., "Exploring the use of functional models as a foundation for biomimetic conceptual design." Las Vegas, NV, United states: American Society of Mechanical Engineers, 2008, pp. 79-92.1.

[53] Cheong, H., Shu, L. H., Stone, R. B., and McAdams, D. A., 2008, "Translating Terms of the Functional Basis Into Biologically Meaningful Keywords," *20th International Conference on Design Theory and Methodology*, Brooklyn, NY.

[54] Kurfman, M. A., Stone, R. B., VanWie, M., Wood, K. L., and Otto, K. N., 2000, "Theoretical underpinnings of functional modeling: preliminary experimental studies," *12th International Conference on Design Theory and Methodology*, Baltimore, Maryland, USA, Sept 10-13, 2000.

[55] Stone, R. B., Wood, K. L., and Crawford, R. H., 2000, "Using quantitative functional models to develop product architectures," *Design Studies,* **21**(3) pp. 239-260.

[56] Caldwell, B. W., Sen, C., Mocko, G. M., Summers, J. D., and Fadel, G. M., 2008, "Empirical examination of the functional basis and design repository," *Third International Conference on Design Computing and Cognition*, Atlanta, USA, June 23-25, 2008.

[57] Sen, C., Summers, J. D., and Mocko, G. M., 2010, "Toward a Formal Representation of the Functional Basis Verbs," *International Symposium on Tools and Methods of Competitive Engineering (TMCE)*, Ancona, Italy, April 12-16, 2010.

[58] Bohm, M. R. and Stone, R. B., 2004, "Representing functionality to support reuse: Conceptual and supporting functions," *24th Computers and Information in Engineering Conference*, Salt Lake City, UT, United States.

[59] Kostovich, V., McAdams, D. A., and Moon, S. K., 2009, "Representing User Activity and Product Function for Universal Design," *14th Design for Manufacturing and the Life Cycle Conference*, San Diego, California, Aug 29 - Sep 2, 2009.

[60] Caldwell, B. W., 2009, *An Evaluation of Function-Based Representations and Information Archival in Engineering Design*, Department of Mechanical Engineering. Clemson University, Clemson, SC.

[61] Kim, Y. S., Lee, S., Park, J. J., Kim, M. K., and Kim, M., 2009, "Study on Personal Characteristics and Affordance Perception: Another Case Study," *17th International Conference on Engineering Design*, Stanford, California, USA, Aug 24-27, 2009.

[62] Maier, J. R. A., Ezhilan, T., and Fadel, G. M., 2007, "The affordance structure matrix - A concept exploration and attention directing tool for affordance based design," *19th International Conference on Design Theory and Methodology*, Las Vegas, NV, United states.

[63] Nagel, R. L., Stone, R. B., Hutcheson, R. S., McAdams, D. A., and Donndelinger, J. A., 2008, "Function Design Framework (FDF): Integrated Process and Function Modeling for Complex Systems," *20th International Conference on Design Theory and Methodology*, Brooklyn, NY, USA, August 3-6, 2008.

[64] Warell, A. V., 1999, "Introducing a Use Perspective in Product Design Theory and Methodology," *11th International Conference on Design Theory and Methodology*, Las Vegas, Nevada, 12-15 Sept 1999.

[65] Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., and Leifer, L. J., "Engineering design thinking, teaching, and learning," 1 ed. vol. 94: American Society for Engineering Education, 2005, pp. 103-119.

[66] Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K., 2011, "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples," *Journal of Mechanical Design,* **133**(8) pp. 081004-11.

[67] Chiu, I. and Shu, L. H., 2008, "Use of opposite-relation lexical stimuli in concept generation," *CIRP Annals - Manufacturing Technology,* **57**(1) pp. 149-152.

[68] Genco, N., D. Johnson, K. Holtta-Otto, and C. C. Seepersad, 2011, "A Study of the Effectiveness of the Empathic Experience Design Creativity Technique," *23rd International Conference on Design Theory and Methodology*, Washington, D.C., Aug 28-31, 2011.

[69] Helms, M., Vattam, S. S., and Goel, A. K., 2009, "Biologically inspired design: process and products," *Design Studies,* **30**(5) pp. 606-622.

[70] Howard, T. J., Culley, S., and Dekoninck, E. A., 2011, "Reuse of ideas and concepts for creative stimuli in engineering design," *Journal of Engineering Design,* **22**(8) pp. 565-581.

[71] Linsey, J. S., Tseng, I., Fu, K., Cagan, J., Wood, K. L., and Schunn, C., 2010, "A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty," *Journal of Mechanical Design*, **132**(4) pp. 041003-12.

[72] Chiu, I. and Shu, L. H., 2007, "Using language as related stimuli for concept generation," *(AI EDAM) Artificial Intelligence for Engineering Design, Analysis and Manufacturing,* **21**(02) pp. 103-121.

[73] Fu, K., Cagan, J., and Kotovsky, K., 2010, "Design Team Convergence: The Influence of Example Solution Quality," *Journal of Mechanical Design*, **132**(11) pp. 111005-11.

