Stakeholder mental model alignment influence on mid-stage performance of new product engineering teams

by

Nathan E. Krehbiel

Bachelor of Science in Mechanical Engineering
Kansas State University, 2004

Master of Business Administration
Texas Christian University, 2009

Submitted to the System Design and Management Program
in Partial Fulfilment of the Requirements for the Degree of

Master of Science in Engineering and Management
at the
Massachusetts Institute of Technology
May 2022

Signature of the author____
Nathan E. Krehbiel
System Design and Management Program
May 5, 2022

Certified by____
Dr. Bryan R. Moser
Academic Director & Senior Lecturer, System Design and Management Program
Thesis Supervisor

Accepted by____
Joan S. Rubin

Executive Director, System Design & Management Program

Page intentionally left blank

Stakeholder mental model alignment influence on mid-stage performance of new product engineering teams

by

Nathan E. Krehbiel

Submitted to the System Design and Management Program on May 5, 2022
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering and Management

Abstract

The engineering of complex systems-of-systems requires management of dependencies and coordination between multidisciplinary teams-of-teams. Coordination strategies and the understanding of supporting mechanisms are critical in the execution. Mental models are cognitive structures used to describe system form and function and predict outcomes. The concept of shared mental models, the influence of shared mental models, and the methods of elicitation and analysis are all relatively recent and active areas of study. This thesis reviews past work in these domains and proposes a treatment to stimulate the development of two attributes of shared mental models in the context of mid-stage project development of R&D teams. This work explores the influence of structured context tools on *sharedness* and *breadth* of team shared mental models and in turn, team shared mental model influence on team performance within the mid-stage R&D context. An experiment was designed placing participants in a team-based, role-play scenario where they were asked to work as a team-of-teams developing a system-of-systems, in particular the verification and validation activity of that development. Information was collected on team mental models utilizing a concept similarity rating elicitation method and analyzed utilizing Pathfinder network and pairwise comparison methods. Performance data was collected via custom software and analyzed utilizing a Pareto fitness ranking method. Trends were detected in the data indicating the importance of shared mental model development. Limitations, recommendations, and future areas of study are provided, including recommendations to: adjust the application of the treatment, include additional instrumentation in the experiment, and explore additional attributes of the teams' mental models.

Thesis Supervisor: Dr. Bryan R. Moser
Title: Academic Director & Senior Lecturer, System Design and Management Program

Acknowledgements

I am deeply grateful to Dr. Bryan Moser, Academic Director and Senior Lecturer, Director of the Global Teamwork Lab, and advisor to this thesis. The motivation, encouragement, support, and insights you provided have been invaluable and created a very cherished experience.

Ignacio Salvador Vazquez Rodarte, Dipti Garg, and Aparna Ravikumar Kulkarni, thank you for the collaboration and support to execute the experiment associated with this work. It was an honor and a pleasure working with you.

To all the volunteers that supported and participated in the experiment, you have my sincere gratitude.

Thank you to all my fellow students and colleagues at GTL and SDM for your support and contributions to the experience and the culture at MIT.

Thank you to all the SDM staff. Your relentless pursuit of perfecting the SDM program is evident in the experience.

To my wife Aileen, my children Madeline, Alexander, and Bryson, this is not possible without your support, your sacrifices, and your love.

Contents

List of Figures

List of Tables

Glossary

Mental model – "… a mental model is a cognitive structure or network of associations between concepts in each individual's mind." (Ward and Reingen, 1990)

Sharedness – "knowledge that is held by all members of a team" Badke-Schaub

Shared mental model – "A team mental model refers to an organized understanding or mental representation of knowledge that is shared by team members" (Cannon-Bowers et al, 1993)

Solution space – The set of all possible designs under the constraints of a model.

Problem space – The set of all possible model performance outcomes achievable from the set of designs within the solution space.

Design Structure Matrix (DSM) – "The DSM is a network modeling tool used to represent the elements comprising a system and their interactions, thereby highlighting the system's architecture (or designed structure)." (Eppinger and Browning, 2012)

Pareto fitness rank – Successive ordinal ranking and removal scheme in which all pareto optimal solutions among a set are assigned a rank before removal for subsequent ranking steps

Pareto optimal – a state in which a solution cannot be improved upon by measure of any single evaluation metric without degrading at least one other evaluation metric

Chapter 1

Intro (Background & Related Work)

Teamwork has become ubiquitous in the workplace. The development of complex systems requires expertise in a multitude of disciplines. The depth of one's understanding of a system only highlights greater complexity. The complexity of the system under development is often paralleled by the complexity of the sociotechnical system executing that development. It is the team that is part of that development system that is the focus of this study. The pervasiveness of teamwork has required understanding of factors that lead to the emergence of team performance and spawned decades of research. The introductory section of this document will provide a brief overview of some of this research to provide context to the research questions under study and the motivation thereof.

First, defining "team" – The definition of a team generally includes three elements: multiple members with interdependencies and a common goal.

In 2015, Eduardo Salas et al. provided what they describe as an "…overarching, practical heuristic of the most critical considerations for teamwork." (Salas et al., 2015). Within this heuristic, they discuss six core processes & emergent states, including cooperation, coordination, cognition, conflict, coaching, and communication. They also discuss three influencing conditions including context, composition, and culture. The heuristic is described as not all-encompassing but rather a consolidated set of key findings and the most critical considerations in the reviewed literature. Salas et al. are clear in their statements that there is no hierarchy or sequence as the factors are highly interdependent.

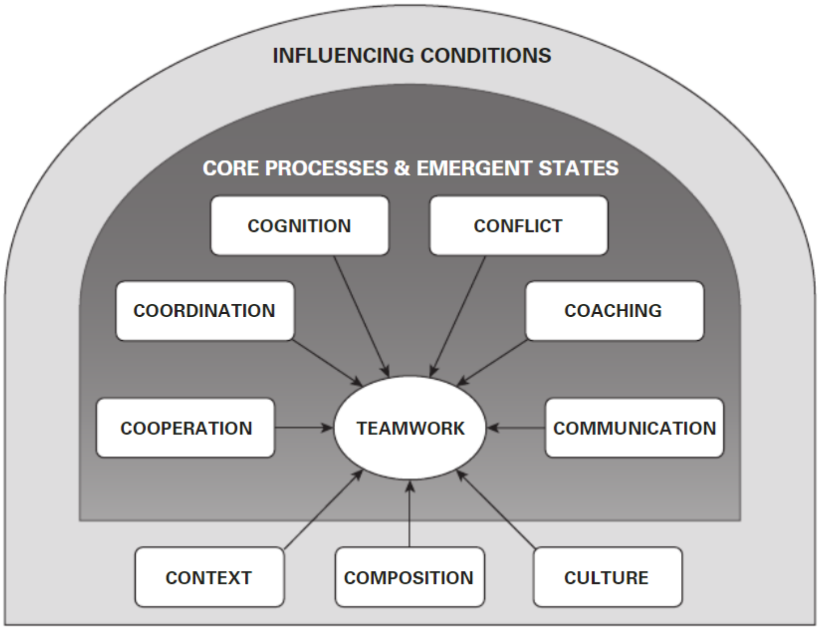


Figure 1: Heuristic of the Critical Considerations of Teamwork (Salas et al., 2015)

While remaining cognizant of these interdependencies, this thesis focuses on the coordination element of this system. Many definitions of coordination have been proposed, including "The enactment of behavioral and cognitive mechanisms necessary to perform a task and transform team resources into outcomes." (Salas et al. 2015), "Coordination involves orchestrating the sequence and timing of interdependent actions" (Marks et al, 2001), and "Coordination is managing dependencies among activities." (Malone and Crowston, 1994). These defining statements underscore commonly held notions of the value of coordination. As the complexity of the task increases, the impact of coordination on team performance also increases (Butchibabu et al., 2016) and coordination is further emphasized in its importance when working within a team-of-teams environment. (Mathieu et al., 2008).

The motivation of this study is driven by both personal interest and a lack of identified comprehensiveness in the existing literature. The majority of literature discovered in the review generally focused on team performance in situations of defined tasks to be executed with certain environmental uncertainties. Though many of the learnings are translatable to the R&D

environment to varying degrees, this has only started to be explored in recent years and adding uncertainty to the task goal is of particular interest.

The uniqueness of the R&D environment in this context is highlighted by Thayer et al. who reviewed the challenges faced by innovative teams (Thayer et al. 2018). They discuss idea creation and idea implementation stages of innovation, and the paradox between the differing knowledge, skills, abilities, and other attributes that are antecedent to performance at each stage. They define the innovation process and the critical considerations of innovative teams to enable navigation between the divergent thinking associated with creative idea generation and the convergent thinking associated with implementation. Thayer et al. discuss the benefits of ambidextrous organizations and leaders navigating the transition between these intertwined and interdependent stages. This transition phenomenon further focuses the subject of this thesis to mid-cycle stages of development.

A topic being explored within much of the reviewed literature and this study is that the underlying phenomenon of team coordination influences performance. In Professor Bryan Moser's Global Teamwork Lab (GTL) at the Massachusetts Institute of Technology such phenomena are studied within teams-of-teams (meso-scale level) working on systems-of-systems. The meso-scale is differentiated from the individual team (micro-scale) level and the total project/portfolio/organization (macro-scale) level. This study focuses on interactions at that meso-scale level and the influence of those interactions on coordination phenomenon. Many factors have been suggested as causal in the emergence of coordination. The motivation for this study draws on the elements of "behavioral and cognitive mechanisms" and "manage dependencies" in the defining statements above. In doing so, the convergence of shared mental models was selected as the focus of investigation. Can a focus on shared mental models more effectively transition a team from divergent idea generation to convergent implementation? Can a team use a structured context tool to converge on a shared mental model more efficiently? Can these items improve team performance in the mid-cycle phases of an R&D project?

Given these motivations, questions, and narrowed focus, the following section investigates current literature in the areas of coordination influence on team performance and shared mental model influence on coordination.

Literature Review

Influence of coordination on team performance

In 1976 Van de Ven et al. proposed three classifications for coordination modes of work activity in an organization (Van de Ven et al. 1976). These categories included impersonal, personal, and group. The impersonal mode (or programming mode) includes plans, schedules, rules, policies, and procedures. The personal mode (or feedback mode) includes adjustments based on feedback conducted by an individual in either vertical or horizontal communication chains. The group mode includes adjustments based on feedback conducted by a group in either formal or informal settings. The study evaluated various factors that may explain the use of alternative mechanisms of coordination. Amongst the findings was that "…as task interdependence increases, more elaborate coordination mechanisms are required…".

As noted above, Malone and Crowston define coordination as the process of managing dependencies among activities (Malone and Crowston, 1994). They promote the value of characterizing dependencies to facilitate selection between a cross-disciplinary set of methods for managing them and emphasize application in complex systems. They highlight the need to adapt coordination methods driven by advancements in enabling technologies. They state that key to the enablement of the proposal is group decision-making and communication with a common language and standards with a goal of common knowledge.

In 1999 Entin et al. tested the adaptive nature of team coordination. Through experimentation, they demonstrated that effective coordination strategies can be trained. The strategies allow teams to maintain performance levels in high-stress situations and draw on shared mental models to shift to implicit communication and reduce coordination overhead.
The elements of the enabling training program included "…(a) preplanning, (b) use of idle periods, (c) favoring information transmission over action/task coordination, (d) anticipation of

information needs (implicit communication), and (e) dynamic redistribution of workload among team members." The training coordination strategies help to exercise shared mental models. This is an example of programming feedback. A set of rules and procedures are used to guide and improve efficiency of information transfer during feedback to facilitate adjustments. Along similar lines, this thesis aims to evaluate the addition of a support tool for coordination similar to the guides described by Entin.

Concurrent engineering requires an overlap of development activities. Dependent processes run in parallel, enabling reduction of overall project timelines but presenting risk. Strategically managing information exchange and the tradeoff between precision and stability of that information can help mitigate the costs incurred by downstream users from work starvation and rework (Terwiesch et al., 2002). Starvation can occur when downstream users must wait on the necessary precision before beginning dependent tasks, and rework can occur due to a lack of stability of the received information. These teamwork effects can be traded off. If precision is offered at the expense of stability, rework risk is increased. The correct strategy for a given situation should be selected based on the relative cost of starvation vs. rework. The quantification of these costs can be the source of great debate and begs several questions. Can specific information exchange be interjected to accelerate multiple parties to agree on the stability vs. precision of a given data set and collectively be more accurate in the assessment? How efficiently can the understanding of the context of the problem from each discipline be aligned/calibrated to collectively establish bounds on the risk and stability of modeling assumptions? A team-of-teams can develop concurrently with structured information exchanges without a wholistic understanding of the system under development. Can structured contextual information effectively motivate a shift from task concurrency to a mode of cross-disciplinary support with implicit communication around system dependencies?

In 2015 Moser et al. explored a broadened view of dependencies in coordination and complex engineering projects (Moser et al. 2015). They expanded beyond the traditional consideration of task sequencing and highlighted the dependence of system consequence on scope, quality, schedule, and cost. Where Malone and Crowston highlighted new coordination methods created by technology, Moser et al. explore the emergence of challenges in team dynamics generated by dispersion of teams and automation of workflow coupled with increasingly complex systems. These challenges can have a negative effect on teams' abilities to anticipate needs within the team and are a result of narrowed focus on structured tasks rather than on interactions across dependencies. Moser et al. present a framework modeling dependence as a constraint on independent performance, e.g., if two tasks are independent, then each can progress without burden from the other. Dependent tasks should generate demands, and if needs are not satisfied, the process should trigger corrective action or "exception handling behavior." Maintaining continuous unconstrained progress requires effective coordination. Coordination requires awareness, attention, allocation, abilities, and experience to be well performed.

The image below from Moser et al.'s "Mechanisms of Dependence in Engineering Projects as Sociotechnical Systems." (Moser et al. 2015) depicts the mechanism for the satisfaction of dependence. Awareness is cited as one of the significant factors determining if and how interaction occurs. The classical means of bringing awareness to demands include tools mainly in the programming mode of coordination like PERT, DSM, IDEF, etc., which can generate too narrow of a focus and misrepresent the interactions. This thesis will explore the supplemental value of shared mental models and the development of a structured context tool. Can such a tool help address some of the challenges presented in this paper?

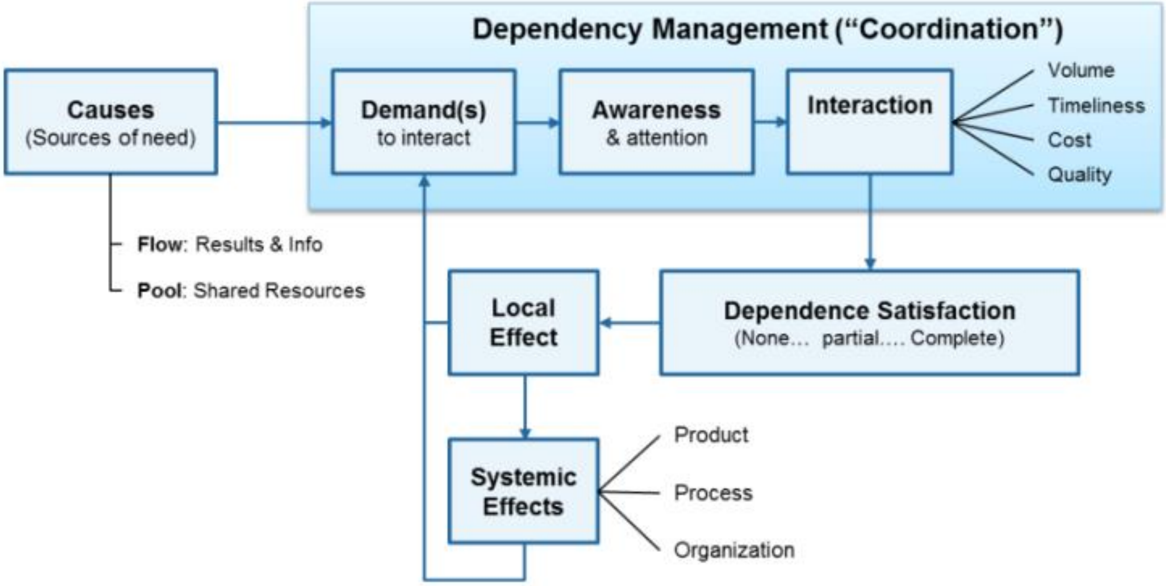


Figure 2: Mechanisms of Dependence from Cause to System Effects (Moser et al., 2015)

Influence of shared mental models on coordination

This review now shifts to explore mental models as a variable for coordination.

Trevelyan states, "…the prominence of coordination strongly suggests that engineering relies on a social process at the microscopic level of individual interactions between people… Coordination also relies on an accurate appreciation of both individual and shared technical knowledge domains and ways to represent technical knowledge in the working context." (Trevelyan, 2007)

Group decision making, common knowledge, shared understanding, shared notions, collective consciousness – the concept of shared mental models, if not explicitly stated, is prevalent in literature on coordination.

Definitions…
Mental model –
"At the simplest level, a mental model is a cognitive structure or network of associations between concepts in each individual's mind." (Ward and Reingen, 1990)

"…to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states." (Rouse and Morris, 1986)

Sharedness –
"…knowledge that is held by all members of a team." (Badke-Schaub et al., 2007)

Shared mental model –
"A team mental model refers to an organized understanding or mental representation of knowledge that is shared by team members" (Cannon-Bowers et al., 1993)

Cannon-Bowers et al. describe shared mental models and their importance on expectations and decision making (Cannon-Bowers et al., 1993). Mental models are described as knowledge structures that enable teams to explain and predict system behavior. They generate value by enabling team members to coordinate and adapt to the task, environment, and team demands. The level of value is related to the ability of the model to generate accurate explanations and expectations of the task and team. Cannon-Bowers et al. do not offer a definitive position on what models are required but compile a taxonomy of common mental models cited as influencing team performance that include: task, equipment, team, and team interaction. These are not independent and the appropriateness of the focus within these or others is likely dependent on the task and situation.

Mental models are constructs capable of running hypothetical scenarios and guiding behavior in active scenarios. Experience alone does not guarantee accurate mental model development (Cannon-Bowers et al., 1993). Training with conceptual models of a system can improve the accuracy of mental models (Mayer, 1989). There is also empirical evidence that teams that engage in high-quality planning form better team interaction shared mental models (Stout, 1999) These findings motivate hypotheses that the process of creating and sharing as well as the resulting artifact of a structured context tool may influence mental models.

When working within a multidisciplinary team on complex systems, there is not likely to be a single mental model but rather a distribution of mental models (Cooke et al., 2000). This is supported by a transactive memory system and team interaction mental model. Transactive memory systems are “the collection of knowledge possessed by each team member and a collective awareness of who knows what” (Mathieu et al., 2008). Multidisciplinary R&D teams working on complex systems commonly start a project with a sizeable conceptual solution space and then narrow it for early parallel development of a few solution sets before converging on a single path. The solution space is influenced by team members’ areas of expertise and experiences. As referenced above, this divergent thinking can be beneficial in the early stages. However, when converging on a single concept architecture, if team members continue to utilize divergent mental models, divergent thinking can generate confusion (Badke-Schaub et al., 2007). Badke-Schaub et al. propose three factors that influence the quality of mental models:

Sharedness, accuracy, and importance. The appropriateness of the focus or weighting is, again, likely situation-specific. For example, tasks with a defined set of optimal strategies are more influenced by accuracy (Edwards et al., 2006) and within novel environments sharedness could be more important (Marks et al., 2000).

Chapter 2

Research Question & Hypotheses

To summarize the findings in Chapter 1:

- Coordination is a critical consideration for team performance.
- Innovative teams need to be able to navigate the transition between idea generation and implementation stages of development.
- Coordination involves managing dependencies.
- The management of these dependencies influences program cost, quality, and timeline.
- Traditional up-front planning tools alone can drive too narrow of a focus.
- Mental models allow a team to explain and predict system behavior.
- Collaborative planning promotes the development of mental models.

