Comparing team feedback in a Team Multiple Errands Test

by

Jamiahus Walton

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Industrial and Manufacturing Systems Engineering

Program of Study Committee:
Stephen Gilbert, Major Professor
Michael Dorneich
Eliot Winer

Iowa State University

Ames, Iowa

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

ABSTRACT

The purpose of this research is to better understand how the privacy and audience component of feedback influence individual and team performance. A modified version of the Multiple Errands Tests (MET), the Team Multiple Errands Test (TMET), was developed for and used to evaluate teams. In the present task, three team members were given individual shopping lists and a team shopping list within a virtual mall. The group had to work together to ensure that all of the items were purchased. An expert system offered feedback to participants during the task, and feedback was given publically or privately (privacy variable) and directed at the group or at the individual (audience variable). Individual performance and team performance scores across four sessions conducted by 10 teams were not significantly impacted by the feedback modality. However, initial analysis of communication patterns over trials and team strategy used suggest trends that would be worth exploring with additional participants. This work does demonstrate the value of the TMET test bed as a new methodology for studying team behavior.

CHAPTER 1 - INTRODUCTION

The Challenge of Giving Appropriate Feedback

Teams have the capability to achieve goals that are difficult to obtain by an individual alone. One way to maintain high performance of a team is to implement effective team training. Effective team training is often demonstrated within sports teams (e.g., soccer or American football). During a pre-season training camp, players receive guidance and instruction on how to properly conduct an action in order to improve their skills. Similarly in education, students need guidance and instruction to improve theirs skill. For students, this guidance can come from instructors, or when appropriate, from software such as intelligent tutoring systems (ITSs) that attempt to simulate a human tutor. ITSs have been successful at instructing students individually (Aleven, Mclaren, Roll, & Koedinger, 2006; Hategekimana, Gilbert, & Blessing, 2008; Koedinger, Aleven, Hockenberry, McLaren, & Heffernan, 2004). However, several challenges arise when an ITS attempts to train a team (R. Sottilare, Holden, Brawner, & Goldberg, 2011).

The distribution of feedback is one of several problems that arise when an ITS attempts to train a team. For example, if student makes a mistake or is unsuccessful in an individual setting, then it is clear that that student will receive feedback on his performance. If a student makes a mistake in a team setting, then it is not clear if feedback should be given solely to the student or to the student and the other team members. This issue raises the question of whether teams would perform more effectively when given individual or team feedback, and under what circumstances.

Another challenge arises regarding feedback distribution. As an example, assume that a student makes a mistake during a task in an individual setting. That student would receive feedback that is given only to her. Assume now that an individual makes a mistake in a team

setting. When the feedback is given to the individual, should the team know what feedback was given, or should the feedback that focuses on an individual's performance be private and only known to the individual? In other words, how might the privacy of the given feedback influence the performance of the team as a whole? To address these questions, it is worth exploring a brief history of ITSs, the contexts in which they've been used, and the forms of feedback that have given.

Brief History of Intelligent Tutoring Systems

Shute and Psotka (1996) give an overview of the history of intelligent tutoring systems. One of the early indications of "intelligent machines" being used for teach can be traced back as far the 1920. Many different forms of ITSs have since then been created for a wide variety of domains.

According to Hartley and Sleeman (1973) an ITS needs to have four educational components in order to be used for teaching purposes. These four components are:

1. A representation of the task being taught

2. A way to represent the student (student model)

3. A set of teaching instructions that can follow once component one and two have been completed (feedback)

4. A set of guidance rules (domain model)

In a team setting, one would complement the student model with a team model as well, representing how the team is performing. Developing the student model has been a challenge since the early days of ITSs (J. A. Self, 1974; J. Self, 1999), and the same challenges (and new ones) apply to developing a team model. One reason why modeling a team is difficult is because the model needs to incorporate the interactions among the multiple members of a team. The

amount of collected data that is needed for these interactions can grow quickly because when one member of a team is both perceiving their teammates and being perceived by their teammates (Eduardo Salas, Fiore, & Letsky, 2012). For example, consider a team with members A, B, and C. Assume that A interacts with B. An ITS will need to observe this interaction and answer two questions: "How is member A being perceived by member B?" and "How is member A perceiving member B?" The same questions will also need to be answer for the interactions between members A and C as well as B and C. There may also be interactions by A with the other two members, B and C, simultaneously, etc. This example illustrates how the number of team interactions can quickly increase. However, these data are important, because they must be used to determine the type of content in feedback given to a team and its members.

The type of team is another factor that helps determine the type of content within feedback given to a team. Two types of teams that are of interests are learning-teams and work-teams. Learning-teams are defined as teams that learn while conducting a task, a problem, while work-teams are teams that effectively use their individual expertise to efficiently perform as a team to complete a given task or problem (Fransen, Weinberger, & Kirschner, 2013). Work-teams are different from learning-teams because in work-teams learning is an added value that may result from the collaboration but it is not the focus. These two team types, as well as other teams, can be described using other parameters as well. For example, teams can be described by characteristics of their structure, e.g., leadership, organization, communication, roles, and location. These characteristics, as well as many others, are described in work presented by Bonner et al. (2014). Effective feedback could help all of these forms of teams increase learning and retain knowledge, but for an ITS to give feedback to a team there needs to be a model of the team's characteristics and how it functions.

Teamwork Model

Salas et al. (2005) attempt to lay the theoretical foundation for using the Big Five personality model in teamwork. According to Salas et al., the five components of teamwork are team leadership, mutual performance monitoring, backup behavior, adaptability, and team orientation. Research results on the importance of team leadership are mixed, however. Fransen et al. (2013), for example, suggest that team leadership is oftentimes not important, while Zaccaro, Rittman, and Marks (2001) concluded that team leadership is important to the effectiveness of a team. This current work agrees with research that supports the hypothesis that team leadership is important because there are certain functions the leadership needs to be capable of in order for a team to be effective (E. Salas et al., 2005). One way to support the team leadership is to provide feedback with content that focuses on the functions that need to be performed. In addition, the members of a team need to be able to trust the team leadership. This theory is a fundamental idea and has been supported by research (Dirks, 2000).

Previous research focuses on team leadership and how it influences team effectiveness (Zaccaro et al., 2001) but little research describes how the leadership in a team should give feedback. Butler and Jaffee (1974) focused on how public and private feedback influences the speech behavior of the team leadership when performing leadership functions. The three variables in that study were incentive (informed about a $5 incentive or not informed about the $5), feedback (positive or negative), and how feedback was given (private or public). They concluded that public or private feedback, when combined with incentive, has some general influence on the content and the amount of the leadership speech behavior. However, in this current study, the ITS will be providing written feedback instead of red and green lights used in

the Butler and Jaffee (1974), which may lead to different results. Also, team performance will be the focus and not the speech behavior of the leadership.

Below, we consider the five components of teamwork specifically and how the feedback parameters explored in this research may affect them. Mutual performance monitoring is the capability of members within a team to track other team members' work while continuing to complete tasks, to ensure that things are functioning as expected and other members are following procedures (Dickinson & McIntyre, 1997; E. Salas et al., 2005). This component of the Big Five is important during a team task but it particularly important when a team and its members are under high levels of stress. The challenge with this component is the ability to measure it. This component is difficult to measure (E. Salas et al., 2005). It is important that teams have support when trying to obtain mutual performance monitoring skills (Albon & Jewels, 2014). One component of support is feedback.

This current work focuses on the privacy and audience level of feedback. The audience characteristic of feedback refers to the feedback content. The audience characteristic has two levels: Direct and Group. Feedback that has direct content will address the performance of a specific individual. Feedback that has group content will address the performance of a group as one entity. Little research has focused on how the privacy or audience of team feedback, pertaining to a team's mutual performance, will influence the team's ability to obtain mutual performance monitoring skills. Since this component is difficult to measure, then it is difficult to provide feedback to a team addressing their mutual performance monitoring. Consequently, it is difficult to study how the privacy or audience of feedback influence a team's mutual performance monitoring skill and a team's overall performance. This research attempts to address that gap. This current work contributed to this area by providing empirical data as to how

the privacy and audience of feedback influence team's performance. It could be the case that one of these components, audience or privacy of feedback, could strengthen mutual performance monitoring. Feedback that strengthens a team's mutual performance monitoring could result in an increase in effective backup behavior.

Backup behavior can be defined as providing resources and support (related to a task) to other team members when it is noticed, by members who can provide backup, that there is imbalance with workload distribution within the team (Porter et al., 2003). There a many different ways team members can provide backup behavior. Three ways to provide backup behavior are for a team member to provide verbal feedback or coaching, for a team member to assist another team member in performing a task (Marks, Mathieu, & Zaccaro, 2001), and for a team member to complete an entire task for a member of the team (Marks et al., 2001). This component of the Big Five has the potential to be easier to measure but still comes with its challenges. For example, it would be easy to create an artificial task that is designed to overload a few members. It would not be too difficult to observe a team's backup behavior (assuming the backup behaviors are clearly defined). Consequently, an ITS would have the ability to analyze the team's behavior accurately and provide feedback with content that helps support a team's backup behavior ability. When distributing this feedback to teams, it is unclear how the privacy and audience of the feedback will influence the team's behavior, i.e., whether feedback given publicly or privately would better encourage team members to exhibit backup behavior. However, creating a task that is intentionally designed to overload members of a team may not be desired by researchers for other experimental reasons. In this case there is a possibility that backup behavior may never occur because there is no need for it. Backup behavior may not be needed because the members of the team have enough skill and experience to handle the

workloads given to them. In this case, feedback may not be needed and no data can be collected on how backup behavior was conducted. Another reason that backup behavior may not occur is because a team and its members are adapting to the task.

Adaptability can be defined as the ability apply knowledge, skills, and attitudes that let members recognize deviations from a predicted action and modify actions appropriately (Priest, Burke, Munim, & Salas, 2002). Marks et al. (2001) describe adaptability as strategy formulation and planning. Marks et al argue that strategy formulation and planning have three dimensions: "deliberate planning, contingency planning, and reactive strategy adjustment" (Marks et al., 2001). Adaptability is perhaps one of the most important components of a teams, especially in complex situations (Uitdewilligen, Waller, & Pitario, 2013). In a simple task there are only a few situations that may occurring during the task. However, in complex tasks there are much more situations that can occur. Consequently, there is no way to plan and account for every possible situation. A team that does not have the ability to adapt to novel situations is likely to struggle completing a complex task. However, teams have great potential to have adaptability with the proper support and training (Burke, Stagl, Salas, Pierce, & Kendall, 2006). According to theoretical considerations presented by Klein and Pierce (2001), feedback is important to internal and external adaptation. The feedback that is distributed to teams must build up a team and pertain information that focuses on the task itself (Klein & Pierce, 2001). For the feedback to be effective, it needs a way to correct wrong behavior or incorrect knowledge. Is it best to provide feedback publicly to teams to help encourage adaptability or should that feedback be private?

Team orientation focuses on the attitudinal aspect of the team members rather than the behavioral aspect. Higher team orientation can improve individual satisfaction, effort and performance (E. Salas et al., 2005), facilitate overall performance (Driskell & Salas, 1992; Eby

& Dobbins, 1997), enhance team cooperation (Eby & Dobbins, 1997). This final component of

the Big Five is important because it has the potential to influence every other component. For

example, if a member of the team has low individual satisfactions towards the team then that

member's motivation to perform well will mostly likely be very low. In turn, that member will

not perform at their highest ability and their individual and team performance will suffer. In fact,

Bell (2007) concluded from a meta-analysis that personality and collectivism are related to team

perform in the field setting. An understanding of the team orientation is needed in order to give

effective feedback. A data collection method (e.g., Wagner and Moch's (1986) individualism

collectivism measure) should be used to develop a better understanding of team's orientation.

Having a better understanding of a team's orientation would allow an ITS to provide feedback

that could either increase team orientation or help maintain team orientation. This component as

well as the other components of the Big Five, are important to this current work because it gives

a better understanding of effective teamwork. As well, the Big Five helps inform this current

work of the possible influences feedback may have on teamwork.

There are many different methods that can be used to train teams to help improve

teamwork. This current work attempts to modify a well-established psychological task called the

Multiple Errands Test (MET). This method will be explained in detail in the following section. A

modified version of this task will be explained after the original version is described.

Research Questions

The work described in this thesis attempts to answer the following questions:

1. How does the privacy component of feedback influence team and individual

   performance?

2. How does the audience of feedback influence the team and individual

performance?

3. Can the Multiple Errands Task (MET) be adapted to a Team MET to differentiate

teams' performance?

Thesis Organization

This current chapter introduces the current problem of how team ITSs should give

feedback. This chapter also describes a brief history of ITSs and different components of team

models. The rest of this thesis outlines the contribution to research on developing intelligent

tutoring systems for teams and a new test domain for teams. Chapter 2 discuses previous

research on developing collaborative ITSs, the multiple errands test (MET), general feedback,

team feedback, and teamwork. Chapter 3 discusses the method used to explore the research

questions mentioned and the new Team Multiple Errands Test (TMET). Chapter 4 presents the

data gathered from this current work. Chapter 5 discusses the implications of the results

presented in Chapter 4 and describes areas of future work. The experimental design and

methodology, as well as a portion of the literature review in this thesis, was published in the

proceedings of the ACM Conference on Computer-Supported Collaborative Work (CSCW 2015)

(Walton et al., 2015). Some ideas from Chapter 2 were also presented in more detail in Walton et

al. (2014). However, the writing in this thesis is entirely original.

CHAPTER 2 - BACKGROUND

An important element of training in any domain is feedback. Feedback has many different functions within training. For example, feedback can guide individuals to a specific action. As well, feedback can highlight future rewards in order to motivate individuals to perform better (Ilgen, Fisher, & Taylor, 1979). When discussing feedback within a team setting it is common for researchers to question if feedback should be given at an individual level or a team level (Tindale, 1989). Some studies conclude that effective team past performance feedback will have a positive influence on team performance (Passos & Caetano, 2005). Other studies suggest that performance is influenced by individual level feedback (Berkowitz & Levy, 1956). The purpose of this chapter is to describe the state of research of team feedback, especially as it relates to the stated Research Questions, and highlight gaps in the research and how this current work address these gaps.

Feedback Characteristics

Modality (audial, visual, tactile)
Privacy (public, private)
Focus level (task, process, self-regulation, "self")
Audience (direct, group)
Timing (delayed, immediate)

Figure 1. *Different characteristics of feedback adapted from* (Walton et al., 2014)

Feedback

Feedback can be defined as information given by a source that deals with the recipient's past performance (Hattie & Timperley, 2007). Feedback is one of the central components of the

learning process (S. W. J. Kozlowski & Ilgen, 2006). As well, feedback also has many different characteristics; see Figure 1, adapted from Walton et al. (2014). Few researchers have investigated the influences of feedback given by an ITS to a group. Feedback in a group setting is different than feedback in an individual setting. There are several elements to consider when studying the influence of feedback on performance. The elements of feedback that need to be considered are the following (Gabelica, Bossche, Segers, & Gijselaers, 2012):

1. **Characteristics of the feedback**

2. **Type of dependent variables the feedback interventions targeted**

3. Individual or team situation and characteristics

4. Perception of feedback

5. Processing feedback

6. **The study design**

7. **The team type**

The considerations of this current work will be narrowed by only focusing on four components of feedback: 1) characteristics of the feedback, 2) type of dependent variables the feedback interventions targeted, 6) the study design, and 7) the team type.

Collaborative or Team-Based Intelligent Tutoring Systems

This current research explores the nature of team feedback distributed by ITSs as opposed to ITSs themselves. However, it is important to understand previous work that has been done on ITSs for teams. Though there are many elements that must be considered when focusing on teams, there have been ITSs that were created to facilitate collaborative learning or team training. For example, the Advanced Embedded Training System (AETS) was designed to

expand tactical training and reduce the amount of human support needed to implement the training for teams (Zachary et al., 1999).

There were number of conclusions pertaining to ITSs for teams discovered as a result from the work with the AETS. First, they concluded that there is no one ITS architecture that can be successfully applied to complex situations that are in the real-world. However, Sottilare and colleagues (2011; 2010), creators of the recent GIFT architecture would likely disagree. Second, Zachary et al. discovered that the AETS "could not rely on only one way to get things done" (p. 257) and could not depend on one source for fully reliable data all the time. The more data sources that were available to use, the better. Lastly, embedded training (ET) research can inform team ITS research. ET can be defined as "a built-in capability of an operational system that enables the operator to use the system in a situation where it was designed for, while that situation is not actually available" (Wedzinga, 2006, p. 73). The hope for this embedded training was that it would produce training that is better and more cost efficient by removing the need for a training area outside of the main location (Zachary et al., 1999).

Another attempt at collaborative or team, training is an ITS architecture created by the Army Research Laboratory (ARL) called the Generalized Intelligent Framework for Tutoring (GIFT). GIFT is an open source tutor architecture that is intended to 1) assist in authoring an ITS, 2) keep track of tutoring instructions, and 3) evaluate the learning and performance of the user(s) being tutored (Sottilare, 2012). Currently, ARL is working towards adding a team component that would allow GIFT tutors to tutor teams. Feedback is an important component that needs to be considered as ARL develops GIFT. A number of these considerations have been outlined in work presented by Walton et al. (2014).

Other types of ITSs were created for team members that work together asynchronously. For example, Suh and Lee (2006) attempted to develop an extensible collaborative learning agent (ECOLA) that was intended to encourage members of a team to work collaboratively with one another and to "activate collaborative learning in an asynchronous text-based collaborative learning environment" (Suh & Lee, 2006, p. 461). This work described the rationale behind the design and development of the ECOLA presented. Suh and Lee also described different challenges that exist with collaborative systems. One challenge is that most collaborative agents are made up of complex educational elements. One complex educational element is feedback and how it is distributed it to teams. This current work focuses on feedback and how it influences the performance of teams.

Team Feedback

Feedback can be defined as a distinct process in which a source gives a message to a receiver that contains content that evaluates the receiver (Ilgen et al., 1979). Feedback is an important element to training that has many different characteristics that can be examined. One of those characteristics is the setting in which the feedback is given. The setting that will be the focus in this current work is the team setting. There are a number of different studies that have focused on feedback within a team setting (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Dominick, Reilly, & Mcgourty, 1997; Peñarroja, Orengo, Zornoza, Sánchez, & Ripoll, 2015). Many of these studies have focused on many different components of feedback.

Individual vs. Team Feedback. Feedback for a team is different than feedback for an individual because of the capacity of team members to do something about the information. The types of feedback that are discussed throughout research are individual feedback and group feedback (Nadler, 1979). There have been a number of studies conducted on the individual and

team feedback. However, a frequently cited study on team feedback was conducted by DeShon et al (2004).

Team Feedback Model. The purpose of the work conducted by DeShon et al (2004) was to present and evaluate a model that describes how goals and performance feedback influences performance and learning. This model consisted of two loops that were connected to the behavior of individuals. These two loops are called the individual loop and the team loop. The model suggests that the behavior choices that are made by individuals within a team setting are driven by the most salient individual or team discrepancies.

To test this model, DeShon et al (2004) developed a number of hypothesis based on their model and then conducted an experiment that had 237 undergraduate psychology students. In this experiment the participants were given a PC-based simulation of radar tracking called TEAMSim (S. Kozlowski & DeShon, 2004). DeShon et al (2004) pointed out a number of different conclusions that was drawn from the results of this experiment. There are two conclusions that are relevant to this current work. The researchers found that 1) participants will focus more on individual performance when they only receive individual (Direct) feedback and 2) participants will focus more on team performance when they only receive team-level (Group) feedback (DeShon et al., 2004). The more an individual focuses on her individual performance, the better the individual performs. As well, the more an individual focuses on his team performance, the higher the team performance. These conclusions suggest that if the content of feedback given to participants has information that focuses on individual performance, then the individual performance will increase. Also, if the feedback content focuses on team performance, then team performance will increase. The question remains, however, how the content of the feedback in terms of focus level will affect performance.

Focus Level of Feedback. According to Hattie and Timperley (2007) there are four focus levels of feedback: 1) feedback can focus on whether or not the task at hand is correct or incorrect, 2) feedback may focus on how (i.e., the process) the receiver completed a task, 3) feedback may focus on how well the receiver is able to self-regulate (e.g., self-evaluate), and 4) feedback may focus on the receiver's "self." Hattie and Timperley (2007) claim that both the process focus level (level number 2) and the self-regulated focus level of feedback (level number 3) are powerful with regards to processing and mastery of tasks. Processing and mastery are important because this will help improve teamwork. The focus level of the feedback presented in this current study focuses mainly on the process (level number 2) used to complete the task (i.e., the TMET tasks described in the Team Multiple Errands Task section).

Modality. Modality describes the method in which feedback is delivered to an individual or team. There are different modalities of feedback such as verbal, text, visual, and tactile. An example of an ITS that gives verbal and text-based feedback is the ITSPOKE (Litman & Silliman, 2004). ITSPOKE is a speech-enabled ITS that engages with students through a spoken English dialogue. The goal of this dialogue is to help correct a user's misunderstandings and draw out more complete and detailed answers or explanations. Another ITS, presented by Fletcher and Harris (1996), incorporates ITS techniques to train Remotely Operated Vehicle (ROV) pilots. This ITS uses audio, visual, and tactile modalities to deliver feedback to the users. The modality used in this current work is text that is displayed to users in the upper left hand corner of a display.

Feedback vs. Schedule of Reinforcement

It is important to note the difference between feedback and schedule of reinforcement. In general, a schedule of reinforcements describes the timing between responses and reinforcing

consequences (Groskreutz, 2013). This is different from feedback because feedback can be altered, dismissed, or accepted and may not have the power to influence or encourage other actions (Hattie & Timperley, 2007).

Gaps in previous work

There are still gaps among research even though there are a number of studies that focus on feedback, feedback in a team setting, and teamwork. Researchers agree on the types of feedback (individual or team) that can be given within a team setting (Nadler, 1979). Research agrees that individual and team feedback have some influence on the focus level of participants (DeShon et al., 2004). However, it is unclear exactly how much feedback, such as team-level feedback, influence the team performance, and which level of feedback promotes the greatest level of performance increase (Gabelica et al., 2012). The purpose of this current work is to better understand how the individual and team level feedback influence team's performance.

Another characteristic of feedback that is important is the degree to which feedback is public (Gabelica et al., 2012). According to the review conducted by Gabelica and colleagues, there needs to be some level of publicness on the individual level feedback. However, there seems to be little research on the level of publicness needed on the team-level feedback. Furthermore, there seems to be little research that examines the combination of the audience of feedback (individual vs. teams) and the privacy (private or public) of feedback. The results from this current work will address these gaps in research. This current work can influence many different domains by addressing these gaps in research to help improve team training.

Predicted Results

Imagine there is a team with three members, named A, B, and C. Individual members A and C are highly skilled at their task, and member B is not as skilled as the other two members.

Consequently, A and C would produce higher performance. However, member B's performance increases when working together with A, B, and C on a conjunctive task. This phenomenon is called the Köhler effect (Kerr, Messe, Park, & Sambolec, 2005). Previous research conducted by Hertel et al. (2000) and Kerr et al. (2005) suggest that when we examine the impact that feedback privacy has on performance, we should expect to find that members of the team should increase their performance depending on the information regarding team member performance that is given to the team. Therefore, the hypothesis for this current work is that the performance of a team will be higher when given feedback that is public and has content that is focused on the team's performance than when given feedback that is public and has content that is focused on individual performance or when given feedback that is private and has content that is focused on individual or team performance.

Multiple Errands Test

In this current work a version of the Multiple Errands Test (MET) will be used to test the current hypothesis. The MET is a method that was originally used by Shallice and Burgess (1991) to investigate the ability of patients with prefrontal brain injuries to carry out different cognitive tasks. The purpose of the MET is to create a situation where it is possible for an unexpected event to occur. The patients are given a total of eight tasks (six simple tasks and two challenging tasks). For the six simple tasks the patients simply bought items in a shopping setting (e.g., brown loaf, packet of cough drops, etc.), and for the 2 challenging tasks the patients would be 1) required to meet at a certain place within fifteen minutes after starting the task and 2) obtain four sets of information. While completing these tasks the patients were given instructions, or rules, that they had to follow. For example, the patients were told to "… spend as

little money as possible (within reason) and take as little time as possible (without rushing

excessively)" (Shallice & Burgess, 1991, p. 739).

Over the years researchers have modified the MET in different ways. The MET -

Hospital Version (MET-HV) (Knight, Alderman, & Burgess, 2002) was designed specifically for

the hospital setting with a focus on patients that were not able to be observed in a public setting.

Later, the MET – Simplified Version (MET – SV) (Alderman, Burgess, Knight, & Henman,

2003) was created to study a wider spectrum of people that one may encounter inside different

hospitals. The Virtual MET (VMET), created by Rand et al (2009), is an altered form of the

MET – HV that was conducted in a Virtual Mall (VMall).

Team Multiple Errands Test (TMET)

When participating in this experiment, participants worked within teams of three.

However, to our knowledge there has not yet been a version of the MET that has been adapted

for teams. In this current research a modified version of the MET was developed for teams. This

version of the MET is called the Team Multiple Errands Test (TMET). In the TMET, teams were

given two sets of shopping lists in a VMall (see Chapter 3 for more VMall details): (1) an

individual list and (2) a team list (see Appendix A). The individual members of the team each

had their own individual lists with items that they were responsible for purchasing in each

shopping session (similarly to the original MET). The individual list was unknown to the other

members of the team unless the member who possesses that list announced the items to the team.

The members of the group were all given an identical team list. The team list had items that the

team is responsible for buying. In other words, no one member of the team was responsible for

buying any specific item on the team list. The team had to work together to buy the items on the

team list.

Part of the MET paradigm is that when buying the items, a participant must follow a set of rules. In the TMET, team members must follow a set of rules that address both individual and team behaviors. When a rule is broken during the duration of the experiment it is a considered an error. The seven following rules were adapted from the original MET rules:

1. Do not spend over your allotted amount of money

2. If you enter a store, you must buy something.

3. You must only pick up one item from each store.

4. You can only visit a store once during the duration of a task.

5. You must buy only items that are on your list.

6. Meet up with your teammates at the fountain by when the timer is at 0:30 (30 seconds remaining) or earlier, after all your items are purchased.

7. Signal the researcher when you are finished.

The scores for both the individual and the team are based on the number of correct items collected. In this current work there are six items on the individual list and 18 items on the team list to collect, and teams have eight minutes to complete the task. The individual and team performance score were calculated based on their items purchases, errors made, and time remaining in a session (see Chapter 3 for details). Higher individual and team performance scores indicate higher performance.

Participants' error rate is calculated similarly to the performance score. The error rate is the total number of times an error was committed divided by the total duration of the task. The total number of errors committed by the members of the team was summed to calculate the error rate for the team. The higher the error rate, the less efficient the team or individual. The performance score and the error rate are the main dependent variables.

The purpose of the TMET is to provide a way to evaluate and assess teams. This test bed was presented by Walton et al (2015). To determine if the TMET is an appropriate platform for evaluating teams, there should be a clear performance difference between high performing individuals and teams and low performing individual and teams.

CHAPTER 3 - METHOD

Introduction

The purpose of this experiment was to explore the hypothesis that the performance of a group will be higher when given feedback that is public and has content that is focused on the team's performance than when given feedback that is public and has content that is focused on individual performance or when given feedback that is private and has content that is focused on individual or team performance. The experimental objective was to better understand how feedback audience (i.e., Direct vs. Group) and privacy (i.e., public vs. private) impact team performance. This knowledge will help improve team training in a variety of contexts. This chapter describes the experimental method that was approved by IRB protocol 15-034.