[74] Wilson, J. O., Rosen, D., Nelson, B. A., and Yen, J., 2009, "The effects of biological examples in idea generation," *Design Studies,* **31**(2) pp. 169-186.

[75] Wodehouse, A. and Ion, W., 2011, "Augmenting the 6-3-5 method with design information," *Research in Engineering Design*, pp. 1-11.

[76] Richardson, J. L., 2010, *Incorporating function structures into morphological charts a user study*, Department of Mechanical Engineering. Clemson University, Clemson, SC.

[77] Smith, G. P., 2007, *Morphological charts: a systematic exploration of qualitative design space*, Department of Mechanical Engineering. Clemson University, Clemson, SC.

[78] Thomas, J., Sen, C., Mocko, G. M., Summers, J. D., and Fadel, G. M., 2009, "Investigation of the Interpretability of Three Function Structure Representations: A User Study," *21st International Conference on Design Theory and Methodology*, San Diego, California, Aug 29 - Sep 2, 2009.

[79] Shah, J. J., Smith, S. M., and Vargas-Hernandez, N., 2003, "Metrics for measuring ideation effectiveness," *Design Studies*, **24**(2) pp. 111-134.

[80] Frey, D. and Dym, C., 2006, "Validation of design methods: lessons from medicine," *Research in Engineering Design*, **17**(1) pp. 45-57.

[81] Thomas, J. E., 2010, *Interpretability analysis of function structures at various levels of abstraction: a user study*, Department of Mechanical Engineering. Clemson University, Clemson, SC.

[82] Stacey, M. and Eckert, C., 2003, "Against Ambiguity," *Computer Supported Cooperative Work (CSCW)*, **12**(2) pp. 153-183.

[83] Linsey, J. S., Wood, K. L., and Markman, A. B., 2008, "Modality and representation in analogy," *(AI EDAM) Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **22**(2) pp. 85-100.

[84] Goel, A. K. and Bhatta, S. R., 2004, "Use of design patterns in analogy-based design," *Advanced Engineering Informatics*, **18**(2) pp. 85-94.

[85] Thebeau, R. E., 2001, *Knowledge management of system interfaces and interactions from product development processes*, System Design and Management Program. Massachusetts Institute of Technology,

[86] Schultz, J. C., Sen, C., Caldwell, B. W., Mathieson, J. L., and Summers, J. D., 2010, "Limitations to Function Structures: A Case Study in Morphing Airfoil Design," *submitted to ASME International Design Engineering Technical Conference*, Montreal, Quebec, Canada, August 15-18, 2010.

[87] Ramachandran, R., 2011, *Understanding the role of functions and interaction in the product design*, Department of Mechanical Engineering. Clemson University, Clemson, SC.

[88] Ramachandran, R., Caldwell, B. W., and Mocko, G. M., 2011, "A User Study to Evaluate the Function Model and Function Interaction Model for Concept Generation," *23rd International Conference on Design Theory and Methodology*, Washington, D.C., Aug 28-31, 2011.

[89] Caldwell, B. W., Mocko, G. M., and Fadel, G. M., 2010, "A Representation of Artefacts and Interactions to Supplement Function," *International Symposium on Tools and Methods of Competitive Engineering (TMCE),* Ancona, Italy, April 12-16, 2010.

[90] Cohen, J., 1960, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement Educational and Psychological Measurement*, **20**(1) pp. 37-46.

[91] Landis, J. R. and Koch, G. G., 1977, "The Measurement of Observer Agreement for Categorical Data," *Biometrics,* **33**(1) pp. 159-174.

[92] Kurfman, M. A., Stock, M. E., Stone, R. B., Rajan, J., and Wood, K. L., 2003, "Experimental studies assessing the repeatability of a functional modeling derivation method," *Journal of Mechanical Design,* **125**(4) pp. 682-693.

[93] Van Wie, M., Bryant, C. R., Bohm, M. R., McAdams, D. A., and Stone, R. B., 2005, "A model of function-based representations," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM,* **19**(2) pp. 89-111.

[94] Blessing, L. T. M. and Chakrabarti, A., 2009, *DRM, a design research methodology* Springer. Dordrecht; London.

[95] R Development Core Team, 2010, *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing. Vienna, Austria.

[96] Torsten Hothorn, Frank Bretz, and Peter Westfall, 2008, "Simultaneous Inference in General Parametric Models," *Biometrical Journal,* **50**(3) pp. 346-363.

[97] Meyer, D., Zeileis, A., and Hornik, K., 2011, *vcd: Visualizing Categorical Data. R package version 1.2-12.*

[98] Pfaffenberger, R. C. and Patterson, J. H., 1987, *Statistical methods for business and economics* Irwin. Homewood, Ill.

[99] Ott, L. and Longnecker, M., 2001, *An introduction to statistical methods and data analysis* Duxbury. Australia; Pacific Grove, CA.