If a team is aligned on how a system is to behave, perhaps confusion around the demand for dependencies can be mitigated, and coordination improved. Perhaps a structured exercise to represent system architecture can accelerate and improve this alignment. Directly stated:

Research Question: Shared mental models are an underlying factor in team coordination that influences team performance. Can discipline-specific, task operational concepts accelerate the development of shared mental models and alignment on execution plans for the middle phases of R&D projects?

In order to test these questions, the following hypotheses are proposed:

H1: Mid-stage performance of R&D teams increases with the **level of sharedness of mental models.**

H2: Mid-stage performance of R&D teams increases with increasing levels of **comprehensive breadth of shared mental models** (i.e., teams' mental models that broadly consider elements of all team member disciplines will perform better).

H3: Teams that develop and socialize specific, **structured context tools** will **converge faster on mental models** than those that socialize concepts in a freeform nature.

H4: Teams that develop and socialize specific, **structured context tools** will develop more **holistic mental models** than those that socialize concepts in a freeform nature.

Chapter 3

Research Method

Context and Definitions

The data collected to test the proposed hypothesis was collected through an empirical study. In order to structure this study, a few essential items required definition/selection: 1) structured context tool, 2) mid-stage, and 3) breadth.

For this study, the structured context tool selected was a combination of an Operational Concept Document (OPSCON) and a context diagram. An OPSCON is "An abstract model of the operations of a specific system or group of systems, usually developed as part of the acquisition process and used throughout the system life cycle." (Rebentisch, 2017) The OpsCon is used to capture stakeholder needs, capture expectations, form requirements, and develop architecture. It is typically developed early in a program to help reveal items that might otherwise be overlooked. The study is not proposing moving this activity to the mid-stages of a program but instead replicating the activity at the lower system level at the mid-stages. Elements mentioned such as needs, expectations, and requirements highlight dependencies present amongst stakeholders that the act of discussing or making explicit may help with the development of mental models. A context diagram highlights interactions with external domains relevant to the system (NASA, 2007).

As mentioned above, the mid-stage of an R&D project is intended to focus on the segment of the project where the team must transition from idea generation to implementation. To add context, the classical V-model is referenced. As indicated in the image below, this model is intended to be recursive. However, the transition from divergent to convergent modes is likely required near the bottom of the V, where a shift is made between project and concept definition to verification and commissioning. This study, therefore, focuses on the V&V planning activity.

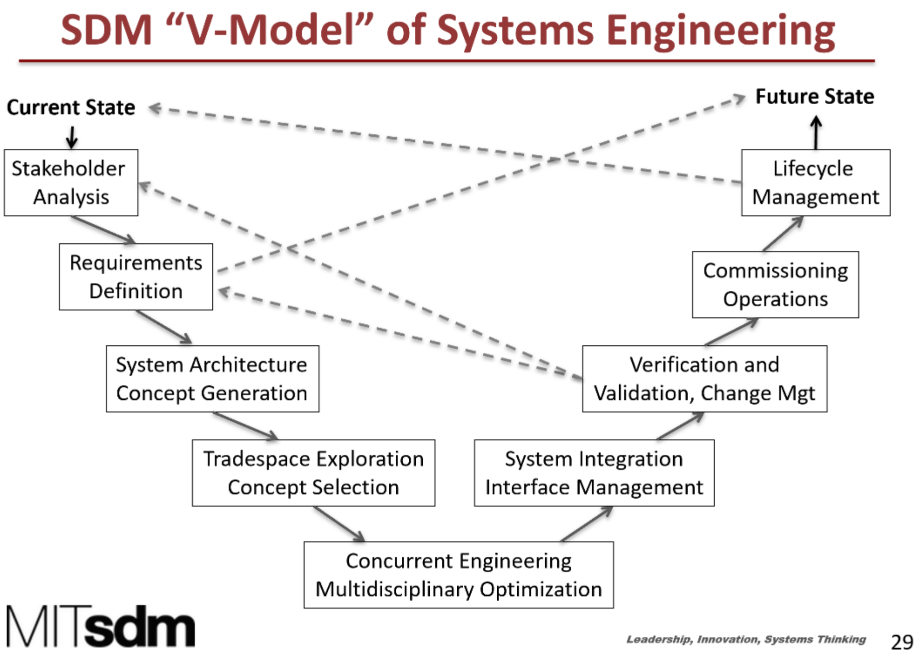


Figure 3: "V-Model" of Systems Engineering (Rebentisch, 2017)

As mentioned above, importance has been identified as a factor influencing the quality of mental models (Badke-Schaub et al., 2007). Badke-Schaub et al. propose centrality, or the number of links an attribute has in a network representation of a mental model, as quantification of importance. In a somewhat related manner, the metric of breadth is being introduced in this study to capture the level of consideration an individual gives across sub-teams in the team-of-teams scenario during development.

Experiment Design

Overview

To test the hypotheses mentioned above, an experiment was designed to enable individual mental model assessment pre- and post-activity and continuous team level performance tracking throughout. The design included a single treatment and placed participants into a team-based, role-play scenario where teams-of-teams were asked to develop a system-of-systems. The experiment was executed in two phases. In phase 1, participants were randomly assigned to teams (separated spatially within the room), assigned an area of focus (discipline) for their team, and asked to work within that context toward stated objectives. The software system with which participants interacted to execute their work structured the decisions for and limited content to the teams' respective areas of focus. In phase 2, participants were reassigned to a team that included members from all disciplines. Objectives for participants during phase 2 were consistent with that of phase 1, except participants were asked to work in the greater context of all disciplines. The software interface was adjusted for phase 2 to include work content for all disciplines. The model problem for the experiment was introduced during phase 1 after teams were formed and just prior to beginning work. The model problem was introduced via pre-recorded video, and all introductory content was also included in packets of information received by participants upon their arrival at the workshop. The packets also included the initial survey, and treatment groups were provided a context tool specific to their assigned discipline within their packets.

The experiment was designed around a generic engineering product development project. The design intent was to separate the focus between sub-teams enough to:

- Provide an appropriate gap requiring convergence between sub-system teams during phase 2, i.e., provide discrete cognitive, team relevant concepts
- Maintain a tractable case for consideration during a workshop timeframe

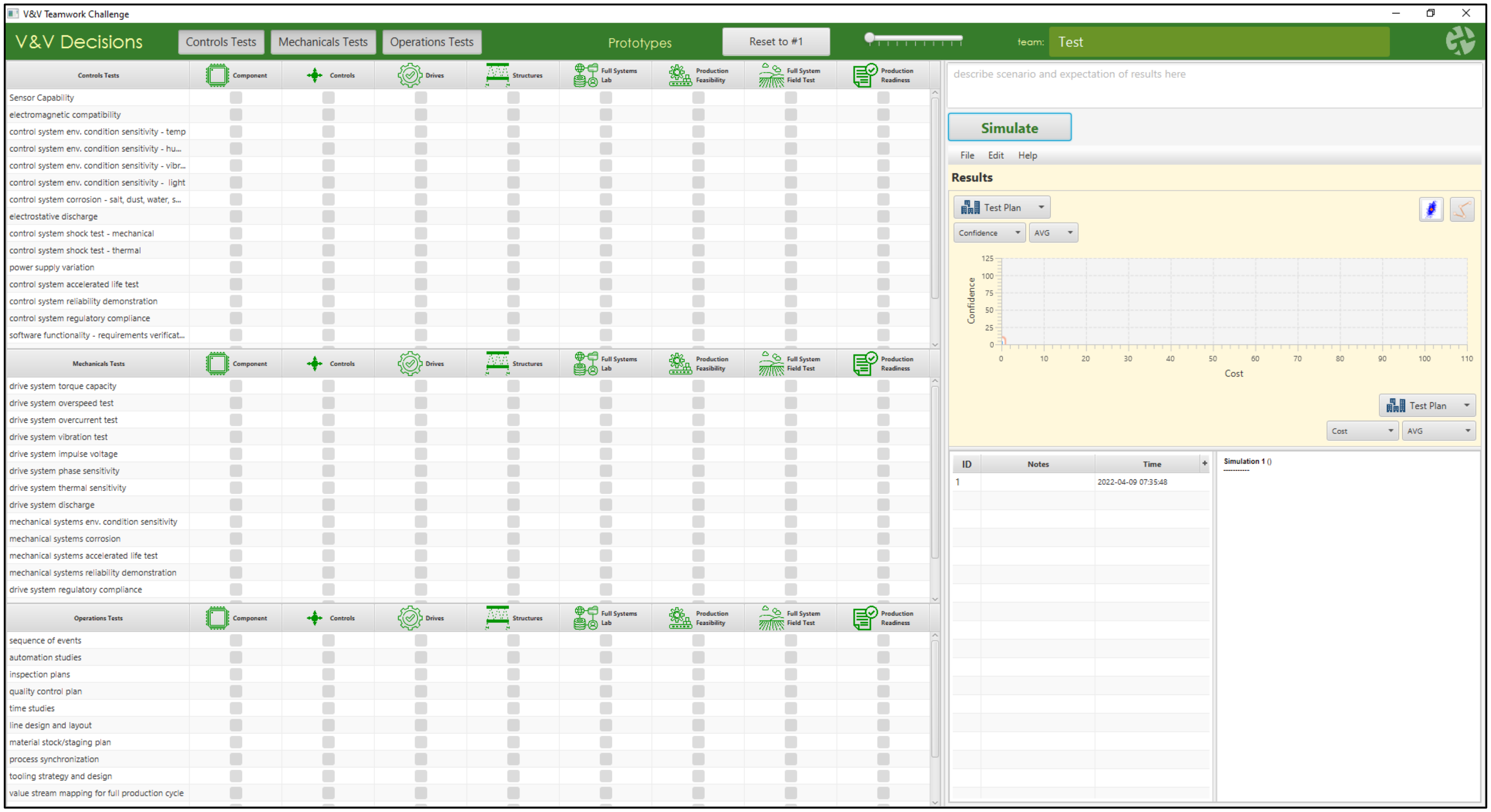
Badke-Schaub et al. warn, "There is a point where full sharedness (of mental models) is not achievable due to complexity of the project and transactive memory systems increase in importance." (Badke-Schaub et al, 2007). This was considered in this work and the design of the experiment. Systems design and systems engineering methods will continue to influence the continuum described by Badke-Schaub et al. The ability of an individual to comprehend the complexity of a system is related to the depth of system layers under evaluation.
Given these objectives and balanced considerations, three disciplines were identified for a team-of-teams scenario and included: mechanical systems, controls systems, and manufacturing operations.

Phase Content

In phase 1, participants were assigned to a sub-system team where they focused on one of three sub-systems. Sub-system teams were tasked with developing a verification plan for the project by allocating verification tests to a pre-determined set of opportunities ranging from early-stage component level bench tests through late-stage production test runs. Each test could be conducted at all opportunities or omitted entirely but could be conducted only once per opportunity. Each test execution resulted in an increase in cost and an increase in product confidence. Non-linearities were introduced into the confidence results based on test weightings and sequence dependencies and introduced into the cost results based on individual test costs and shared fixed costs of selected opportunities. The cost and confidence values associated with each individual selection were unknown to the participants prior to the selection. Any single configuration of test selections constituted a test plan. Each plan could be simulated via software, with results displayed immediately. Participants were provided 30 minutes in phase 1 to execute the sub-system level test planning and simulation runs. After sub-system level test planning was complete, participants were reallocated to form full project teams to develop a project level test plan in a similar fashion.

Software

The software utilized for the experiment consisted of the system model and a customized interface displaying the problem space and solution space, as well as a data log of all previous simulations within the session. A visual of the software interface is shown below in Figure 4.



describe scenario and expectation of results here

Controls Tests	Component	Controls	Drives	Structures	Full Systems Lab	Production Feasibility	Full System Field Test	Production Readiness
Sensor Capability	☐	☐	☐	☐	☐	☐	☐	☐
electromagnetic compatibility	☐	☐	☐	☐	☐	☐	☐	☐
control system env. condition sensitivity - temp	☐	☐	☐	☐	☐	☐	☐	☐
control system env. condition sensitivity - hu...	☐	☐	☐	☐	☐	☐	☐	☐
control system env. condition sensitivity - vibr...	☐	☐	☐	☐	☐	☐	☐	☐
control system env. condition sensitivity - light	☐	☐	☐	☐	☐	☐	☐	☐
control system corrosion - salt, dust, water, s...	☐	☐	☐	☐	☐	☐	☐	☐
electrostative discharge	☐	☐	☐	☐	☐	☐	☐	☐
control system shock test - mechanical	☐	☐	☐	☐	☐	☐	☐	☐
control system shock test - thermal	☐	☐	☐	☐	☐	☐	☐	☐
power supply variation	☐	☐	☐	☐	☐	☐	☐	☐
control system accelerated life test	☐	☐	☐	☐	☐	☐	☐	☐
control system reliability demonstration	☐	☐	☐	☐	☐	☐	☐	☐
control system regulatory compliance	☐	☐	☐	☐	☐	☐	☐	☐
software functionality - requirements verificat...	☐	☐	☐	☐	☐	☐	☐	☐

Mechanicals Tests	Component	Controls	Drives	Structures	Full Systems Lab	Production Feasibility	Full System Field Test	Production Readiness
drive system torque capacity	☐	☐	☐	☐	☐	☐	☐	☐
drive system overspeed test	☐	☐	☐	☐	☐	☐	☐	☐
drive system overcurrent test	☐	☐	☐	☐	☐	☐	☐	☐
drive system vibration test	☐	☐	☐	☐	☐	☐	☐	☐
drive system impulse voltage	☐	☐	☐	☐	☐	☐	☐	☐
drive system phase sensitivity	☐	☐	☐	☐	☐	☐	☐	☐
drive system thermal sensitivity	☐	☐	☐	☐	☐	☐	☐	☐
drive system discharge	☐	☐	☐	☐	☐	☐	☐	☐
mechanical systems env. condition sensitivity	☐	☐	☐	☐	☐	☐	☐	☐
mechanical systems corrosion	☐	☐	☐	☐	☐	☐	☐	☐
mechanical systems accelerated life test	☐	☐	☐	☐	☐	☐	☐	☐
mechanical systems reliability demonstration	☐	☐	☐	☐	☐	☐	☐	☐
drive system regulatory compliance	☐	☐	☐	☐	☐	☐	☐	☐

Operations Tests	Component	Controls	Drives	Structures	Full Systems Lab	Production Feasibility	Full System Field Test	Production Readiness
sequence of events	☐	☐	☐	☐	☐	☐	☐	☐
automation studies	☐	☐	☐	☐	☐	☐	☐	☐
inspection plans	☐	☐	☐	☐	☐	☐	☐	☐
quality control plan	☐	☐	☐	☐	☐	☐	☐	☐
time studies	☐	☐	☐	☐	☐	☐	☐	☐
line design and layout	☐	☐	☐	☐	☐	☐	☐	☐
material stock/staging plan	☐	☐	☐	☐	☐	☐	☐	☐
process synchronization	☐	☐	☐	☐	☐	☐	☐	☐
tooling strategy and design	☐	☐	☐	☐	☐	☐	☐	☐
value stream mapping for full production cycle	☐	☐	☐	☐	☐	☐	☐	☐

ID	Notes	Time
1		2022-04-09 07:35:48

Figure 4: VnV Challenge Software User Interface

The image above shows the software configured for input from all disciplines, each discipline displayed in one of the segmented portions on the left. The software can also be configured for a single sub-system focus. The rows on the left correspond to a predetermined set of tests and the columns correspond to a predetermined set of prototypes. Each prototype represents an opportunity to run each test. A test is selected for execution on a prototype by simply selecting the box corresponding to the intersecting row and column.

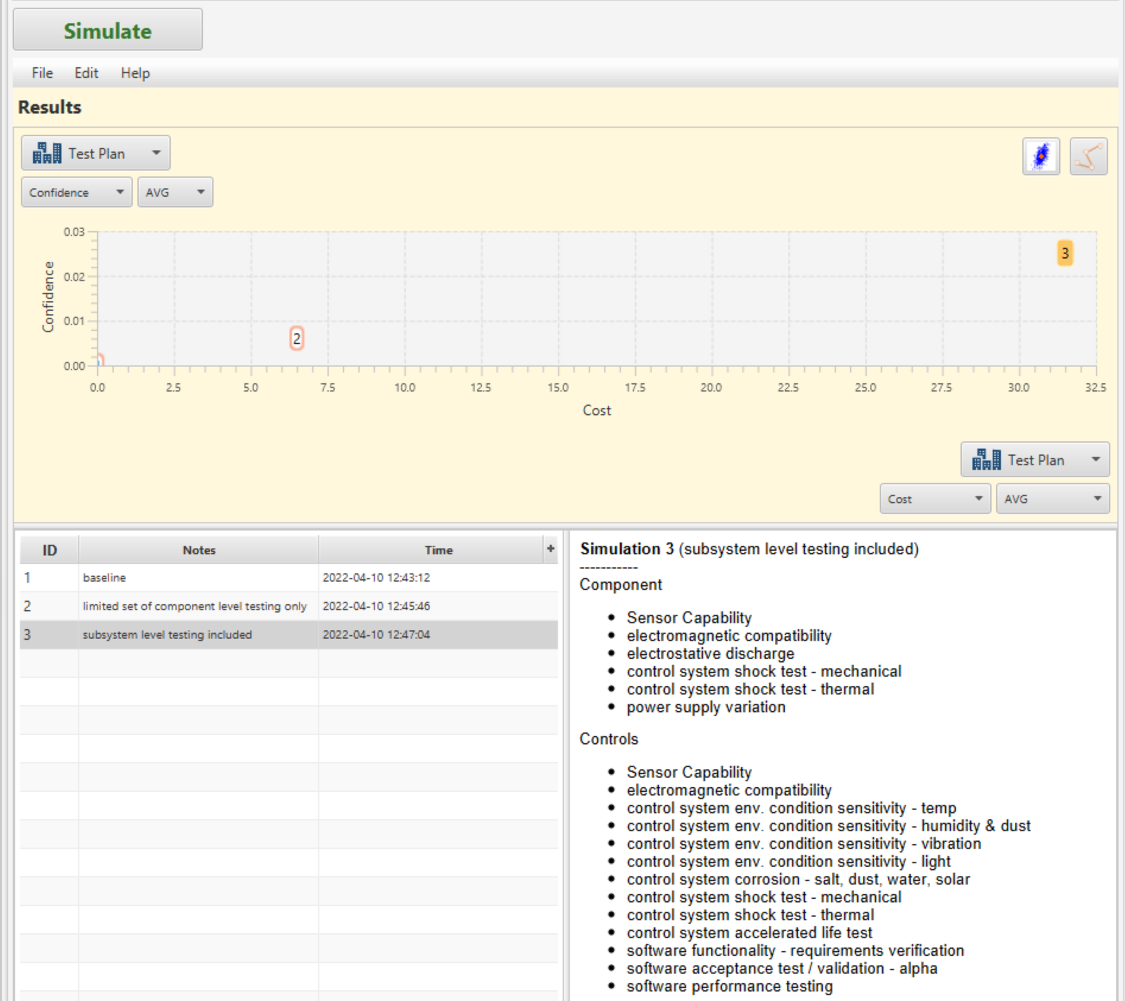
Tests and prototypes were selected to correspond to industry or organizational standards testing practices. Test selections are binary, they can be executed or omitted, but they can only be run once per prototype for a given test plan.

Test lists for each discipline were targeted to be comprehensive at a similar level of abstraction. The list was targeted to be extensive enough to drive focus on technical content and avoid "gaming" while remaining short enough for full consideration during the allotted workshop time. The number of possible test plan configurations is simply $2^{(\text{\# of selectable test, prototype pairs})}$. The table below shows the total number of possible test plan configurations by phase by team.