Participants

Since this research is focused on team performance in a somewhat generic shopping task (see Chapter 2 for a description of the TMET task), participants could be drawn from a broad population. In this case, participants were sampled from the undergraduate and graduate student population at a large Midwestern university. Confederates were used when needed to ensure that groups of participants contained three people. Participants were excluded from this experiment if they were not able to use the Body Media sensor for electrodermal skin activity (EDA) or had any known metal allergies as per the Body Media manual (the metal on the Body Media sensor may cause redness or irritation). EDA data were gathered as part of this study but were not the focus of this current research, and will not be discussed in this document.

Procedure

Prior to arriving to the testing location the participants completed an online informed consent form. This consent form outlined the risks of participating in this experiment. When the

participants arrived they were told what they would do in the experiment but were not initially told the specific purpose of the experiment or that there might be confederates. After the introduction to the study, the members of the team participated in a training session within the virtual environment that lasted until they were comfortable and had no further questions, ranging 5-8 minutes. The purpose of this training session was to ensure that each team member had at least the same level of familiarity with the virtual environment and a similar level of understanding on how to interact with objects within the environment. During the training the teams were shown how to control their avatar and were told the rules that they must abide by for each session.

Once the training was completed, the members of the team were each calibrated to their own Body Media sensor (SenseWear Armband Model MF) by finding the baseline for each member, a 10-minute activity. The EDA data were later normalized for each participant. Once the calibration was complete, the team began their first session (each session was capped at eight minutes). Immediately following the first session, the members of the team completed a NASA TLX (approximately 2-3 minutes). After the participants completed the NASA TLX, they completed a post-session survey (approximately 3-4 minutes). After that, the participants had five minutes to debrief the session with each other in free form conversation away from the VMall. Then, the participants began the second session. There were four sessions total. After the all of the sessions were completed (i.e., after the fourth post-session survey was completed), the participants completed an overall survey (approximately five minutes), and then they were debriefed. During the debriefing the participants were told the specific purpose of the experiment and that there may have been confederates involved and were asked not to describe any of the

details of the experiment to anyone. An overview of the procedure in the experiment is described

in Table 1. A total team's session required approximately 1.75 hours.

Table 1. Outline of experimental procedure

Procedure
1. Online Consent Form and Pre-Survey
2. Introduction to Experiment
3. Training Session
4. Calibrate Body Media sensor
5. Session 1
6. NASA TLX Survey
7. Post Session Survey
8. Team Reflection
9. Session 2
10. NASA TLX Survey
11. Post Session Survey
12. Team Reflection
13. Session 3
14. NASA TLX Survey
15. Post Session
16. Session 4
17. NASA TLX Survey
18. Post Session Survey
19. Overall Post Session Survey
20. Debriefing

Participating in a Session

This section will describe the experience of someone participating in a session. Before

the experiment began each participant was assigned a player number (i.e., Player 1, Player 2, or

Player 3). The participants were asked to refer to one another by the name assigned to them.

Each participant received a new individual and team shopping list before the start of every

session. Each participant's individual shopping list was different from the other participants', and

these were different in each session. Each team shopping list was the same for all participants in

a given session, but they were different for each session. Lists were designed to be equal

difficulty each session. The store inventories for each store was reset (same items in same store).

After each session, the team was allowed to discuss with each other how they performed in the previous session and the strategy for the upcoming session.

Surveys

The surveys that were used in this experiment were a post session survey (Appendix B), overall-post session survey (Appendix C), and NASA TLX. Each survey was given online using Qualtrics.

Pre-survey

All of the participants completed a pre-survey before conducting this experiment. The purpose of this survey was to gather some basic demographic information about the participants (e.g., age, degree pursuing, corrected vision, etc.). This survey also asked open ended questions about how often participants worked with teams and the type of teams with which the participants have had experience. The purpose of these questions are to better understand how familiar participants are with teams how often they work with teams to see if this has any influence on the participants performance.

NASA TLX

The NASA TLX is a measure of workload that was a result of a three year long research effort (Hart & Staveland, 1988). The NASA TLX was originally designed with aviation in mind. Today this index has been utilized in many different experiments in many different domains. This index is has six subscales: Mental, Physical, and Temporal Demands, Frustration, Effort, and Performance. The theory behind the NASA TLX is that the workload can be represented by some combination of these six dimensions (Hart, 2006). In this experiment the NASA TLX was given to the participants immediately after all four sessions.

Post-session survey and overall post-session survey

Participants were given a post-session survey and an overall post-session survey in this experiment. The post-session survey was given to the participants after each session, four sessions in total. The purpose of this survey was to gain an understanding of how the participants perceived their performance in the previous session, as well as to gauge their immediate perception of what they believe caused their level of performance. Also, this survey attempted to uncover any hidden influences that may otherwise not have been discovered by observation by asking open-ended questions pertaining to the different types of feedback given. Once all of the sessions were completed, the participants were given an overall-post session survey to complete. The purpose of this survey was to get an understanding of the participant's confidence in her own performance. This survey also attempted to uncover any major influence during the experiment, which may not have been initially considered, by asking a few open-ended questions to each of the participants. The questions pertained to the type of feedback given in the different sessions and the overall experience of the participants.

Confederates

It is common, especially in social psychology, to use confederates in an experimental design (Goethals & Reckman, 1973; Kuhlen & Brennan, 2013; Lockridge & Brennan, 2002). A confederate is a participant who is privy to the purpose of the experiment and the entire experimental process. In this current work the goal of the confederate was to be "neutral." In other words, the confederate would attempt to be an average player that would not encourage specific team strategies, would go along with any plan that the participants (non-confederate) devised during the experiment, would not hinder any team plan that was devised. If the team did not create or discuss a plan, then the participant was given a script of general instructions to follow (see Appendix D).

Feedback

During each session the participants had the chance to receive either direct or group feedback. An example of group feedback is "Team, remember not to spend more money than budgeted." An example of direct feedback is "Player 2, remember to always buy an item from every store you enter." All of the feedback messages that the teams received were relevant to the rule that was broken (Table 2). The first two feedback sentences were used the first time the related rule was broken by an individual within a team. The third and fourth rule in the Related Feedback box were used when the related rule was broken 3 or more times. The privacy of feedback and the audience of feedback had one of four combinations as shown in Figure 2. If participants received multiple feedback responses then the feedback would stack above one another visually. The latest feedback was always at the top of the stack (see Figure 9 for a visual example).
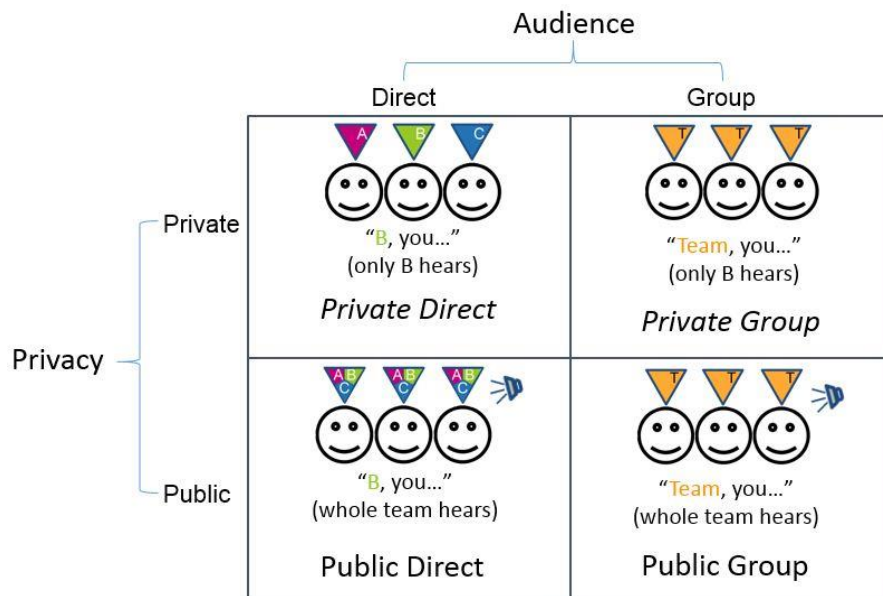
Figure 2. *Experimental conditions. Used with permission from Walton et al (2015). Created by student interns Hud, Mater, and Walker.*

Table 2. List of rules and the related feedback

Rule	Related Feedback
1. Do not spend over your allotted amount of money.	• Player X, remember not to spend more money than budgeted. • Team, remember not to spend more money than budgeted. • Player X, you have spent more money than budgeted. • Team, at least one of your members has spent more money than budgeted.
2. If you enter a store, you must buy something	• Player X, remember to always buy an item from every store you enter. • Team, remember to always buy an item from any store you enter. • Player X, you have forgotten Y times to buy something from a store you enter. • Team, your team has forgotten to buy something from a store you enter Y times.
3. You must only pick up one item from each store.	• Player X, remember you may buy only 1 item from each store. • Team, remember you may buy only 1 item from each store.
4. You can only visit a store once during the duration of a session.	• Player X, remember that you can only visit a store once during the duration of this session. • Team, remember that you can only visit a store once during the duration of this session. • Player X, Y times you have visited a store more than once. • Team, Y times members of your team have visited a store more than once.
5. You must buy only items that are on your list	• Player X, remember to buy only items that are on your list. • Team, remember to buy only items that are on the list. • Player X, Y times now you have bought an item that was not your list. • Team, Y times members of your team have bought an item that was not on the list.

6. Meet up with your teammates at the fountain by when the timer is at 0:30 (30 seconds remaining) or earlier, after all your items are purchased	• Player X, remember to meet at the fountain with at least 30 seconds remaining. • Team, remember to meet at the fountain with at least 30 seconds remaining • Player X, you have X minutes remaining to return to the fountain. • Team, you have X minutes remaining to return to the fountain.

Training

Before conducting the experiment the participants underwent training to ensure that every participant had the same amount of familiarity with the virtual environment. The user interface was explained to the members of the team. They were told how to move their avatar, use the information on the user interface, how to interact with the different stores, and how to access their own inventory of items they had bought. The participants were told about the two shopping lists they would be given during each session. As well, they were also given the seven rules they must follow during every session. At the end of the training, the participants were told how to complete a NASA TLX workload scale properly.

Testing Environment

This experiment was conducted in the User Experience Laboratory at the Virtual Reality Application Center (VRAC) at Iowa State University. Each participant had his or her own desk, laptop computer, and space during the session. The following is the list of hardware and software used in this experiment:

- 3 laptop computers, for participants (Figure 3)
- 1 desktop computer (for researcher)
- 3 Logitech Web Cameras (Figure 4)
- Noldus - Media recorder
- Dividers (x 2) for the participants (Figure 3)
- Unity 4

- Photon Unity Network (PUN)
- 3 Body Media Sensors (Figure 5)

Figure 3*. Experimental setting for three participants.*

Figure 4*. Logitech web camera used to record audio and video of each participant.*

Figure 5*. Body Media sensor*

The fourth computer for the researcher was used to control the clock within the VMall. The three Logitech Web Cameras were used to record the facial expression and communication of the team during each session. The software that was used to record the participants in parallel was Noldus Media Recorder. The dividers were used to separate the participants so they could not see one another. The participants were able to talk to one another during each session by speaking out loud. The number of utterances that were spoken by the participants during each session were recorded and counted. The Unity game engine (v4) was used to create the entire virtual environment for the participants. The Photon Unity Network (PUN) is a free network script, found in the Unity Asset Store, which was used to handle the networking for the teams. This script was used instead of using stock Unity network commands.

Virtual Mall Environment

The virtual mall environment (VMall) that was used in this experiment was first developed and created by National Science Foundation-funded SPIRE-EIT REU student interns Kelsey Walker, Anton Hud, and Samantha Mater in the summer of 2014 (see http://wordpress.vrac.iastate.edu/REU2014/ and video https://youtu.be/mbltv_P_GDE). This REU research was supported in part by National Science Foundation Grant CNS-1156841. The VMall was created with the Unity Game engine (v4) (Figure 6) using C#. The REU student interns set the groundwork for this environment to be possible. They created the logic for participants to enter and exit the stores, buy items, populate the store, display prices for each item. They also designed the stores themselves, combining shapes and objects included in Unity with banners and background imagery found on the Internet. Reuse of existing commercial brands was avoided (Figure 7). However, due to time constraints, the REU students were not

able to produce a complete version of the virtual environment. Their excellent work was continued in the present research to develop a more complete version of the virtual environment.

The author of this work addressed bugs in the virtual environment and added new features to support the present study. Specifically, the interface was adapted to display the amount of money the user had left to spend accurately, display an accurate game timer that was controlled only by the researcher, display the correct player name (e.g., "Player 1") for the participant, and display feedback via text in the upper left hand corner. Shopping items were adjusted to fit the theme of the store (e.g., food store would have food). The author created logic that checked whether the items bought by the individuals matched the items on the lists (individual and team) and checked the actions of the participant to verify that a rule was not broken. Logic was created that removed items bought by one team member from all team members' stores so that items could be bought only once. Scripts were created that handled the networking for the team so the members could see their respective avatars moving within the same world. Name tags were added above each avatar that displayed the assigned name for the entire team to see. A specific interaction between the stores and the users was also created.

To move around the world the participant used the A, W, S, and D or four arrow keys, whichever was more comfortable. It was important that the participants were comfortable with the controls because the avatar did not move exactly like avatars in customary commercial games. To enter a store the user needed to move the avatar towards the store until text appeared on the left side of the screen that told the users to press "M" to enter the respective store. When a user entered a store, that user was able to see all of the items that were available to buy. There were two different inventories, the user inventory (what the player possessed) and the store inventory (available items to purchase). If the user wanted to know what she bought from the

store then she could check her own inventory by pressing "i." To buy an item the user needed to

double click on an item desired for purchase. An item that was bought in a store by a specific

user was not available to other members of the team (i.e., it was removed from the store

inventory). To leave the store, the user moved away from the store until the store inventory

disappeared (Figure 8).

Figure 6. *Overview of the entire VMall with its 12 stores.*

Figure 7. Store within the VMall. To enter, a player approaches the black double doors.

Figure 8. *Example of user interface with game timer, amount of money left, player name (upper right corner), and example of feedback given (upper left corner).*

Figure 9. *Example of user interface and stacked feedback (upper left)*

Experimental overview

This experiment is a 2 x 2 within-subjects factorial design. The two independent variables in this experiment are **privacy** of feedback (Public vs. Private) and **audience** of feedback (Group vs. Direct). The dependent variables are task performance, accuracy, and perception of team performance. Electrodermal skin activity (EDA) was also recorded for the purpose of future data exploration. Other measures were gathered as well in an effort to help establish whether they might be correlated with team performance (See Table 3). Because the same participants experience four consecutive sessions, Session can be explored as an additional main effect that might impact performance. To exploring communication as a predictor of team or individual performance, participants' utterances were counted. An utterance was defined as the collection of words spoken with a single intent, typically divided by a breath. E.g., "Hey should we all go

together?" [breath] "Did you hear me? Should we all go together?" would be considered two

utterances.

Table 3. Details for dependent variables

Construct	Measure	Method of Data Collection
Individual Items Collected	Number of correct individual list items collected over the number of possible correct items (6) X 100	Software Log
Individual and Team Time Remaining	The amount of time remaining at the end of each session over the total amount of time given (480 secs) X 100	Software Log
Individual Errors	The sum of errors committed by an individual in a session	Software Log
Team Items Collected	Number of correct team list items collected over the number of possible correct items (6) X 100	Software Log
Team Errors	The sum of errors committed by the entire team in a session	Software Log
Accuracy	Correct items collected over items collected	Software Log
Error Rate	Rules broken per second	Software Log
Workload	Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration	NASA TLX Survey
Perception of feedback	Likert scale	Surveys
Communication	Number of utterances by players/team	Manually counted within audio recordings
Team Strategy	Any particular strategy team members used during the task.	Observation / Video playback

This study was counterbalanced by ensuring that teams did not experience each condition

in the same order. The four conditions are private direct, public direct, private group, public

group (Figure 2). It was expected that the public team condition would produce higher

performance than the other conditions due to the Köhler Effect (Kerr et al., 2005).

Data Analysis Plan

The three dependent variables for this experiment were the number of items collected (team and individual), time remaining (team and individual), and errors (team and individual). Each of these measures was an important part of overall performance. An excellent team would succeed in all three of these measures and a poor team would produce poor results in all three of these areas. First, a high performing team would have a high percentage of items collected, and a low performing team would have a low percentage. Second, a high performing team would commit few or no errors, and a low performing team would commit a high number of errors. Lastly, a high performing team would have a high percentage of time left, and a low performing team would have a low percentage of time left.

The percentage for items collected was calculated as a percentage based on the number of correct items collected over the total number of items that needed to be collected. The total number of items on an individual list was six, and the total number of items on a team list was 18. These items were separate from one another. In other words, the items that were on a players individual list was not on the team list, and vice versa. Consequently, the percentage of individual items collected had no influence on the percentage of the team items collected. The highest items collected score was 100 and the lowest items collected score is zero.

The timing remaining variable was calculated as a percentage by taking the amount of time remaining at the end of a session and dividing it by the total amount of time given to complete a task. For each session a team was given eight minutes (480 seconds) to complete the shopping task. Theoretically, a team could have, at most, 480 seconds remaining at the end of the session if they did not need any time to complete the task. On the other hand, a team could use up all of the time available and have zero seconds remaining at the end of the session. Therefore, the amount of time remaining ranged from zero to 100. The time remaining score was the same

for the team and for the individual. In the individual analysis, the same time remaining score was given to each participant. In the team analysis, the time remaining score was assigned to the team as a whole. For example, imagine a team had 20% of time remaining at the end of a session. For the team analysis a score of 20% was given to that team as a whole for that session. For the individual analysis, Players 1, 2, and 3 were each given a score of 20%.

The errors variable was calculated by adding up the total number of errors committed within a session. There were two different errors that were calculated for each session. One error was the total number of individual errors committed by a respective team member. The other error score was the team error that is the sum of all the errors committed by every member of the team. The error score, unlike the other dependent variables (Items Collected and Time Remaining), had no maximum value since five out of the seven rules could be broken multiple times. However, individuals and teams could receive an error score of zero if no errors were committed during a session.

It was impossible to combine all three measurements together because it was difficult to normalize the error score since five out of the seven rules could be broken multiple times. Other measurements were gathered during this experiment. Separate videos and audio of each participant was recorded in parallel. These videos were used to manually count the number of utterances for each participant. Table 3 gives an overview of all of the dependent variables collected during this study.

Limitations / Assumptions

When conducting this experiment there were some assumptions that were made and some limitations within this experiment. First, it was assumed that each similar session had a similar level of difficulty, based on results from pilot studies. If the sessions were of different difficulty,

then it would not be possible to compare the three main dependent variables (items collected, time remaining, and errors) in different sessions cleanly because it would depend on the difficulty of the session. Second, this current work is limited by the small number of teams that participated, which limited the statistical power of the analysis. Third, the participants were not told explicitly that they would be scored based on the number of times a rule was broken on the amount of time remaining at the end of each session. The participant might not have internalized the idea of following the rules. The participants may have behaved differently if they knew they were being scored on those additional two measures. Fourth, the participants did not have a pen to use during any of the sessions. The absence of a pen during each session may have increased participants' workload. The participants were not given a pen because the original MET study design did not report participants having a pen during their shopping experience. Therefore, it was assumed that the workload for participants in this current work was similar to the workload of participants in the original MET. However, the teaming aspect of this task, in which participants have to remember which team items have been purchased by others, may have altered that design. This factor can be explored in the future.

CHAPTER 4 – RESULTS

The results and data analysis will be discussed in this chapter. As noted in Chapter 3, each participant experienced four sessions that were counterbalanced with feedback in each of the four configurations:

	Private	Public
Direct	**Private Direct**	**Public Direct**
Group	**Private Group**	**Public Group**

Participants

There were originally 11 teams (33 participants) but due to a technical issue in the data logging system, some data for one entire team was lost, and therefore the entire team was removed from the analysis. Thus there were 10 teams of three with 26 real participants and four confederates used in this study. Four teams had one confederate, each played by four different research personnel trained in the confederate role.

In the demographic data reported below, confederates are not included. They are included in the analyses below since they influenced the teams. The 26 non-confederate participants were made up of 16 males (62%), 10 females (38%) (Table 4). The age range of the participants was as follows: 62% (16) were 18-21 years old, and 38% (10) were 22-30 years old (Table 5). The participant weekly gaming experience was as follows: 27% (7) played zero hours per week, 27% (7) played zero to one hour per week, 4% (1) played one to two hours per week, 23% (6) played two to five hours per week, 12% (3) played five to ten hours per week, 4% (1) played 10 to 15 hours per week, and 4% (1) played 15 to 20 hours per week (Table 6). The highest educational degrees received by participants were as follows: 57% (17) had a High school degree, 7% (2)

had an Associate's degree, 7% (2) had a Bachelor's degree, 13% (4) had a Master's degree, and

3% (1) had a PhD (

Table 7). The team experience for participants was as follows: 31% (8) worked in teams daily,

50% (13) work in teams once or twice a week, 4% (1) worked in teams once or twice every two

weeks, 4% (1) worked in teams once a month, 4% (1) worked in teams once or twice every year,

and 8% (2) rarely worked in teams (Table 8).

Table 4. Participant Gender Demographics

Gender	Count	Percentage
Male	16	62%
Female	10	38%

Table 5. Participant Age Demographics

Age Range	Count	Percentage
"18-21"	16	62%
"22-30"	10	38%

Table 6. Participant Weekly Game Experience

Weekly Game Experience (Hours per week)	Count	Percentage
0 (I don't play video games)	7	27%
0-1	7	27%
1-2	1	4%
2-5	6	23%
5-10	3	12%
10-15	1	4%
15-20	1	4%

Table 7. Highest degree for participant

Highest Degree	Count	Percentage
High school	17	65%
Associate's	2	8%
Bachelor's	2	8%

Master's	4	15%
PhD	1	4%

Table 8. Participant team experience

Team Experience	Count	Percentage
Daily	8	31%
Once or twice a week	13	50%
Once or twice every two weeks	1	4%
Once a month	1	4%
Once or twice every year	1	4%
I rarely work in teams	2	8%

Overview of Confederates and the Three Independent Variables

Below we provide a visualization of all the dependent variables for both individuals and teams. The hope is to give a sense of the variability and growth across sessions. Also, to explore whether the status of being a confederate might have affected the results, the four confederates used in this experiment are displayed with stripes. Because the confederates and confederate teams are somewhat randomly distributed, rather than grouped consistently, we conclude that confederates might have had a notable effect on the results (Figure 10 - Figure 33). As a result, the confederate data was removed from some of the data analysis: data points for all confederates were removed from the individual data analysis, but kept for the team data analysis. The reason the confederate data was kept for the team data analysis because it would be difficult to compare teams of two to teams of three.

Individual Items Collected by Session

The following four figures show individuals' percentage of items collected in each session. The figures were sorted from highest performing participants to lowest performing, to illustrate the overall change in performance over sessions. The items collected percentage for

individuals is the correct number of individual items collected over the total number of correct individual items possible (6). The red striped bars were confederates.
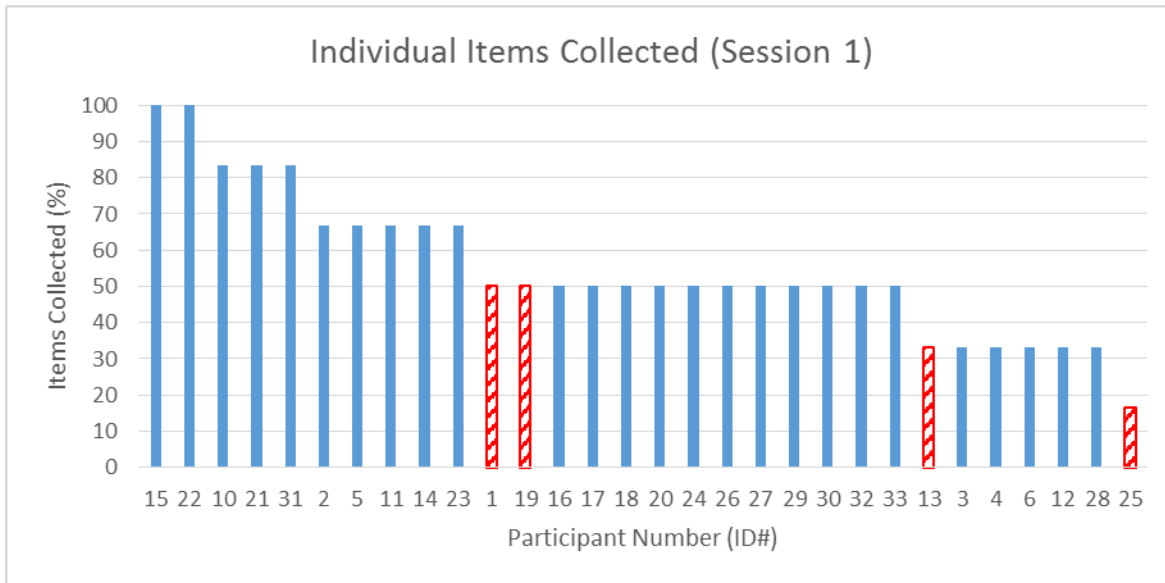
Figure 10. Individual items collected for each participant (Session 1)
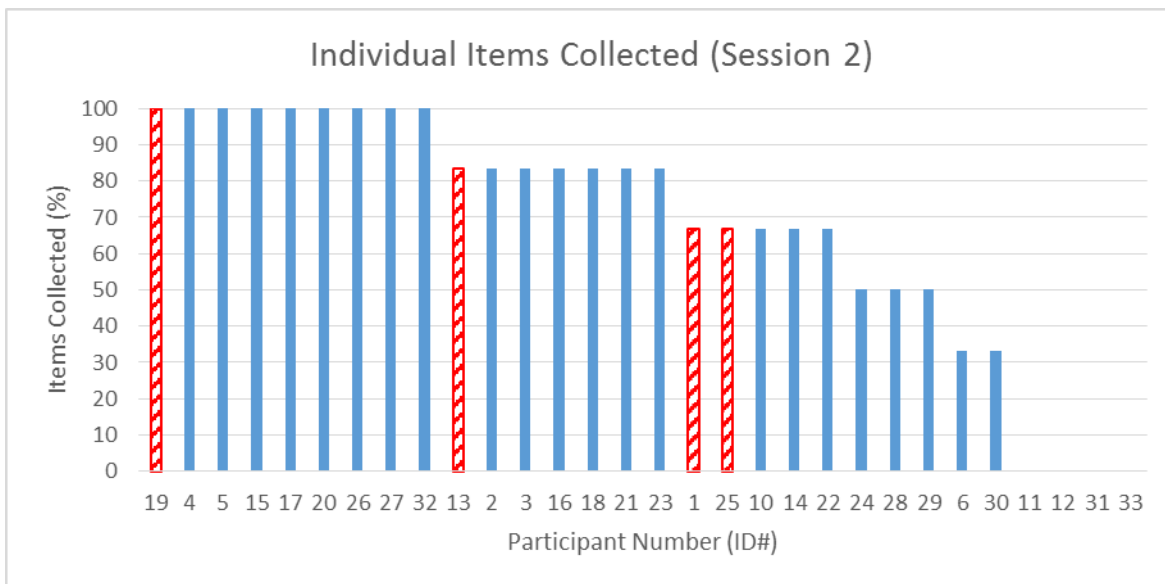
Figure 11. Individual items collected for each participant (Session 2)
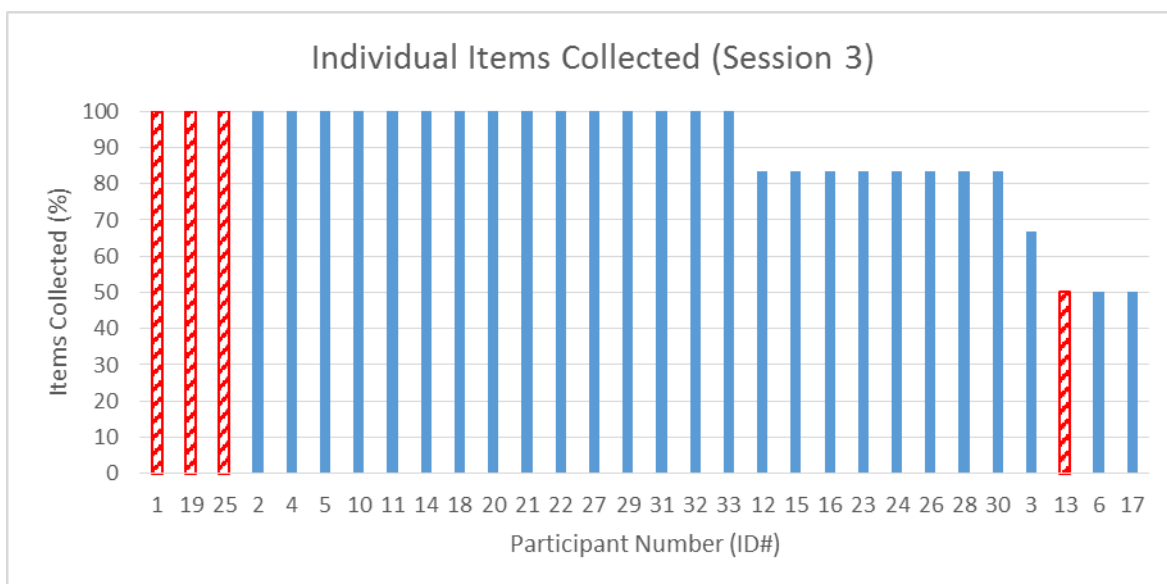
Figure 12. *Individual items collected for each participant (Session 3)*
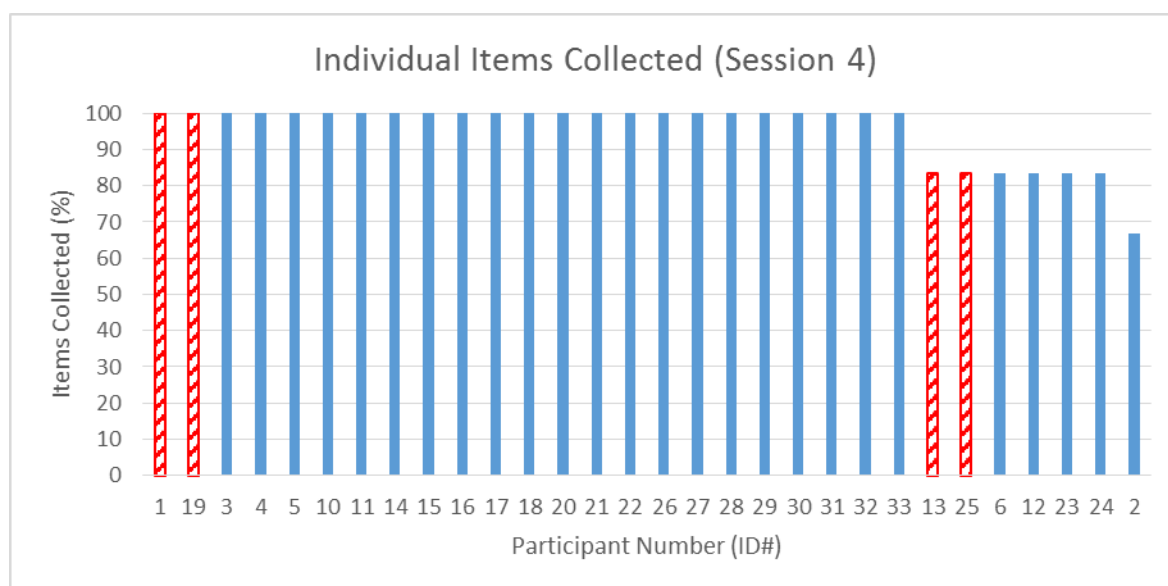
Figure 13. *Individual items collected for each participant (Session 4)*

