2019

Evaluating the effect of displaying team vs. individual metrics on team performance

Jamiahus Walton
Iowa State University

Evaluating the effect of displaying team vs. individual metrics on team performance

by

Jamiahus Walton

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human-Computer Interaction

Program of Study Committee:
Stephen Gilbert, Major Professor
Michael Dorneich
Jonathan Kelly
Larysa Nadolny
Eliot Winer

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

DEDICATION

I dedicate this dissertation to my amazing, wonderful, strong, and extraordinary wife, Dana. Without your encouragement, sacrifice, and support, this work would not be possible. The love you show me inspires me to be a better man every day. Thank you!

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

ACKNOWLEDGMENTS

I would like to thank my advisor, Stephen Gilbert. Thank you for supporting and advising me throughout my graduate career. Also, thank you for allowing me to teach a small section during one of your classes. Without your support and belief in my work, this dissertation would not be possible. Thank you to my committee for viewing my work through a critical lens with the goal of increasing the strength of the work. I would also like to thank the many undergraduate research assistants who all supported this work in various ways. Thank you everyone!

ABSTRACT

Teams have the potential to display high performance or low performance, depending on how well team members interact with one another. Training is commonly used to maintain or enhance the performance of various team types (e.g., sport or work teams). Intelligent Tutoring Systems (ITSs) have been used for years in multiple domains to tutor individuals. However, challenges arise when attempting to develop an Intelligent Team Tutoring System (ITTS). This current work focuses on the challenge of delivering effective feedback to teams via an ITTS designed to improve team performance. This research specifically focuses on how the assessment basis of the feedback (based on individual vs. team performance metrics) affects the team.

Specifically, this research examines how feedback displaying individual vs. team performance metrics influences team performance across multiple factors. The participants in this study performed a modified version of a classic shopping task designed to test cognition known as the Multiple Errands Test (MET). The researcher created a three-person team version of the MET called the Team Multiple Errands Test (TMET) within a virtual world on desktop computers. In three different feedback conditions, teams received performance feedback with information about individual performance, team performance, or both. Dependent measures included: performance (individual and team scores), items collected (correct and incorrect), errors, time remaining, collection time per item, and task strategy. Results were analyzed at the team and individual level. The analysis was conducted in three phases. First, the researcher analyzed the influence the feedback intervention had on main performance metrics: performance (individual and team scores), items collected (correct and incorrect), errors,

and time remaining. Second, the researcher analyzed how the influence of the feedback intervention on the dependent measures depended on the strategy that teams implemented. The third analysis explored the team's perception of performance by examining the correlation between performance and how participants viewed their teammates' performance and their own performance

The results from the first analysis suggested that the time remaining for participants in the Team feedback condition was significantly higher than in the Individual and Team (I&T) feedback condition, suggesting that feedback containing only Team-based feedback reduced the time teams spent in a session. The results from the second analysis suggested that the time remaining for participants on teams that used a specific strategy in which team members mostly stayed close to each other (designated "Go Together"), after gaining experience with the task, was significantly greater in the Team condition than in the Individual and I&T conditions. The analysis also revealed that the average collection time per item for participants in teams that used Go Together was significantly lower in the Team condition than in the Individual condition. The analysis also showed that frustration for participants on teams that used Go Together was significantly greater in the Individual condition than in the Team and I&T conditions. The third analysis demonstrated that participants in the Team condition consistently have a correct perception of their own performance and their team's performance. The results also showed that the influence of the feedback condition on participant's perception of individual or team performance changed depending on task experience.

This research offers a unique contribution to the growing body of research on Intelligent Team Tutoring Systems by exploring the benefits of offering real-time

feedback based on individual assessment vs. team assessment. This research also demonstrates that under some team circumstances, providing more information to teams is less effective than providing less.

CHAPTER 1. INTRODUCTION

Teams can achieve more than an individual alone. Teams win sports championships, design and build aircraft that safely transport hundreds of people across the world, and design and build spacecraft that can explore the unknown depths of deep space. In a world that is becoming increasingly complex, it is essential that team members interact effectively with one another. One way to support team interaction is to implement effective team training that improves both team and task skills. Training can come from a human instructor or, when possible, from software that attempts to mimic a human instructor. While Intelligent Tutoring Systems (ITS) have successfully instructed individual students via automated software (Aleven, Mclaren, Roll, & Koedinger, 2006; Graesser, Hu, & Sottilare, 2018; Hategekimana, Gilbert, & Blessing, 2008; Koedinger, Aleven, Hockenberry, McLaren, & Heffernan, 2004), there are few successful examples of Intelligent Team Tutoring Systems (ITTSs), i.e., software designed to coach teams.

A barrier to developing an effective ITTS is understanding the best method of giving feedback to team members. Generally, three important dimensions of feedback that impact the effectiveness of team feedback are Assessment (is the feedback based on the team vs. individual performance), Audience ("Player A, you…" Vs. "Team, you…"), and Privacy (public to the whole team vs. private to an individual). These three dimensions are components of what researchers have more generally labeled individual-level feedback vs. team-level feedback (Gabelica, Bossche, Segers, & Gijselaers, 2012). All three dimensions are important to feedback effectiveness, but the author limited the scope of this current work to the Assessment dimension: how the basis of assessment for feedback (individual score vs. team score) affects team performance and team dynamics. The purpose of this study is to

provide insights into how the assessment variable of feedback affects team training. The following section explores this challenge in more detail.