Phase	Team	Possible Configurations
1	Mechanical Systems	3.4028×10^{38}
	Controls Systems	2.2301×10^{43}
	Operations	2.0282×10^{31}
2	Full Team	1.5391×10^{113}

Table 1: Test Plan Configuration Possibilities by Phase and Team

The solution space is displayed on the righthand side of the screen. Each time a proposed plan is simulated, the corresponding cost and confidence values are plotted in the results pane. The user can input notes for each simulation run for future reference. The solution space also includes a running list of all simulations in the lower right-hand portion of the screen. The table populates with a sequential ID number for each simulation, the user notes, and a timestamp for the run. When a historical run is selected (either in the graph or table in the solution space), the run is highlighted on the graph, in the table, and a summary of selections is listed to the right of the table. See the solution space example below in Figure 5.



ID	Notes	Time
1	baseline	2022-04-10 12:43:12
2	limited set of component level testing only	2022-04-10 12:45:46
3	subsystem level testing included	2022-04-10 12:47:04

Simulation 3 (subsystem level testing included)

Component

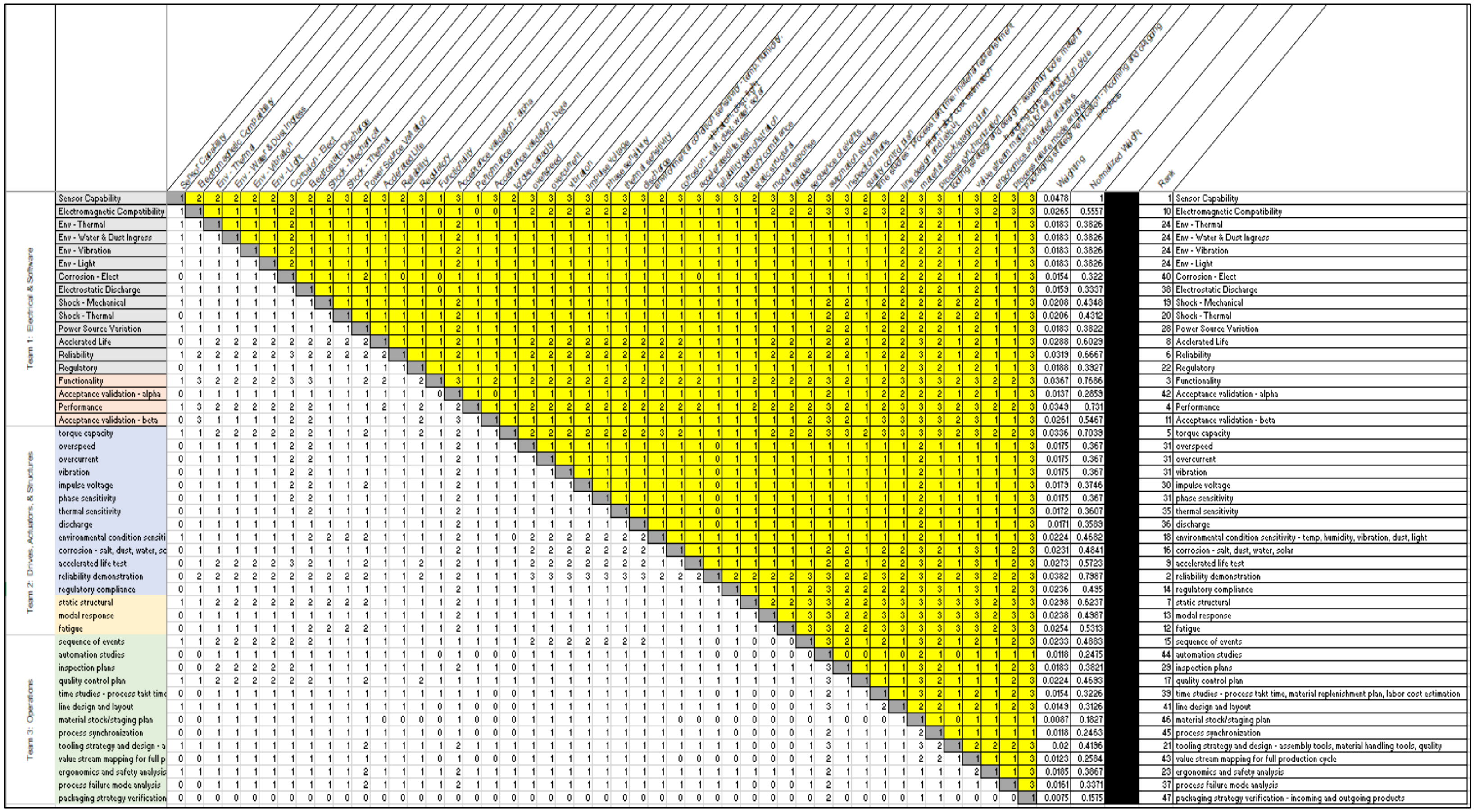
- Sensor Capability
- electromagnetic compatibility
- electrostative discharge
- control system shock test - mechanical
- control system shock test - thermal
- power supply variation

Controls

- Sensor Capability
- electromagnetic compatibility
- control system env. condition sensitivity - temp
- control system env. condition sensitivity - humidity & dust
- control system env. condition sensitivity - vibration
- control system env. condition sensitivity - light
- control system corrosion - salt, dust, water, solar
- control system shock test - mechanical
- control system shock test - thermal
- control system accelerated life test
- software functionality - requirements verification
- software acceptance test / validation - alpha
- software performance testing

Figure 5: VnV Challenge Software Solution Space

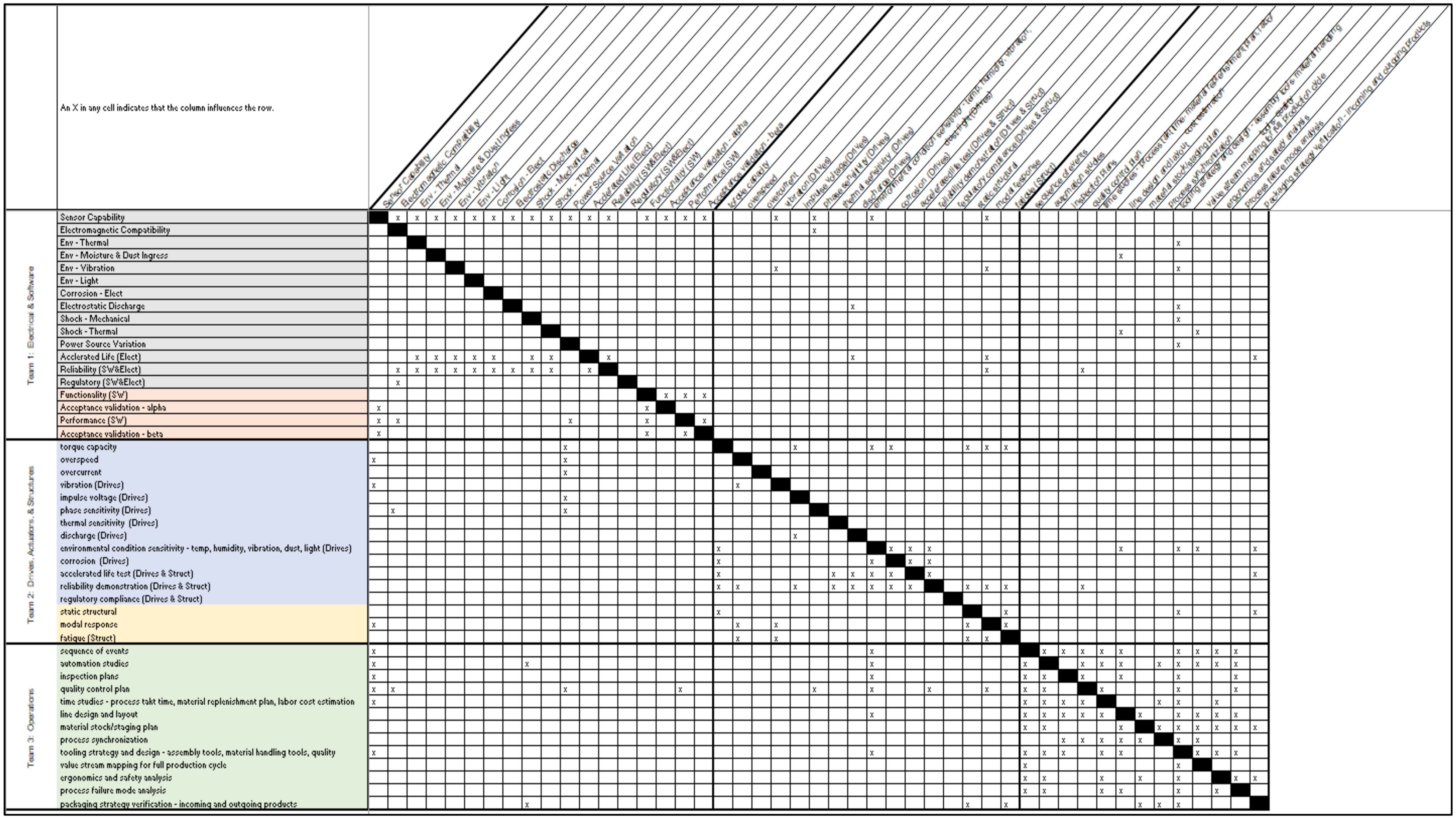
The system model was constructed to run in the background and generate cost and confidence values for each simulation. Confidence values were generated via a pairwise comparison of all tests. The resulting weightings were scaled to a maximum achievable system confidence of 0.9.



Team	Weighting	Normalized Weight	Rank	
Team 1: Electrical & Software	0.0478	1	1	Sensor Capability
	0.0265	0.5557	10	Electromagnetic Compatibility
	0.0183	0.3826	24	Env - Thermal
	0.0183	0.3826	24	Env - Water & Dust Ingress
	0.0183	0.3826	24	Env - Vibration
	0.0183	0.3826	24	Env - Light
	0.0154	0.322	40	Corrosion - Elect
	0.0159	0.3337	38	Electrostatic Discharge
	0.0208	0.4348	19	Shock - Mechanical
	0.0206	0.4312	20	Shock - Thermal
	0.0183	0.3822	28	Power Source Variation
	0.0288	0.6029	8	Acclerated Life
	0.0319	0.6667	6	Reliability
	0.0188	0.3927	22	Regulatory
	0.0367	0.7686	3	Functionality
	0.0137	0.2859	42	Acceptance validation - alpha
	0.0349	0.731	4	Performance
	0.0261	0.5467	11	Acceptance validation - beta
Team 2: Drives, Actuators, & Structures	0.0336	0.7039	5	torque capacity
	0.0175	0.367	31	overspeed
	0.0175	0.367	31	overcurrent
	0.0175	0.367	31	vibration
	0.0179	0.3746	30	impulse voltage
	0.0175	0.367	31	phase sensitivity
	0.0172	0.3607	35	thermal sensitivity
	0.0171	0.3589	36	discharge
	0.0224	0.4682	18	environmental condition sensitivity - temp, humidity, vibration, dust, light
	0.0231	0.4841	16	corrosion - salt, dust, water, solar
	0.0273	0.5723	9	accelerated life test
	0.0382	0.7987	2	reliability demonstration
	0.0236	0.495	14	regulatory compliance
	0.0298	0.6237	7	static structural
	0.0238	0.4987	13	modal response
	0.0254	0.5313	12	fatigue
Team 3: Operations	0.0233	0.4883	15	sequence of events
	0.0118	0.2475	44	automation studies
	0.0183	0.3821	29	inspection plans
	0.0224	0.4693	17	quality control plan
	0.0154	0.3226	39	time studies - process takt time, material replenishment plan, labor cost estimation
	0.0149	0.3126	41	line design and layout
	0.0087	0.1827	46	material stock/staging plan
	0.0118	0.2463	45	process synchronization
	0.02	0.4196	21	tooling strategy and design - assembly tools, material handling tools, quality
	0.0123	0.2584	43	value stream mapping for full production cycle
	0.0185	0.3867	23	ergonomics and safety analysis
	0.0161	0.3371	37	process failure mode analysis
	0.0075	0.1575	47	packaging strategy verification - incoming and outgoing products

Figure 6: VnV Challenge Software Model Confidence Weighting

Test interactions were identified in an asymmetric DSM where selected cells indicate the test associated with the respective column influences the test associated with the respective row. The system model identifies situations where these dependencies occur, and if the influencing test is not conducted at or before the time of the influenced test, a confidence penalty is incurred. Penalties increase in weight progressively through the time series of prototypes.



An X in any cell indicates that the column influences the row.

		Sensor Capability	Electromagnetic Compatibility	Env - Thermal	Env - Moisture & Dust Ingress	Env - Vibration	Env - Light	Corrosion - Elect	Electrostatic Discharge	Shock - Mechanical	Shock - Thermal	Power Source Variation	Acclerated Life (Elect)	Reliability (SW&Elect)	Regulatory (SW&Elect)	Functionality (SW)	Acceptance validation - alpha	Performance (SW)	Acceptance validation - beta	torque capacity	overspeed	overcurrent	vibration (Drives)	impulse voltage (Drives)	phase sensitivity (Drives)	thermal sensitivity (Drives)	discharge (Drives)	environmental condition sensitivity - temp, humidity, vibration, dust, light (Drives)	corrosion (Drives)	accelerated life test (Drives & Struct)	reliability demonstration (Drives & Struct)	regulatory compliance (Drives & Struct)	static structural	modal response	fatigue (Struct)	sequence of events	automation studies	inspection plans	quality control plan	time studies - process takt time, material replenishment plan, labor cost estimation	line design and layout	material stock/staging plan	process synchronization	tooling strategy and design - assembly tools, material handling tools, quality	value stream mapping for full production cycle	ergonomics and safety analysis	process failure mode analysis	packaging strategy verification - incoming and outgoing products
Team 1: Electrical & Software	Sensor Capability	■	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x				x		x			x						x														
	Electromagnetic Compatibility		■																						x																							
	Env - Thermal			■																																								x				
	Env - Moisture & Dust Ingress				■																																				x							
	Env - Vibration					■																	x											x										x				
	Env - Light						■																																									
	Corrosion - Elect							■																																								
	Electrostatic Discharge								■																		x																	x				
	Shock - Mechanical									■																																		x				
	Shock - Thermal										■																														x				x			
	Power Source Variation											■																																x				
	Acclerated Life (Elect)			x	x	x	x	x		x	x		■	x													x							x														x
	Reliability (SW&Elect)		x	x	x	x	x	x	x	x	x		x	■																				x					x									
	Regulatory (SW&Elect)		x												■																																	
	Functionality (SW)															■	x	x	x																													
	Acceptance validation - alpha	x														x	■																															
	Performance (SW)	x	x									x				x		■	x																													
	Acceptance validation - beta	x														x		x	■																													
Team 2: Drives, Actuators, & Structures	torque capacity											x								■				x				x	x				x	x	x													
	overspeed	x										x									■																											
	overcurrent											x										■																										
	vibration (Drives)	x																			x		■																									
	impulse voltage (Drives)											x												■																								
	phase sensitivity (Drives)		x									x													■																							
	thermal sensitivity (Drives)																									■																						
	discharge (Drives)																						x			■																						
	environmental condition sensitivity - temp, humidity, vibration, dust, light (Drives)																			x								■	x	x	x										x			x	x			x
	corrosion (Drives)																			x								x	■	x	x																	
	accelerated life test (Drives & Struct)																			x						x	x	x	x	■	x																	x
	reliability demonstration (Drives & Struct)																			x	x			x		x	x	x	x	x	■		x	x	x				x									
	regulatory compliance (Drives & Struct)																															■																
	static structural																			x													■		x									x				x
	modal response	x																			x		x										x	■	x													
	fatigue (Struct)																				x		x										x	x	■													
Team 3: Operations	sequence of events	x																									x									■	x	x	x	x	x			x	x	x	x	
	automation studies	x								x																	x									x	■		x	x	x		x	x	x	x	x	
	inspection plans	x																									x									x	x	■	x		x			x			x	
	quality control plan	x	x									x						x							x			x			x			x		x	x		■	x				x			x	
	time studies - process takt time, material replenishment plan, labor cost estimation	x																																		x	x	x	x	■			x	x		x		
	line design and layout																										x									x	x	x	x	x	■	x		x	x	x	x	
	material stock/staging plan																																			x	x				x	■	x	x	x	x	x	x
	process synchronization																																					x	x	x	x	x	■	x	x			
	tooling strategy and design - assembly tools, material handling tools, quality	x																									x									x	x	x		x	x			■	x	x	x	
	value stream mapping for full production cycle																																			x								x	■			
	ergonomics and safety analysis																																			x	x			x		x		x		■	x	x
	process failure mode analysis																																			x	x			x	x			x		x	■	
	packaging strategy verification - incoming and outgoing products									x																							x		x							x	x	x				■

Figure 7: VnV Challenge Software Model Confidence Interactions

Incremental cost values were established for each instance of a test and fixed cost values were established for each prototype. Incremental costs were incurred for each box selected. Fixed costs were only incurred when one or more boxes in the respective prototype column were selected, but the cost incurred was the same regardless of the number of boxes selected in that column.

The resulting potential performance outcomes (problem space) were encompassed by black dashed lines in the figure below.

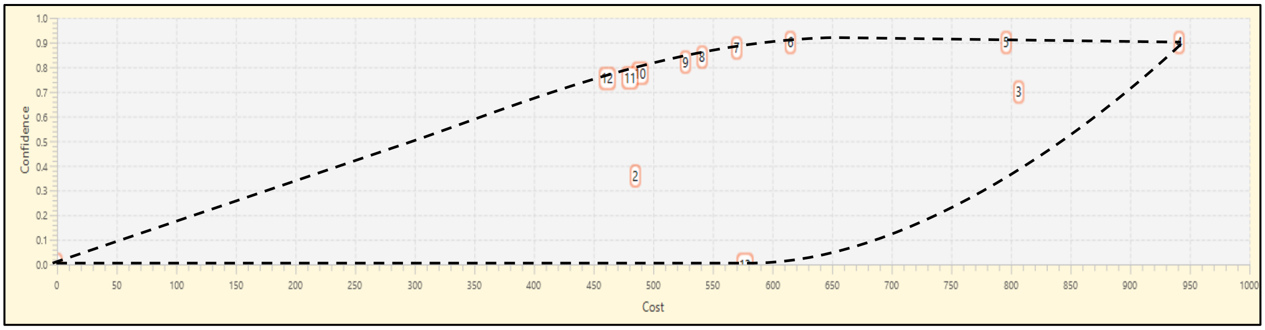


Figure 8: VnV Challenge Software Problem Space Outlined

Data Collection

As stated in the hypotheses, the attributes of interest include the sharedness and comprehensive breadth of mental models and team performance.

Data collection activities were conducted by multiple means. A similarity ratings approach was used to elicit data associated with mental models. Participants were requested to individually complete surveys just after the introduction of the project scenario and again just prior to completion of the workshop. The surveys presented participants with seven concepts associated with the project work. The term “concept” in this context refers to elements of the cognitive structure used to describe system behavior. Participants were asked to assess the level of dependency between concept pairs and the relative importance of concept pairs by populating the upper half of N^2 matrices with provided Likert scale values. See Appendix A below for visuals of the surveys. These survey results were utilized to measure the sharedness and breadth of mental models for analysis and testing of H1 and H2, respectively. Throughout the workshop’s working sessions, each time a team ran a software simulation on a proposed verification plan the plan details, team label, and performance outcomes were timestamped and stored. The simulation and survey data were analyzed for testing H3 and H4. Performance in the context of this study is measured by the level of advancement of product concepts (reduction of uncertainty) while consuming the fewest amount of resources.

In 2000, Mohammed et al. compared and critiqued methods of eliciting and analyzing mental models in an attempt to promote the advancement of empirical research into team mental models (Mohammed et al., 2000). A lack of agreement on methods and growing demand for empirical work to support a relatively large body of conceptual work was cited as motivation. Mohammed et al. explain the importance of a clear specification of the phenomenon to be tested in order to select a measurement technique. They make subtle distinctions between knowledge and belief structures and structure and process. The phenomenon of interest in this thesis could be categorized according to definitions by Schneider and Angelmar, (1993) as related to the cognitive structure or “representations of knowledge that contain and organize information” (p. 349) rather than a cognitive process which is how “knowledge is selected, organized, transformed, stored, and utilized” (p. 351). And further refined according to definitions by Mohammed et al., (2000) as a knowledge structure “descriptive states of nature that one knows to be true“(p. 3) rather than beliefs which are defined as “desired states that one prefers, expects, or demands” (p. 3). These distinctions helped guide the elicitation method to seek symmetric relationships of abstractions of the test planning categories.