Individual Time Left by Session

The following four figures show individuals' percentage of time left in each session. The figures were sorted from highest performing participants to lowest performing, to illustrate the overall change in performance over sessions. The red striped bars were confederates.
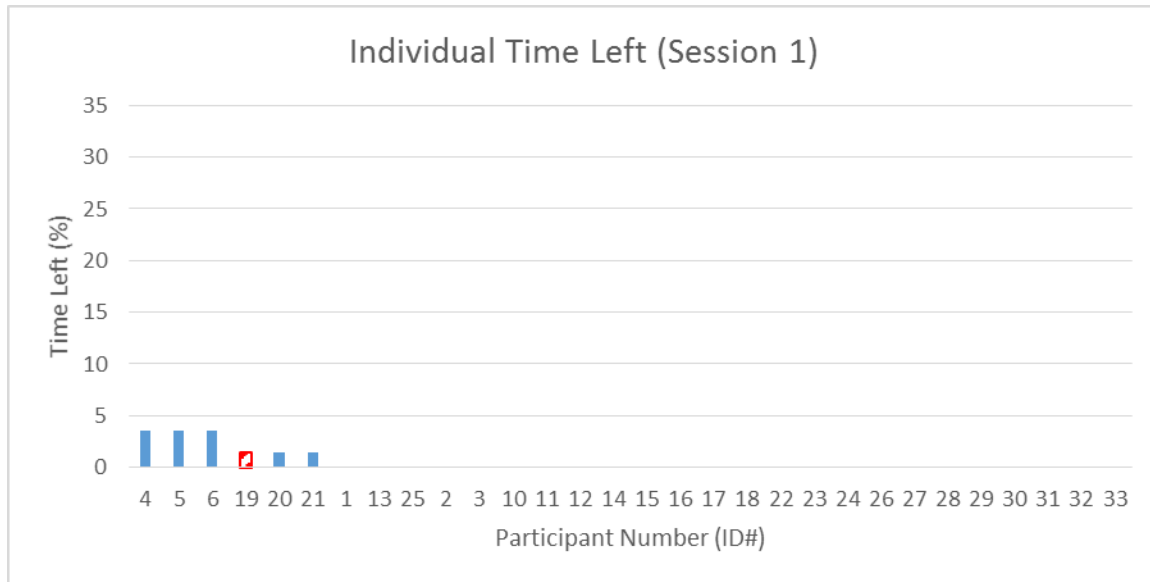
Figure 14. *Time Left for each team (Session 1)*.
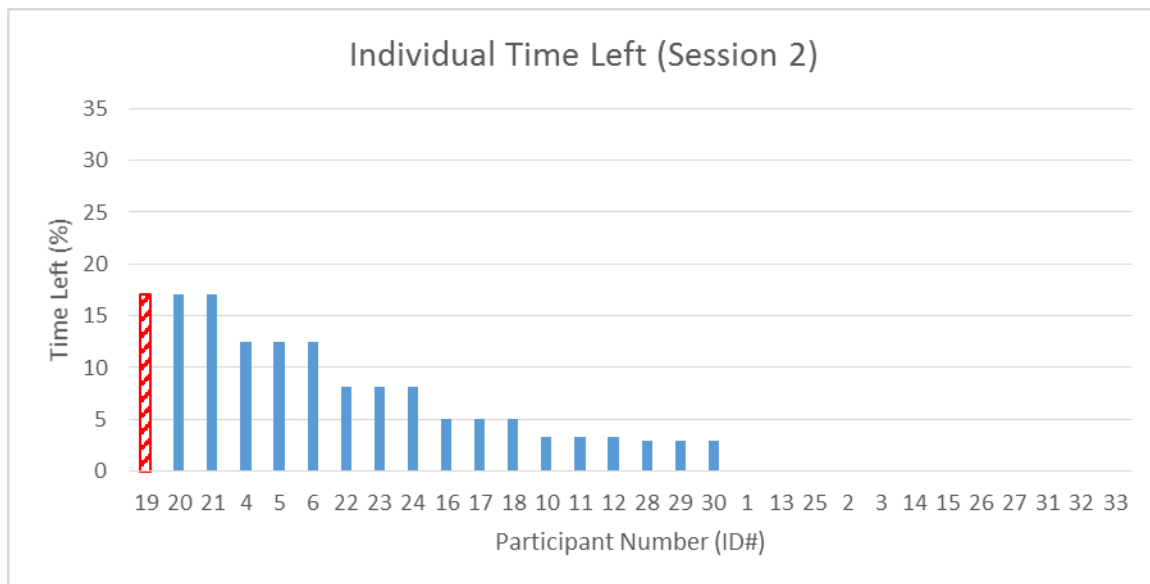
Figure 15. *Time Left for each team (Session 2)*.
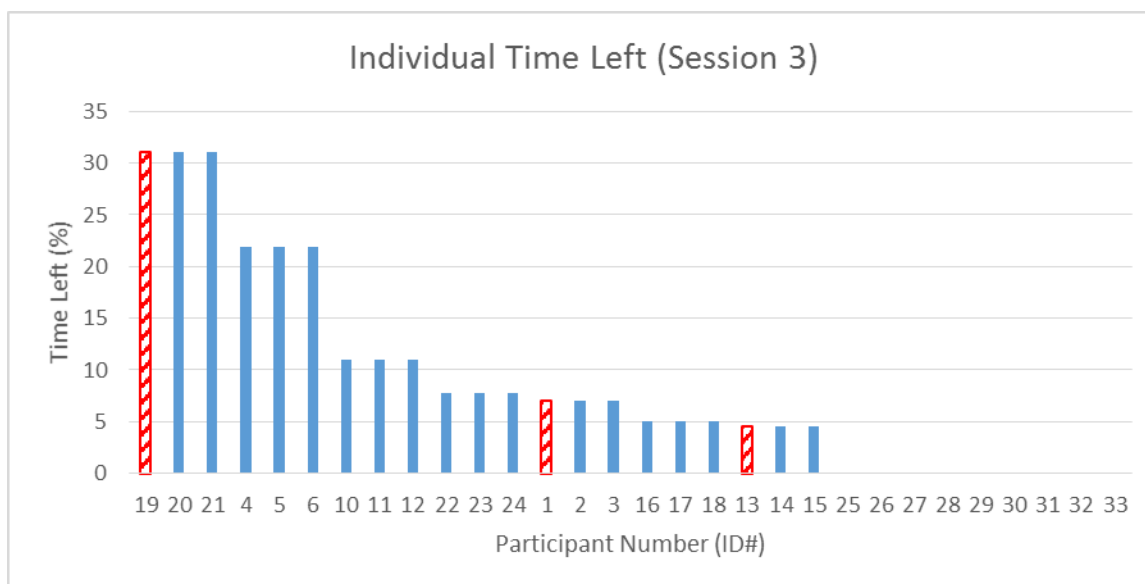
Figure 16*. Time Left for each team (Session 3).*
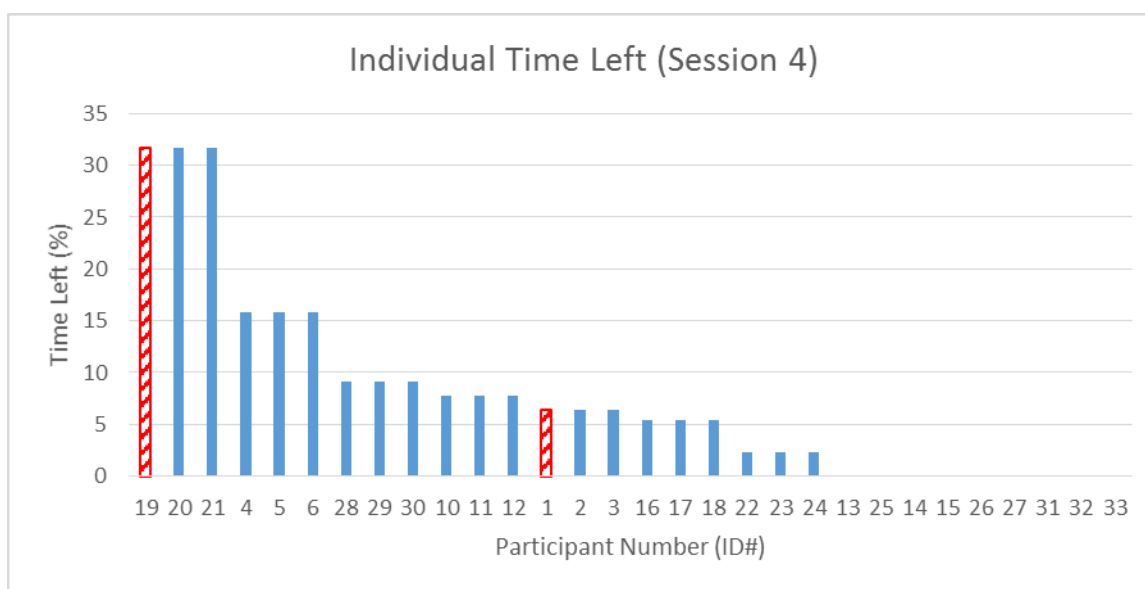
Figure 17*. Time Left for each team (Session 4).*

Individual Error by Session

The following four figures show the sum of individuals' errors in each session. The

figures were sorted from highest sum of errors committed to the lowest number of errors

committed, to illustrate the overall change in performance over sessions. The red striped bars were confederates.
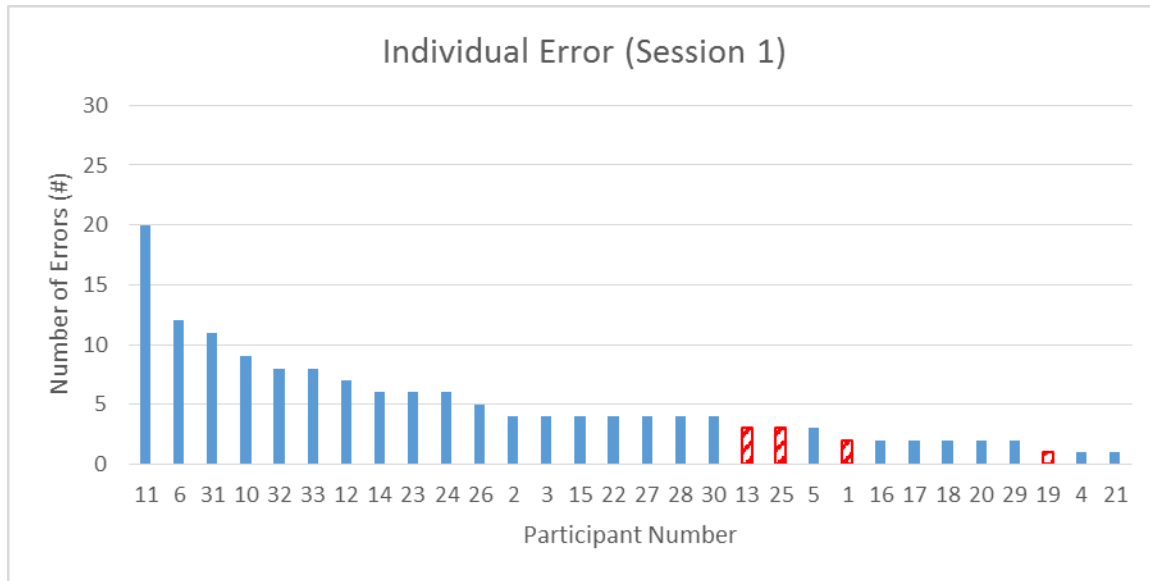
Figure 18. *Individual errors for each participant (Session 1).*
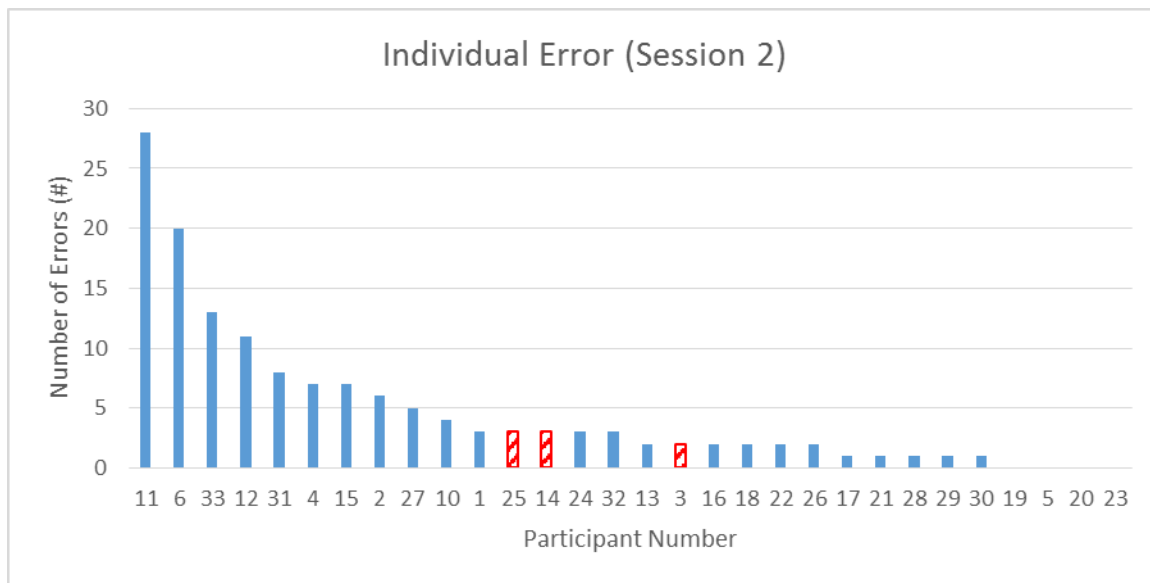
Figure 19. *Individual errors for each participant (Session 2).*
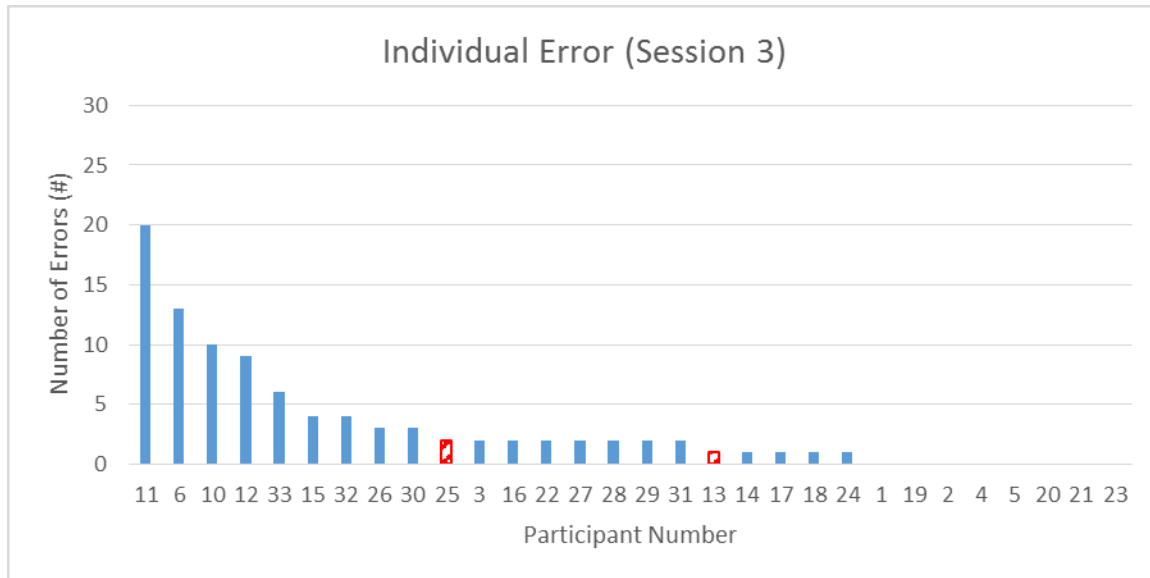
Figure 20. *Individual errors for each participant (Session 3).*
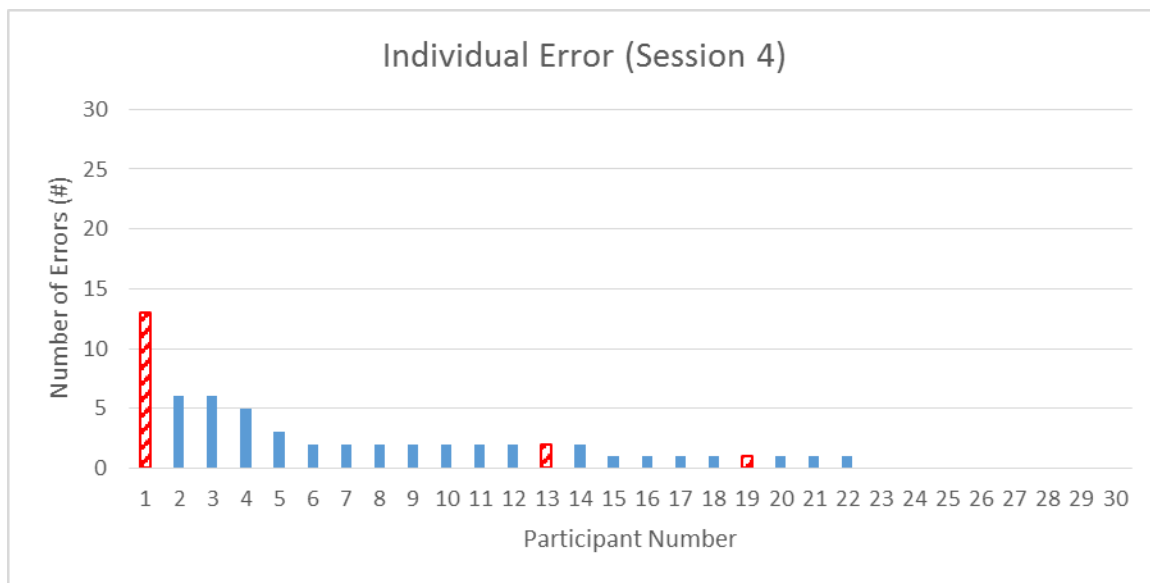
Figure 21. *Individual errors for each participant (Session 4).*

Team Items Collected by Session

The following four figures show teams' percentage of items collected in each session. The figures were sorted from highest performing teams to lowest performing, to illustrate the overall change in performance over sessions. The items collected percentage for teams are the

correct number of teams items collected over the total number of correct team items possible

(18). The red striped bars were teams that had one confederate members.
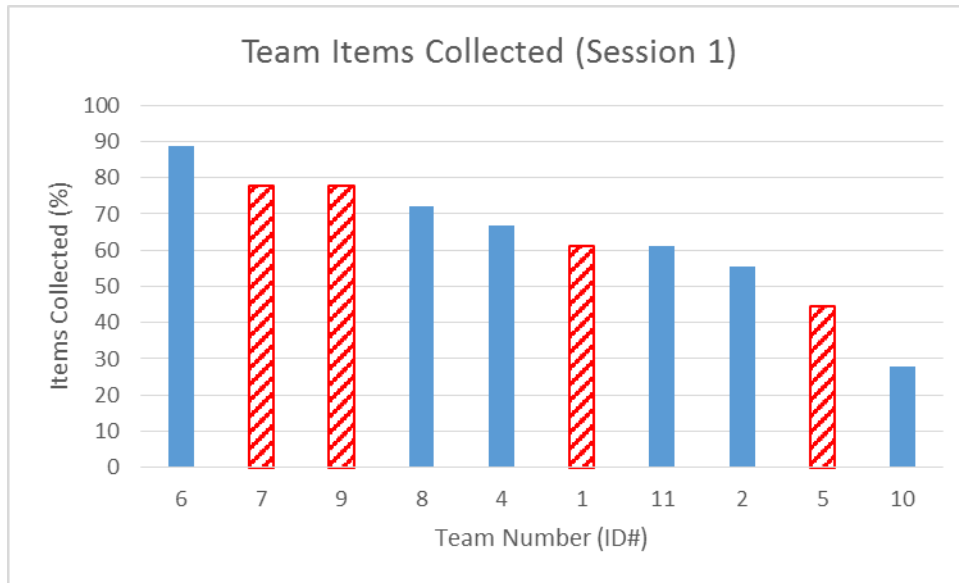
Figure 22. *Team items collected for each participant (Session 1)*
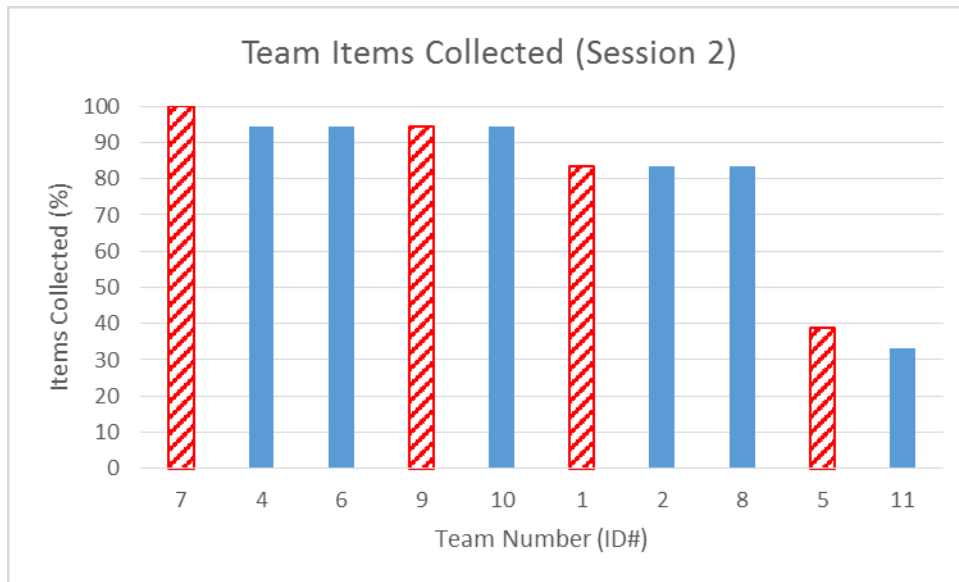
Figure 23. *Team items collected for each participant (Session 2)*
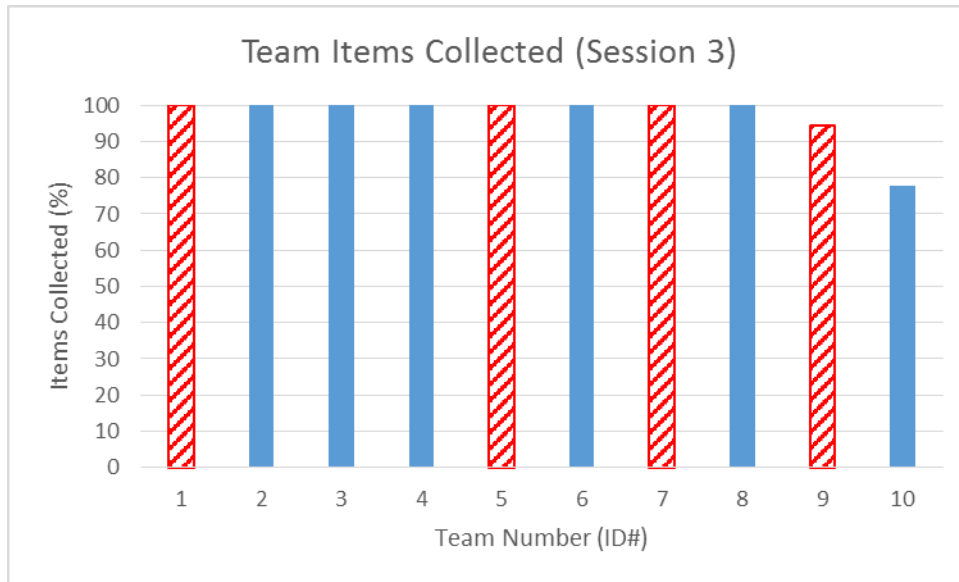
Figure 24. *Team items collected for each participant (Session 3)*
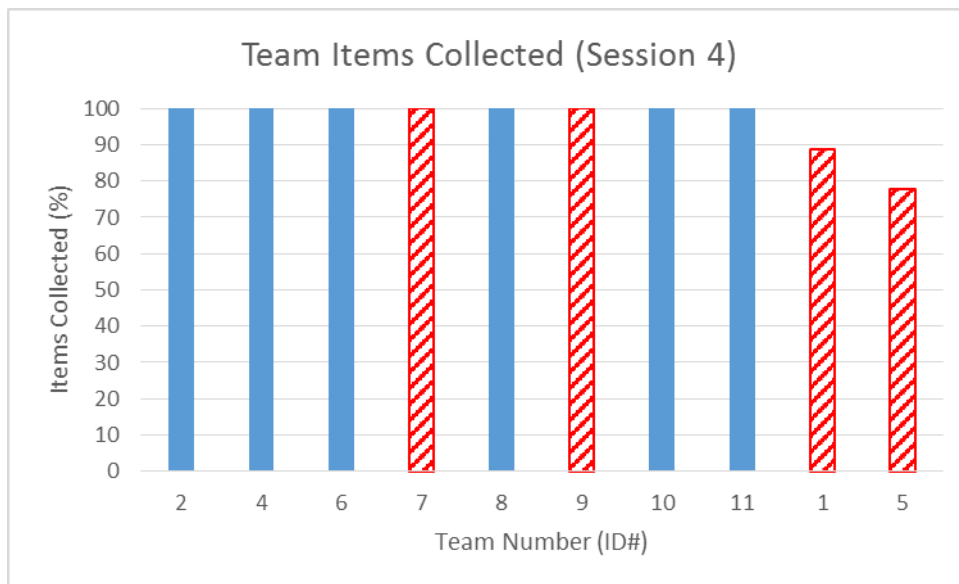
Figure 25. *Team items collected for each participant (Session 4)*