Chapter 4

Quasi-Experiment

A quasi-experiment is an experiment “in which the investigator cannot randomly assign units or participants to conditions, cannot generally control or manipulate the independent variable, and cannot limit the influence of extraneous variables” (APA, n.d.)

A natural experiment is an “observational study in which an event or a situation that allows for the random or seemingly random assignment of study subjects to different groups is exploited to answer a particular question” (Messer, 2016).

Given the developmental nature of this experiment and the population count, the empirical study was conducted as a quasi-experiment. It was conducted to determine if a signal could be detected from the selected methods/sensors in a small-scale academic setting to provide insight and guidance to further study. Given the potential population count driven by the complexity of the problem, it is not feasible to scale at this stage. It is the intent that the method and reflection enable further refinement of this area of study for added scaling. See the Chapter 8 for recommendations on full experiments.

Workshop

Participants for the study were identified through the Massachusetts Institute of Technology Gordon Engineering Leadership program. Twenty-two individuals participated in the 2-hour event. Upon arrival, participants were provided an individually assigned number (IAN) and randomly allocated to teams based on the sequence of their arrival. Figure 9 shows participant assignments for Phase 1.

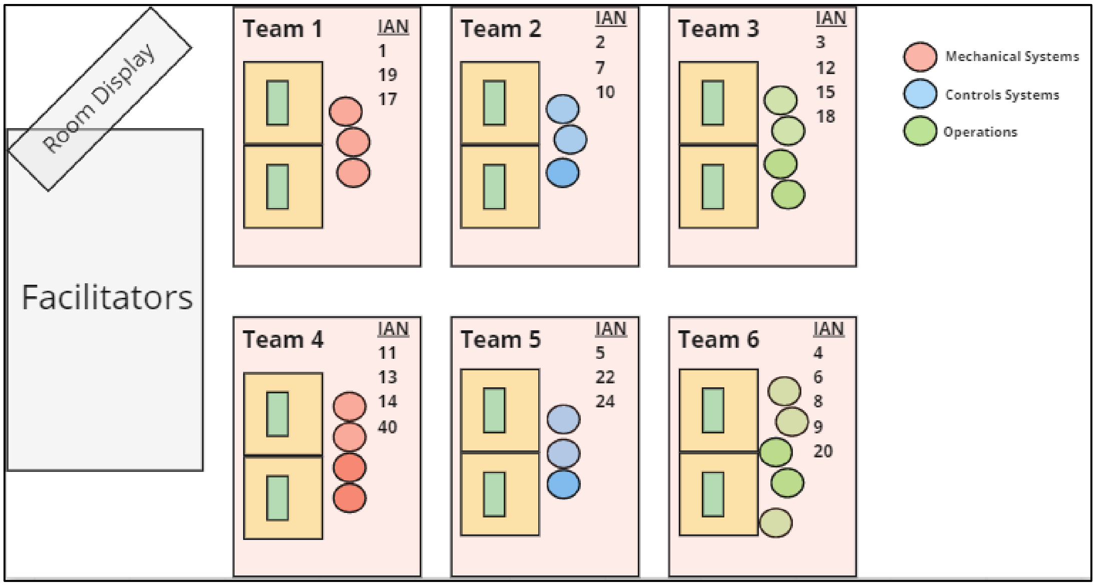


Figure 9: Workshop Participant Assignments Phase 1

Once teams were formed, the participants were introduced to the model problem – development of a verification and validation plan of a pick-and-place machine used in the production of printed circuit boards. The pick-and-place machine was selected as the system of focus due to the relative ease of describing its base function and high-level sub-system interactions. The introduction was given with the intent of level setting contextual knowledge and motivations across participants. It presented an overview of the product under development, the company for which it was being developed, and the verification and validation activity as well as the following detail which provided context on the project state:

- A customer need has been identified that presents a viable business opportunity.
- Primary value drivers have been identified, but not all stakeholders and use cases have been determined.
- A basic concept architecture is selected in alignment with stakeholder needs and the business’s core competencies.
- Detailed design is in progress, but many design decisions have yet to be made.
- The project requires completion on a tight timeline.
- You have been selected to the project team and asked to draft a verification plan.

- The project team has collectively identified various prototype opportunities and a set of required tests.

A software demonstration was then completed to avoid any confusion associated with the interface or objectives and help ensure equal time available across teams for verification plan development.

Participants were then asked to complete the initial set of surveys. After survey completion, teams were provided 30 minutes to work on phase 1 verification plan development.

This concluded phase 1 of the workshop and participants were provided a break. Workstation computers were reset over the break and configured for visualization and interface with all subsystems. Upon returning from break, the participants were re-allocated to multi-disciplinary project teams as shown in Figure 10 below. Participants were then asked to build a full program verification plan, including test allocations for all disciplines. System teams were provided 30 minutes to build the full program plans and were informed of a minimum viable confidence target of 0.7.

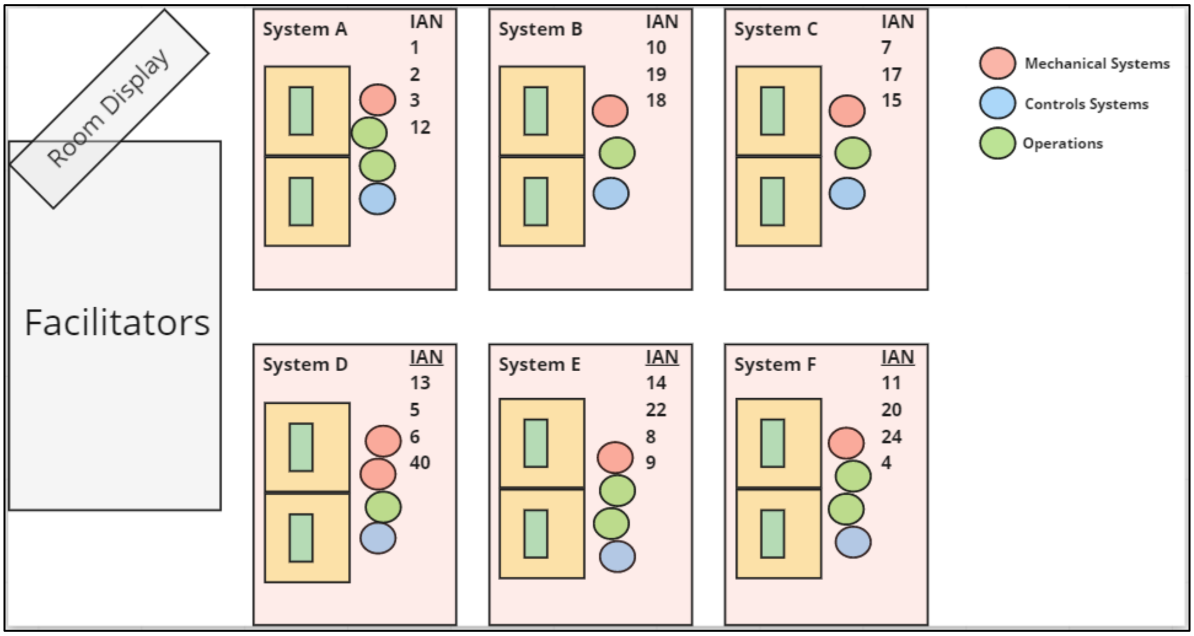


Figure 10: Workshop Participant Assignments Phase 2

At the conclusion of the 30-minute allocation, participants were asked to stop working with the software. Surveys identical to the initial surveys were then distributed to participants and requested to be completed. This marked the conclusion of phase 2 of the workshop. The workshop concluded with a short debrief with an open discussion on participant feedback. The workshop was followed by a brief lecture from Dr. Bryan Moser, Director of the Global Teamwork Lab, on the experiment's objectives and the mission of the Global Teamwork Lab.

Chapter 5

Analysis Method

In order to test the hypotheses, data were collected in an attempt to quantify individual mental model states and team performance. Individual mental model information was then used to assess two metrics associated with team shared mental models: sharedness and breadth. In accordance with the hypotheses, the analysis seeks to detect the influence of the treatment (structured context tool) on the state of the mental models (sharedness and breadth) and, in turn, the state of the shared mental model influence on team performance.

In the 2000 study previously referenced by Mohammed et al., they evaluated four analysis methods based on the following metrics (Mohammed et al., 2000):

- Treatment of content
- Treatment and evaluation of the structure
- Standardization and acceptance level of the approach
- Evidence of reliability
- Ability to capture mental models
- Availability of special features
- Utilization method for team-level analysis

Given the desire for structured representation, the efficiency of paired comparison of fixed concepts, and identification of symmetric relationships, the evaluation by Mohammed et al. guided method selection toward either Pathfinder or Multidimensional Scaling (MDS).

In 2010 DeChurch et al. completed a meta-analysis to examine three aspects of shared mental model measurement as potential moderators to team process and performance (DeChurch et al., 2010). They observed similar positive relationships between similarity in mental models and team performance across elicitation methods, structure representation approaches, and representation of emergence. Their work included data from 23 independent, empirical studies and included Pathfinder and MDS analyses amongst the methods of structure representation and representation of emergence. It also included similarity ratings amongst the methods of elicitation.

Given the findings cited above and its prevalence in recent literature, the Pathfinder analysis technique was selected for use. Pathfinder is a network modeling and scaling technique based on graph theory utilizing proximity data (Schvaneveldt, 1990). Schvaneveldt offers the abbreviated overview of the method, "Essentially, Pathfinder networks are determined by identifying the proximities that provide the most efficient connections between the entities by considering the indirect connections provided by paths through other entities." (Schvaneveldt, 1990). For the application within this study, the entities correspond to the concepts in the user surveys. These are the nodes in the network. The patterns of proximities are used to generate the links in the network. For this application the proximities are derived from participant survey responses on similarities between concepts. Two parameters are utilized to create the network - q and r. The r-parameter (Minkowski r-metric) is used to calculate the distance between nodes where the weight of a path is calculated as:

$$W(P) = \left(\sum_{i=1}^{k} w_i^r\right)^{1/r}$$

The q-parameter limits the number of links in the path examined during the construction of the network. The algorithm then finds the minimum weight path(s) with no more than q links.
For this application, the data was elicited as relative paired comparisons, so the weight matrix is symmetric, and all network links are undirected (symmetric) between entities. The derived networks can then be compared between participants to determine similarity (sharedness of the mental model). An example of a network diagram output for this study is shown below.

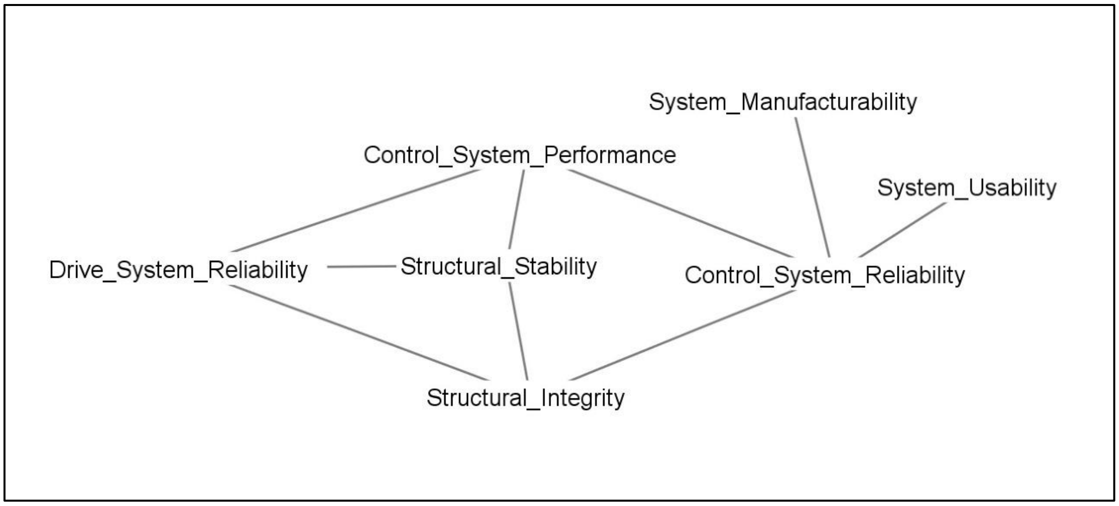
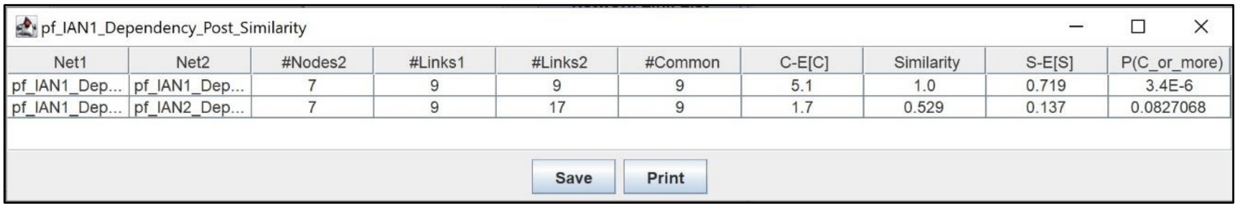


Figure 11: Pathfinder Network Diagram Example Output

Similarity was used as a quantitative index to determine the sharedness of mental models in the evaluation of H1 and H3. The survey question related to the level of dependency between concepts was utilized for assessment. Similarity is determined by the correspondence of links in the two networks (Schvaneveldt, n.d.). It is calculated as the total number of links in common divided by the total number of unique links. An example output of the quantitative indices is shown below.



Net1	Net2	#Nodes2	#Links1	#Links2	#Common	C-E[C]	Similarity	S-E[S]	P(C_or_more)
pf_IAN1_Dep...	pf_IAN1_Dep...	7	9	9	9	5.1	1.0	0.719	3.4E-6
pf_IAN1_Dep...	pf_IAN2_Dep...	7	9	17	9	1.7	0.529	0.137	0.0827068

Figure 12: Pathfinder Network Quantitative Indices Example Output

Data analysis was completed with JPathfinder software, and a minimum links approach was used with r = ∞ and q = n-1 (Schvaneveldt, n.d.)

Mental model breadth was determined by evaluating the distribution of participants' responses on the relative importance of concepts. Hypotheses H2 and H4 attempt to expand on whether team members have similar mental models and evaluate to what degree team members consider all elements of the model. This is related to Badke-Schaub et al.'s discussion on importance as a major factor influencing the quality of a mental model (Badke-Schaub et al., 2007). They propose the centrality of a node in a network as a measure of its importance due to the number of links the node has within the overall network. Rather than reference the number of influences a node has as a measure of its importance, this study directly inquires about participants' perceived weightings of concepts and evaluated the distribution of those weightings as a measure of breadth. A smaller standard deviation of concept weightings was used to indicate more even consideration of concepts in the mental model. A broad distribution was used as an indication of a more narrowly focused model. Weights were derived from survey responses on the "relative to importance" related question using an averaging over normalized columns procedure.

Performance data was collected in process as indicated above. Each time a configuration was simulated, the associated configuration, sequence, and results data was stored. Performance was measured using the Pareto fitness ranking method. This method allows the data to be analyzed without assigning weights to individual objectives. The method progressively ranks and removes data points. For the first iteration of analysis, all non-dominated points are assigned a 0 ranking and then removed from consideration during future iterations. The second iteration considers the remaining points and assigns all non-dominated points a 1 ranking and so on. A non-dominated point, in this case, is one that is in a position where no other point would be considered superior by both cost and confidence metrics.

Chapter 6

Results

Twenty-two individuals participated in the event. Not all participants completed all survey questions in the entirety, and the table below shows the resulting sample sizes available for full analysis for each hypothesis.

Hypothesis	**Sample Sizes**
H1, H3	21 individuals, 5 teams, 28 teammate pairs *2 control teams, 10 teammate pairs* *3 treatment teams, 18 teammate pairs*
H2, H4	13 individuals, 3 teams *1 control team* *2 treatment teams*

Table 2: Workshop Sample Sizes

Performance results were compiled for all teams with the exception of team B due to a lack of available survey data. The chart below shows how generative each of the applicable teams was.

Team	**Datapoints Generated**
A	34
B	N/A – Incomplete Survey Data
C	33
D	51
E	19
F	79

Table 3: Workshop Simulation Count by Team

The following images illustrate the “design walk” taken by each team. The design walk is the ordinal sequence of solutions generated by a team. It illustrates the path of exploration taken by that team in the problem space. The lines in the charts connect sequential iterations, and the line weight increases with each step. These graphs were generated for context and to give insights into team behaviors. The first graph provides a comparison of all teams, and subsequent graphs provide a zoomed-in view of the later stages of the exploration for each team.