Team Time Left by Session

The following four figures show teams' percentage of time left in each session. The

figures were sorted from highest performing teams to lowest performing, to illustrate the overall

change in performance over sessions. The red striped bars were teams that had one confederate

members.
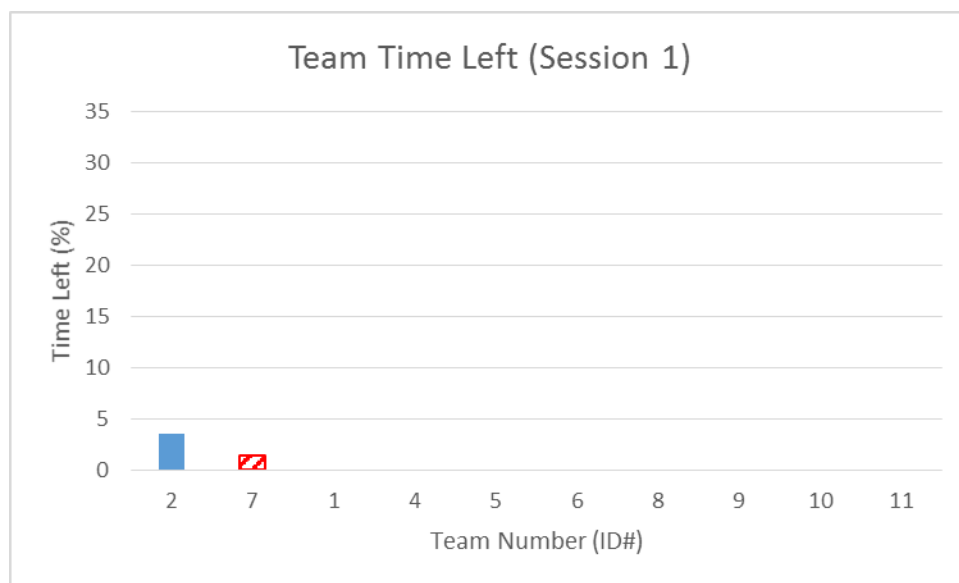
Figure 26. Time left for each team (Session 1).
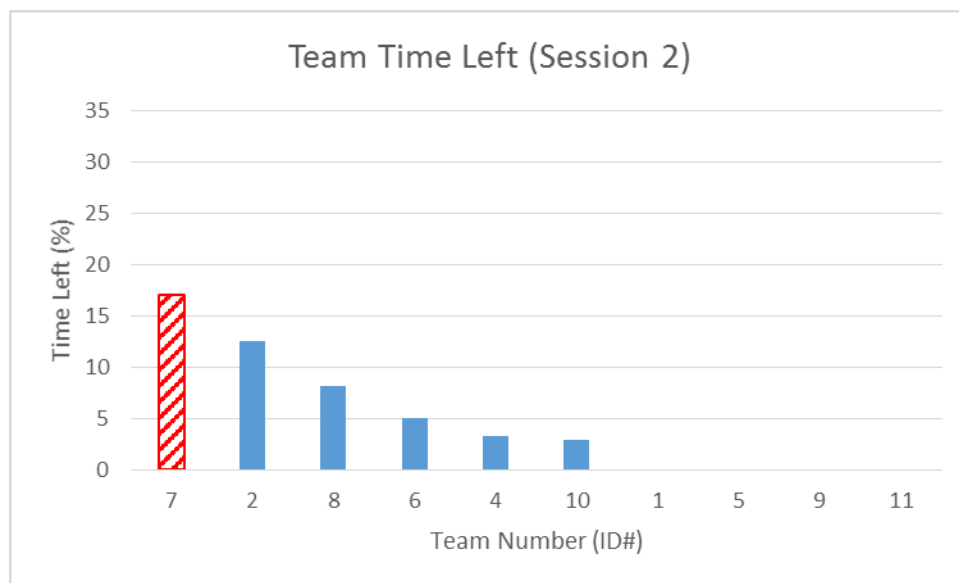
Figure 27. Time left for each team (Session 2).
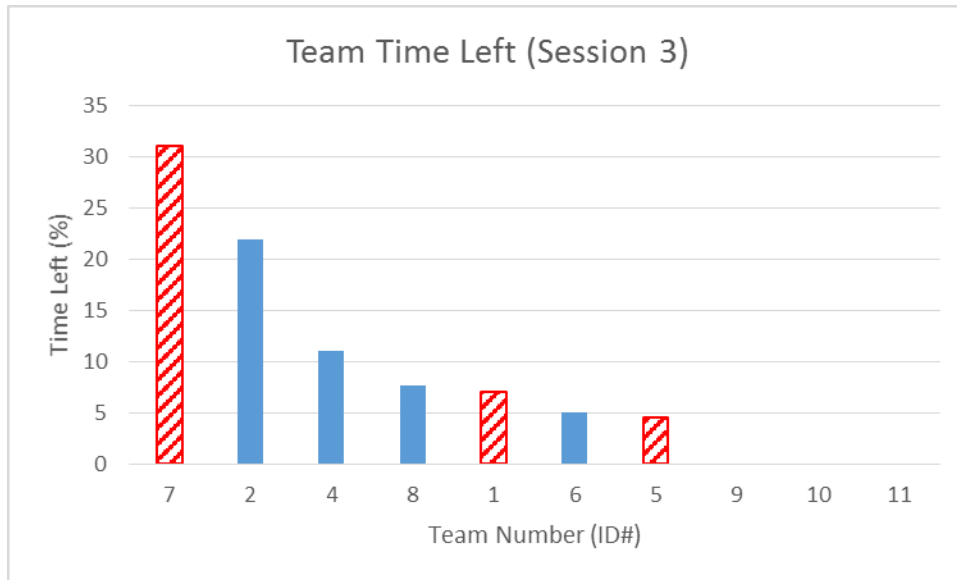
Figure 28. Time left for each team (Session 3).
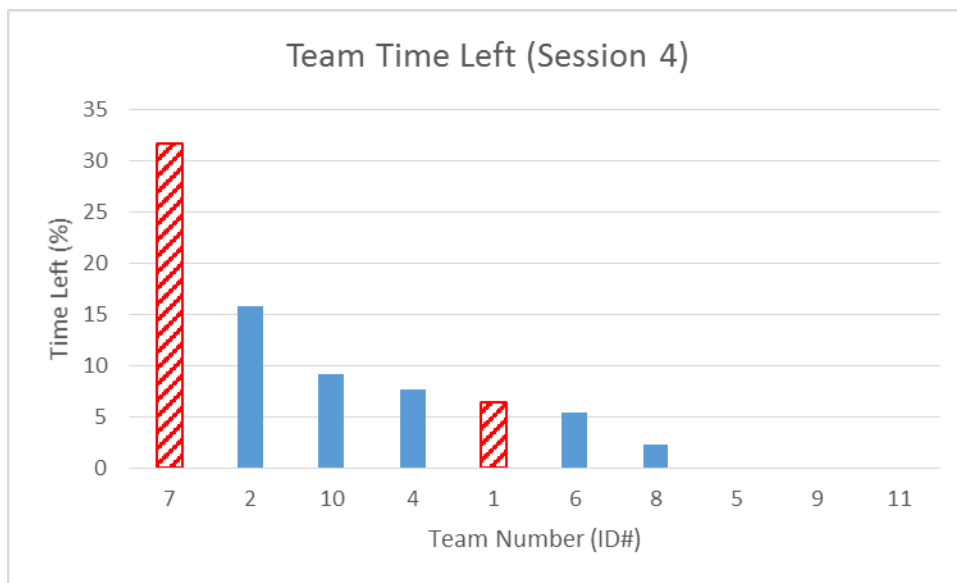
Figure 29. Time left for each team (Session 4).

Team Error by Session

The following four figures show the sum of teams' errors in each session. The figures were sorted from highest sum of errors to lowest sum of errors, to illustrate the overall change in performance over sessions. The red striped bars were teams that had one confederate members.
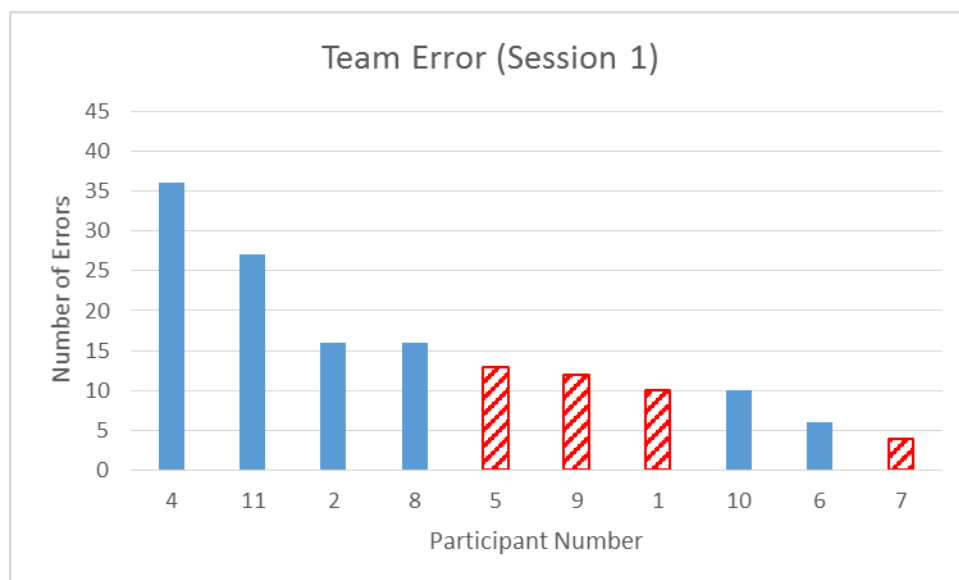
Figure 30. *Error for each team (Session 1).*
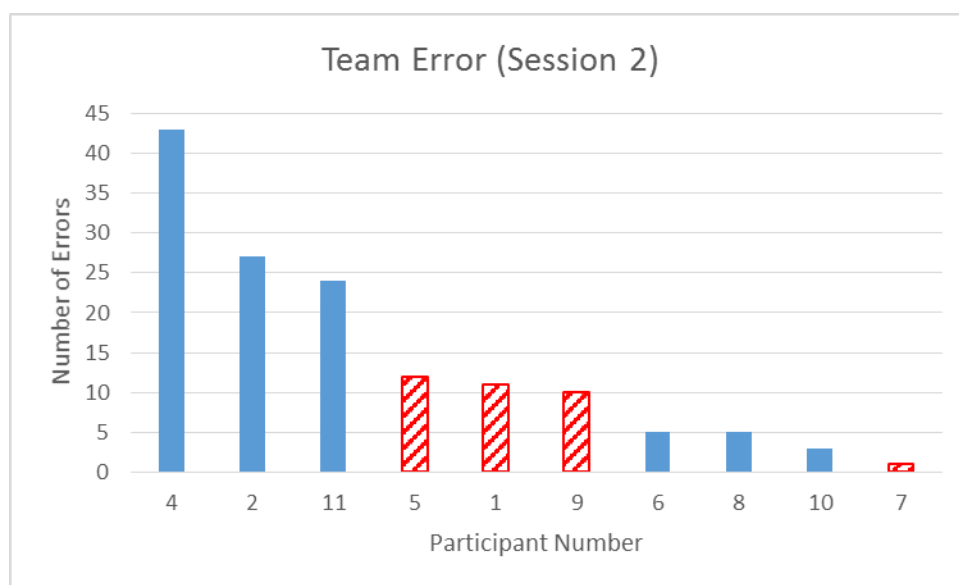
Figure 31. *Error for each team (Session 2).*
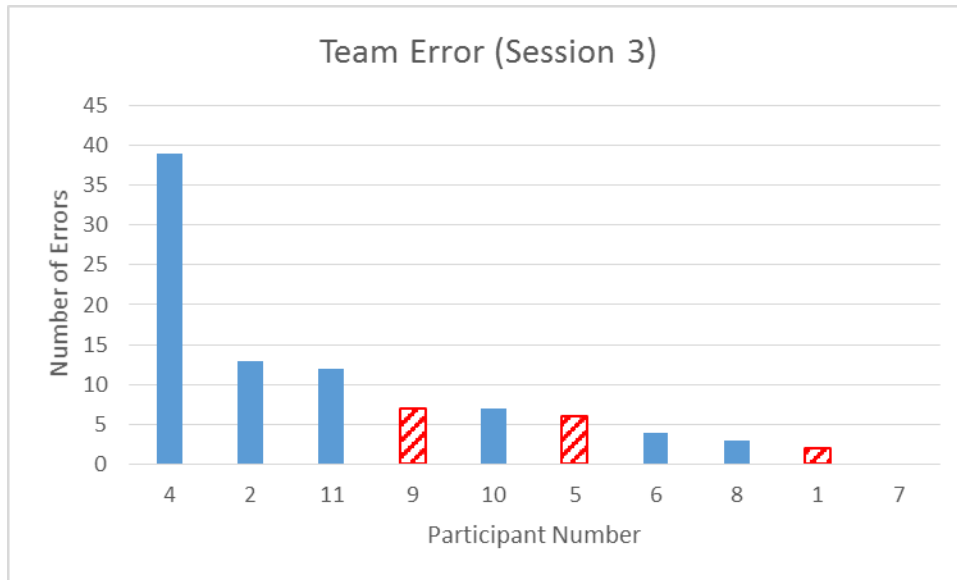
Figure 32*. Error for each team (Session 3).*
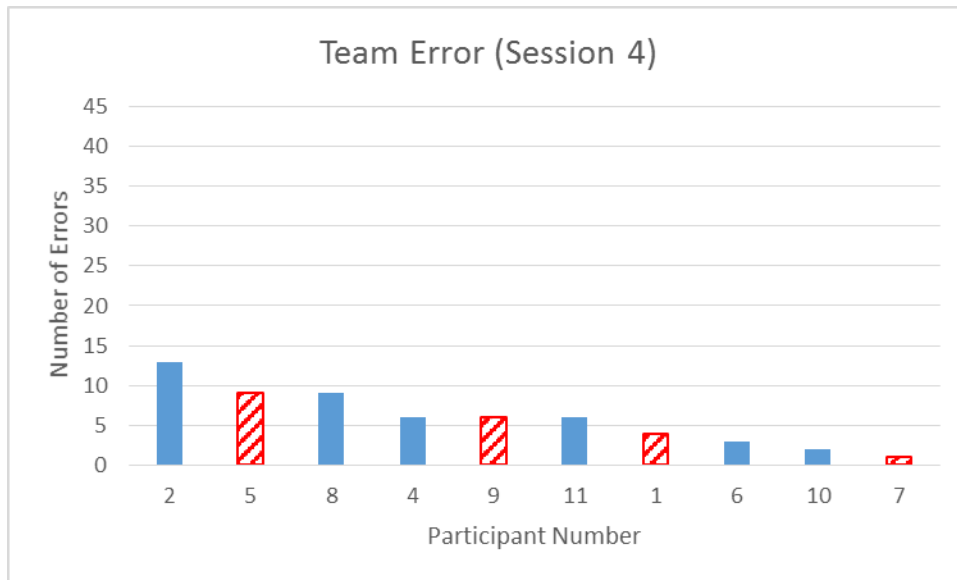
Figure 33*. Error for each team (Session 4).*

Impact of Feedback on Items Collected for Individual and Team

We report the items collected data obtained as described in Data Analysis Plan section in Chapter 3, and explore whether it was affected by the privacy of feedback (private vs. public) or the audience (team or individual).

Individual – Items Collected

It is important to note that the confederate data was not included in the following analysis. The following section describes the results of the individual items collected for participants. The sections are separated by session order. Each session describes the privacy of feedback, audience of feedback, and the interaction of privacy and audience of feedback with the statistical significance. The data in the following section address Research Questions 1 and 2.

Graphically, there appeared to be an interaction between the privacy and audience independent variables (Figure 34). A within-subjects ANOVA was used because each participant experienced all both levels of both independent variables. After running the ANOVA analysis, there was no statistical significance for the privacy of feedback ($F_{(1,96)} = 2.821$, $p = .096$, $\eta^2 = .008$), the audience of feedback ($F_{(1,96)} = 1.056$, $p = .306$, $\eta^2 = .006$), or their interaction ($F_{(1,96)} = .124$, $p = .725$, $\eta^2 = .004$).

More investigation was needed in order to better understand the data collected. Upon further investigation it was discovered that 70% of the participants strongly agree that their own individual performance improved over time. Analysis showed that there was a learning curve for the participants over time (Figure 35). This result led to the analysis of the individual items collected with respect to session order. After running the one-way ANOVA analysis, there was a statistical significance ($F_{(1,100)} = 6.583$, $p = .011$) among the session orders. Specifically, Session 1 and Session 2, Session 1 and Session 4, Session 2 and Session 3, and Session 2 and Session 4.

Figure 34. *Average individual items collected (all sessions). Error bars represent S.E.M. Differences are not statistically significant.*

Figure 35. *Average items collected of teams by session. Error bars represent S.E.M. Time left across sessions is significantly different.*

Due to learning curve differences by session, the impact of feedback audience and privacy on items collected errors was analyzed by session.

<u>Session 1</u>. The privacy of feedback ($F_{(1,18)} = 1.104$, $p = .307$, $\eta^2 = .001$), the audience of feedback ($F_{(1,18)} = .275$, $p = .607$, $\eta^2 < .001$), and the interaction ($F_{(1,18)} = .005$, $p = .946$, $\eta^2 = .028$) of the privacy and audience of feedback had no significant difference (within-subjects ANOVA). Overall, participants in the Private Group condition produced the highest average items collected, $61.11\% \pm 7.00\%$. The participants in the Public Group condition produced the lowest average items collected, $50.00\% \pm 7.41\%$. Participants in the Private Direct and the Public Direct produced average items collected of $57.69\% \pm 9.30\%$ and $56.25\% \pm 0.00\%$, respectively (Table 9 and Figure 36).

Table 9. Average individual items collected (Session 1, *n* = 26)

	Private	Public
Direct	$57.69\% \pm 9.30\%$	$56.25\% \pm 0.00\%$
Group	$61.11\% \pm 7.00\%$	$50.00\% \pm 7.41\%$

Figure 36. *Average individual items collected (Session 1). Error bars represent S.E.M. Differences were not significant.*

Session 2. The audience of feedback ($F_{(1,18)} = .014$, $p = .905$, $\eta^2 = .002$) and the interaction ($F_{(1,18)} = .341$, $p = .566$, $\eta^2 = .035$) of the privacy and audience of feedback had no significant difference (within-within-subjects ANOVA). The privacy of feedback ($F_{(1,18)} = 10.922$, $p = .003$, $\eta^2 = .001$) had a significant difference. Overall, participants in the Private Group condition produced the highest average items collected, 70.37% ± 13.73%. The participants in the Public Direct condition produced the lowest average items collected, 58.33% ± 17.22%. Participants in the Private Direct and the Public Group produced average items collected of 66.03% ± 9.53% and 64.29% ± 16.67%, respectively (Table 10 and Figure 37).

Table 10. Average individual items collected (Session 2, *n* = 26)

	Private	Public
Direct	66.03% ± 9.53%	58.33% ± 17.22%
Group	70.37% ± 13.73%	64.29% ± 16.67%

Figure 37. *Average individual items collected (Session 2). Error bars represent S.E.M. Differences was significant for some variables but not others.*

Session 3. The privacy of feedback ($F_{(1,18)} = .001$, $p = .979$, $\eta^2 = .040$), the audience of

feedback ($F_{(1,18)} = .002$, $p = .966$, $\eta^2 = .033$), and the interaction ($F_{(1,18)} = .241$, $p = .629$, $\eta^2 = $

.052) of the privacy and audience of feedback had no significant difference (within-subjects

ANOVA). Overall, participants in the Private Direct condition produced the highest average

items collected, 100.00% ± 0.00. The participants in the Private Group condition produced the

lowest average items collected, 86.67% ± 9.72%. Participants in the Public Direct and the Public

Group produced average items collected of 87.50% ± 4.17% and 87.50% ± 6.10%, respectively

(Table 11 and Figure 38).

Table 11. Average individual items collected (Session 3, $n = 26$)

	Private	Public
Direct	100.00% ± 0.00	87.50% ± 4.17%
Group	86.67% ± 9.72%	87.50% ± 6.10%
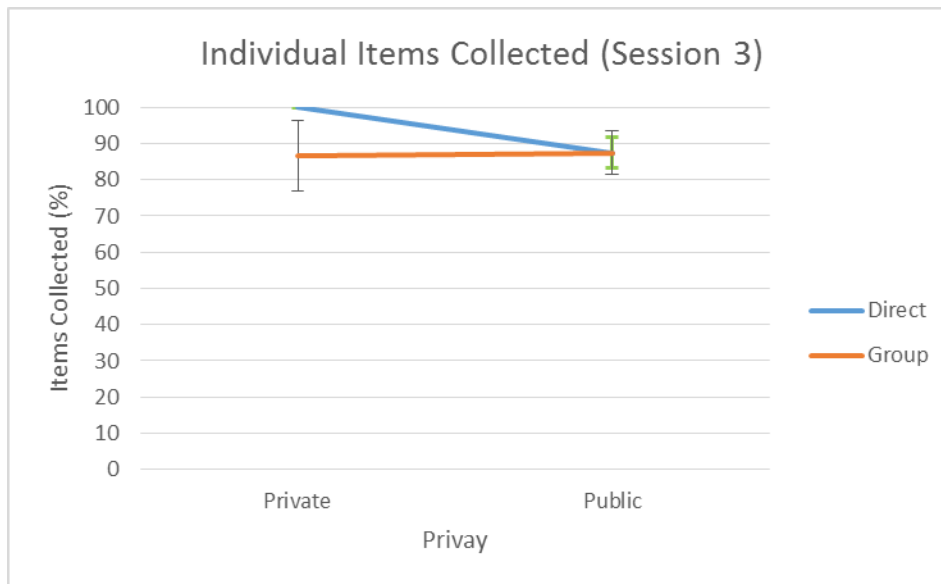
Figure 38. *Average individual items collected (Session 3). Error bars represent S.E.M.*
Differences were not significant.

Session 4. The privacy of feedback ($F_{(1,18)} = 1.635$, $p = .217$, $\eta^2 = .005$), the audience of

feedback ($F_{(1,18)} = 2.276$, $p = .149$, $\eta^2 = .015$), and the interaction ($F_{(1,18)} = .263$, $p = .614$, $\eta^2 = $

.105) of the privacy and audience of feedback had no significant difference (within-subjects

ANOVA). Overall, participants in the Public Direct condition produced the highest average

items collected, 100.00% ± 2.08%. The participants in the Private Group condition produced the

lowest average items collected, 91.67% ± 0.00%. Participants in the Private Direct and the Public

Group condition produced average items collected of 96.15% ± 3.73% and 97.92% ± 4.76%,

respectively (Table 12 and Figure 39).

Table 12. Average individual items collected (Session 4, $n = 26$)

	Private	Public
Direct	96.15% ± 3.73%	100.00% ± 2.08%
Group	91.67% ± 0.00%	97.92% ± 4.76%

Figure 39. *Average individual items collected (Session 4). Error bars represent S.E.M.*
Differences were not significant.

Team – Items Collected

It is important to note that the confederate data was included in the following analysis. As well, small parts of the data analysis is missing (i.e., marked "N/A") because of the small team sample size. Graphically, there appeared to be a main effect for both the privacy and audience independent variables (Figure 40). After running this ANOVA, there was no statistical significance for the privacy of feedback ($F_{(1,32)} = .006$, $p = .936$, $\eta^2 = .042$), the audience of feedback ($F_{(1,32)} = .295$, $p = .590$, $\eta^2 = .097$), or their interaction ($F_{(1,32)} = 3.775$, $p = .060$, $\eta^2 = .002$). It is important to note that the interaction of the privacy and audience of feedback was trending towards significance.

More investigation was needed in order to better understand the data collected. Upon further investigation it was discovered that 77% of the participants strongly agree that the team's performance improved over time. Further analysis showed that there appeared to be a learning curve for the teams over time (Figure 41). After running the one-way ANOVA analysis, there was no statistical significance ($F_{(1,36)} = 2.791$, $p = .103$) among the session orders.
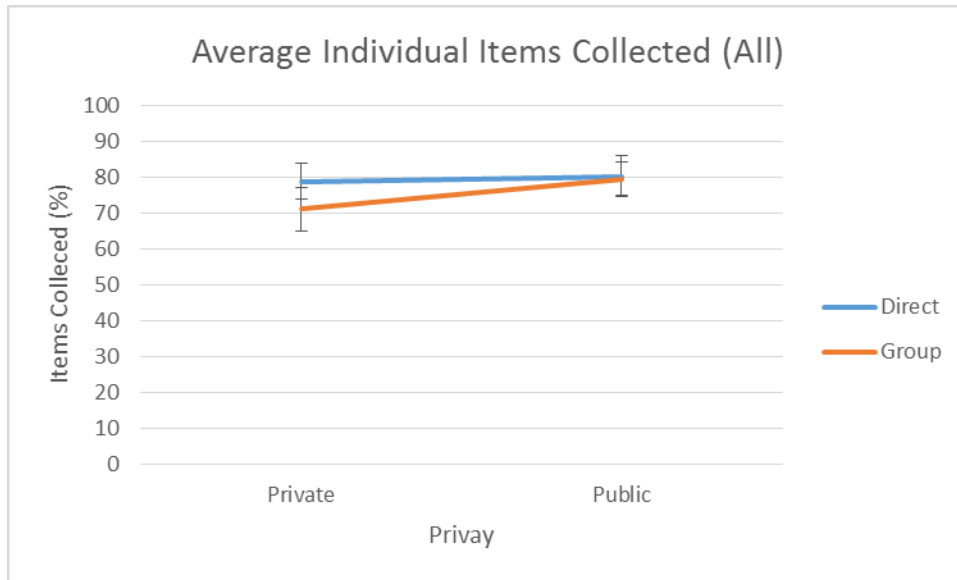

Figure 40. *Average team items collected (all sessions). Error bars represent S.E.M. Differences are not statistically significant.*
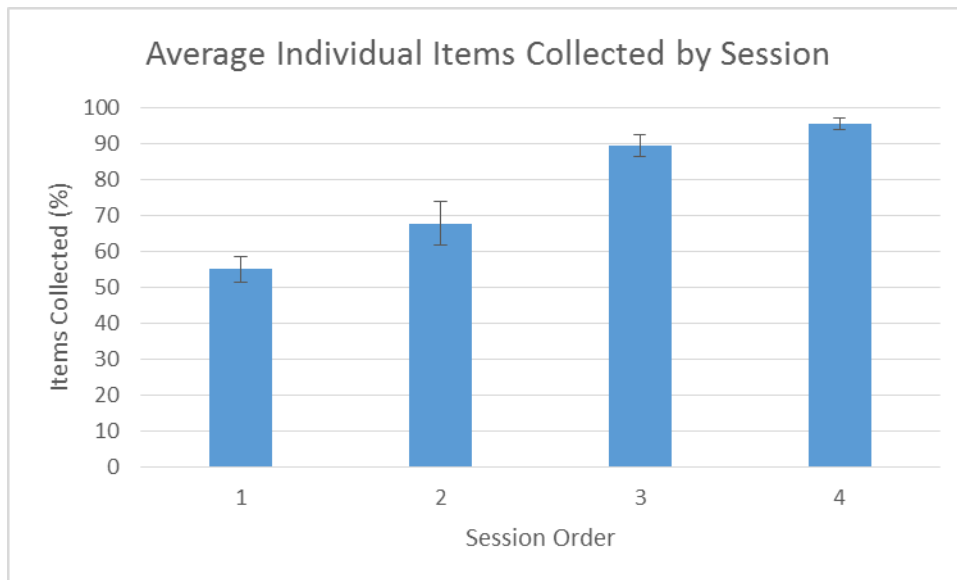
Figure 41*. Average items collected of teams by session. Error bars represent S.E.M. Items collect across sessions was not significantly different.*

Due to learning curve differences by session, the impact of feedback audience and privacy on items collected was analyzed by session.