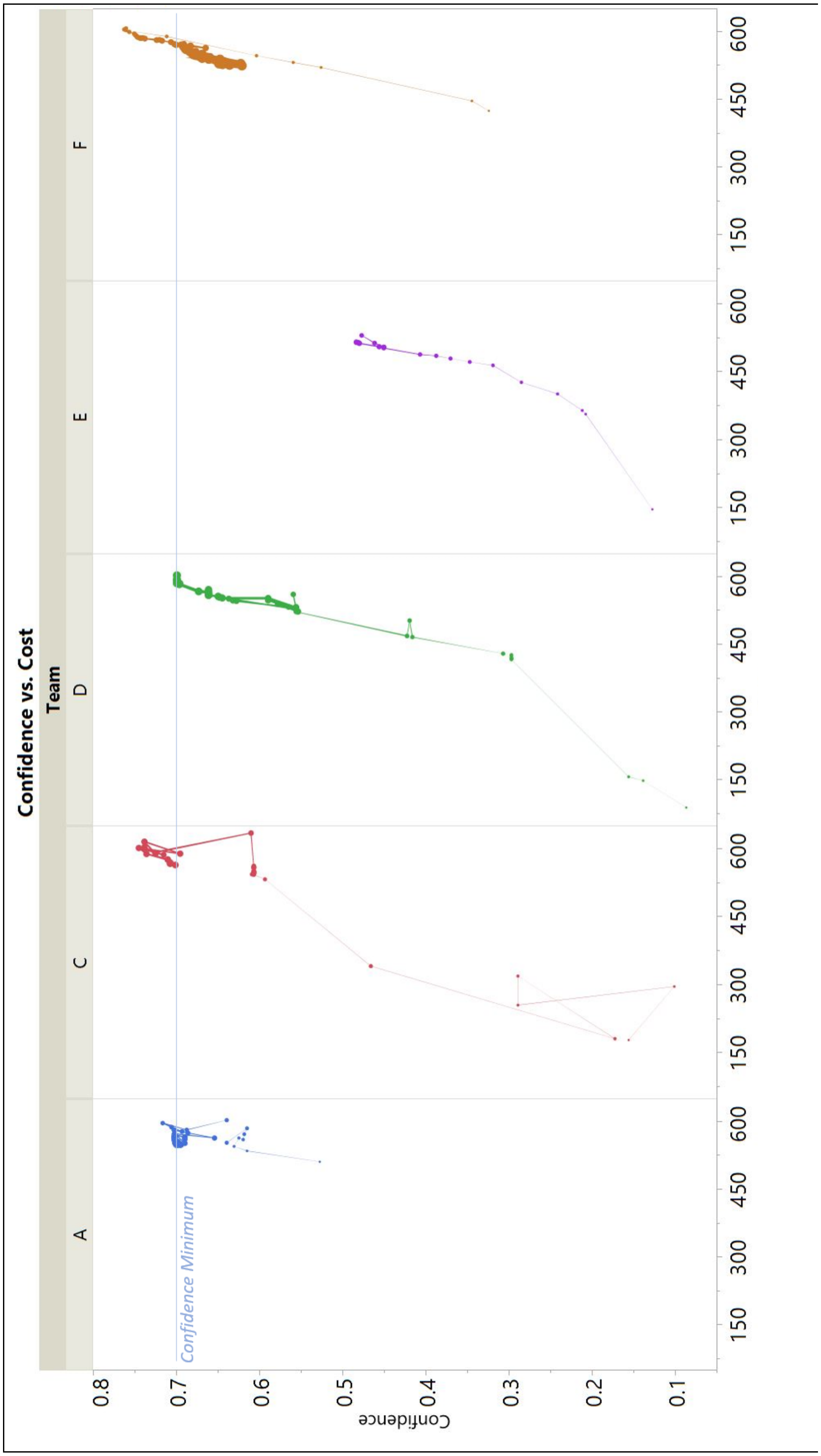


Figure 13: Workshop Participant Team Design Walks (All Teams)

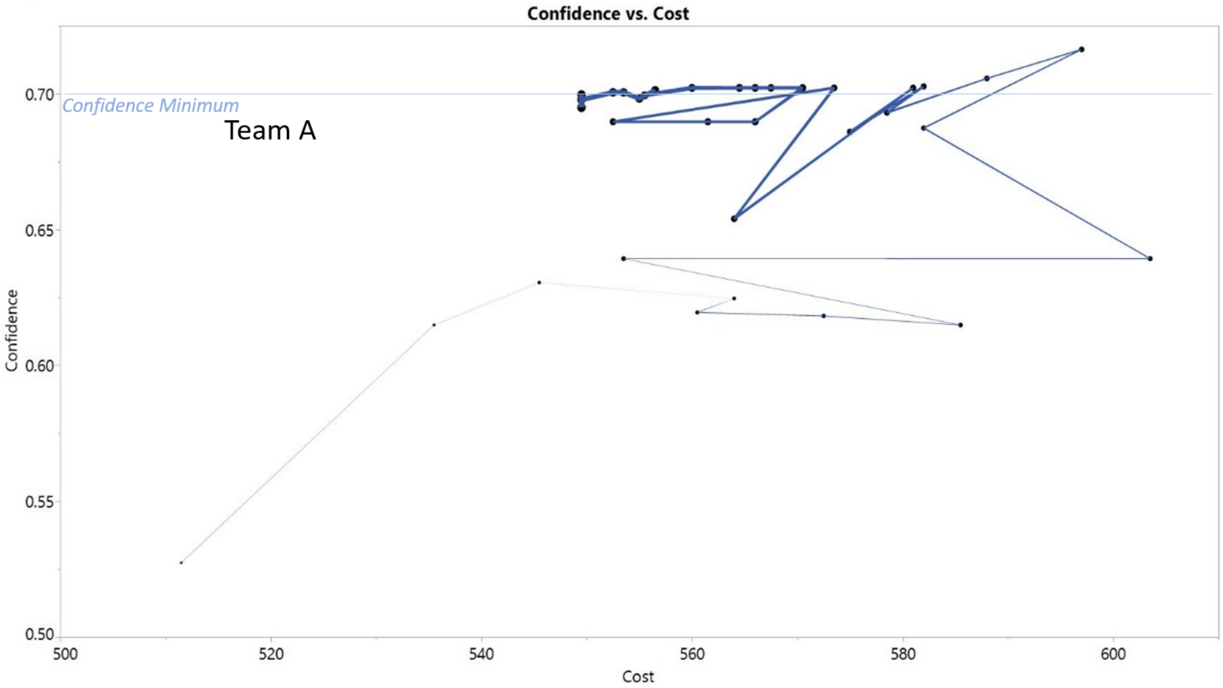


Figure 14: Workshop Participant Team Design Walk - Team A

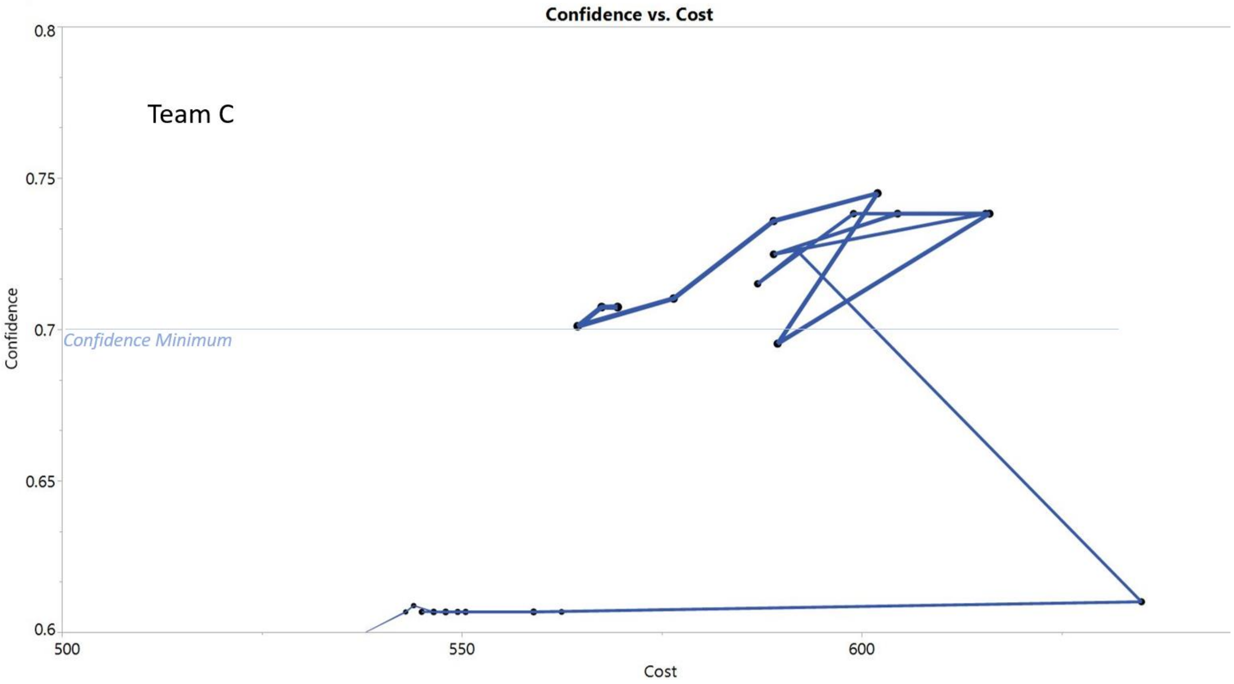


Figure 15: Workshop Participant Team Design Walk - Team C

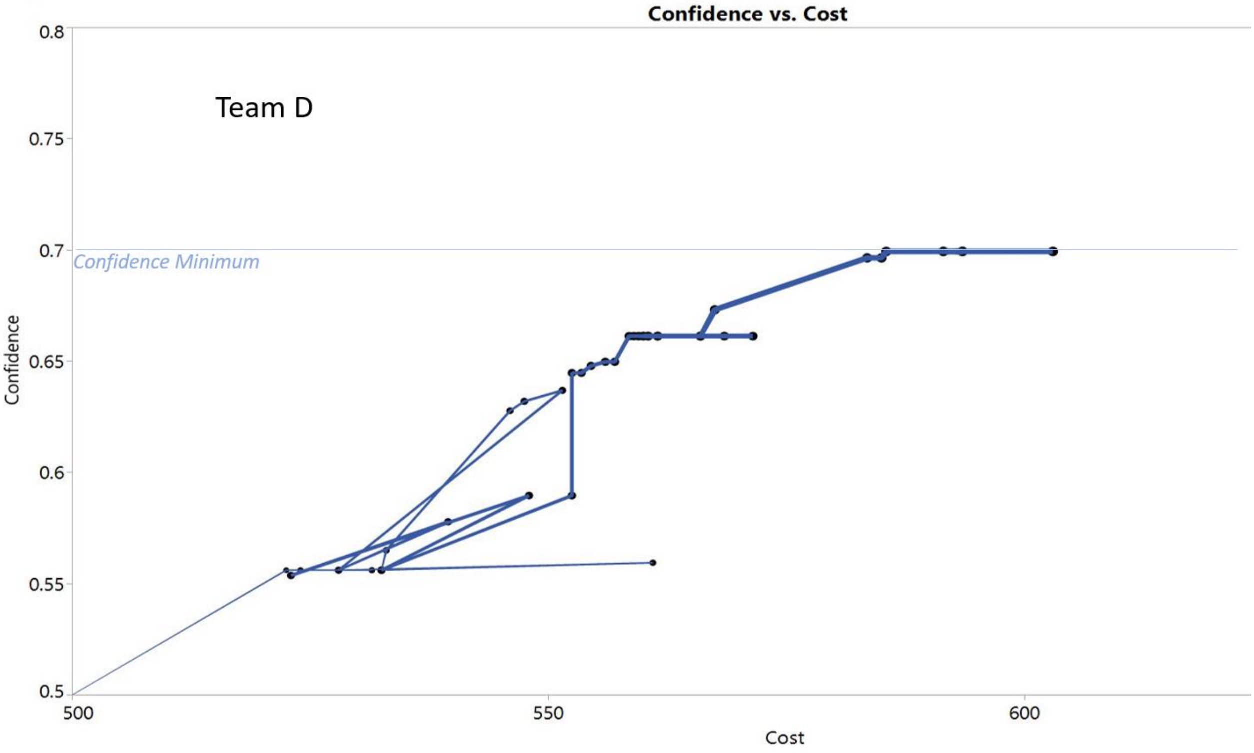


Figure 16: Workshop Participant Team Design Walk - Team D

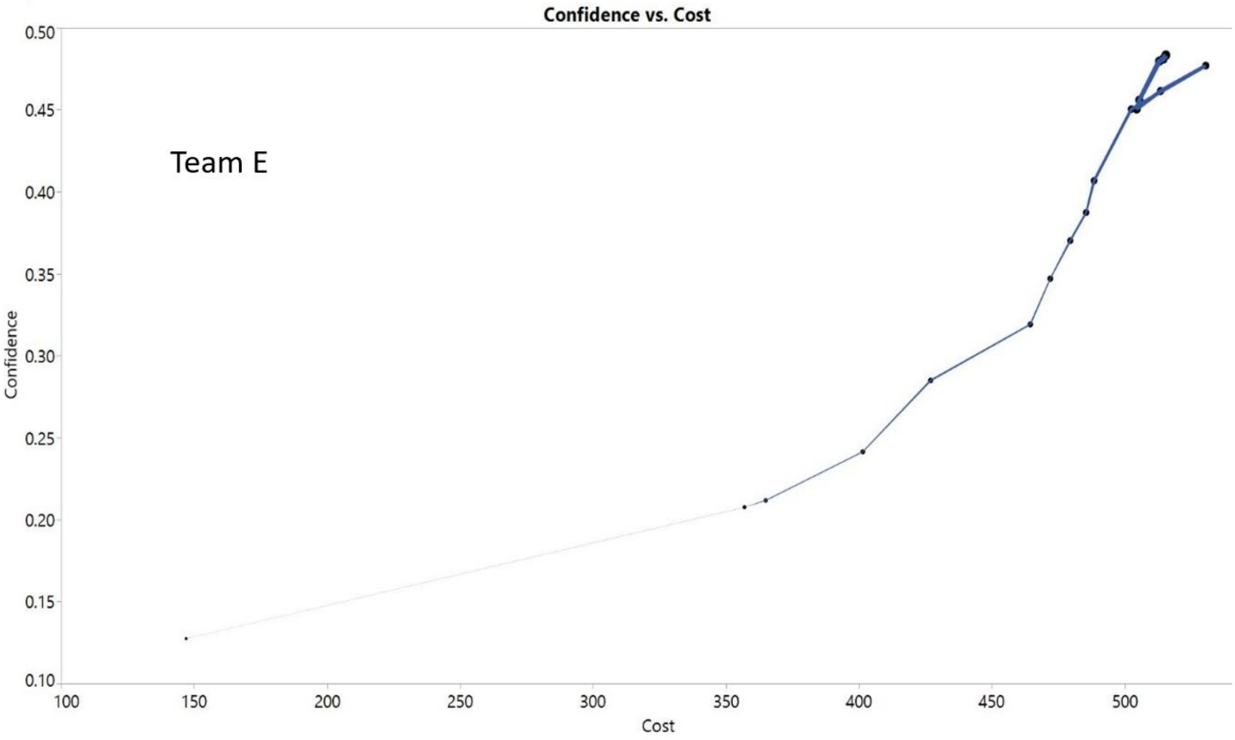


Figure 17: Workshop Participant Team Design Walk - Team E

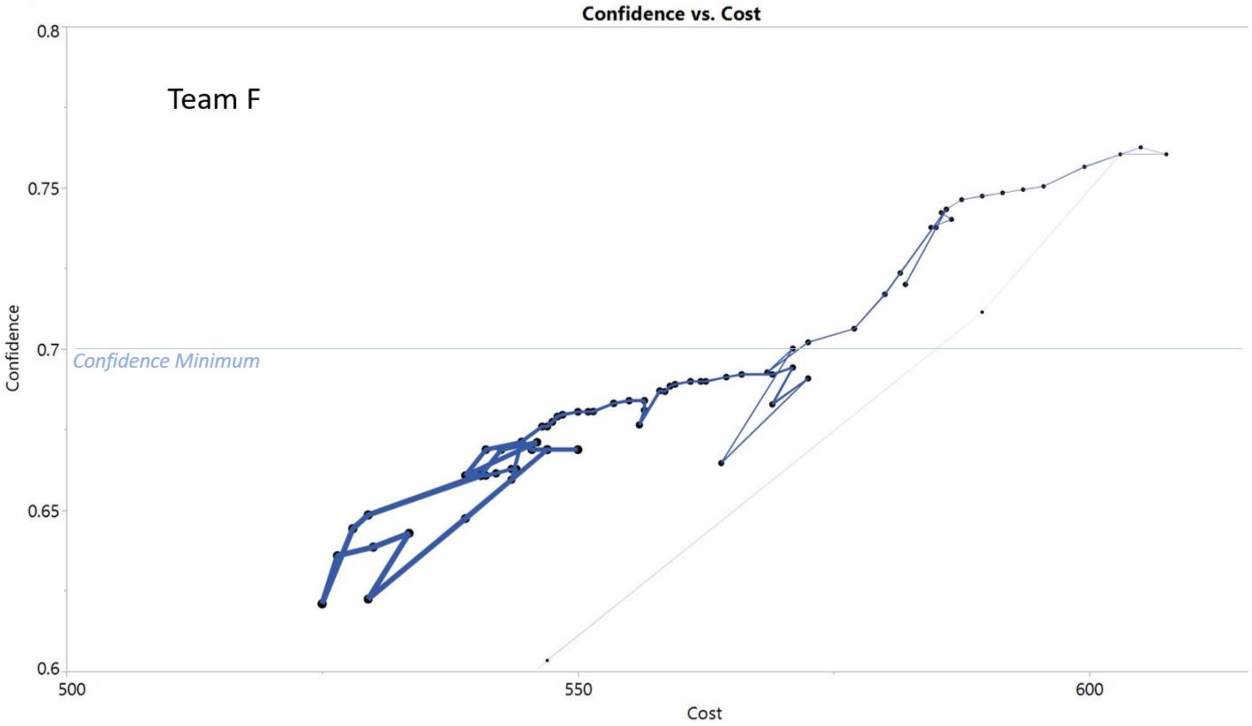


Figure 18: Workshop Participant Team Design Walk - Team F

Figure 19 shows Pathfinder similarity data for all participants from both pre- and post-activity survey questions related to the level of dependency of concept pairs. The control group pre- and post-activity data are plotted in the first and third column respectively and the treatment group in the second and fourth columns, respectively. Higher correlation values are being evaluated as an indication of greater sharedness of mental models.

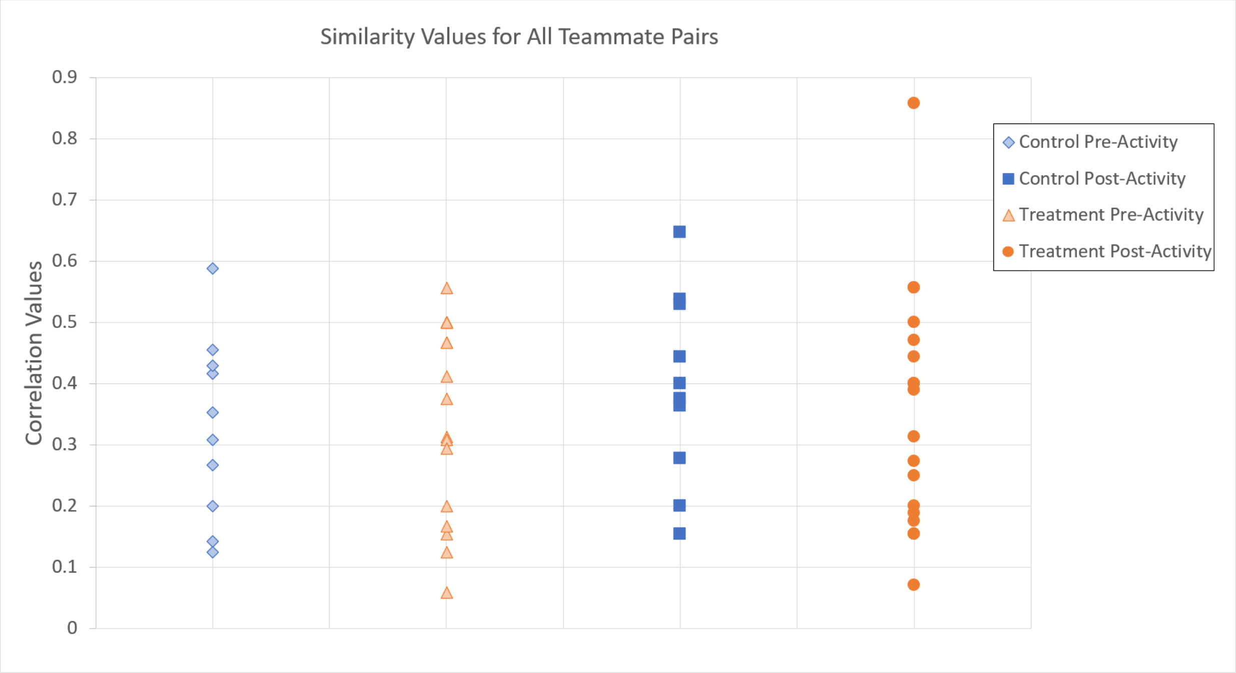


Figure 19: Workshop Team Mental Model Similarity for All Teammate Pairs

Figure 20 shows pairwise comparison results for all participants from pre- and post-activity survey questions related to the relative importance of concept pairs. The plotting scheme is similar to Figure 19 above. Lower standard deviation values are being evaluated as an indicator of greater breadth in mental models. For reference, the analysis method enables a range of between 0 and approximately 0.496 for the standard deviation.