Session 1. The privacy of feedback ($F_{(1,3)} = .075$, $p = .802$, $\eta^2 = .153$) and the audience of feedback ($F_{(1,3)} = 1.701$, $p = .283$, $\eta^2 = .117$) had no significant difference (within-subjects ANOVA). Overall, teams in the Public Direct condition produced the highest average items collected, 88.89% ± (N/A). The S.E. for the Public Direct condition was not possible to calculate because of the low sample size. The teams in the Private Group condition produced the lowest average items collected, 57.41% ± 15.16%. Teams in the Private Direct and the Public Group produced average items collected of 61.11% ± 9.62% and 62.96% ± 4.90%, respectively (Table 13 and Figure 42).

Table 13. Average team items collected (Session 1, *n* = 10)

	Private	Public
Direct	61.11% ± 9.62%	88.89% ± (N/A)
Group	57.41% ± 15.16%	62.96% ± 4.90%

Figure 42. *Average Team items collected (Session 1). Error bars represent S.E.M. Differences were not significant.*

Session 2. The privacy of feedback ($F_{(1,3)} = .034$, $p = .865$, $\eta^2 = .037$) and the audience of feedback ($F_{(1,3)} = .190$, $p = .692$, $\eta^2 = .611$) had no significant difference (within-subjects ANOVA). Overall, teams in the Public Direct condition produced the highest average items collected, 96.30% ± 1.85%. The teams in the Public Group condition produced the lowest average items collected, 38.89% ± (N/A). The S.E. for the Public Direct condition was not possible to calculate because of the low sample size. Teams in the Private Direct and the Private Group produced average items collected of 90.74% ± 3.70% and 66.67% ± 16.67%, respectively (Table 14 and Figure 43).

Table 14. Average team items collected (Session 2, *n* = 10)

	Private	Public
Direct	90.74% ± 3.70%	96.30% ± 1.85%
Group	66.67% ± 16.67%	38.89% ± (N/A)
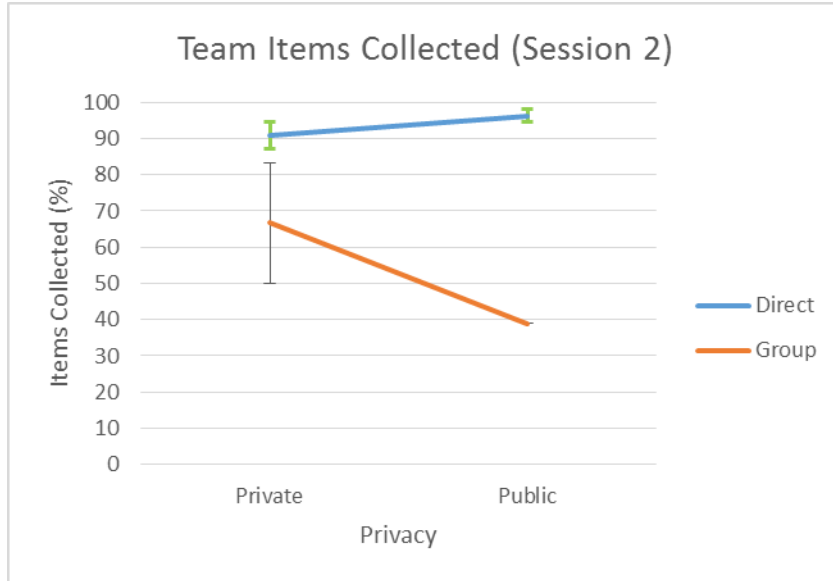
Figure 43. *Average Team items collected (Session 2). Error bars represent S.E.M. Differences were not significant.*

Session 3. The privacy of feedback ($F_{(1,2)} = 4.248$, $p = .175$, $\eta^2 = .161$), the audience of

feedback ($F_{(1,2)} = .081$, $p = .803$, $\eta^2 = .223$), and the interaction ($F_{(1,2)} = 1.638$, $p = .329$, $\eta^2 =$

.161) of the privacy and audience of feedback had no significant difference (within-subjects

ANOVA). Overall, teams in the Private Direct and the Public Direct condition produced the

highest average items collected, $100.00\% \pm 0.00\%$. The teams in the Private Group condition

produced the lowest average items collected, $88.89\% \pm 11.11\%$. Teams in the Public Group

produced average items collected of $98.15\% \pm 1.85\%$ (Table 15 and Figure 44).

Table 15. Average team items collected (Session 3, *n* = 10)

|	Private	Public
Direct	$100.00\% \pm 0.00\%$	$100.00\% \pm 0.00\%$
Group	$88.89\% \pm 11.11\%$	$98.15\% \pm 1.85\%$

Figure 44*. Average Team items collected (Session 3). Error bars represent S.E.M. Differences were not significant.*

  <u>Session 4</u>. The privacy of feedback ($F_{(1,2)}$ = .180, *p* = .713, η^2 = .152), the audience of

feedback ($F_{(1,2)}$ = .003, *p* = .964, η^2 = .029), and the interaction ($F_{(1,2)}$ = .000, *p* = .987, η^2 =

.019) of the privacy and audience of feedback had no significant difference (within-subjects

ANOVA). Overall, teams in the Private Direct and the Private Group conditions produced the

highest average items collected, 100.00% ± 0.00%. The teams in the Public Direct condition

produced the lowest average items collected, 92.59% ± 7.41%. Teams in the Public Group

produced average items collected of 96.30% ± 3.70% (Table 16 and Figure 45).

Table 16. Average team items collected (Session 4, *n* = 10)

	Private	Public
Direct	100.00% ± 0.00%	92.59% ± 7.41%
Group	100.00% ± 0.00%	96.30% ± 3.70%

Figure 45. *Average Team items collected (Session 4). Error bars represent S.E.M. Differences were not significant.*

Impact of Feedback on Errors for Individual and Team

Second we report the error data obtained as described in Data Analysis Plan section in Chapter 3, and explore whether it was affected by the privacy of feedback (private vs. public) or the audience (team or individual).

Individual - Errors

The following section describes the results of the individual errors for participants. The sections are separated by session order. Each session describes the privacy of feedback, audience of feedback, and the interaction of privacy and audience of feedback with the statistical significance. The data in the following section address Research Questions 1 and 2.

It is important to note that the confederate data was not included in the following analysis. Graphically, there appeared to be an interaction between the privacy and audience independent variables (Figure 46). A within-subjects ANOVA was used because each participant

experienced all levels of both independent variables. After running the ANOVA analysis, there was no statistical significance for the privacy of feedback ($F_{(1,96)} = .093$, $p = .761$, $\eta^2 = .001$), the audience of feedback ($F_{(1,96)} = .056$, $p = .814$, $\eta^2 = .009$), or their interaction ($F_{(1,96)} = .000$, $p = .987$, $\eta^2 < .001$).

More investigation was needed in order to better understand the data collected. Upon further investigation it was discovered that 70% of the participants strongly agree that their own individual performance improved over time. Graphically it appeared there was a learning curve for the participants over time (Figure 47). This result led to the analysis of the individual error with respect to session order. After running the one-way ANOVA analysis, there was no statistical significance ($F_{(1,100)} = 1.456$, $p = .23$) among the session orders.
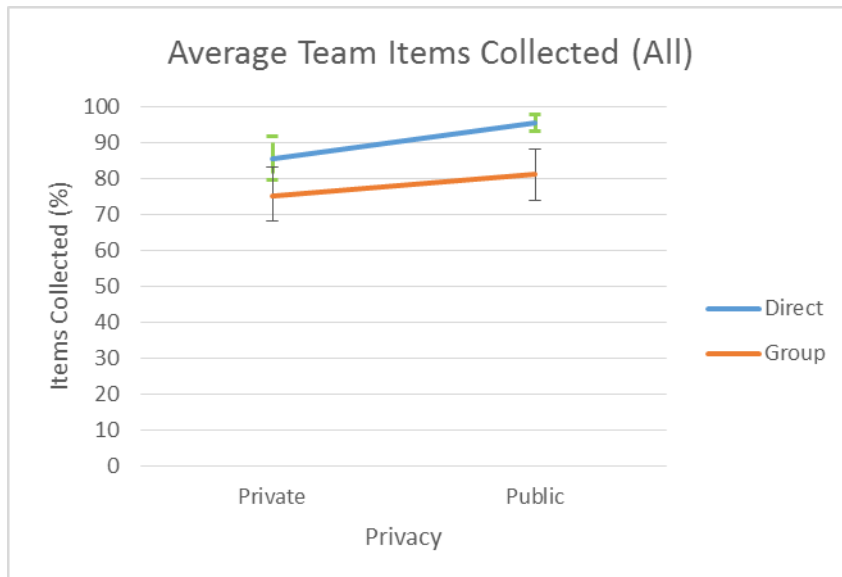

Figure 46. *Average individual error (all sessions). Error bars represent S.E.M. Differences are not statistically significant.*
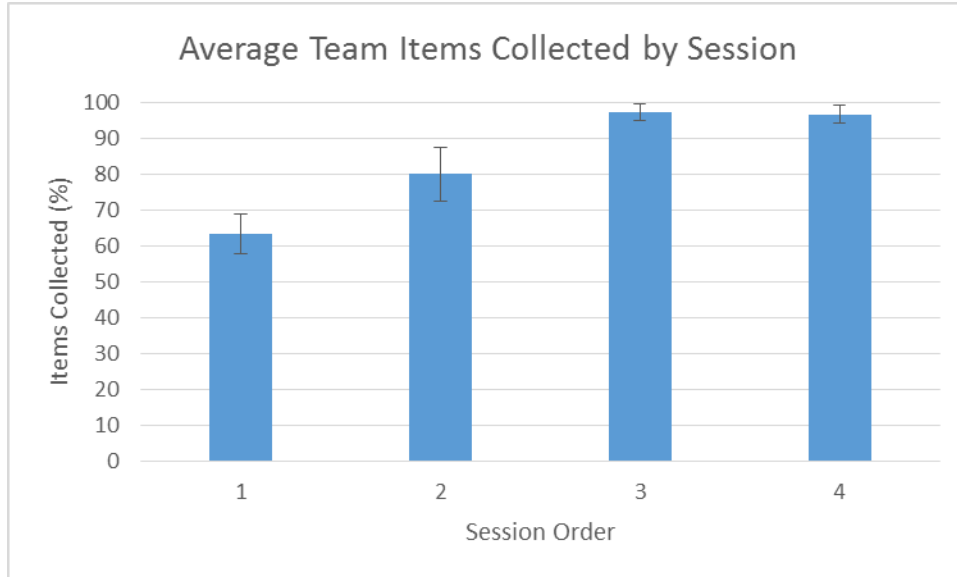
Figure 47. *Average error of teams by session. Error bars represent S.E.M. Error across sessions was not significantly different.*

Due to learning curve differences by session, the impact of feedback audience and privacy on errors was analyzed by session.

Session 1. The privacy of feedback ($F_{(1,18)} = 1.794$, $p = .197$, $\eta^2 = .004$) and the interaction ($F_{(1,18)} = .048$, $p = .830$, $\eta^2 = .029$) of the privacy and audience of feedback had no significant difference (within-within-subjects ANOVA). The audience of feedback ($F_{(1,18)} = 6.910$, $p = .017$, $\eta^2 = .104$) did have a significant difference on the errors committed during each session. Overall, participants in the Public Group condition produced the highest average errors, 6.56 ± 1.20. The participants in the Public Direct condition produced the lowest average errors, 2.00 ± 0.00. Participants in the Private Direct and the Private Group produced average errors of 4.50 ± 0.34 and 6.13 ± 2.20, respectively (Table 17 and Figure 48).

Table 17. Average individual errors (Session 1, *n* = 26)

	Private	Public
Direct	4.50 ± 0.34	2.00 ± 0.00
Group	6.13 ± 2.20	6.56 ± 1.20
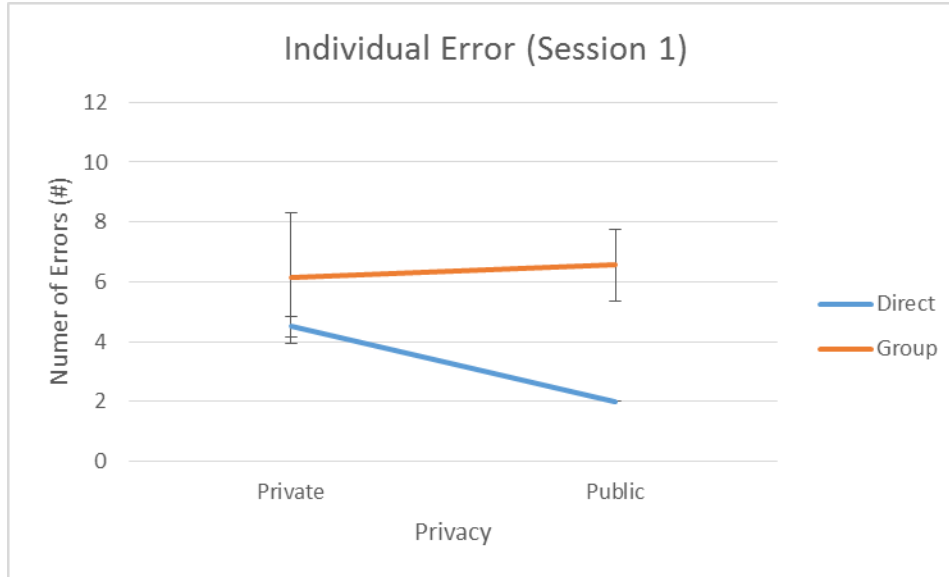
Figure 48. *Average Individual Errors (Session 1). Error bars represent S.E.M. Differences were significant for some variables but not others.*

Session 2. The privacy of feedback ($F_{(1,18)} = 3.728$, $p = .069$, $\eta^2 = .032$), the audience of feedback ($F_{(1,18)} = 1.269$, $p = .274$, $\eta^2 < .001$), and the interaction ($F_{(1,18)} = .259$, $p = .617$, $\eta^2 = .009$) of the privacy and audience of feedback had no significant difference (within-subjects ANOVA). It is important to note that the privacy of feedback had a trend towards significance. Overall, participants in the Public Direct condition produced the highest average errors, 7.29 ± 3.71. The participants in the Private Direct condition produced the lowest average errors, 3.89 ± 2.12. Participants in the Private Group and the Public Group produced average errors of 4.63 ± 1.49 and 5.00 ± 2.00, respectively (Table 18 and Figure 49).

Table 18. Average individual errors (Session 2, *n* = 26)

|	Private	Public
Direct	3.89 ± 2.12	7.29 ± 3.71
Group	4.63 ± 1.49	5.00 ± 2.00

Figure 49. *Average Individual Errors (Session 2). Error bars represent S.E.M. Differences were not significant.*

Session 3. The privacy of feedback ($F_{(1,18)}$ = .841, p = .371, η^2 = .007) and the interaction ($F_{(1,18)}$ = .188, p = .669, η^2 = .034) of the privacy and audience of feedback had no significant difference (within-subjects ANOVA). The audience of feedback ($F_{(1,18)}$ = 10.553, p = .004, η^2 = .126) had a statistical significance. Overall, participants in the Public Group condition produced the highest average errors, 6.00 ± 2.36. The participants in the Public Direct condition produced the lowest average errors, 1.50 ± 0.38. Participants in the Private Direct and the Private Group produced average errors of 2.40 ± 1.17 and 3.60 ± 2.46, respectively (Table 19 and Figure 50).

Table 19. Average individual errors (Session 3, *n* = 26)

	Private	Public
Direct	2.40 ± 1.17	1.50 ± 0.38
Group	3.60 ± 2.46	6.00 ± 2.36
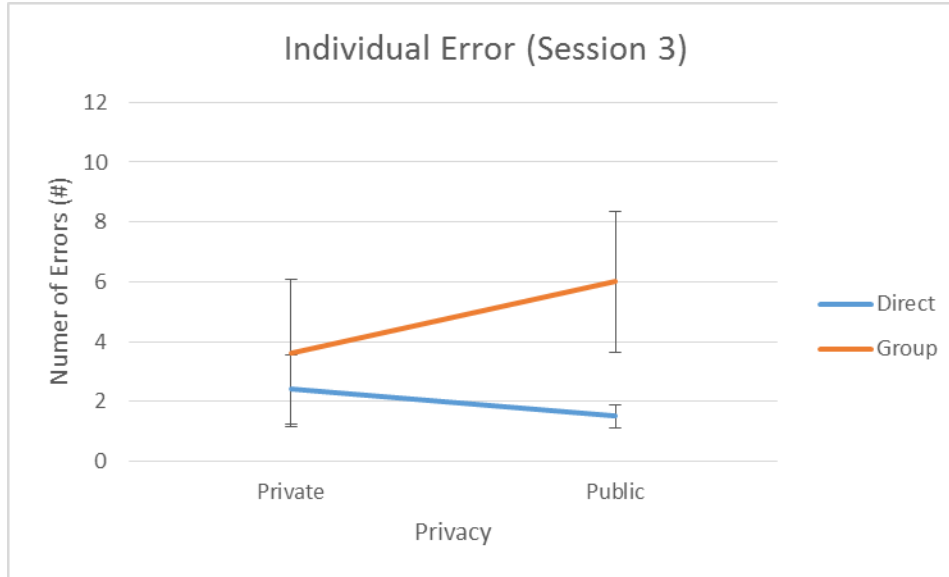
Figure 50. *Average Individual Errors (Session 3). Error bars represent S.E.M. Differences were significant for some variables but not others.*

Session 4. The privacy of feedback ($F_{(1,18)}$ = .819, p = .377, η^2 < .000), the audience of

feedback ($F_{(1,18)}$ = .818, p = .378, η^2 = .112), and the interaction ($F_{(1,18)}$ = .023, p = .881, η^2 =

.014) of the privacy and audience of feedback had no significant difference (within-subjects

ANOVA). Overall, participants in the Public Direct condition produced the highest average

errors, 3.25 ± 1.50. The participants in the Public Group condition produced the lowest average

errors, 0.86 ± 0.40. Participants in the Private Direct and the Private Group produced average

errors of 2.50 ± 1.15 and 1.40 ± 0.24, respectively (Table 20 and Figure 51).

Table 20. Average individual errors (Session 4, *n* = 26)

	Private	Public
Direct	2.50 ± 1.15	3.25 ± 1.50
Group	1.40 ± 0.24	0.86 ± 0.40
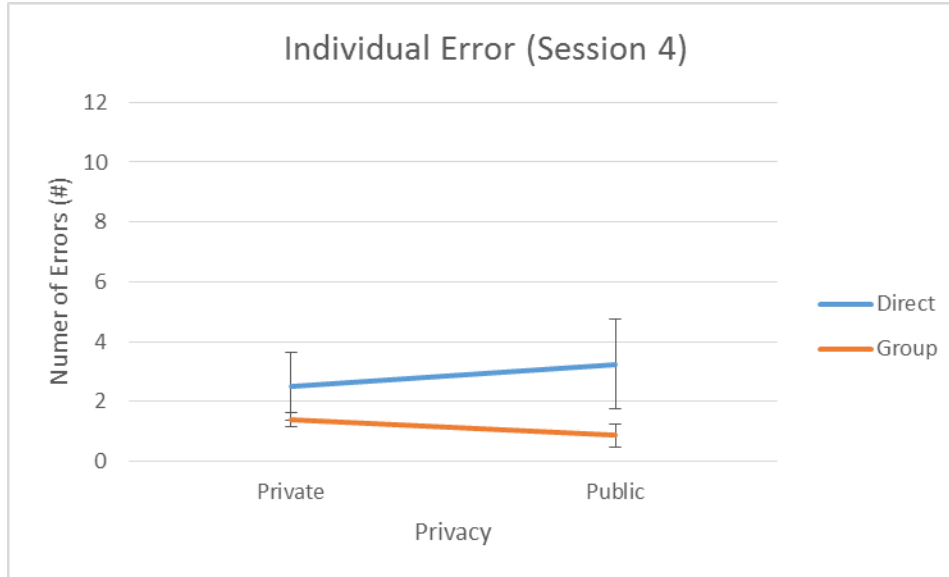
Figure 51. *Average Individual Errors (Session 4). Error bars represent S.E.M. Differences were not significant.*

Team – Errors

It is important to note that the confederate data was included in the following analysis. As well, some parts of the data analysis are missing (i.e., marked "N/A") because of the small team sample size. Graphically, there appeared to be a main effect for both the privacy and audience independent variables (Figure 52). After running this ANOVA, there was no statistical significance for the privacy of feedback ($F_{(1,32)} = .003$, $p = .957$, $\eta^2 = .001$), the audience of feedback ($F_{(1,32)} = .051$, $p = .823$, $\eta^2 = .014$), or their interaction ($F_{(1,32)} = .001$, $p = .975$, $\eta^2 < .001$).

More investigation was needed in order to better understand the data collected. Upon further investigation it was discovered that 77% of the participants strongly agree that the team's performance improved over time. Graphically it appeared there was a learning curve for the participants over time (Figure 53). After running the one-way ANOVA analysis, there was no statistical significance ($F_{(1,36)} = .863$, $p = .359$) among the session orders.
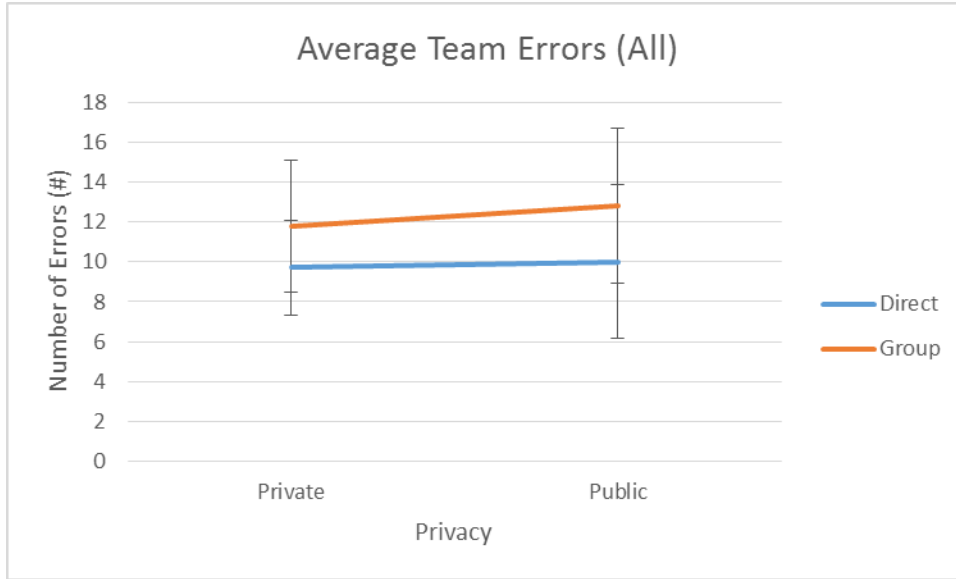
Figure 52. *Average team error (all sessions). Error bars represent S.E.M. Differences are not statistically significant.*
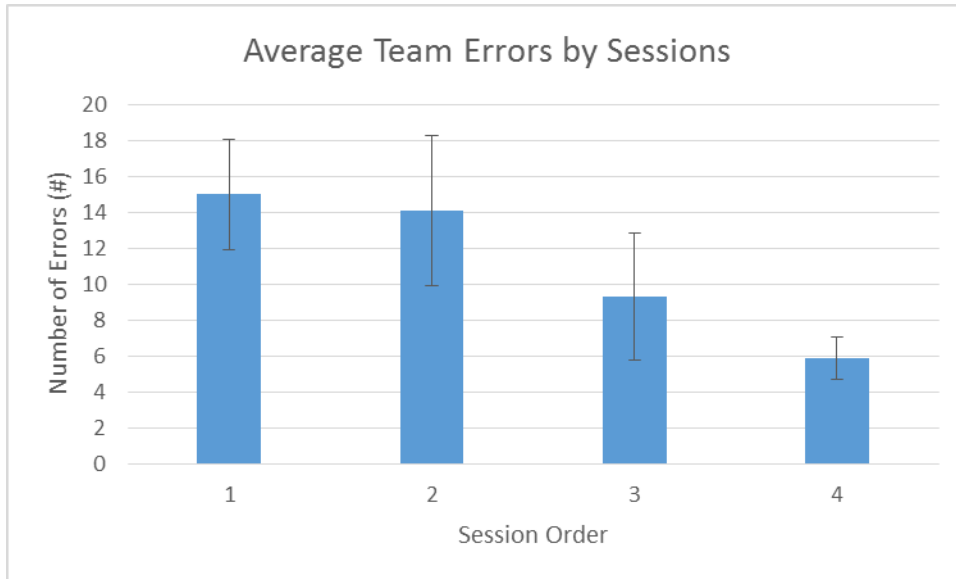
Figure 53. *Average error of teams by session. Error bars represent S.E.M. Differences were not statistically significant.*

Due to learning curve differences by session, the impact of feedback audience and privacy on errors was analyzed by session.

Session 1. The privacy of feedback ($F_{(1,3)} = .288$, $p = .629$, $\eta^2 < .001$) and the audience of feedback ($F_{(1,3)} = 3.465$, $p = .160$, $\eta^2 = .174$) had no significant difference (within-subjects ANOVA). Overall, teams in the Public Group condition produced the highest average errors, 19.67 ± 3.67. The teams in the Public Direct condition produced the lowest average errors, $6.00 \pm$ (N/A). The S.E. for the Public Direct condition was not possible to calculate because of the low sample size. Teams in the Private Direct and the Private Group produced average errors of 11.67 ± 0.88 and 16.67 ± 9.82, respectively (Table 21 and Figure 54).

Table 21. Average team errors (Session 1, $n = 10$)

	Private	Public
Direct	11.67 ± 0.88	$6.00 \pm$ (N/A)
Group	16.67 ± 9.82	19.67 ± 3.67
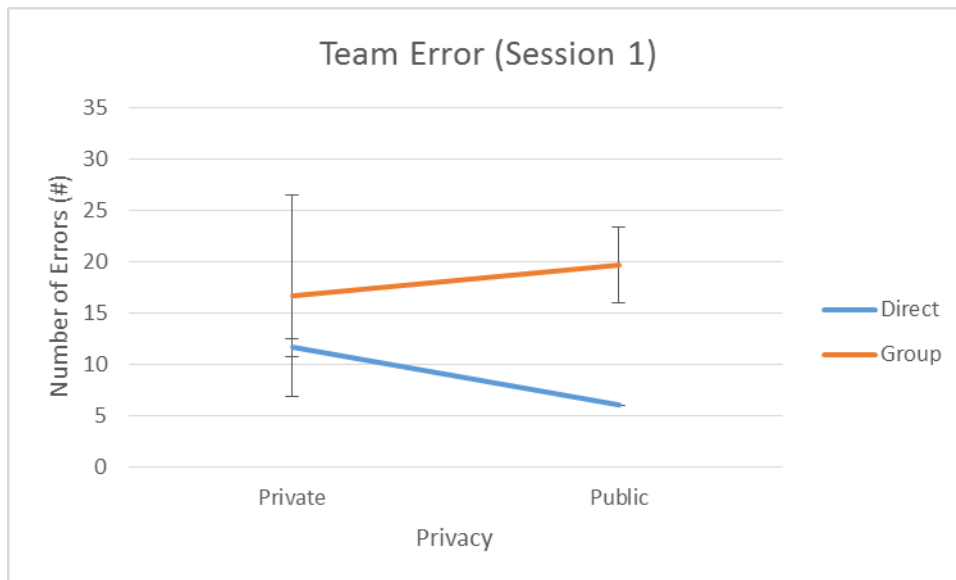
Figure 54. Average Team Errors (Session 1). Error bars represent S.E.M. Differences were not significant.

Session 2. The privacy of feedback ($F_{(1,3)} = 2.953$, $p = .184$, $\eta^2 = .020$) and the audience of feedback ($F_{(1,3)} = 1.072$, $p = .377$, $\eta^2 = .001$) had no significant difference (within-subjects

ANOVA). Overall, teams in the Public Direct condition produced the highest average errors,

18.00 ± 12.77. The teams in the Private Direct condition produced the lowest average errors,

11.67 ± 7.69. Teams in the Private Group and the Public Group produced average errors of 13.33

± 5.61 and 12.00 ± (N/A), respectively (Table 33 and Figure 55). The S.E. for the Public Direct

condition was not possible to calculate because of the low sample size.

Table 22. Average team errors (Session 2, n = 10)

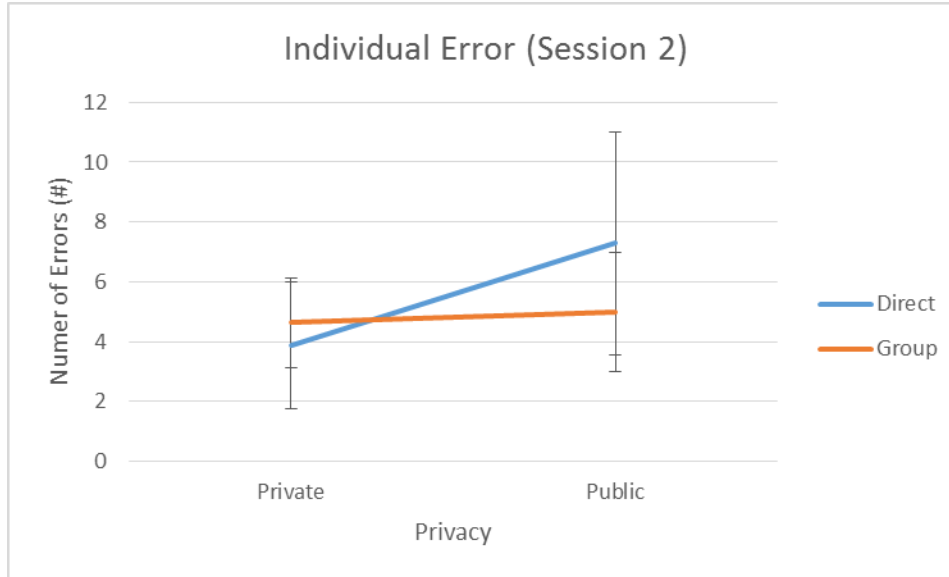|	Private	Public
Direct	11.67 ± 7.69	18.00 ± 12.77
Group	13.33 ± 5.61	12.00 ± (N/A)
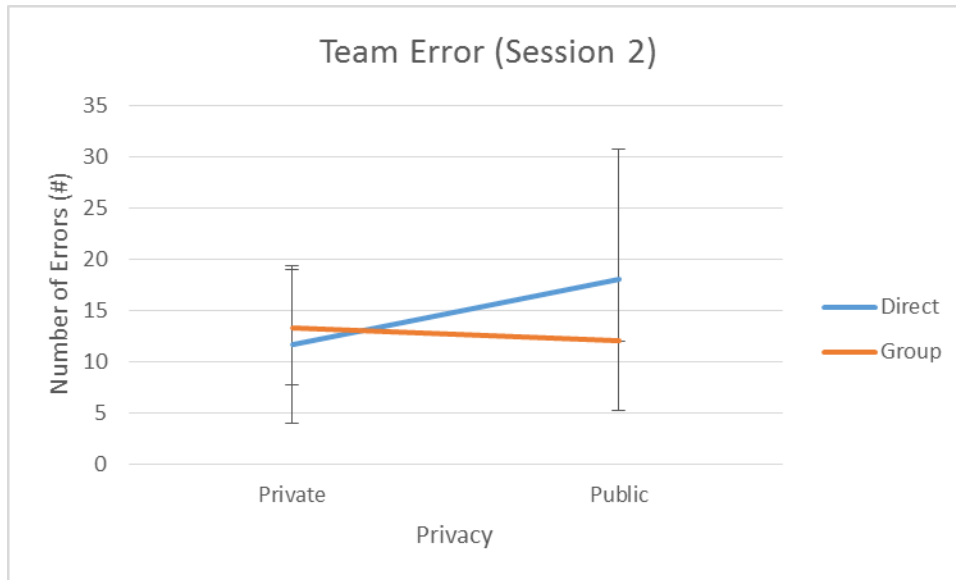
Figure 55. *Average Team Errors (Session 2). Error bars represent S.E.M. Differences were not significant.*

Session 3. The privacy of feedback ($F_{(1,2)}$ = .067, p = .820, η^2 = .018), the audience of

feedback ($F_{(1,2)}$ = 4.111, p = .180, η^2 = .190), and the interaction ($F_{(1,2)}$ = .061, p = .828, η^2 =

.055) of the privacy and audience of feedback had no significant difference (within-subjects

ANOVA). Overall, teams in the Public Group condition produced the highest average errors,

16.67 ± 11.20. The teams in the Public Direct condition produced the lowest average errors, 4.00

± 1.53. Teams in the Private Direct and the Private Group condition produced average errors of

6.00 ± 6.00 and 9.50 ± 3.50, respectively (Table 23 and Figure 56).

Table 23. Average team errors (Session 3, *n* = 10)

	Private	Public
Direct	6.00 ± 6.00	4.00 ± 1.53
Group	9.50 ± 3.50	16.67 ± 11.20
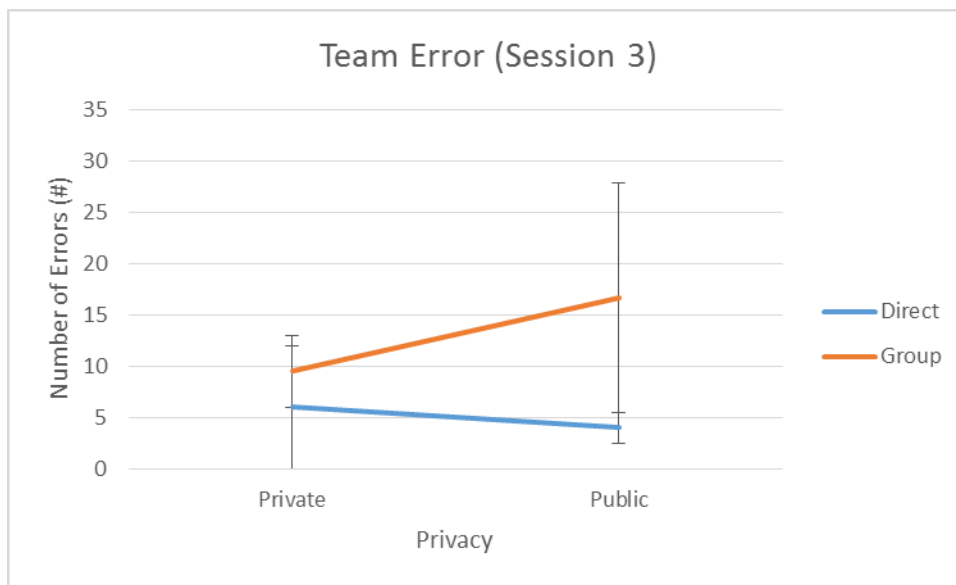
Figure 56. *Average Team Errors (Session 3). Error bars represent S.E.M. Differences were not significant.*

Session 4. The privacy of feedback ($F_{(1,2)}$ = 7.489, *p* = .111, η^2 = .001) and the interaction

($F_{(1,2)}$ = .409, *p* = .588, η^2 = .200) of the privacy and audience of feedback had no significant

difference (within-subjects ANOVA). The audience of feedback ($F_{(1,2)}$ = 23.564, *p* = .039, η^2 =

.655) had a statistical significance. Overall, teams in the Public Direct condition produced the

highest average errors, 9.33 ± 2.03. The teams in the Public Group condition produced the lowest

average errors, 2.33 ± 0.88. Teams in the Private Direct and the Private Group produced average

errors of 7.50 ± 1.50 and 4.50 ± 1.50, respectively (Table 24 and Figure 57).

Table 24. Average team errors (Session 4, *n* = 10)

	Private	Public
Direct	7.50 ± 1.50	9.33 ± 2.03
Group	4.50 ± 1.50	2.33 ± 0.88
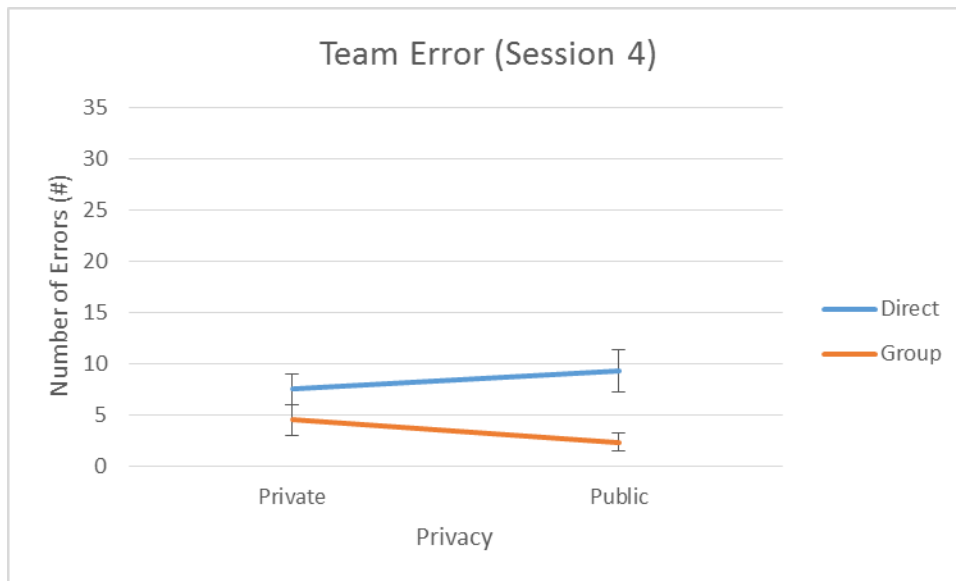
Figure 57. *Average Team Errors (Session 4). Error bars represent S.E.M. Differences were significant for some variables but not others.*

Impact of Feedback on Time Left for Individual and Team

First we report time remaining data obtained as described in Data Analysis Plan section

in Chapter 3, and explore whether it was affected by the privacy of feedback (private vs. public)

or the audience of feedback (team or individual).

Individual - Time Left

It is important to note that the confederate data was not included in the following

analysis. The following section describes the results of the individual time left for participants.

The sections are separated by session order. Each section describes the privacy of feedback, audience of feedback, and the interaction of privacy and audience of feedback with the statistical significance. The data in the following section address Research Questions 1 and 2.

Graphically, there appeared to be an interaction between the privacy and audience independent variables (Figure 58). A within-subjects ANOVA was used because each participant experienced all both levels of both independent variables. After running the ANOVA analysis, there was no statistical significance for the privacy of feedback ($F_{(1,96)} = .076, p = .784, \eta^2 < .001$), the audience of feedback ($F_{(1,96)} = .058, p = .810, \eta^2 < .001$), or their interaction ($F_{(1,96)} = .579, p = .448, \eta^2 = .007$).

More investigation was needed in order to better understand the data collected. Upon further investigation it was discovered that 70% of the participants strongly agree that their own individual performance improved over time. Analysis showed that there was a learning curve for the participants over time (Figure 59). This result led to the analysis of the individual time left with respect to session order. After running the one-way ANOVA analysis, there was a statistical significance ($F_{(1,100)} = 10.84, p = .001$) among the session orders. Specifically, there is a significant difference between Sessions 1 and 3 ($p < .001$) and Sessions 1 and 4 ($p < .002$).

Figure 58*. Average individual time left (all sessions). Error bars represent S.E.M. Differences are not statistically significant.*

Figure 59*. Average time left of individuals by session. Error bars represent S.E.M. Time left across sessions is significantly different.*

Due to learning curve differences by session, the impact of feedback audience and privacy on time left was analyzed by session.

Session 1. The audience of feedback ($F_{(1,18)} = .268$, $p = .611$, $\eta^2 = .088$) and the interaction ($F_{(1,18)} = .260$, $p = .616$, $\eta^2 = .029$) of the privacy and audience of feedback had no significant difference (within-subjects ANOVA). The privacy of feedback ($F_{(1,18)} = 53.069$, $p < .001$, $\eta^2 = .060$) did have a significant difference on the time left at the end of each session. Overall, participants in the Public Group condition produced the highest average time left, 1.180% ± .590%. The participants in the Private Direct and the Public Direct condition produced the lowest average time left, 0.00% ± 0.00%. Participants in the Private Group produced an average time left of .364% ± .238% (Table 25 and Figure 60).

Table 25. Average individual time left (Session 1, *n*=26)

	Private	Public
Direct	0.00% ± 0.00%	0.00% ± 0.00%
Group	.36% ± .23%	1.18% ± .59%

Figure 60. *Average Individual Time Left (Session 1). Error bars represent S.E.M. Differences was significant for some variables but not others.*

Session 2. The privacy of feedback ($F_{(1,18)} = .201$, $p = .659$, $\eta^2 = .013$) and the interaction ($F_{(1,18)} = .003$, $p = .958$, $\eta^2 = .011$) of the privacy and audience of feedback had no significant difference (within-subjects ANOVA). The audience of feedback ($F_{(1,18)} = 4.918$, $p = .0397$, $\eta^2 = .156$) did have a significant difference on the time left at the end of each session. Overall, participants in the Private Direct condition produced the highest average time left, 6.80% ± 1.45%. The participants in the Public Group condition produced the lowest average time left, 0.00% ± 0.00%. Participants in the Public Direct and Private Group condition produced an average time left of 6.31% ± 2.83% and 3.04% ± 1.48% respectively (Table 26 and Figure 61).

Table 26. Average individual time left (Session 2, $n = 26$)

	Private	Public
Direct	6.80% ± 1.45%	6.31% ± 2.83%
Group	3.04% ± 1.48%	0.00% ± 0.00%

Figure 61. *Average Individual Time Left (Session 2). Error bars represent S.E.M. Differences was significant for some variables but not others.*

Session 3. The audience of feedback ($F_{(1,18)} = 4.039$, $p = .059$, $\eta^2 = .011$) had no

significance but has a strong trend towards significance (within-subjects ANOVA). The privacy

of feedback ($F_{(1,18)} = 116.115$, $p < .001$, $\eta^2 = .195$) and the interaction ($F_{(1,18)} = 96.900$, $p < .001$,

$\eta^2 = .001$) of the privacy and audience of feedback did have a significant difference on the time

left at the end of each session. Overall, participants in the Private Group condition produced the

highest average time left, 14.96% ± 4.24%. The participants in the Public Direct condition

produced the lowest average time left, 4.66% ± 1.37%. Participants in the Private Direct and

Public Group condition produced an average time left of 12.42% ± 7.60% and 6.02% ± 1.64%

respectively (Table 27 and Figure 62).

Table 27. Average individual time left (Session 3, $n = 26$)

	Private	Public
Direct	12.42% ± 7.60%	4.66% ± 1.37%
Group	14.96% ± 4.24%	6.02% ± 1.64%

Figure 62*. Average Individual Time Left (Session 3). Error bars represent S.E.M. Differences was significant for some variables but not others.*

Session 4. The privacy of feedback ($F_{(1,18)} = .003$, $p = .954$, $\eta^2 = .138$), audience of feedback ($F_{(1,18)} = .198$, $p = .661$, $\eta^2 = .085$), and the interaction ($F_{(1,18)} = .109$, $p = .745$, $\eta^2 = .117$) of the privacy and audience of feedback had no significant difference on the time left at the end of each session (within-subjects ANOVA). Overall, participants in the Public Group condition produced the highest average time left, 14.82% ± 4.37%. The participants in the Private Group condition produced the lowest average time left, 3.25% ± 1.33%. Participants in the Private Direct and Public Direct condition produced an average time left of 5.00% ± 1.21% and 5.94% ± 2.90% respectively (Table 28 and Figure 63).

Table 28. Average individual time left (Session 4, *n* = 26)

	Private	Public
Direct	5.00% ± 1.21%	5.94% ± 2.90%
Group	3.25% ± 1.33%	14.82% ± 4.37%
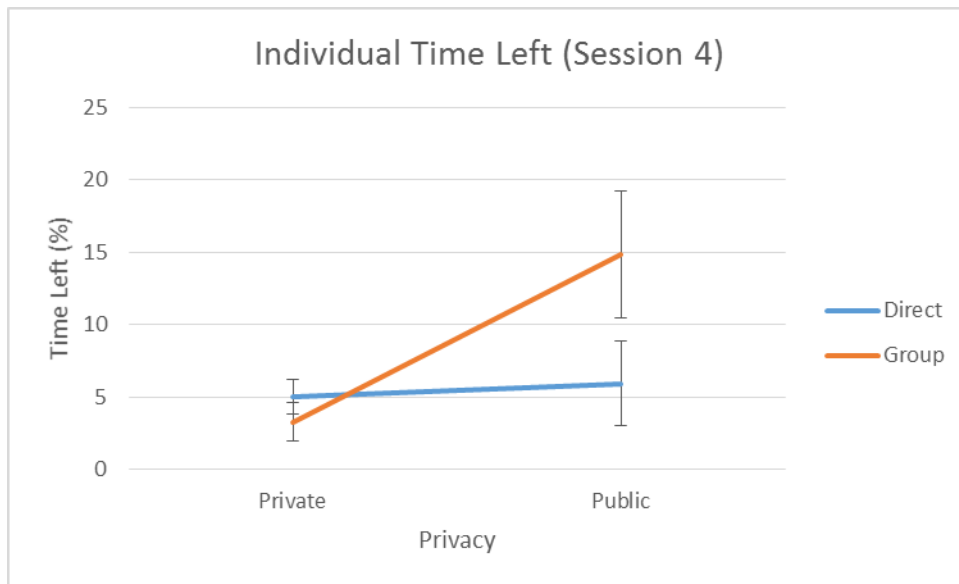
Figure 63. *Average Individual Time Left (Session 4). Error bars represent S.E.M. Differences were not significant.*

Team – Time Left

It is important to note that the confederate data was included in the following analysis. As well, small parts of the data analysis is missing (i.e., marked "N/A") because of the small team sample size. Graphically, there appeared to be a main effect for both the privacy and audience independent variables (Figure 64) on time left for teams. After running this ANOVA, however, there was no statistical significance for the privacy of feedback ($F_{(1,32)} = .002.$, $p = .965$, $\eta^2 = .002$), the audience of feedback ($F_{(1,32)} = .016$, $p = .901$, $\eta^2 < .001$), or their interaction ($F_{(1,32)} = .103$, $p = .750$, $\eta^2 = .012$).

More investigation was needed in order to better understand the data collected. Upon further investigation it was discovered that 77% of the participants strongly agree that the team's performance improved over time. Further analysis showed that there was a learning curve for the participants over time (Figure 65). After running the one-way ANOVA analysis, there was no statistical significance ($F_{(1,36)} = 3.057$, $p = .088$) among the session orders. However, it is important to note that there is a trend toward significance.
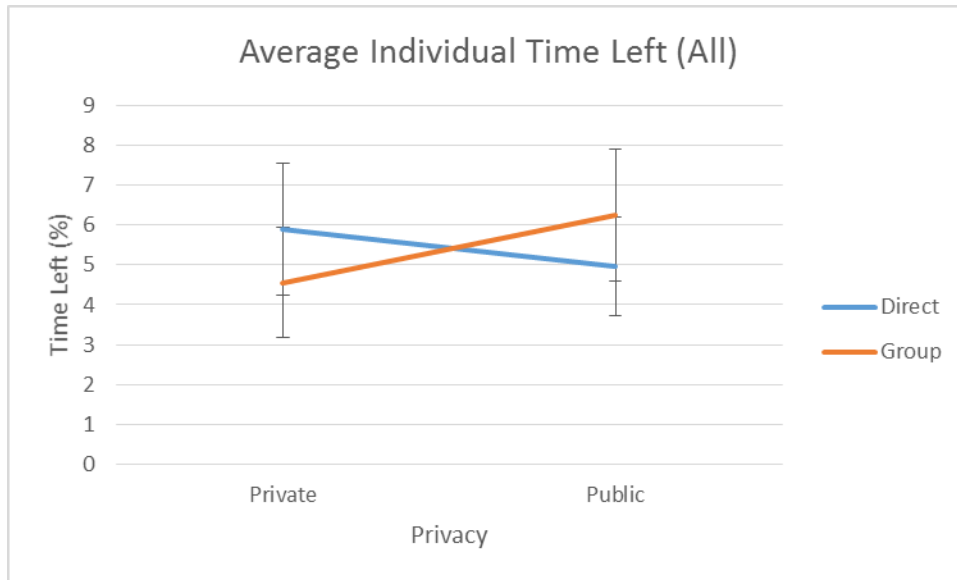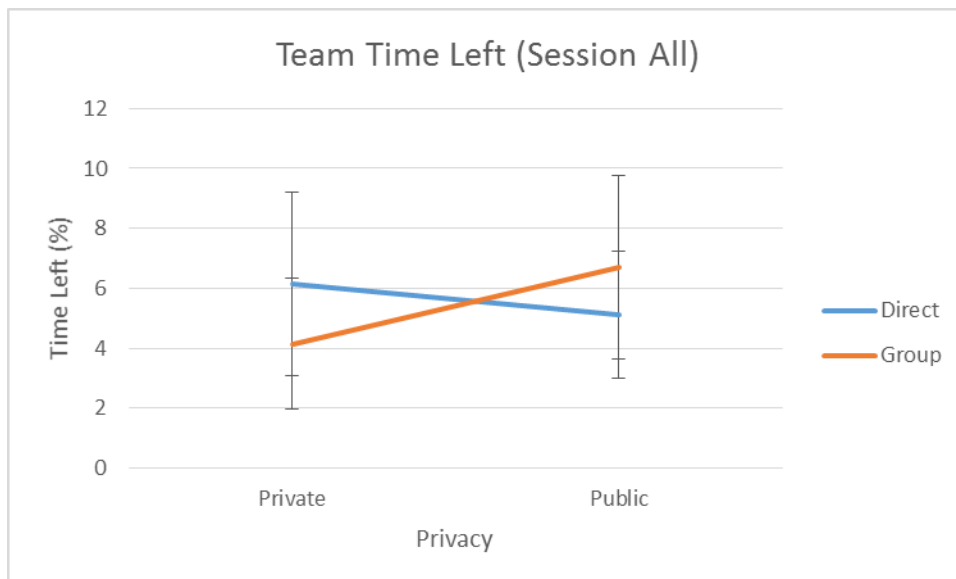

Figure 64*. Average team time left (all sessions). Error bars represent S.E.M. Differences are not statistically significant.*
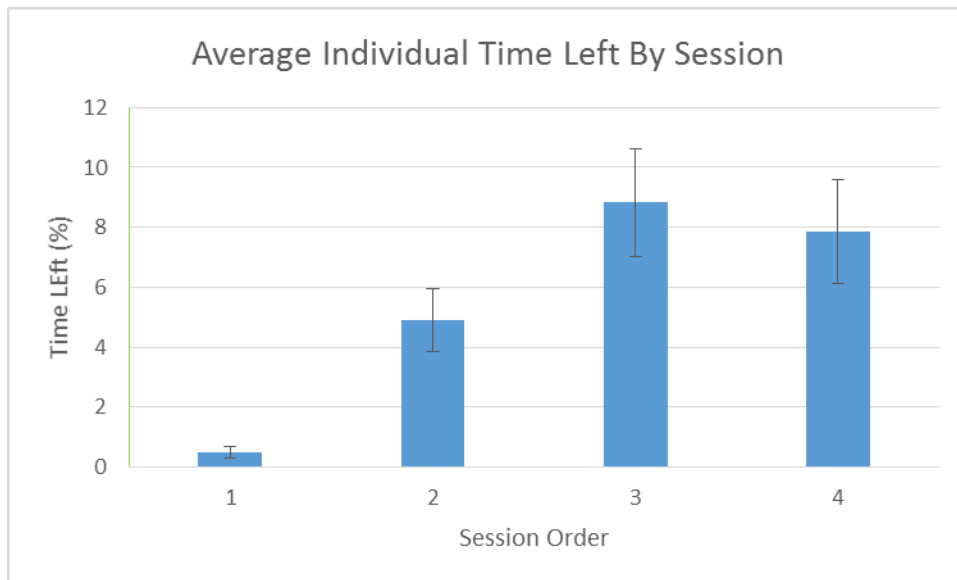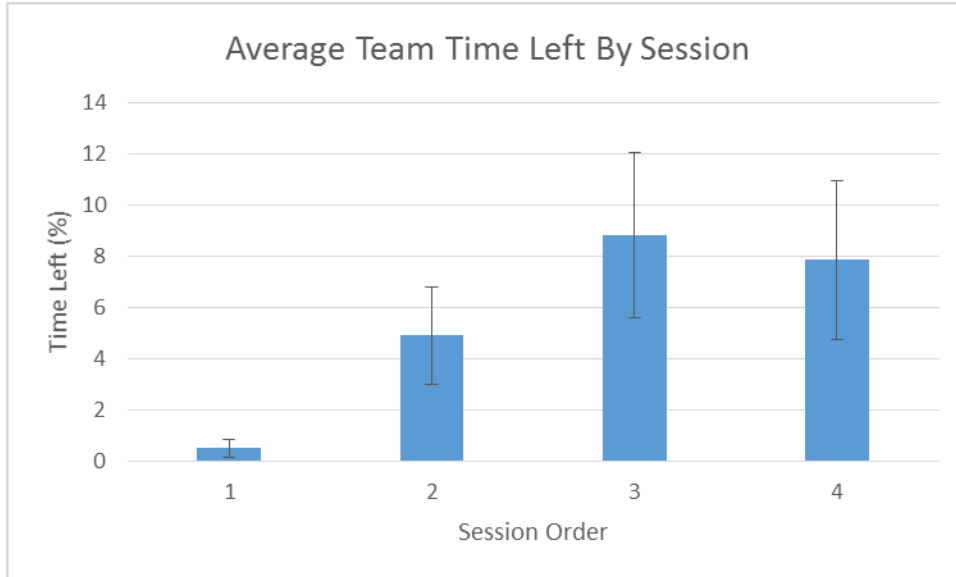
Figure 65. *Average time left of teams by session. Error bars represent S.E.M. Time left across sessions was not significantly different.*

The following section describes the impact of feedback configuration on time left by teams broken out by session. The data in the following section address Research Questions 1 and 2. As a reminder, there were 10 teams. Because of the relatively small number of teams, not every feedback condition was well-represented in each session. For example, in Session 1, only one team received feedback in the Public Direct condition.