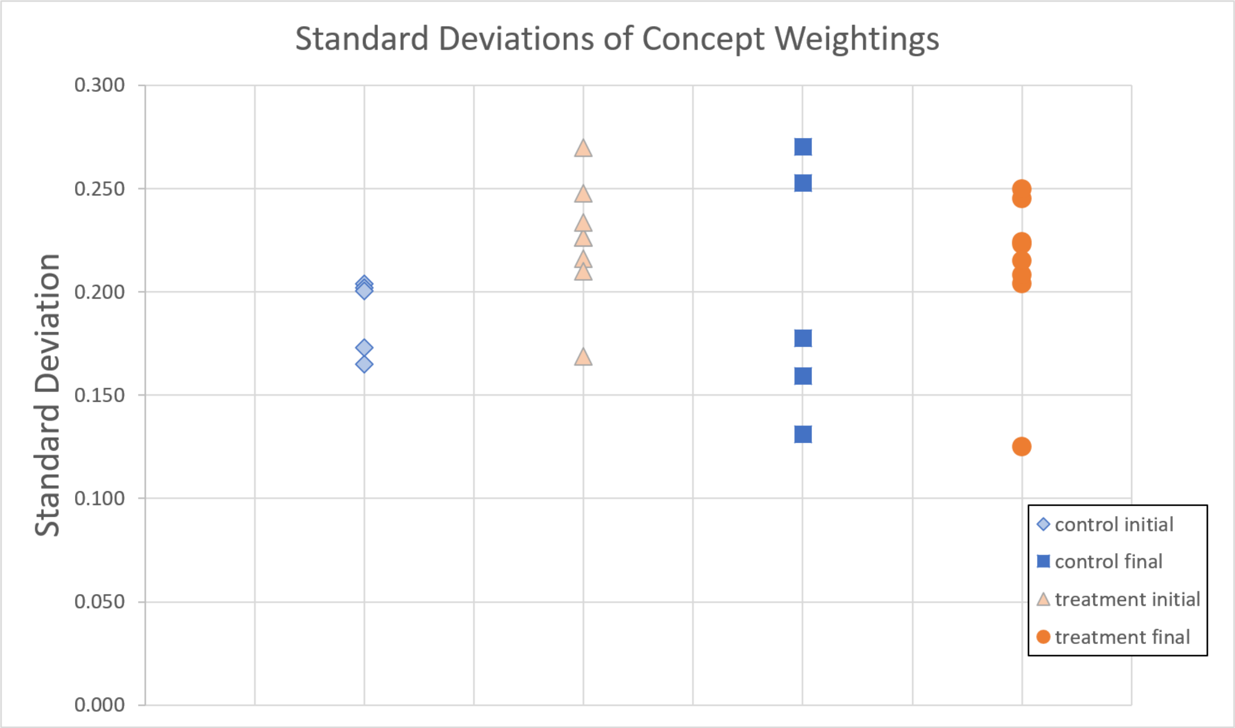


Figure 20: Workshop Team Mental Model Deviation of Concept Weightings

Figure 21 shows results from all simulation runs from all teams, again with the exception of team B.

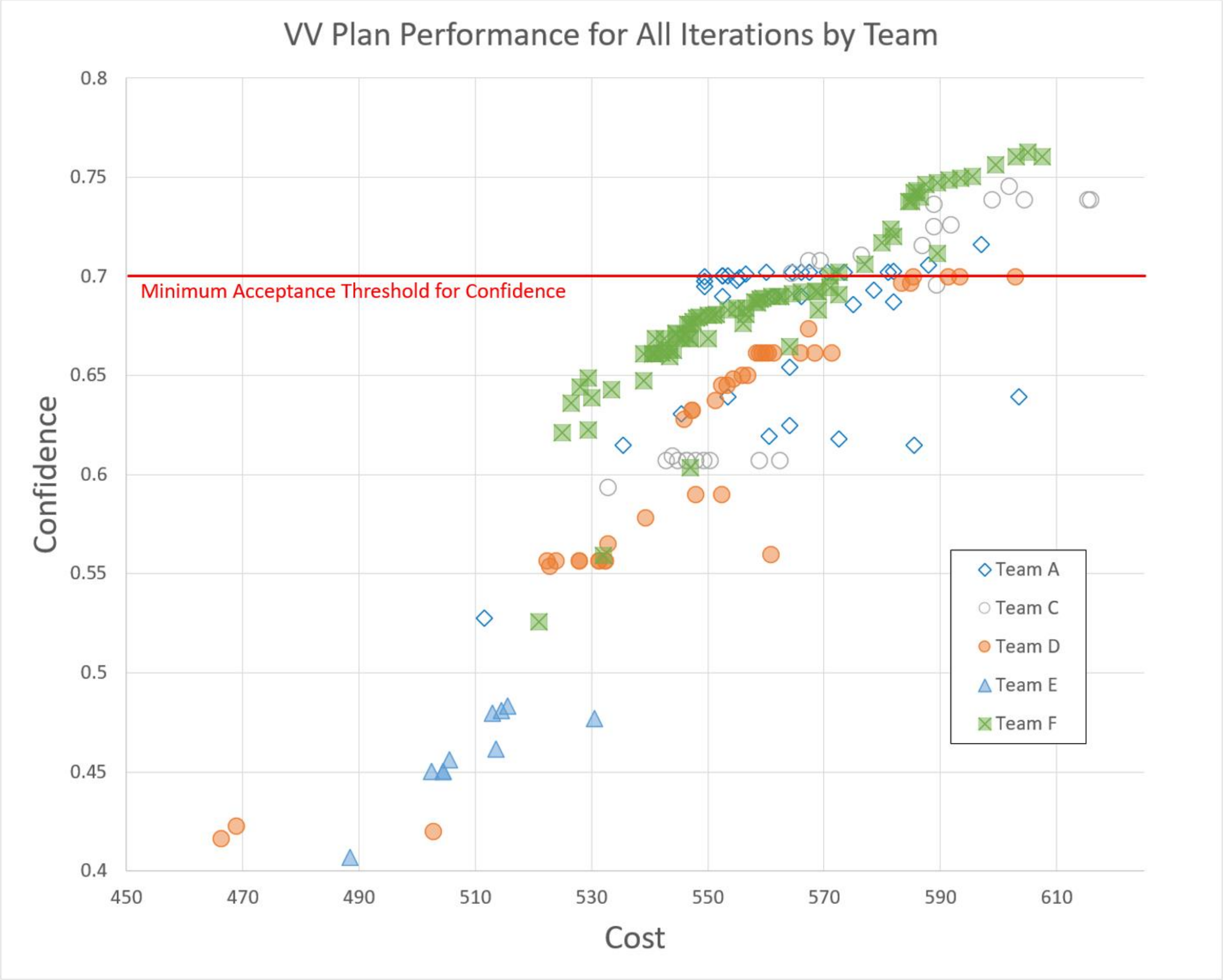


Figure 21: Workshop Team Plan Performance for All Simulations – Teams with filled markers represent treatment teams, those without fill represent control teams

Figure 22 is a zoomed-in version of Figure 21. In this figure vertical and horizontal lines extend from sample points as a visual representation of Pareto optimal points in the Pareto fitness ranking method. If there were any other points in the upper left quadrant extending from a given point, then that given point would be considered dominated.

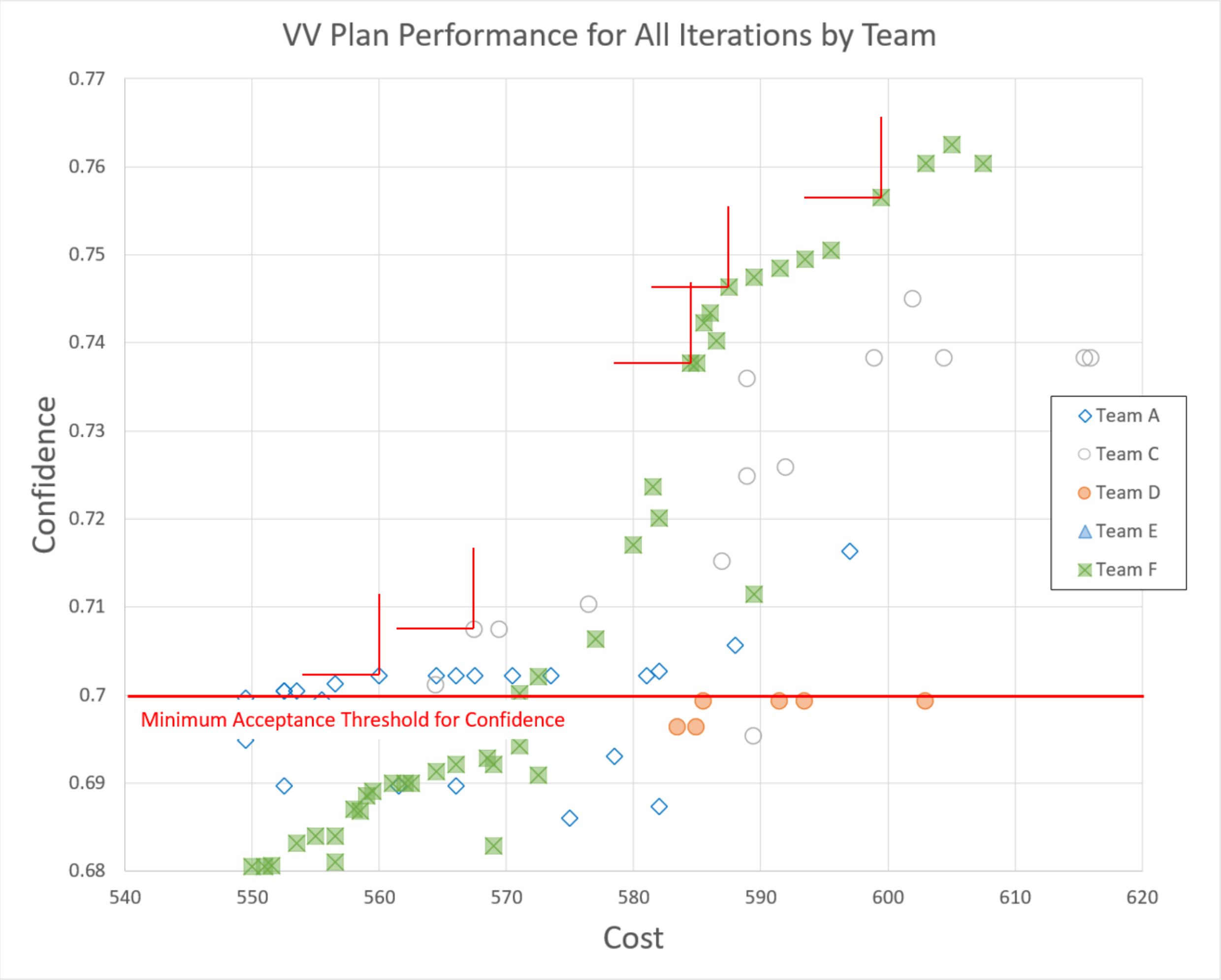


Figure 22: Workshop Team Plan Performance for Simulations Meeting Acceptance Criteria – The lines extending up and to the left of various points provide examples of Pareto optimality. If there were any other points in the upper left quadrant extending from a given point, then that given point would be considered dominated.

Table 4 summarizes the Pareto fitness ranking results. The column label indicates the ranking iteration in sequential order. The row label indicates the team. For example, team A had 12 points that were nondominated in the first round ("0" round) of ranking and 5 points that were nondominated in the second round ("1" round). Remember, all nondominated points from the "0" round from all teams are removed from consideration during ranking in the "1" round.

Team	**Sum of Simulation Results by Pareto Rank**									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A	12	5	3	5	2	4	2	1	1	
B	5	4		2		5	2	1	2	
C	9	2	6	6	2	3	3	1		1
D	5	9	6	12	12	4		1	1	1
E	1	10	8							
F	30	20	20	5	3		1			

Table 4: Pareto Ranking of Simulation Results by Team

Figure 23 shows the Pareto Fitness Ranking from the table above plotted by team against the team's average Pathfinder network similarity score. Each team has a column of plotted points. The lowest point corresponds to Pareto scores from the "0" round only. The next point up corresponds to the "0" round plus the "1" round and so on until rounds "0" through "3" are all included in the uppermost point.

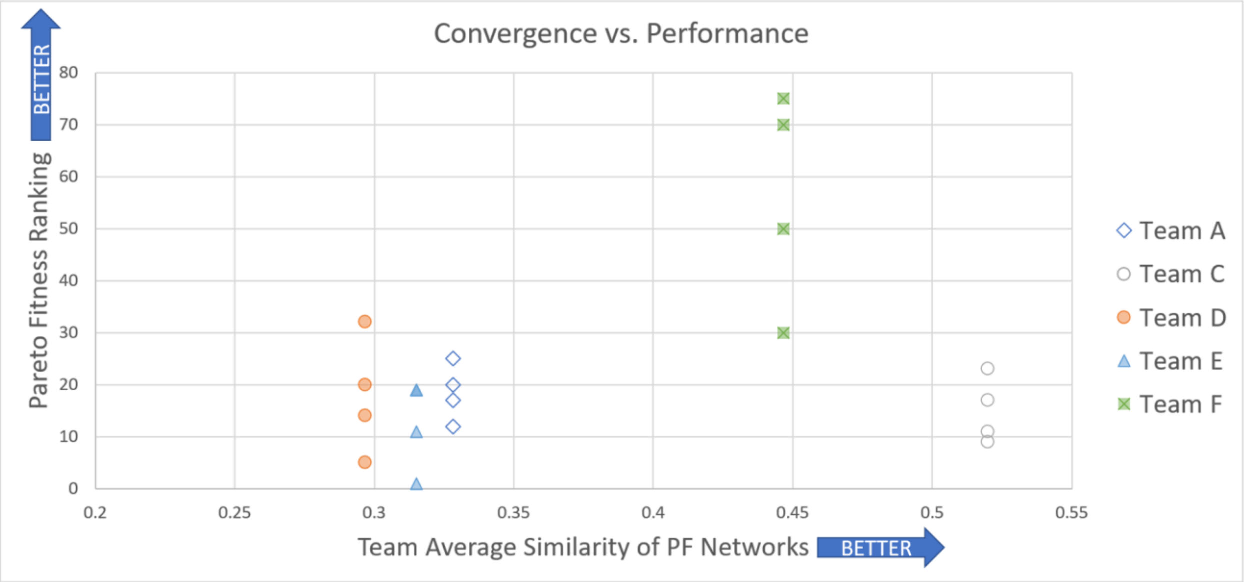


Figure 23: Workshop Team Mental Model Convergence vs. Performance Results

Figure 24 is similar to Figure 23 except it shows the Pareto Fitness Ranking plotted against the team's average standard deviation of concept weights.

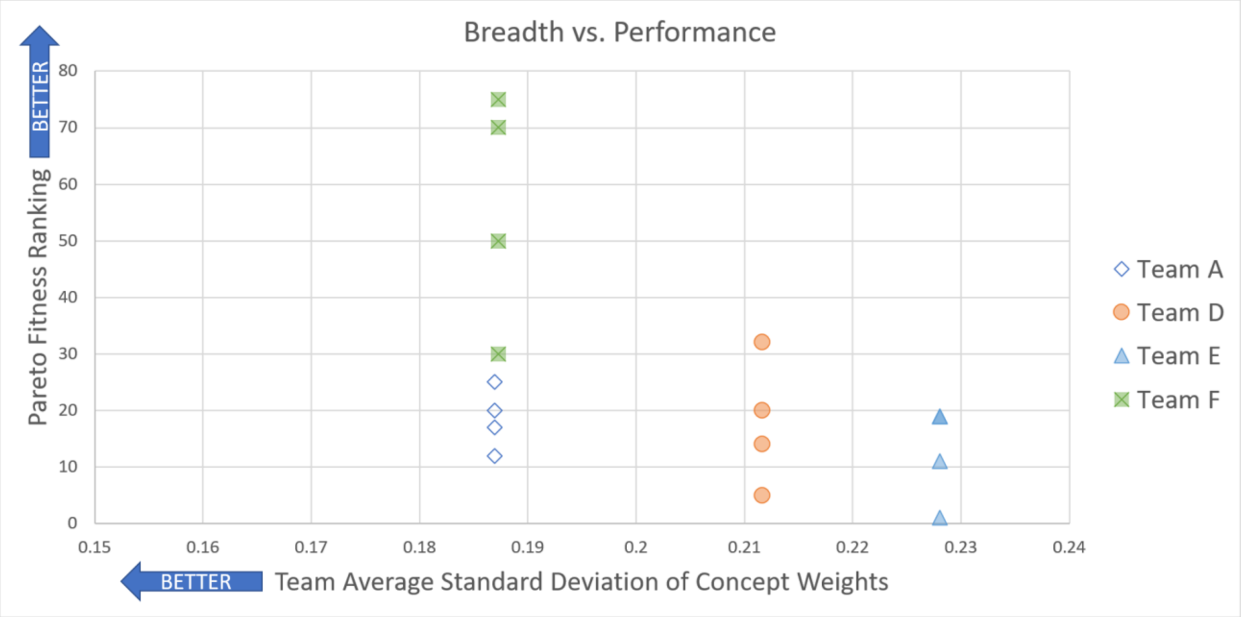


Figure 24: Workshop Team Mental Model Breadth vs. Performance Results

Chapter 7

Discussion/Interpretation

The following section revisits the hypotheses in light of the results shared above.

H1: Mid-stage performance of R&D teams increases with the level of sharedness of mental models.

Figure 23 shows a generally positive association between increasing team average similarity of Pathfinder networks and team performance. This becomes more dramatic as additional iterations of Pareto ranks are included. However, Team C (a control team) stands out as having the highest average similarity yet low performance. The design walk does indicate Team C has amongst the most thorough exploration above the confidence minimum. They did, however, generate a relatively low number of results. Their design walk was somewhat chaotic and trended toward the utopian (low cost, high confidence) direction which led to more dominated points. This was an interesting general trend for all control teams. In contrast, the treatment teams' design walks had more tendency toward iso-performance type exploration or exploration parallel to the Pareto curve.

Team F was a clear, strong performer in this analysis. They were highly generative and had broad exploration of the area near the Pareto front. It should be noted that Team F generally had high similarity amongst team members. However, the team did include two individuals with the highest single pairing score in the experiment (approximately 2.7 standard deviations above the mean).

H2: Mid-stage performance of R&D teams increases with increasing levels of comprehensive breadth of shared mental models.

The analysis results for H2 are not clear but do show a general trend toward supporting H2, especially within the treatment group. As the team average standard deviation of concept weights goes down, the performance tends to increase, again however with a control group outlier. Team A had consistently low standard deviation across the group but low performance. They had fairly chaotic exploration of the solutions space as well and, at one point, began to reduce cost in small increments as they maintained approximately 0.7 confidence. This walk toward lower cost while maintaining confidence led to many dominated points.

H3: Teams that develop and socialize specific, structured context tools will converge faster on mental models than those that socialize concepts in a freeform nature.

Analysis results for H3 indicate very limited if any movement in the sharedness of mental models over the course of the experiment.

H4: Teams that develop and socialize specific, structured context tools will develop more holistic mental models than those that socialize concepts in a freeform nature.

Analysis results for H4 also show only a slight trend toward increased breadth for treatment groups.

See Chapter 8 for further discussion.

Chapter 8

Conclusions, limitations, and considerations for future work

This thesis has empirically explored the application of relatively recent concepts and analysis techniques in team dynamics to a focused situation. The situation of mid-stage development of R&D teams is a fundamental situation that has challenged many organizations. How can a complex engineering development project be efficiently transitioned between critical phases? Various tactics are often employed, including phase gates, shifts in governance, structured entry points for additional teams into the team-of-teams, layered metrics, and handoffs between completely separate teams of like disciplines to list only a few. Often the efforts are heavily focused in the domain of process controls. This thesis aims to extend the exploration of detection and feedback of behavioral phenomena and enhance awareness and response to factors in the domain of industrial and organizational psychology.

The quasi-experiment was designed for execution in a student workshop scenario. It had the architectural elements of the situation described above and was carried out in a relatively rapid fashion. This enabled insights into elements of the phenomenon being studied as well as elements of the design of the experiment, both of which are discussed in this section for future consideration.

Data collected during this experiment indicate trends supporting the hypothesis that mid-stage performance of R&D teams increases with both increasing convergence and increasing comprehensive breadth of shared mental models. Based on these findings, it is concluded that there is significant potential for focused development of shared mental models within R&D teams to support critical mid-phase transitions.

However, the experiment was unable to identify support for the hypothesized method by which mental models can be developed. This could be indicative of the limitations of the experiment. In order to investigate further, additional sensors are proposed for future experiments. The analysis was conducted with the treatment as a binary input. Teams were either in the control group or the treatment group, but there were no quantitative means of assessing the teams' level of interaction with the treatment materials. Future experiments should consider instrumenting this interaction. They should also consider eliminating the tension between time demands on activities. In the existing structure of this experiment, participants had dedicated time blocks to develop test plans. It was at the discretion of the team how this time was divided between activities such as information review, strategizing, software familiarization/experimentation, reflection/analysis of results, etc. Future experiment design should consider dedicated, instrumented blocks of time for these activities where treatments could be implemented in a controlled fashion, e.g., requesting teams generate deliverables such as structured context tools from early sessions.