Session 1. The privacy of feedback ($F_{(1,3)}$ = 8.506, p = .0617, η^2 = .043) and the audience of feedback ($F_{(1,3)}$ = .076, p = .801, η^2 = .108) had no significant difference (within subject ANOVA). No F or p value for the interaction of the privacy and audience of feedback because of the low sample size. Overall, teams in the Public Group condition produced the highest time left, 1.18% ± 1.18%. The teams in the Private Direct and Public Direct condition produced the lowest time left, 0.00% ± 0.00% and 0.00% ± (N/A) respectively. The S.E. for the Public Direct condition was not possible to calculate because of the low sample size. Teams in the Private

Group condition produced an average time left of 0.49% ± 0.49% respectively (Table 29 and

Figure 66).

Table 29. Average team time left (Session 1, *n* = 10)

	Private	Public
Direct	0.00% ± 0.00%	0.00% ± (N/A)
Group	0.49% ± 0.49%	1.18% ± 1.18%
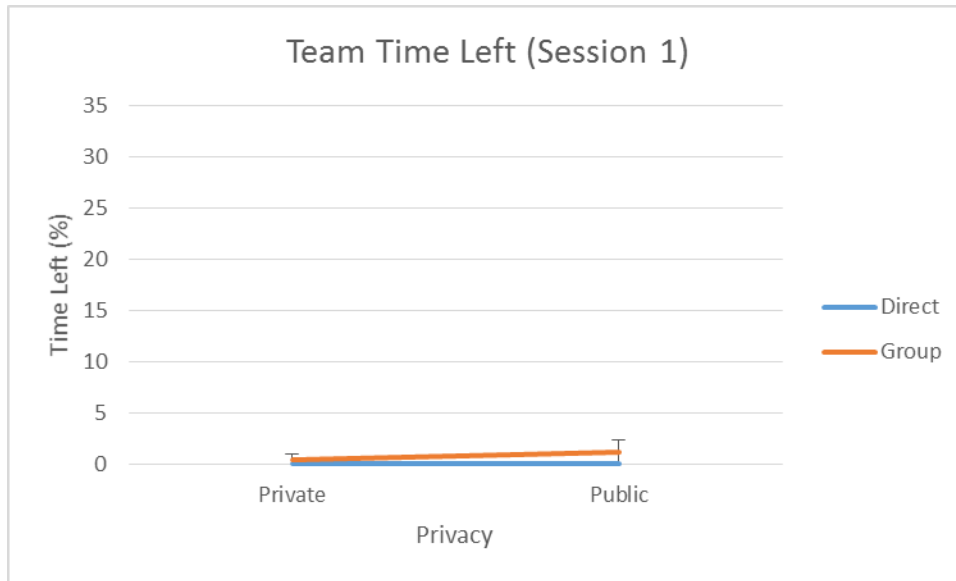
Figure 66*. Average Team Time Left (Session 1). Error bars represent S.E.M. Differences were not significant.*

Session 2. The privacy of feedback ($F_{(1,3)}$ = .007, *p* = .94, η^2 = .006) and the audience of

feedback ($F_{(1,3)}$ = .869, *p* = .42, η^2 = .178) (within subject ANOVA). No F or p value for the

interaction of the privacy and audience of feedback because of the low sample size. Overall,

teams in the Private Direct and Public Direct condition produced the highest average time left,

6.81% ± 2.91% and 6.81% ± 5.23% respectively. The teams in the Public Group condition

produced the lowest average time left, 0.00% ± (N/A). The S.E. for the Public Group condition

was not possible to calculate because of the low sample size. Teams in the Private Group

condition produced an average time left of 2.71% ± 2.71% (Table 30 and Figure 67).

Table 30. Average team time left (Session 2, *n* = 10)

	Private	Public
Direct	6.81% ± 2.91%	6.81% ± 5.23%
Group	2.71% ± 2.71%	0.00% ± (N/A)
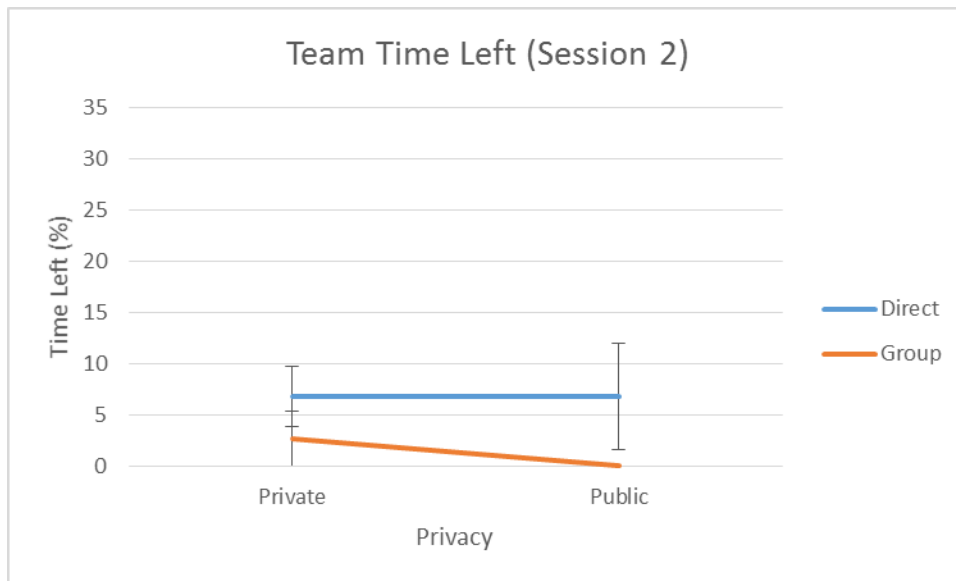
Figure 67. *Average Team Time Left (Session 2). Error bars represent S.E.M. Differences were not significant.*

Session 3. The audience of feedback ($F_{(1,2)}$ = .669, *p* = .499, η^2 = .001) and the

interaction ($F_{(1,2)}$ = 17.184, *p* = .053, η^2 = .005) of the privacy and audience of feedback had no

significant difference (within subject ANOVA). However, it is important to note that the

interaction had a trend towards significance. The Privacy of feedback ($F_{(1,2)}$ = 28.524, *p* = .033,

η^2 = .219) did have a significant difference on the time left at the end of each session. Overall,

teams in the Private Direct condition produced the highest average time left, 15.52% ± 15.52%.

The teams in the Public Direct condition produced the lowest average time left, 4.93% ± 2.47%.

Teams in the Private Group and the Public Group condition produced an average time left of

13.23% ± 8.65% and 5.35% ± 3.19% respectively (Table 31 and Figure 68).

Table 31. Average team time left (Session 3, *n* = 10)

	Private	Public
Direct	15.52% ± 15.52%	4.93% ± 2.47%
Group	13.23% ± 8.65%	5.35% ± 3.19%
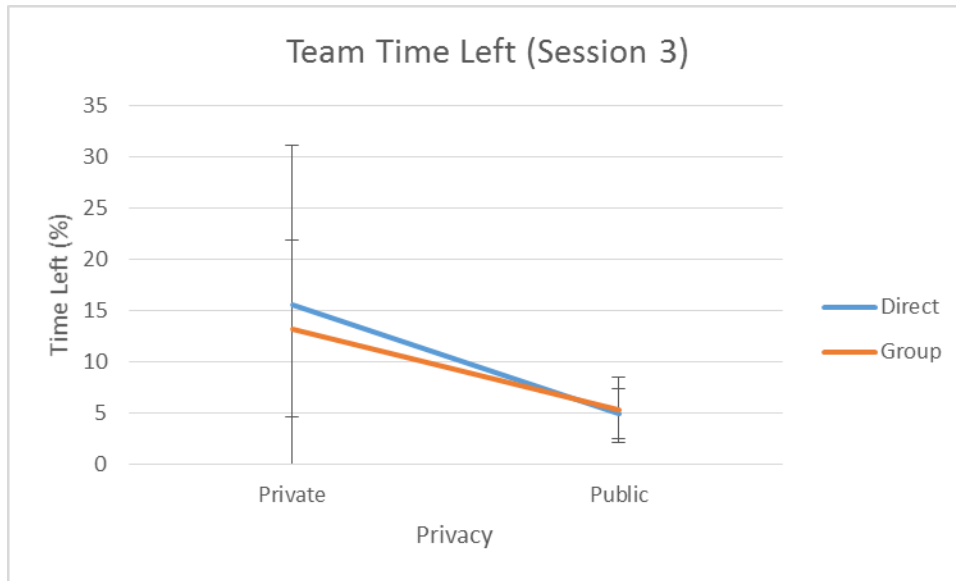
***Figure 68**. Average Team Time Left (Session 3). Error bars represent S.E.M. Differences was significant for some variables but not others.*

Session 4. The privacy of feedback ($F_{(1,2)} = .002$, $p = .965$, $\eta^2 = .156$), the audience of

feedback ($F_{(1,2)} = .018$, $p = .905$, $\eta^2 = .111$), and their interaction ($F_{(1,2)} = .011$, $p = .928$, $\eta^2 =$

.145) had no significant difference (within subject ANOVA). Overall, teams in the Public Group

condition produced the highest average time left, 15.76% ± 7.99%. The teams in the Private

Group condition produced the lowest average time left, 2.71% ± 2.71%. Teams in the Private

Direct and Public Direct condition produced an average time left of 5.00% ± 2.71% and 5.28% ±

5.28% respectively (Table 32 and Figure 69).

Table 32. Average team time left (Session 4, *n* = 10)

	Private	Public
Direct	5.00% ± 2.71%	5.28% ± 5.28%
Group	2.71% ± 2.71%	15.76% ± 7.99%
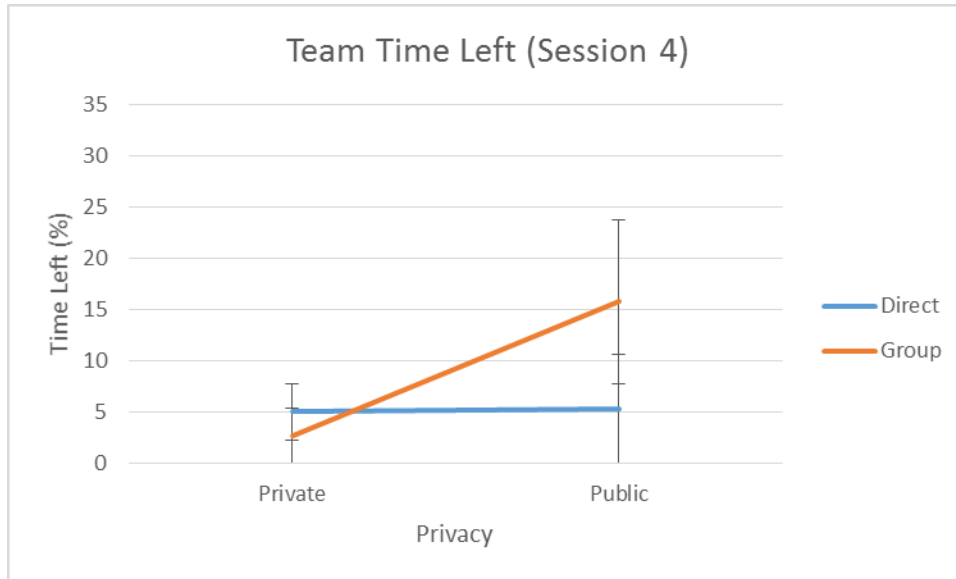
Figure 69. *Average Team Time Left (Session 4). Error bars represent S.E.M. Differences were not significant.*

Observed Team Strategies

There were two different team strategies that were observed during this experiment. These two strategies were designated Stay Together and Go Alone. A team using the Stay Together strategy stayed together as they moved around the mall. When the team arrived at a store, one member of the team, usually the member with the most money, entered the store and read all of the items out loud to the other members of the team. This allowed the other members to read their respective lists to see if any items in the store were also on the individual or team lists. If there were items that were needed, then the members of the team would enter the store and buy the desired items, one-by-one if needed in order to satisfy the rule to buy only one item

in a store. The advantage to using this method is that all of the members of the team were in the same location, focused on the same store, and everyone was confident that an item they needed was in the store. Furthermore, the team as whole would be able to keep track of who has visited which store. A disadvantage to this strategy is that it may require more time because of the heavy communication requirement: a member needed to read all of the items in the store. This is followed by all of the other members checking the items available in the store against their respective list, which also takes time.

Another strategy that teams used to complete the experiment was the Go Alone strategy. In the Go Alone strategy, members of the team spread out and visited different stores at different times. As the members of the team visited the stores, they would communicate to the team when they bought a team item or when a team item was in a store. An advantage to this strategy is that teams are able to cover more ground in a shorter amount of time. In addition, members were responsible for their own items on their personal list and only needed to communicate with the team if there was a team item that needed to be purchased. A major disadvantage to this strategy was the difficulty for team members to keep track of the stores visited. For example, when the teams that used this strategy were reviewing the items collected near the end of the session, sometimes there would be an item that no one collected. It was difficult to know the stores that members already visited. Consequently, it was difficult to know which stores members of the team should check. A team was classified using the Stay Together or Go Alone strategy if they were using a particular strategy for more than 50% of the total time used. For example, imagine a team used six minutes to complete a session. If the team used two minutes planning their strategy and then used the Go Alone strategy for four minutes to complete the task, then that team was classified as Go Alone. Table 33 shows the evolution of strategy for each team.

The Stay Together and Go Alone strategies seem different but they are actually the same at core. In both strategies the members of the team communicate with one another about the different items in the different stores. The major difference between these two strategies is the working memory load. The Go Alone strategy requires a higher cognitive load than the Stay Together strategy, because each member of the team had to remember the items in the store over time. The Stay Together strategy requires little to no cognitive load over time.

Table 33. Evolution of each team's strategy. For the strategy column: 0 = "No Strategy" (No color), 1 = "Stay together" (Yellow), 2 = "Go-alone" (Green).

Team	Session	Strategy		Team	Session	Strategy
1	1	2		7	1	2
1	2	2		7	2	1
1	3	1		7	3	1
1	4	1		7	4	1
2	1	2		8	1	2
2	2	2		8	2	1
2	3	2		8	3	1
2	4	2		8	4	1
4	1	2		9	1	2
4	2	1		9	2	2
4	3	1		9	3	2
4	4	1		9	4	2
5	1	2		10	1	2
5	2	2		10	2	1
5	3	2		10	3	1
5	4	2		10	4	1
6	1	1		11	1	2
6	2	1		11	2	2
6	3	1		11	3	2
6	4	1		11	4	1

The following section will describe how these two strategies were used. In this experiment there were 40 different sessions (10 teams with four sessions each). In 19 (48%)

different sessions the Stay together strategy was used, and in 21 (53%) different sessions the Go Alone method was used. Either the Stay Together strategy or Go Alone strategy was used for every session (i.e., no other strategy was apparent). The choice of strategy was not predictive of overall Team Performance, but it is worth noting that the Stay Together strategy could be viewed as a potentially superior one to Go Alone in that by Session 4, seven (70%) of teams had decided to use Stay Together (see **Table 34**). No team chose the Stay Together strategy and then reverted to Go Alone. Because Stay Together requires more communication and coordination than Go Alone, only one team used Stay Together in Session 1.

Table 34: Strategy used by teams by Session 4 and the session in which their final strategy was adopted.

Team	Strategy Used in Session 4	Session that Strategy Was Adopted
1	Stay Together	3
2	Go Alone	1
4	Stay Together	2
5	Go Alone	1
6	Stay Together	1
7	Stay Together	2
8	Stay Together	2
9	Go Alone	1
10	Stay Together	2
11	Stay Together	4

Communication

During participation, audio of participant interactions was recorded. From the recordings, utterances by each participant were counted. An utterance was defined as the collection of words spoken with a single intent, typically divided by a breath. E.g., "Hey should we all go together?" [breath] "Did you hear me? Should we all go together?" would be considered two utterances. For several teams, there were technical issues with audio recording, and the audio was lost.

It was hypothesized that the number of utterances might be predictive of the three team performance measures, or that total team utterances would decrease by Session 4 as teams became more experienced. However, there were too few data points to evaluate this statistically. A chart of the total number of team utterances by session is shown in Figure 70. Teams in this chart are ordered by an aggregate team performance score in Session 4 (correct team items collected+0.05*team time remaining-0.05*team errors) from lowest to highest, and tagged with their strategy used. Team 8, 2, and 7, for example, do show a pattern of an initial burst of communication that decreases by Session 4, but that pattern is not consistent across all teams. Strategy used does not seem to be related to communications.
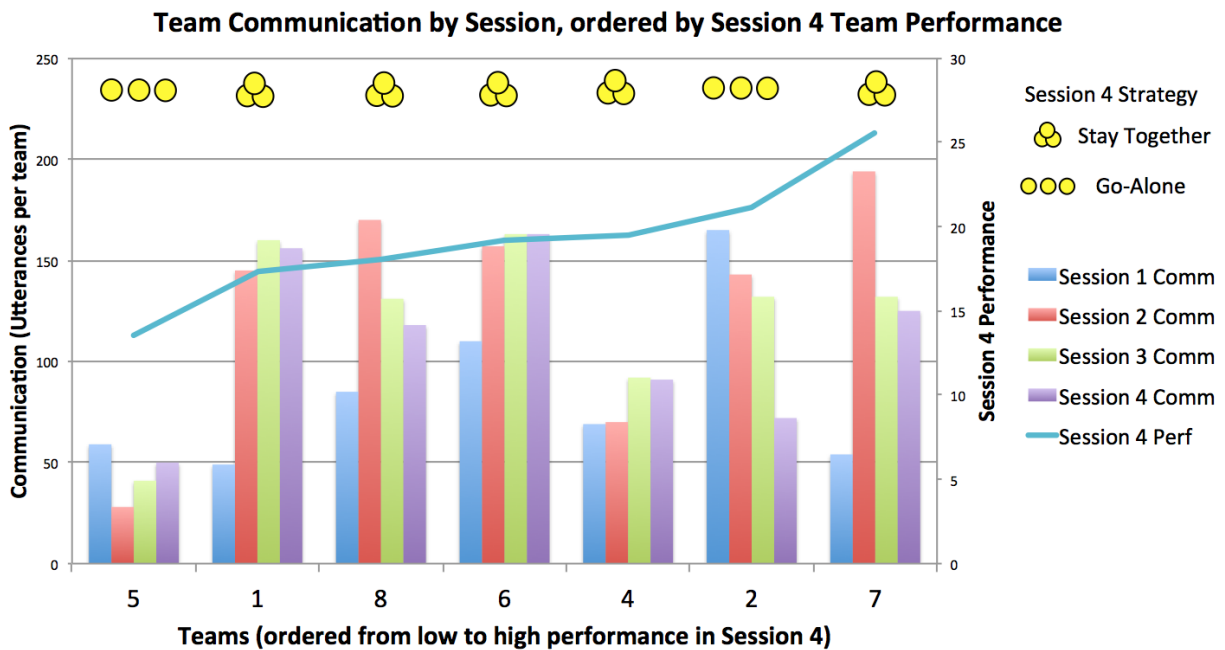
Figure 70. Team communication (number of utterances) by session, with teams ordered by low-high performance in Session 4. Team Strategy used also noted. Some teams missing due to incomplete audio communication data.

CHAPTER 5 – DISCUSSION AND CONCLUSIONS

As noted in the introduction, feedback is an important element of training. The purpose of this research was to better understand how the audience of feedback and the privacy of feedback influence the performance of teams and individuals. This chapter will discuss the conclusions drawn from the results presented in the previous chapter and how they relate to the research questions.

Prediction Outcomes

It was expected that the team public condition would be most effective condition due to the Köhler effect that produces a motivational gain. Overall, the results showed that no feedback condition is consistently better than another feedback condition. The individual and team performance are discussed in the following sections. Finally, limitations and future directions are discussed at the end of this chapter.

Time Left

The time left for individuals and teams is one of the main dependent variables in this experiment. The time left for individual and team was based on the amount of time left at the end of a session over the amount of time given for each session (480 seconds).

Individual – Time Left

In the data presented in the previous chapter there was no significance or trend toward significance for the privacy of feedback, audience of feedback, or the interaction of privacy and audience (regardless of session order). The results showed Group feedback should be given publicly because it produces the highest individual time left overall.

Most of the participants strongly agreed that their own individual performance improved over time. This claim is supported by the fact that there was a statistical significance between the

session orders. Specifically, there was a significance between sessions 1 and 3, and sessions 1 and 4. The time left scores were then grouped by session order.

In Session 1 there was no statistically significant difference between the audience of feedback and the interaction of the privacy and audience of feedback. There was a statistically significant difference for the privacy of feedback. This would indicate that the privacy of feedback did have some influence. The data indicate that the percentage of time left was highest when feedback was given publicly and had a group audience, which suggests that this condition increases efficiency.

In Session 2 there was no statistically significant difference for the privacy of feedback or the interaction of privacy and audience of feedback. There was a significant difference for the audience of feedback. This would indicate that the audience of feedback had a significant influence. The data indicated that feedback that was Direct produced the highest time left when given publicly or privately.

In Session 3 the audience of feedback had no statistical significance. The privacy of feedback and the interaction of privacy and audience of feedback did have some statistical significance. The data suggested that Private feedback produced the highest time left when given to a Group audience. Feedback that was Public produced the highest time left when given to a Group audience. The data also suggested that the time left was higher overall when Feedback was given privately. In Session 4 there was no significance for any of the independent variables. As a result, no conclusion could be drawn.

Team – Time Left

In the data presented in the previous chapter there was no significance or trend toward significance for the privacy of feedback, audience of feedback, or the interaction of privacy and

audience (regardless of session order). The results showed Group feedback should be given publicly because it produces the highest team time left. The results also showed that Direct feedback should be given privately because it produces the highest team time left.

Most of the participants strongly agreed that the team's performance improved over time. This suggests that the session order had a significant influence on the team's performance, but there was no statistical evidence to support this claim. The time left scores were grouped by session order.

In Session 1, Session 2, and Session 4 there was no statistical significance for any of the independent variables, so no conclusion could be drawn. In Session 3 there was a significant for the privacy of feedback. The data suggested that generally feedback should be given privately because it produces a high percentage of time left.

Errors

The error count for individuals and teams is one of the main dependent variables in this experiment. The error count was based on the number of errors committed by an individual or the sum of team members' errors during a session.

Individual – Errors

In the data presented in the previous chapter there was no significance or trend toward significance for the privacy of feedback, audience of feedback, or the interaction of privacy and audience (regardless of session order). The results showed Direct feedback should be given privately because it produces the lowest individual error. The results also showed that Group feedback should be given privately because it produces the lowest individual error.

Most of the participants strongly agreed that their own individual performance improved over time. However, there was no statistical significance among the session orders. The errors scores were then grouped by session order.

In Session 2 and Session 4 there was no significant difference between the two independent variables. In Session 1 there was no statistically significant difference between the privacy of feedback and the interaction of the privacy and audience of feedback. There was a statistically significant difference for the audience of feedback. This would indicate that the audience of feedback did have some influence. The data indicated that the amount of errors committed is lowest when feedback is given publicly and has a direct audience, which suggests that this condition helps reduce errors during Session 1.

In Session 3 there was no statistically significant difference between with the privacy of feedback or the interaction of the privacy and audience of feedback. There was a statistically significant difference for the audience of feedback. This would indicate that the audience of feedback did have some influence. The data indicated that the amount of errors committed is lowest when feedback is given publicly and has a direct audience.

Team – Errors

In the data presented in the previous chapter, there was no significance or trend toward significance for the privacy of feedback, audience of feedback, or the interaction of privacy and audience (regardless of session order). The results showed Direct feedback should be given privately because it produces the lowest team errors. The results also showed that Group feedback should be given privately because it produces the lowest team errors.

Most of the participants strongly agreed that the team's performance improved over time. This suggests that the session order had a significant influence on the team's performance, but

there was no statistically significant difference between session orders. The error scores were grouped by session order.

In Session 1, Session 2, and Session 3 there was no statistical significance for any of the independent variables so no conclusion could be drawn. In Session 4, audience of feedback had significant difference on the errors committed as a team. The data suggested that generally fewer errors are committed when the audience of the feedback is Group.

Items Collected

The items collected for individuals and teams is one of the main dependent variables in this experiment. The items collected was based on the number of correct items collected during a shopping session.

Individual – Items Collected

In the data presented in the previous chapter there was no significance or trend toward significance for the privacy of feedback, audience of feedback, or the interaction of privacy and audience (regardless of session order). The results showed Direct feedback should be given publicly because it produces the highest correct items collected. The results also showed that Group feedback should be given publicly because it produces the highest correct items collected.

Most of the participants strongly agree that their own individual performance improved over time. This suggests that most participants improved their items collected score over time. However, there was no statistical significance among the session orders. The items collected scores were then grouped by session order.

In Session 1, Session 3, and Session 4 there was no statistical significance for any of the independent variables so no conclusion could be drawn. In Session 2 there was a significant

difference for the privacy of feedback. The data shows that the highest items collected score occurs when the Group feedback is given privately.

Team – Items Collected

In the data presented in the previous chapter there was no significance or trend toward significance for the privacy of feedback, audience of feedback, or the interaction of privacy and audience (regardless of session order). The results showed Direct feedback should be given publicly because it produces the highest correct items collected. The results also showed that Group feedback should be given publicly because it produces the highest correct items collected.

Most of the participants strongly agree that the team's performance improved over time. This suggests that the session order had a significant influence on the team's item collected score but there was no statistically significant difference between session orders. The items collected scores were grouped by session order. In Session 1, Session 2, Session 3, and Session 4 there was no statistical significance for any of the independent variables, so no conclusion could be drawn.

A possible reason why different results appear in different session may be due to the fact that influence of privacy and audience of feedback had a difference influence based on a team's experience or skill level. A highly skilled team may not have needed much guidance and therefore the feedback had little to no influence on the team. A less skilled team may have needed more guidance, and therefore the feedback would have had a higher influence on the team.

Team strategies of Go Alone and Stay Together were noted, and though most teams evolved to use Stay Together, strategy was not predictive of performance. Strategies do seem worth further study, however.

Communication as measured by spoken utterances was measured but not found to be predictive of performance. Future research might code the content of the communication to explore whether certain types of communication might be related to performance or strategy.

In regards to the Research Question 3, the TMET is an acceptable platform to study team performance teams. First, this platform will allow researchers or trainers to modify the cognitive workload for participants. This can be done by (1) varying the number items on the individual shopping list, (2) varying the number of items on the team shopping lists, or (3) by increasing or decreasing the amount of time given to perform the task at hand. Second, there was a wide variation of individual and team performance (i.e., few ceiling or floor effects). This is ideal because it allows us to get a better understanding of how performance is influenced. Lastly, the performance appeared to improve consistently over time (i.e., from session to session). These three characteristics would suggest that TMET is a good platform for studying a team's performance under different types of cognitive loads. However, more validation is still needed in the future to validate that the environment is not driving the results.

Conclusion

Overall, there is no strong conclusion about feedback modality that could be drawn from the data gathered. Over time (i.e., across the session order) the influence of the privacy and audience of feedback seem to change. There was no specific feedback condition that had a consistent influence on the performance for both teams and individuals. An explanation for this finding may be that some of the participants may not have attended to the feedback given to them in the upper left-hand corner of the screen. Though there was no significant overall findings in the data set presented, there are still interesting implications.