Additional Questions

The data that was collected from the experiment also highlighted aspects that raise additional questions.

When teams had high average mental model sharedness but high standard deviation (resulting from a few individuals with very highly aligned mental models), how did this affect team dynamics? Were other elements from the Salas et al. heuristic highlighted, and how did this affect the development of the team shared mental model?

There appeared to be a treatment-specific trend in the design walks where treatment teams exhibited a tendency toward iso-performance type exploration vs. a more chaotic exploration from the control groups that generally trended toward the utopia point. What were the differences in strategies that led to this behavior? Were the control teams attempting to "game" the software? How might this be further instrumented for insights?

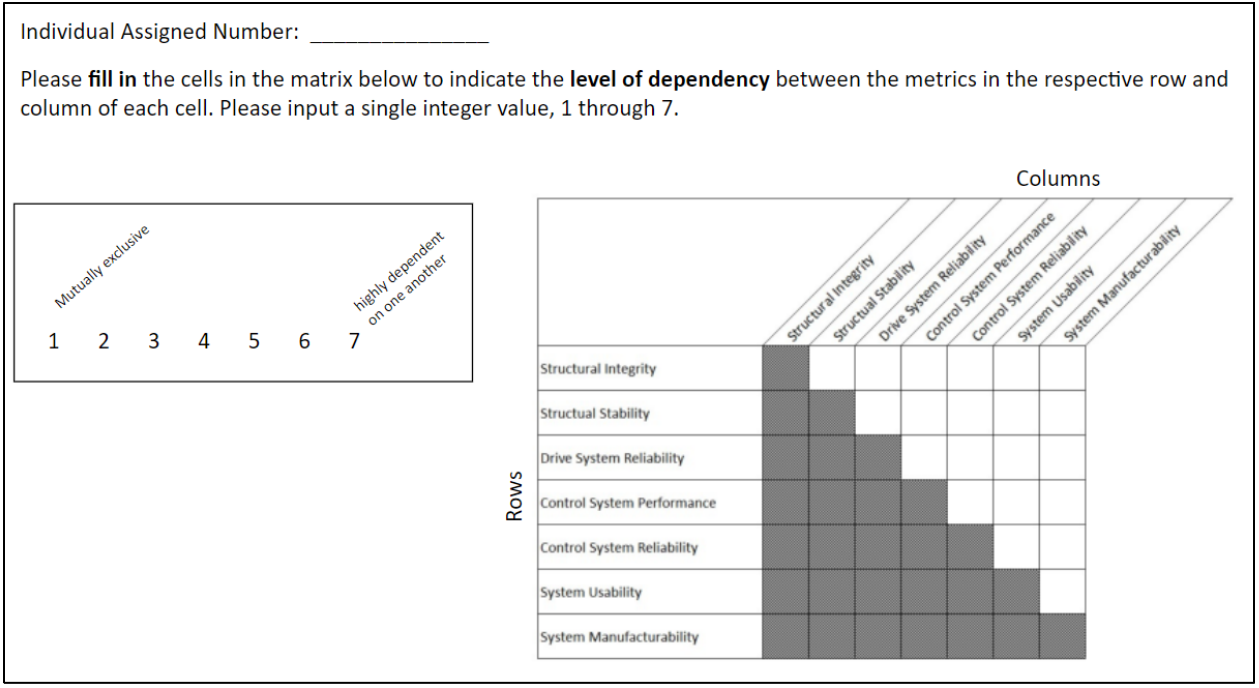
Could there be additional ways to interpret the patterns of exploration as somehow indicative of treatment? For example, were treatment teams exploring solution elements within a single or subset of concepts that might indicate exploration of some aspect of concept links/proximities? Could the selections themselves indicate team perception of different concepts (nodes)?

Is there an indication of what might have led to the control group outliers?   These teams were assessed as having aligned mental models but low performance. Did they have aligned but inaccurate mental models? How might an exploration of the teams' perception of systemic criticality of the presence of nodes and the topological significance of specific proximities be altered, i.e., is the generalization of average similarity of network representations too structured or too abstract to draw meaningful conclusions? Can integration of feedback within teamwork enable identification of a need to not just align but possibly re-architect team mental models?

How does the variation in perceived relatedness (presence or absence of specific links) affect the topology of the remaining concepts that are perceived dependent? Can differing mental models be measured as more or less complimentary based on the topological variations, i.e., can differing constructs compensate for known differences?

These concept questions represent potential next steps in further developing the teamwork experiments in this domain.

Appendix A – Surveys



Individual Assigned Number: _______________

Please **fill in** the cells in the matrix below to indicate the **level of dependency** between the metrics in the respective row and column of each cell. Please input a single integer value, 1 through 7.

Mutually exclusive

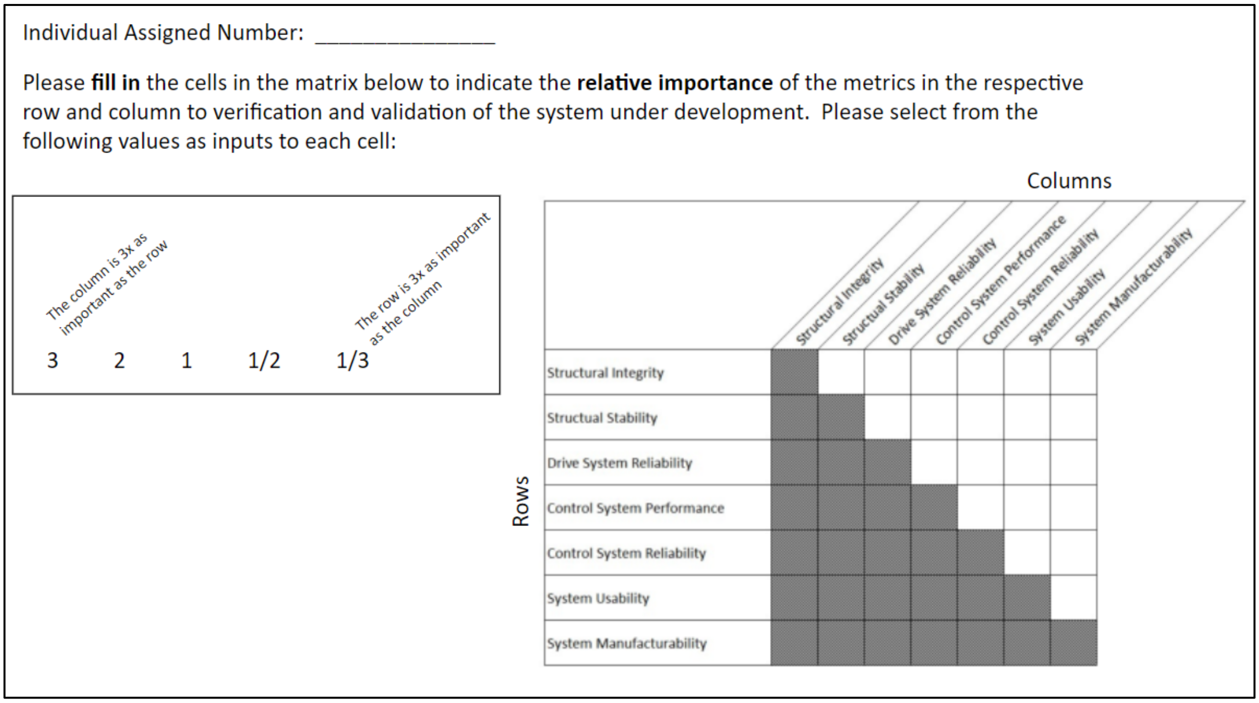
highly dependent on one another

1 2 3 4 5 6 7

Columns

Rows

	Structural Integrity	Structual Stability	Drive System Reliability	Control System Performance	Control System Reliability	System Usability	System Manufacturability
Structural Integrity	▨						
Structual Stability	▨	▨					
Drive System Reliability	▨	▨	▨				
Control System Performance	▨	▨	▨	▨			
Control System Reliability	▨	▨	▨	▨	▨		
System Usability	▨	▨	▨	▨	▨	▨	
System Manufacturability	▨	▨	▨	▨	▨	▨	▨



Individual Assigned Number: _______________

Please **fill in** the cells in the matrix below to indicate the **relative importance** of the metrics in the respective row and column to verification and validation of the system under development. Please select from the following values as inputs to each cell:

The column is 3x as important as the row

The row is 3x as important as the column

3 2 1 1/2 1/3

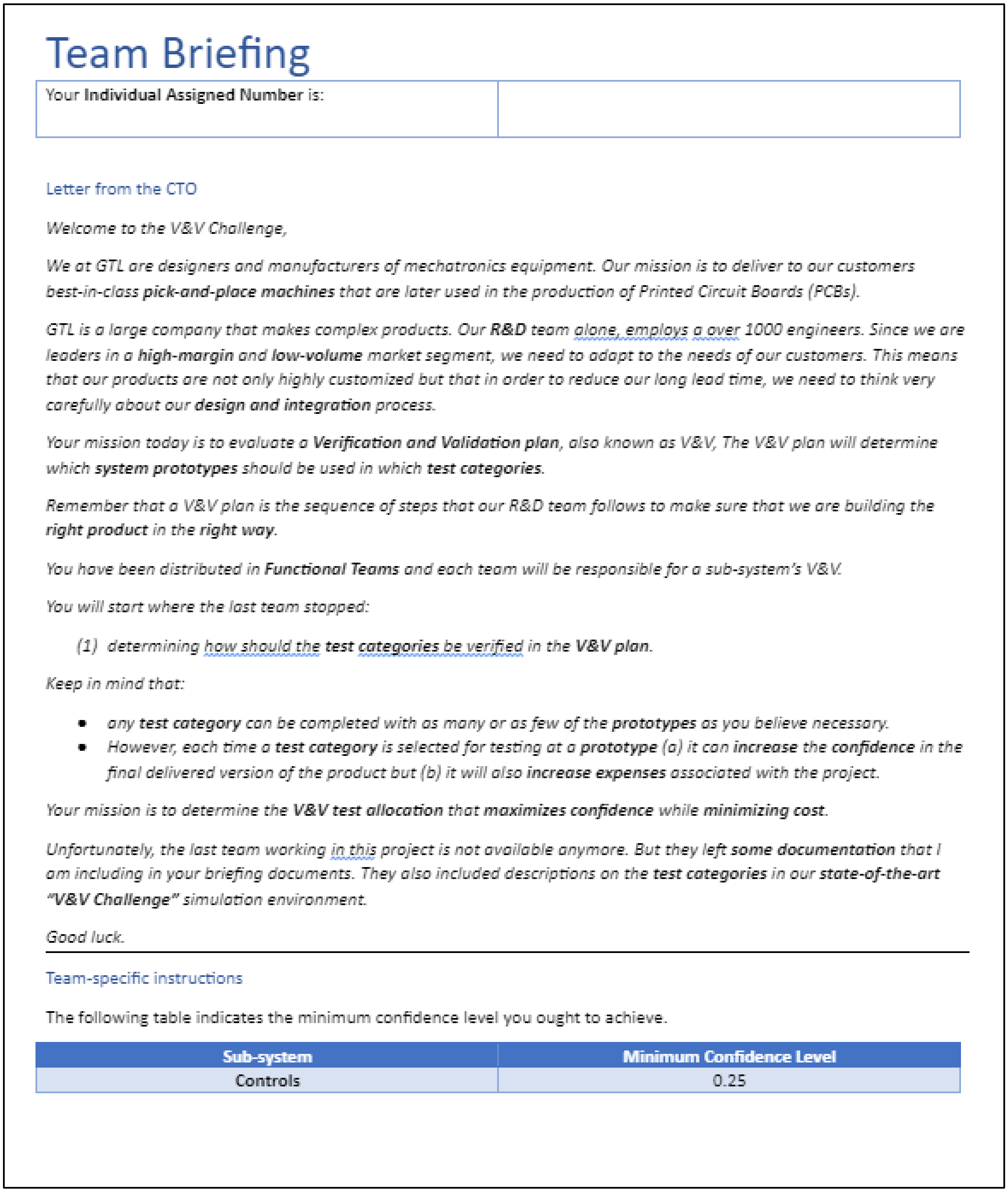
Columns

Rows

	Structural Integrity	Structual Stability	Drive System Reliability	Control System Performance	Control System Reliability	System Usability	System Manufacturability
Structural Integrity	▨						
Structual Stability	▨	▨					
Drive System Reliability	▨	▨	▨				
Control System Performance	▨	▨	▨	▨			
Control System Reliability	▨	▨	▨	▨	▨		
System Usability	▨	▨	▨	▨	▨	▨	
System Manufacturability	▨	▨	▨	▨	▨	▨	▨

Appendix B – Workshop Materials

Controls Team Member Content



Team Briefing

Your **Individual Assigned Number** is:	

Letter from the CTO

Welcome to the V&V Challenge,

*We at GTL are designers and manufacturers of mechatronics equipment. Our mission is to deliver to our customers best-in-class **pick-and-place machines** that are later used in the production of Printed Circuit Boards (PCBs).*

*GTL is a large company that makes complex products. Our **R&D** team alone, employs a over 1000 engineers. Since we are leaders in a **high-margin** and **low-volume** market segment, we need to adapt to the needs of our customers. This means that our products are not only highly customized but that in order to reduce our long lead time, we need to think very carefully about our **design and integration** process.*

*Your mission today is to evaluate a **Verification and Validation plan**, also known as V&V, The V&V plan will determine which **system prototypes** should be used in which **test categories**.*

*Remember that a V&V plan is the sequence of steps that our R&D team follows to make sure that we are building the **right product** in the **right way**.*

*You have been distributed in **Functional Teams** and each team will be responsible for a sub-system's V&V.*

You will start where the last team stopped:

*(1) determining how should the **test categories** be verified in the **V&V plan**.*

Keep in mind that:

- *any **test category** can be completed with as many or as few of the **prototypes** as you believe necessary.*
- *However, each time a **test category** is selected for testing at a **prototype** (a) it can **increase** the **confidence** in the final delivered version of the product but (b) it will also **increase expenses** associated with the project.*

*Your mission is to determine the **V&V test allocation** that **maximizes confidence** while **minimizing cost**.*

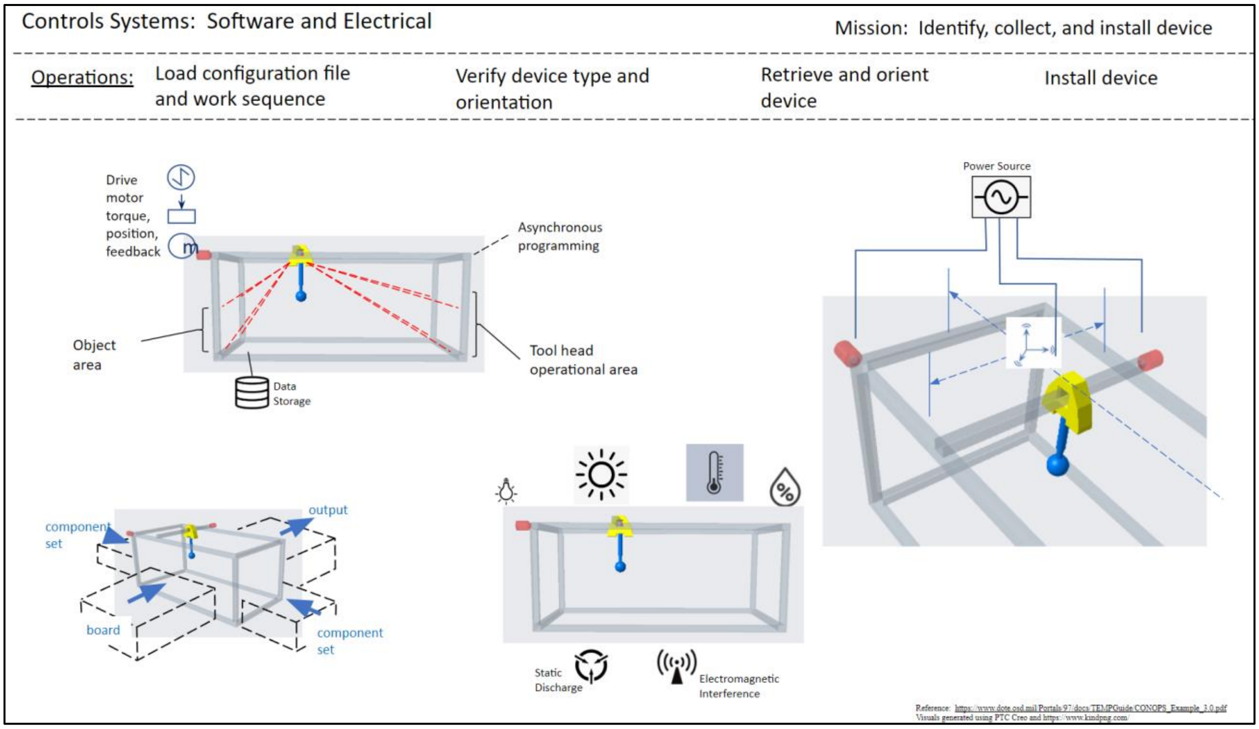
*Unfortunately, the last team working in this project is not available anymore. But they left **some documentation** that I am including in your briefing documents. They also included descriptions on the **test categories** in our **state-of-the-art "V&V Challenge"** simulation environment.*

Good luck.

Team-specific instructions

The following table indicates the minimum confidence level you ought to achieve.