One implication is that over time the influence that privacy and audience of feedback has on performance changes. This could be due to the fact that the performance of teams and individuals improved over time. As the skills increase, the different types of feedback may become more valuable than other types of feedback. Another implication is that the influence of privacy and audience of feedback is different for the team performance when compared to the influence on individual performance. The reason for this conclusion could be explained by the audience level of feedback. The feedback that has an individual audience level may in turn influence the individual performance more than the team performance. As well, feedback that has content that focuses on the team performance may influence the team's performance more than the individual performance.

Limitations

This experiment does have some limitations. First, only 26 participants (10 teams) participated in this study. More teams and participants are needed in order to produce more significant results. Second, the familiarity of team members was not controlled during this experiment. A team's performance could have been higher than other teams if the members of that team were familiar with one another. Also, a team with less familiarity among its members may have had lower performance.

Future Work

There are several of opportunities for future research. Future studies should focus on how the privacy and target of feedback influences teams whose members are very familiar with one another. If the members of the team are familiar with one another and have previously worked together before, the positive or critical effect of team public feedback could have a different impact on performance. Future work should use a familiarity metric to help better understand if

familiarity influences performance. To measure familiarity, some studies use surveys (McDougall, Curry, & de Bruijn, 1999) and other studies develop a numerical value (Dunlop & Levine, 2012). Ultimately it depends on the overall focus of the experiment. Also, researchers should focus on the modality and timing of the feedback given to participants. In this current work the feedback was given to participants only when an error occurred. Future studies could explore whether the modality and timing of the feedback given to participants influence the overall performance of participants. Finally, research should seek to further validate the team modified version of the Multiple Errands Test (TMET) presented in this current work. If this method is found to be useful when studying teams, then this could provide a common metric that can be used to study the complexities of teams. Finally, an exploration of the whether these findings generalize to real-world tasks would be worthwhile. Researchers could focus on implementing team public feedback into different types of real training.

REFERENCES

Albon, R., & Jewels, T. (2014). Mutual Performance Monitoring: Elaborating the Development of a Team Learning Theory. *Group Decision and Negotiation*, *21*(1), 149–164. doi:10.1007/s10726-012-9311-9

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society : JINS*, *9*, 31–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12570356

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, *16*, 101–130. doi:10.1.1.121.9138

Bell, S. T. (2007). Deep-Level Composition Variables as Predictors of Team Performance: A Meta-Analysis. *The Journal of Applied Psychology*, *92*(3), 595–615. doi:10.1037/0021-9010.92.3.595

Berkowitz, L., & Levy, B. I. (1956). Pride in group performance and group-task motivation. *Journal of Abnormal Psychology*, *53*(3), 300–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13376233

Bonner, D., Gilbert, S., Dorneich, M., Burke, S., Walton, J., Ray, C., & Winer, E. (2014). Taxonomy of Teams, Team Tasks, and Tutors. In *Proceedings of the 2nd Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium*. Army Research Laboratory.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding Team Adaptation: A Conceptual Analysis and Model. *The Journal of Applied Psychology*, *91*(6), 1189–207. doi:10.1037/0021-9010.91.6.1189

Butler, R. P., & Jaffee, C. L. (1974). Effects of incentive, feedback, and manner of presenting the feedback on leader behavior. *Journal of Applied Psychology*, *59*(3), 332–336. doi:10.1037/h0036655

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A Multiple-Goal, Multilevel Model of Feedback Effects on the Regulation of Individual and Team Performance. *The Journal of Applied Psychology*, *89*(6), 1035–56. doi:10.1037/0021-9010.89.6.1035

Dickinson, T. L., & McIntyre, R. M. (1997). A Conceptual Framework for Teamwork Measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance*

Assessment and Measurement: Theory, Methods, and Applications (pp. 19–43). Mahwah, NJ: Lawrence Erlbaum Associates.

Dirks, K. (2000). Trust in Leadership and Team Performance: Evidence From NCAA Basketball. *Journal of Applied Psychology*, *85*(6). doi:10.1037//0021-90I0.85.6.1004

Dominick, P. G., Reilly, R. R., & Mcgourty, J. W. (1997). The Effects of Peer Feedback on Team Member Behavior. *Group & Organization Management*, *22*(4), 508–520. doi:10.1177/1059601197224006

Driskell, J. E., & Salas, E. (1992). Collective Behavior and Team Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(3), 277–288. doi:10.1177/001872089203400303

Dunlop, M., & Levine, J. (2012). Multidimensional pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking. *Proceedings of CHI 2012 ACM Press*. doi:10.1145/2207676.2208659

Eby, L. T., & Dobbins, G. H. (1997). Collectivistic orientation in teams : an individual and group-level analysis. *Journal of Organizational Behavior*, *18*, 275–295.

Fletcher, B., & Harris, S. (1996). Development of a virtual environment based training system for ROV pilots. In *OCEANS '96. MTS/IEEE. Prospects for the 21st Century. Conference Proceedings* (Vol. 1). doi:10.1109/oceans.1996.572546

Fransen, J., Weinberger, A., & Kirschner, P. a. (2013). Team Effectiveness and Team Development in CSCL. *Educational Psychologist*, *48*(1), 9–24. doi:10.1080/00461520.2012.747947

Gabelica, C., Bossche, P. Van Den, Segers, M., & Gijselaers, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review*, *7*(2), 123–144. doi:10.1016/j.edurev.2011.11.003

Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, *9*(6), 491–501. doi:10.1016/0022-1031(73)90030-9

Groskreutz, M. (2013). Schedule of Reinforcement. In (F. Volkmar, Ed.)*Encyclopedia of Autism Spectrum Disorders*. Springer New York. doi:10.1007/978-1-4419-1698-3_1169

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, pp. 904–908). doi:10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Resuls of Empirical and Theoretical Research. *Advances in Psychology*, *52*, 139–83.

Hartley, J. R., & Sleeman, D. H. (1973). Towards More Intelligent Teaching Systems. *International Journal of Man-Machine Studies*, *5*(2), 215–236.

Hategekimana, C., Gilbert, S., & Blessing, S. (2008). Effectiveness of using an intelligent tutoring system to train users on off-the-shelf software. In *Society for Information Technology & Teacher Education International Conference* (pp. 414–419). Retrieved from http://www.editlib.org/p/27199/

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. doi:10.3102/003465430298487

Hertel, G., Kerr, N. L., & Messé, L. A. (2000). Motivation gains in performance groups: paradigmatic and theoretical developments on the Köhler effect. *Journal of Personality and Social Psychology*, *79*(4), 580–601. doi:10.1037/0022-3514.79.4.580

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of Individual Feedback on Behavior in Organizations. *Journal of Applied Psychology*, *64*(4), 349–371. doi:10.1037//0021-9010.64.4.349

Kerr, N. L., Messe, L. A., Park, E. S., & Sambolec, E. J. (2005). Identifiability, Performance Feedback and the Kohler Effect. *Group Processes & Intergroup Relations*, *8*(4), 375–390. doi:10.1177/1368430205056466

Klein, G., & Pierce, L. (2001). Adaptive Teams. In *6th International Command and Control Research and Technology Symposium*. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA467743

Knight, C., Alderman, N., & Burgess, P. W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation*, *12*(3), 231–255. doi:10.1080/09602010244000039

Koedinger, K. R., Aleven, V., Hockenberry, M., McLaren, B. M., & Heffernan, N. (2004). Opening the door to non-programmers: Authoring intelligent tutor behavior by demonstration. In *Proceedings of the Seventh International Conference of Intelligent Tutoring Systems*. Maceio, Brazil. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-30139-4_16

Kozlowski, S., & DeShon, R. (2004). A Psychological Fidelity Approach to Simulation-Based Training: Theory, Research, and Principles. In E. Salas, L. R. Elliott, S. G. Schflett, & M. D. Coovert (Eds.), *Scaled Worlds: Development, Validation, and applications* (pp. 75–99). Burlington, VT: Ashgate Publishing.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the Effectiveness of Work Groups and Teams. *Psychological Science in the Public Interest*, *7*(3), 77–124. doi:10.1111/j.1529-1006.2006.00030.x

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72. doi:10.3758/s13423-012-0341-8

Litman, D. J., & Silliman, S. (2004). ITSPOKE: An intelligent tutoring spoken dialogue system. In *Proceedings of the human language technology conference: 3rd meeting of the North American chapter of the association of computational linguistics (HLT-NAACL)* (pp. 5–8). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1614025.1614027

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*(3), 550–557. doi:10.3758/BF03196312

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.

McDougall, S. J. P., Curry, M. B., & de Bruijn, O. (1999). Measuring symbol and icon characteristics: norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 487–519. doi:10.3758/BF03200730

Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance*, *23*(3), 309–338. doi:10.1016/0030-5073(79)90001-1

Passos, A. M., & Caetano, A. (2005). Exploring the effects of intragroup conflict and past performance feedback on team effectiveness. *Journal of Managerial Psychology*, *20*(3/4), 231–244. doi:10.1108/02683940510589028

Peñarroja, V., Orengo, V., Zornoza, A., Sánchez, J., & Ripoll, P. (2015). How team feedback and team trust influence information processing and learning in virtual teams : A moderated mediation model. *Computers in Human Behavior*, *48*, 9–16.

Porter, C. O. L. H., Hollenbeck, J. R., Ilgen, D. R., Ellis, A. P. J., West, B. J., & Moon, H. (2003). Backing up behaviors in teams: the role of personality and legitimacy of need. *The Journal of Applied Psychology*, *88*(3), 391–403.

Priest, H. a., Burke, C. S., Munim, D., & Salas, E. (2002). Understanding Team Adaptability: Initial Theoretical and Practical Considerations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(3), 561–565. doi:10.1177/154193120204600372

Rand, D., Basha-Abu Rukan, S., Weiss, P. L. T., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitationehabilitation*, *19*(4), 583–602. doi:10.1080/09602010802469074

Salas, E., Fiore, S. M., & Letsky, M. P. (Eds.). (2012). Gaining Insight Into Team Processes on Cognitive Tasks. In *Theories of Team Cognition: Cross-Disciplinary Perspectives* (1st ed.). New York, NY: Routledge.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555–599. doi:10.1177/1046496405277134

Self, J. (1999). The defining characteristics of intelligent tutoring systems research : ITSs care , precisely. *International Journal of Artifical Intelligence in Education*, *10*, 350–364.

Self, J. A. (1974). Student Models in Computer-aided Instruction. *International Journal of Man-Machine Studies*, *6*, 261–276.

Shallice, T., & Burgess, P. W. (1991). Deficits in Strategy Application Following Frontal Lobe Damage in Man. *Brain*, *114*, 727–741.

Shute, V. J., & Psotka, J. (1996). Intelligent tutoring systems: Past, Present, and Future. In D. Jonasse (Ed.), *Handbook of Research on Educational Communications and Technology*. Scholastic Publications. doi:10.1126/science.228.4698.456

Sottilare, R. (2010). Toward the Development of an Intelligent Tutoring System for Distributed Team Training through Passive Sensing. In *Intelligent Tutoring Systems* (pp. 411–413). Springer Berlin Heidelberg.

Sottilare, R. A. (2012). Considerations in the Development of an Ontology for a Generalized Intelligent Framework for Tutoring. In *Proceedings of the International Defense and Homeland Security Simulation Workshop*.

Sottilare, R., Holden, H., Brawner, K., & Goldberg, B. (2011). Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training. *U.S. Army Research Laboratory, Human Research and Engineering Directorate, Orlando, Florida*.

Suh, H., & Lee, S. (2006). Collaborative Learning Agent for Promoting Group Interaction. *ETRI Journal*, *28*(4), 461–474. doi:10.4218/etrij.06.0105.0235

Tindale, R. S. (1989). Group vs Individual Information Processing : The Effects of Outcome Feedback on Decision Making. *Organizational Behavior and Human Decision Processes*, *44*, 454–473.

Uitdewilligen, S., Waller, M. J., & Pitario, A. H. (2013). Mental Model Updating and Team Adaptation. *Small Group Research*. doi:10.1177/1046496413478205

Wagner, J. A., & M. K. Moch. (1986). Individualism-collectivism: Concept and measure. *Group & Organization Management*, *11*(3), 280 – 304.

Walton, J., Bonner, D., Walker, K., Mater, S., Dorneich, M., Gilbert, S., & West, R. (2015). The Team Multiple Errands Test : A Platform to Evaluate Distributed Teams. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (pp. 247–250).

Walton, J., Dorneich, M. C., Gilbert, S., Bonner, D., Winer, E., & Ray, C. (2014). Modality and Timing of Team Feedback : Implications for GIFT. In *Proceedings of the 2nd Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium*. Army Research Laboratory.

Wedzinga, G. (2006). E-CATS: First time demonstration of embedded training in a combat aircraft. *Aerospace Science and Technology*, *10*, 73–84. doi:10.1016/j.ast.2005.08.005

Zaccaro, S. J., Rittman, A. L., & Marks, M. a. (2001). Team Leadership. *The Leadership Quarterly*, *12*(4), 451–483. doi:10.1016/S1048-9843(01)00093-5

Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. (1999). The Advanced Embedded Training System (AETS): An Intelligent Embedded Tutoring System for Tactical Team Training. *International Journal of Artificial Intelligence in Education*, 257–277.

APPENDIX

Appendix A - Shopping List

Session (1 of4)

Team

Items	Description
Glue	For the sticky, inner four year old in you.
Water colors	Really hard to work with.
Pencil Sharpener	Good for sharpening pencils.
T-shirt	A plain white shirt.
Dress Shirt	Nice and crisp, very professional.
Cowboy Boots	Giddy up!!
Blue Present	If you shake it, you can hear a faint rattling.
Armageddon	You know, like the movie.
The Hieroglyph Bridge	A magical bridge leads to Egypt. How will our young protagonists use this?
Tissues	These have lotion in them so they don't irritate your nose.
Coca-Cola	A delicious soft drink.
First Aid Kit	Better hope you don't need this.
Corn	A staple food!
Peanut Butter	Crunchy.
Paper Towels	Perfect for cleaning up other people's messes.
Bucket	Weren't you looking for one of these?
Tongs	You use them to grasp things.
Box of Matches	Light up your world.

Player 1

Items	Description
Napkins	Here, let me wipe your face.
Webcam	Old school style.
Safety Goggles	Remember kids, safety first!
Red Envelopes	Put letters in them!
Table	You can put stuff on it.
Ice Cream Cone	A perfect vanilla swirl.

Player 2

Items	Description
Balloons	Breathe out.
Cupcake Pan	It's the start of something new.
Nails	Let's hang stuff on walls!
White Invitations	For when you want people to come over.
Video Camera	For recording all those memories on an antiquated system.
Coffee	It's not an addiction.

Player 3

Items	Description
Party Hat	Now it's a PARTY.
Book of Matches	Light up your world.
Lighter Fluid	Be careful with this.
Green Invitations	For when you want people to come over.
Toaster	Don't forget to clean out the crumbs occasionally.
Pound Cake	With icing drizzled on top.

Session (2 of 4)

Team

Items	Description
Colored Pencils	Ooooh, pretty colors!
Streamers	I once covered my friend's room with this stuff.
Tape	Fixes most things.
Red Present	Wow you must like your friend a lot.
Newspaper	What's happening today?
Hawaiian Shirt	A pattern so brilliant it burns your eyes.
Invisible Queen	If no one can see her, no one can usurp her.
Beard of Soul	Beards are magical.
The Quest Into the Sixth Age	The future is now!
7-up	A delicious soft drink.
Eraser	Everybody makes mistakes, everybody has those days.
Milk	Strong bones.
Condiments	Ketchup AND mustard.

Pepsi	An average soft drink.
Ice Cream	A tub of chocolate, perfect for sharing.
Hard Hat	Keep that noggin safe.
Tongs	You use them to grasp things.
Bucket	Weren't you looking for one of these?

Player 1

Items	Description
Napkins	Here, let me wipe your face.
Cupcake Pan	It's the start of something new.
Box Cutter	Good for opening boxes.
Pencil	Good for drawing things.
Toaster	Don't forget to clean out the crumbs occasionally.
Cake Pan	We've got a cake to bake!

Player 2

Items	Description
Cardboard Box	You can put stuff in it.
Harmonica	The best way to annoy your brother
Nails	Let's hang stuff on walls!
Red Envelopes	Put letters in them!
Table	You can put stuff on it.
Caramels	So soft and chewy and delicious.

Player 3

Items	Description
Green Candle	An emerald candle.
Book of Matches	Light up your world.
Tape Measurer	Good for up to 300 yards.
White Invitations	For when you want people to come over.
Printer	For when you need hard copies of things.
Carmel Apple	What a treat!

Session (3 of 4)

Team

Items	Description
Streamers	I once covered my friend's room with this stuff.
Pencil Sharpener	Good for sharpening pencils.
Glue	For the sticky, inner four year old in you.
T-shirt	A plain white shirt.
Newspaper	What's happening today?
Hawaiian Shirt	A pattern so brilliant it burns your eyes.
The Hieroglyph Bridge	A magical bridge leads to Egypt. How will our young protagonists use this?
Invisible Queen	If no one can see her, no one can usurp her.
Armageddon	You know, like the movie.
7-up	A delicious soft drink.
First Aid Kit	Better hope you don't need this.
Milk	Strong bones.
Cheesecake	Good for birthday parties.
Tomato Soup	Tastes like my childhood.
Pepsi	An average soft drink.
Box of Matches	Light up your world.
Hard Hat	Keep that noggin safe.
Tongs	You use them to grasp things.

Player 1

Items	Description
Balloons	Breathe out.
Batteries	Guaranteed to last for up to six hours!
Hammer	This one doesn't belong to the captain.
Envelopes	Put letters in them!
Printer	For when you need hard copies of things.
Spam	Ham in a can.

Player 2

Items	Description
Red Candle	A ruby candle.
Crayons	Just don't let your kids write on the walls.
Crowbar	Will help you break things. Or open them. Whatever.
Pencil	Good for drawing things.
Microwave	Because you're too lazy to actually cook.

Pound Cake	With icing drizzled on top.

Player 3

Items	Description
Etch-a-sketch	Guaranteed to make you extremely frustrated.
Cupcake Pan	It's the start of something new.
Drill	You never know when you'll need one.
Red Envelopes	Put letters in them!
Table	You can put stuff on it.
Mint Tea	Refreshing, and caffeine free!

Session (4 of 4)

Team

Items	Description
Tape	Fixes most things.
Water colors	Really hard to work with.
Colored Pencils	Ooooh, pretty colors!
Cowboy Boots	Giddy up!!
Red Present	Wow you must like your friend a lot.
Dress Shirt	Nice and crisp, very professional.
Beard of Soul	Beards are magical.
Blue Present	If you shake it, you can hear a faint rattling.
The Quest Into the Sixth Age	The future is now!
Tissues	These have lotion in them so they don't irritate your nose.
Coca-Cola	A delicious soft drink.
Eraser	Everybody makes mistakes, everybody has those days.
Corn	A staple food!
Peanut Butter	Crunchy.
Paper Towels	Perfect for cleaning up other people's messes.
Bucket	Weren't you looking for one of these?
Hard Hat	Keep that noggin safe.
Box of Matches	Light up your world.

Player 1

Items	Description
Red Candle	A ruby candle.
Harmonica	The best way to annoy your brother
Crowbar	Will help you break things. Or open them. Whatever.
Green Invitations	For when you want people to come over.
Video Camera	For recording all those memories on an antiquated system.
Carmel Apple	What a treat!

Player 2

Items	Description
Napkins	Here, let me wipe your face.
Webcam	Old school style.
Tape Measurer	Good for up to 300 yards.
Envelopes	Put letters in them!
Toaster	Don't forget to clean out the crumbs occasionally.
Superdog	You don't know what you're missing.

Player 3

Items	Description
Balloons	Breathe out.
Batteries	Guaranteed to last for up to six hours!
Anvil	You never know when you're going to need one.
Pencil	Good for drawing things.
Microwave	Because you're too lazy to actually cook.
Coffee	It's not an addiction.

Appendix B - Post session survey

Attachment
G

Post Session

Post Session

Please move the slider to answer the following questions that pertain to the difficulty of the session.

	Very Difficult	Difficult	Somewhat Difficult	Neutral	Somewhat Easy	Easy	Very Easy
	0	1	2	3	4	5	6
How difficult was this shopping experience overall?							
How difficult was it to meet up with your team member at the correct time?							
How difficult was it to collect the different items from the team list?							
How difficult was it to collect the different items from the individual list?							
How difficult was it to collect the correct items from the virtual stores?							

Are there any comments that you would like to make about the difficulty of this experience?

Please mark the extent to which you agree these statements that pertain to the feedback in this session.
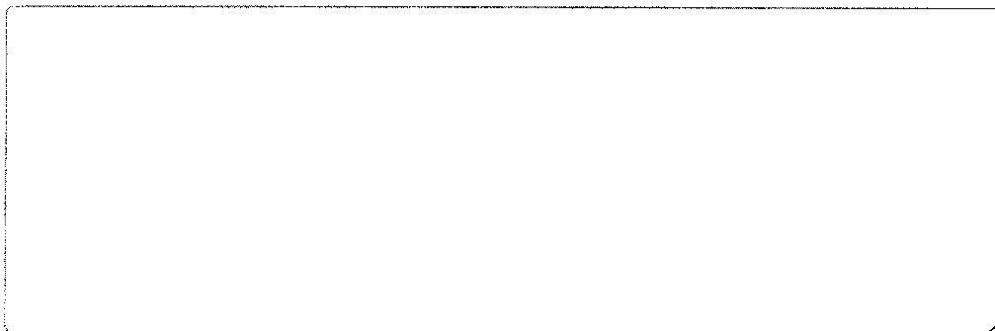
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
	0	1	2	3	4
I noticed when new feedback appeared.					
The feedback was helpful.					
The feedback was well timed.					
I got the feedback I needed to perform well.					

Are there any comments that you would like to make about the feedback in this session?

Did you receive feedback that mentioned other team members?

○ Yes

○ No

If you did receive feedback that mentioned other team members, was it helpful? If so, why?

Did you receive feedback that was displayed to other team members?

○ Yes

○ No

If you did receive feedback that was displayed to other team members, how did feel about it?

Are there any comments that you would like to make about how feedback was communicated in this session?

Appendix C - Overall-post session survey

Attachment

Overall Post Survey

Post Survey

Mark the extent to which you agree with these statements:

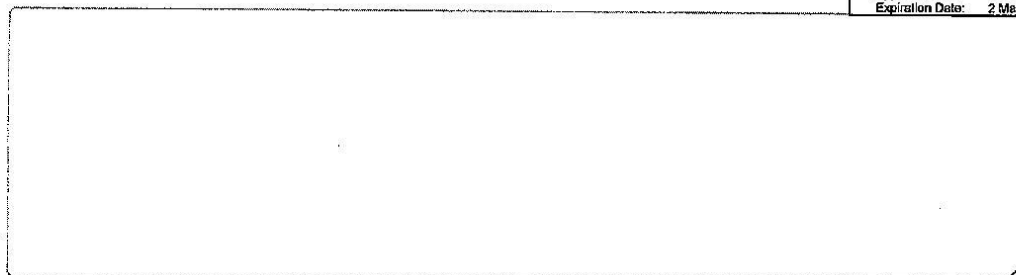|	Strongly Disagree 0	Disagree 1	Neither Agree nor Disagree 2	Agree 3	Strongly Agree 4
I thought my own performance improved over time.					
I thought our performance as a team improved over time.					
I could have done better if I had had better team members.					
We would have done better as a team if I had done a better job myself.					
It was better when the feedback was addressed to the team, not to me.					
It was better when the feedback was announced to the whole team instead of just to me.					

Did you feel like your performance, as an individual, changed over time? If so, why and in what way did your performance change? If not, why did your performance not change?

Did you feel like your performance as a team changed over time? If so, why and in what way did your team performance change? If not, why did your team performance not change?
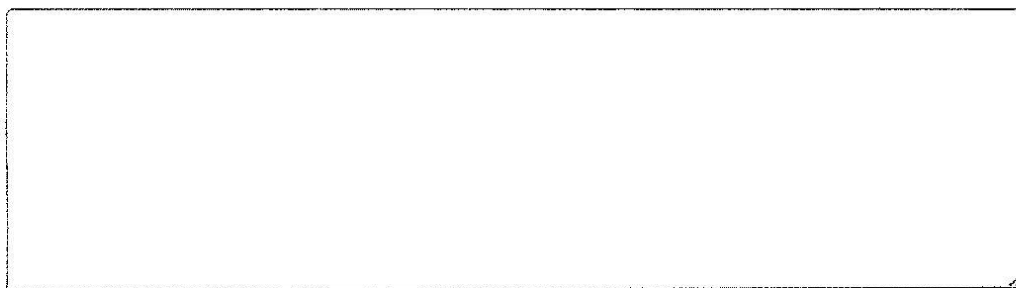
You participated in several sessions. You may have enjoyed one more or less than others. Please list below 3 or more characteristics of session(s) that you enjoyed. What made a session a better experience?

Now please list 3 or more characteristics of session(s) that you didn't enjoy as much. What made a session a worse experience?
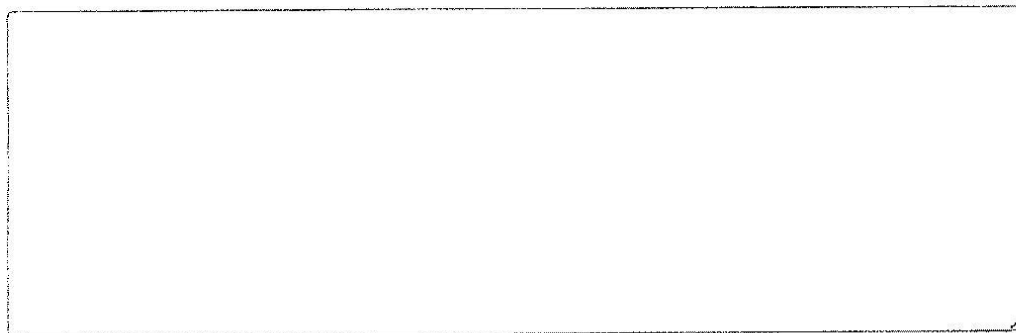
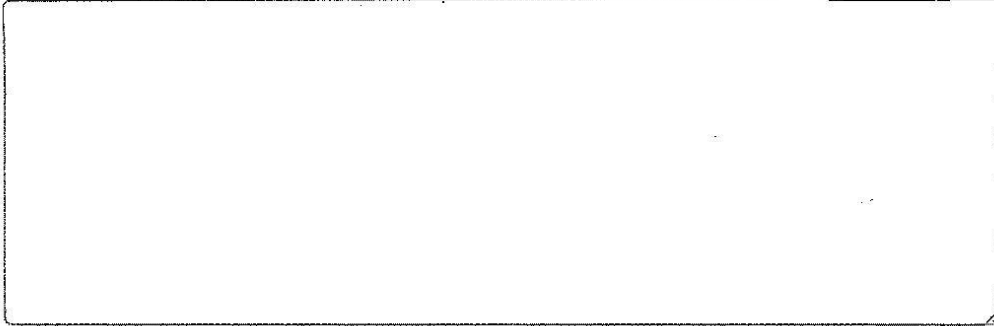Do you feel as though the feedback you received throughout this experience was useful? Why or why not?

How would you change the feedback to make it more effective?

Are there other comments you might like to share?

ISU IRB # 1	15-034
Approved Date:	3 March 2015
Expiration Date:	2 March 2016

Appendix D - Confederate script

1 Confederate

- Always be player 1

- Follow the group

    o If the group forms a plan, follow that plan (as long as it does not conflict with any

       of the other instructions given to you)

- Only speak when spoken to directly by one of your teammates

- Only respond to incoming vocal communication and do not initiate vocal communication

- Buy an item every minute

    o Buy only three items from the individual and team list first. Buy the remaining

       items second.