Sub-system	Minimum Confidence Level
Controls	0.25



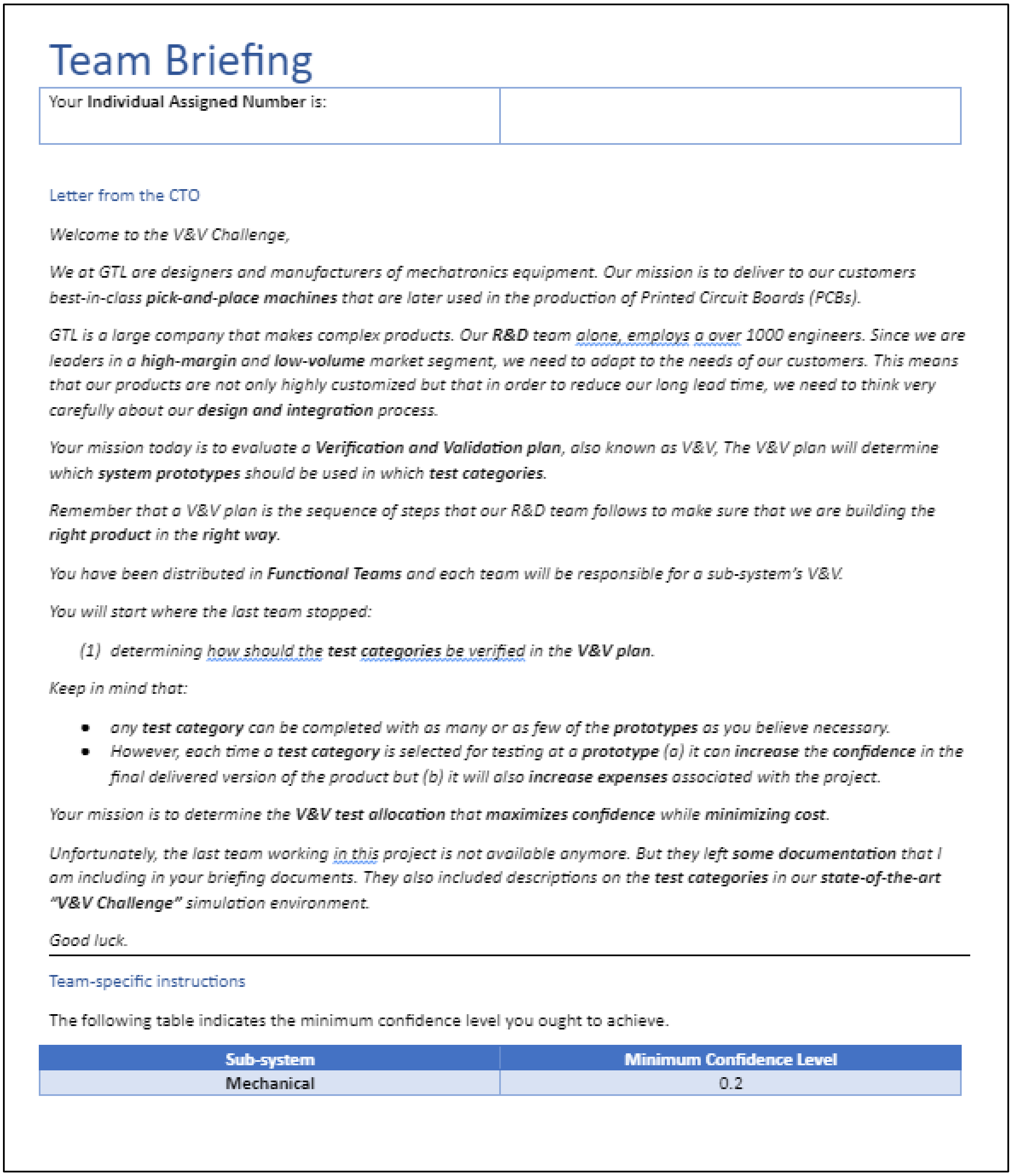
Controls Systems: Software and Electrical

Mission: Identify, collect, and install device

Operations: Load configuration file and work sequence | Verify device type and orientation | Retrieve and orient device | Install device

Reference: https://www.dote.osd.mil/Portals/97/docs/TEMPGuide/CONOPS_Example_3.0.pdf
Visuals generated using PTC Creo and https://www.kindpng.com/

Mechanical Systems Team Member Content



Team Briefing

Your **Individual Assigned Number** is:	

Letter from the CTO

Welcome to the V&V Challenge,

We at GTL are designers and manufacturers of mechatronics equipment. Our mission is to deliver to our customers best-in-class ***pick-and-place machines*** *that are later used in the production of Printed Circuit Boards (PCBs).*

GTL is a large company that makes complex products. Our ***R&D*** *team alone, employs a over 1000 engineers. Since we are leaders in a* ***high-margin*** *and* ***low-volume*** *market segment, we need to adapt to the needs of our customers. This means that our products are not only highly customized but that in order to reduce our long lead time, we need to think very carefully about our* ***design and integration*** *process.*

Your mission today is to evaluate a ***Verification and Validation plan****, also known as V&V, The V&V plan will determine which* ***system prototypes*** *should be used in which* ***test categories****.*

Remember that a V&V plan is the sequence of steps that our R&D team follows to make sure that we are building the ***right product*** *in the* ***right way****.*

You have been distributed in ***Functional Teams*** *and each team will be responsible for a sub-system's V&V.*

You will start where the last team stopped:

(1) determining how should the ***test categories*** *be verified in the* ***V&V plan****.*

Keep in mind that:

- *any* ***test category*** *can be completed with as many or as few of the* ***prototypes*** *as you believe necessary.*
- *However, each time a* ***test category*** *is selected for testing at a* ***prototype*** *(a) it can* ***increase*** *the* ***confidence*** *in the final delivered version of the product but (b) it will also* ***increase expenses*** *associated with the project.*

Your mission is to determine the ***V&V test allocation*** *that* ***maximizes confidence*** *while* ***minimizing cost****.*

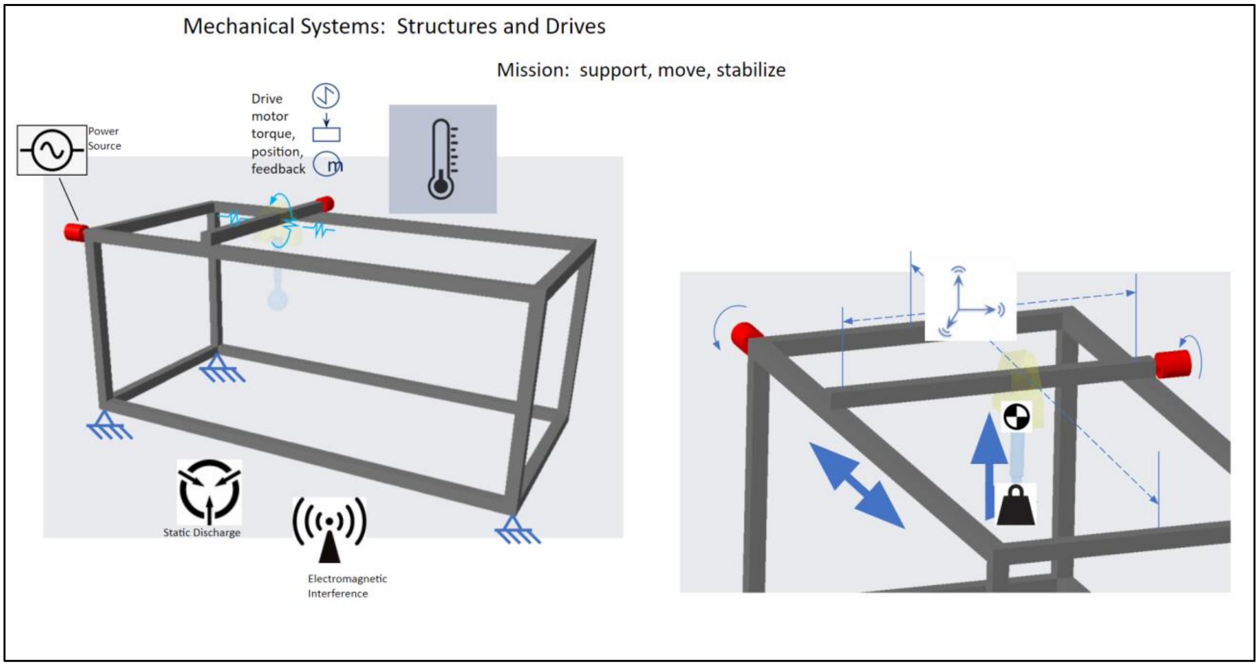
Unfortunately, the last team working in this project is not available anymore. But they left ***some documentation*** *that I am including in your briefing documents. They also included descriptions on the* ***test categories*** *in our* ***state-of-the-art "V&V Challenge"*** *simulation environment.*

Good luck.

Team-specific instructions

The following table indicates the minimum confidence level you ought to achieve.

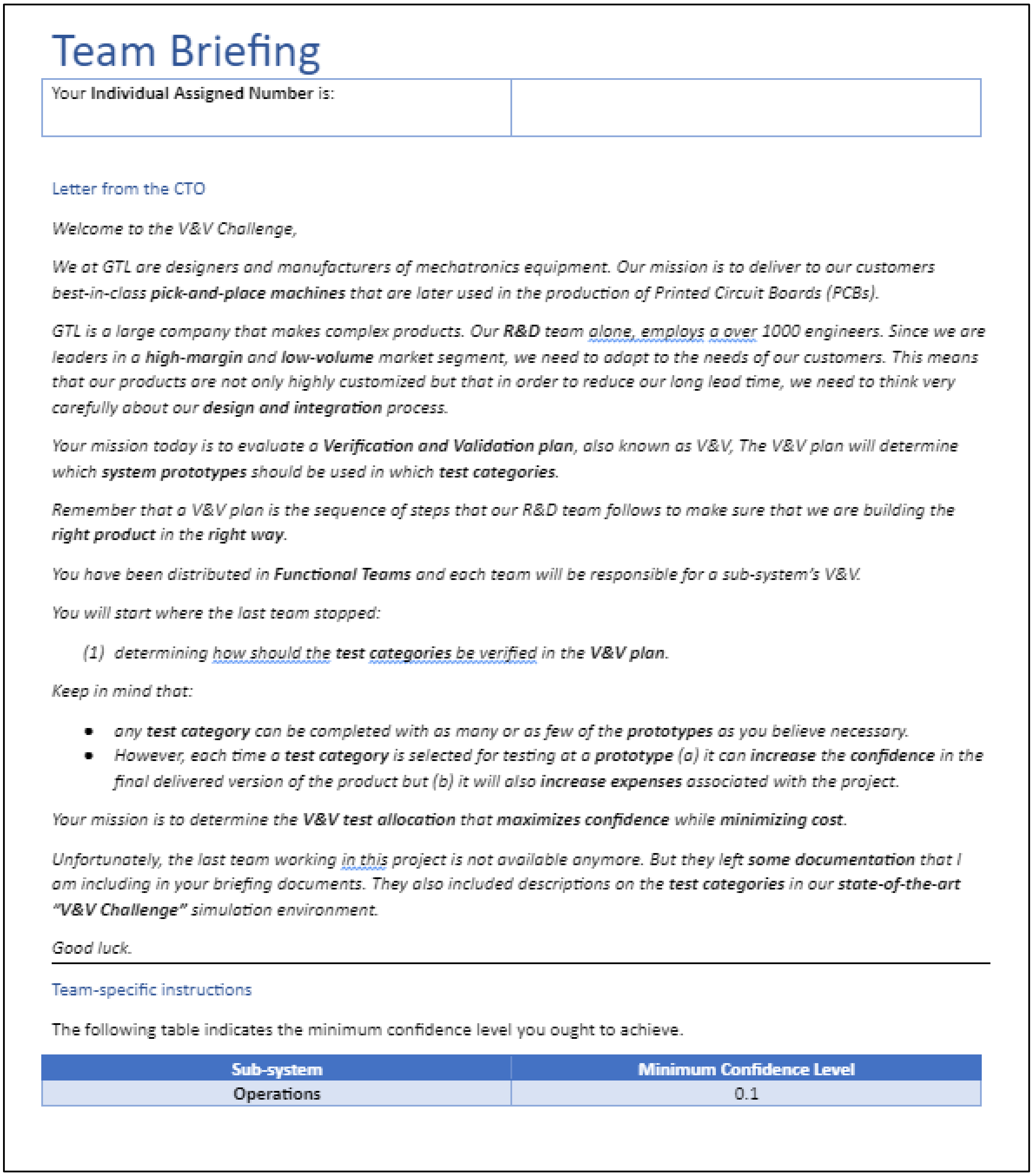
Sub-system	Minimum Confidence Level
Mechanical	0.2



Mechanical Systems: Structures and Drives

Mission: support, move, stabilize

Operations Team Member Content



Team Briefing

Your **Individual Assigned Number** is:	

Letter from the CTO

Welcome to the V&V Challenge,

*We at GTL are designers and manufacturers of mechatronics equipment. Our mission is to deliver to our customers best-in-class **pick-and-place machines** that are later used in the production of Printed Circuit Boards (PCBs).*

*GTL is a large company that makes complex products. Our **R&D** team alone, employs a over 1000 engineers. Since we are leaders in a **high-margin** and **low-volume** market segment, we need to adapt to the needs of our customers. This means that our products are not only highly customized but that in order to reduce our long lead time, we need to think very carefully about our **design and integration** process.*

*Your mission today is to evaluate a **Verification and Validation plan**, also known as V&V, The V&V plan will determine which **system prototypes** should be used in which **test categories**.*

*Remember that a V&V plan is the sequence of steps that our R&D team follows to make sure that we are building the **right product** in the **right way**.*

*You have been distributed in **Functional Teams** and each team will be responsible for a sub-system's V&V.*

You will start where the last team stopped:

*(1) determining how should the **test** categories be verified in the **V&V plan**.*

Keep in mind that:

- *any **test category** can be completed with as many or as few of the **prototypes** as you believe necessary.*
- *However, each time a **test category** is selected for testing at a **prototype** (a) it can **increase** the **confidence** in the final delivered version of the product but (b) it will also **increase expenses** associated with the project.*

*Your mission is to determine the **V&V test allocation** that **maximizes confidence** while **minimizing cost**.*

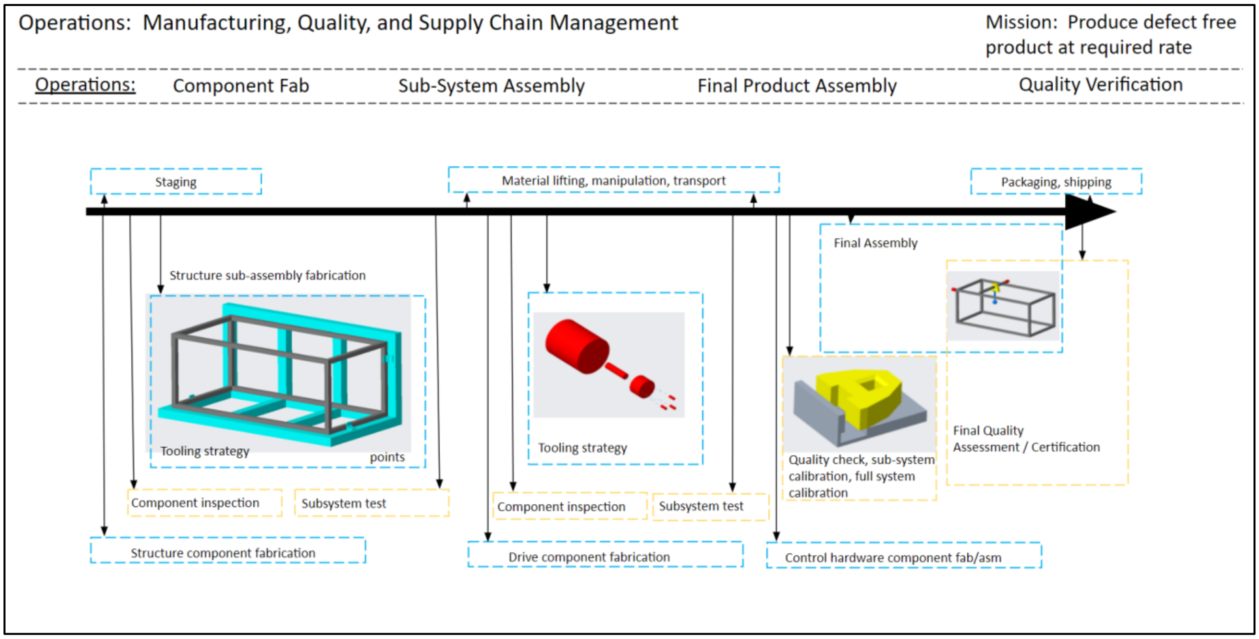
*Unfortunately, the last team working in this project is not available anymore. But they left **some documentation** that I am including in your briefing documents. They also included descriptions on the **test categories** in our **state-of-the-art "V&V Challenge"** simulation environment.*

Good luck.

Team-specific instructions

The following table indicates the minimum confidence level you ought to achieve.

Sub-system	Minimum Confidence Level
Operations	0.1


Operations: Manufacturing, Quality, and Supply Chain Management
Mission: Produce defect free product at required rate
Operations:
Component Fab
Sub-System Assembly
Final Product Assembly
Quality Verification
Staging
Material lifting, manipulation, transport
Packaging, shipping
Final Assembly
Structure sub-assembly fabrication
Tooling strategy
points
Tooling strategy
Quality check, sub-system calibration, full system calibration
Final Quality Assessment / Certification
Component inspection
Subsystem test
Component inspection
Subsystem test
Structure component fabrication
Drive component fabrication
Control hardware component fab/asm

Bibliography

American Psychological Association (APA), *APA Dictionary of Psychology*. Retrieved April 30, 2022, from https://dictionary.apa.org/quasi-experimental-research

Badke-Schaub, Petra et al. “Mental Models in Design Teams: a Valid Approach to Performance in Design Collaboration?” CoDesign 3.1 (2007): 5–20. Web.

Butchibabu, Abhizna et al. “Implicit Coordination Strategies for Effective Team Communication.” Human factors 58.4 (2016): 595–610. Web.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.),Individual and group decision making (pp. 221- 246). Hillsdale, NJ: Lawrence Erlbaum.

Cooke, N.J., Salas, E., Cannon-Bowers, J.A. and Stout, R.J., Measuring team knowledge. Human Factors, 2000, 42(1), 151 – 173.

Edwards, B.D., Day, E.A., Arthur, W. and Bell, S.T., Relationships among team ability composition, team mental models, and team performance. J. Appl. Psychol., 2006, 91(3), 727 – 736.;

Entin, Elliot E, and Daniel Serfaty. “Adaptive Team Coordination.” Human factors 41.2 (1999): 312–325. Web.

Eppinger, Steven D., and Tyson R. Browning. Design Structure Matrix Methods and Applications. Cambridge, Mass: MIT Press, 2012. Print.

Malone, Thomas, and Kevin Crowston. “The Interdisciplinary Study of Coordination.” ACM computing surveys 26.1 (1994): 87–119. Web.

Marks, Michelle A, Stephen J Zaccaro, and John E Mathieu. “Performance Implications of Leader Briefings and Team-Interaction Training for Team Adaptation to Novel Environments.” Journal of applied psychology 85.6 (2000): 971–986. Web.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. Academy of Management Review, 26(3), 356–376.

Mathieu, J. E., Maynard, M. T., Rapp, T., & Gilson, L. L. (2008). Team effectiveness 1997–2007: A review of recent advancements and a glimpse into the future. Journal of Management, 34(3), 410–476.)

Mayer, R. E. (1989). Models for understanding. Review of Educational Research, 59, 43-64.

Messer, Lynne C.. "natural experiment". *Encyclopedia Britannica*, 7 Oct. 2016, https://www.britannica.com/science/natural-experiment. Accessed 30 April 2022.

Mohammed, S., Klimoski, R. and Rentsch, J.R., The measurement of team mental models: we have no shared schema. Organiz. Res. Meth., 2000

Moser, Bryan R., William Grossmann, Phillip Starke. “Mechanisms of Dependence in Engineering Projects as Sociotechnical Systems.” ISPE CE 2015: 142-151

NASA Systems Engineering Handbook. NASA, 2007. Print.

Rebentisch, E. 2017, Foundations of System Design and Management I, lecture notes, Systems Design and Management EM.411, delivered 4Oct2017

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100, 349-363.

Salas, Eduardo et al. “Understanding and Improving Teamwork in Organizations: A Scientifically Based Practical Guide.” Human resource management 54.4 (2015): 599–622. Web.

Schneider, S. C., & Angelmar, R. (1993). Cognition in organizational analysis: Who’s minding the store? Organization Studies, 14(3), 347-374.

Schvaneveldt, Roger W. Pathfinder Associative Networks : Studies in Knowledge Organizations. Norwood, N.J: Ablex Pub. Corp., 1990. Print.

Schvaneveldt, Roger W. Pathfinder Networks. Retrieved 22Feb2022, from https://research-collective.com/PFWeb/index.html

Stout, Renee J et al. “Planning, Shared Mental Models, and Coordinated Performance: An Empirical Link Is Established.” Human factors 41.1 (1999): 61–71. Web.

Terwiesch, Christian, Christoph H Loch, and Arnoud De Meyer. “Exchanging Preliminary Information in Concurrent Engineering: Alternative Coordination Strategies.” Organization science (Providence, R.I.) 13.4 (2002): 402–419. Web.

Thayer, Amanda L, Alexandra Petruzzelli, and Caitlin E McClurg. “Addressing the Paradox of the Team Innovation Process: A Review and Practical Considerations.” The American psychologist 73.4 (2018): 363–375. Web.

Trevelyan, James. “Technical Coordination in Engineering Practice.” Journal of engineering education (Washington, D.C.) 96.3 (2007): 191–204. Web.

Van De Ven, Andrew H., Andre L. Delbecq and Richard Koenig, Jr. (1976). Determinants of Coordination Modes within Organizations. American Sociological Review , Apr., 1976, Vol. 41, No. 2 (Apr., 1976), pp. 322-338

Ward, J. C., & Reingen, P. H. (1990). Sociocognitive analysis of group decision making among consumers. Journal of Consumer Research, 17, 245-262.