Intelligent Tutoring Systems (ITSs)

Researchers and practitioners have used ITSs in various domains for various uses, such as training users on off-the-shelf software (Hategekimana et al., 2008) or helping users develop better help-seeking strategies (Aleven et al., 2006). According to Shute and Psotka (1996), the earliest evidence of using "intelligent machines" can be traced back to around 1920. Different domains have used ITSs over the years, and many of them have similar essential components. According to Hartley and Sleeman (1973), an ITS needs to have four major components to be used for teaching: a representation of the task, a representation of the student (student model), teaching instructions based on the task and student representation (feedback), and a set of guidance rules (domain model). Generally, the goal is to develop an ITS that communicates the knowledge that it possesses effectively. This goal does not mean that the ITS needs to convey information in a way that mimics human tutors (Shute & Psotka, 1996). When developing a tutor for any domain, it is important to focus more on the ITS's effectiveness and less on how similar it is to an expert human tutor in a specific domain. Though all four components of the ITS are important, this current work focused on the feedback component (i.e., the teaching instructions).

An important question to answer when designing an ITTS is what characteristics are important and should be considered when developing the feedback distribution. In general, ITSs are useful because they can provide instructions to students in ways that are difficult to mimic by traditional teaching. For example, an ITS or a computer-based system can give immediate feedback (Azevedo & Bernard, 1995), provide immediate feedback to a specific responses (Larreamendy-Joerns, Leinhardt, & Corredor, 2005; Morris, 2001), encourage

active involvement (P. A. Cohen & Dacanay, 1992), encourage self-correction (VanLehn, 2011), and provide more opportunities for students to practice and receive instruction (Martin, Klein, & Sullivan, 2007). Based on previous research (Ma, Adesope, Nesbit, & Liu, 2014; Sosa, Berger, Saw, & Mary, 2011), one could argue that the instructional advantages afforded by ITSs are a significant reason why they are associated with higher achievement. ITS success with individual learners motivates the question of whether it is possible to develop an ITS that effectively trains a team to increase team performance. Several challenges arise when researchers attempt to develop an ITTS (Bonner et al., 2016). Feedback is an important element of learning (Gabelica et al., 2012; Mory, 2004) and the design of the feedback component is an ongoing challenge when developing an ITTS (Gilbert, Dorneich, Walton, & Winer, 2018; Gilbert, Slavina, et al., 2018).

The Need to Evaluate the Impact of Team Feedback

Previous studies have focused on the influences of various components of feedback, such as how goal and performance feedback regulate performance (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004), how feedback generally influences learning and achievement (Hattie & Timperley, 2007), how feedback influences individual behaviors (Ilgen, Fisher, & Taylor, 1979), or the modality and timing of feedback (Walton et al., 2014). Understanding the effect feedback has on individual vs. team performance is a complex issue, and there is still a need for better insight into the influence of feedback on team performance (Gonzalez-Mulé, Courtright, DeGeest, Seong, & Hong, 2016). This study seeks to address the need for more insight into how the assessment basis of feedback delivered by an ITTS influences team performance. The next section offers brief background information on ITTSs.

Feedback Considerations When Developing an ITTS

When developing an Intelligent Team Tutoring System (ITTS), there are unique challenges that arise that need to be addressed (Sottilare, Holden, Brawner, & Goldberg, 2011). One unique challenge is determining how to deliver feedback to a team. Feedback is an important design consideration that must be addressed before implementing an ITTS (Bonner et al., 2016). People often make feedback design considerations in the context of other design decisions. Imagine designing a training environment that improves the coordination and task performance of a military squad. A team leader must have the ability to identify behaviors, individual and team behaviors, that reduces the team's performance and use his or her expertise to adapt his or her instructions in a way that advances the team's performance. How should feedback be given to this team? In the use case just described, the feedback given by the ITTS will need to display content that is important to the leader.

According to Johnston, Burke, Milham, Ross, and Salas (2018), there are at least three considerations when developing an ITTS. First, to assist the team leader, the ITTS must understand the complexities of team dynamics and the leader's role within those dynamics. Understanding the leader's role is complex because the leader's functions can change over time (Burke, Georganta, & Hernandez, 2017). Second, as well as understanding the dynamics of teams, the ITTS must understand how to measure a team's knowledge, skill, and abilities such as attitude, mental processes, and behavior. Finally, the ITTS needs to use its understanding of team dynamics and team measurements to determine an effective team training strategy that will have a significant positive influence on team performances (Johnston et al., 2018). These three considerations are complex concepts that are often difficult for a human tutor or coach to understand and apply. How then are we to design and embed our limited understanding of the complexity of teams and their interaction into an

ITTS in a way that effectively supports team training? This question becomes more daunting when one considers the need to understand individuals as well.

When giving feedback to individuals, an ITTS also needs to understand how an individual may receive feedback depending on his or her affect state (e.g., frustration). Previous research indicates that feedback etiquette strategies significantly influence performance, motivation, and confidence (Yang & Dorneich, 2016). However, this research cautioned that although different etiquette strategies were used to influences individuals' performance, motivation, and confidence, this outcome does not indicate that one particular etiquette strategy improved all three areas. The results suggested that further studies are needed to understand how the interaction of etiquette strategies influences students. This conclusion suggests that even the tone of the feedback delivered to an individual may have a significant influence on individual performance, which may, in turn, influence the team performance overall. For example, one feedback etiquette strategy may positively benefit individual A's performance, but that same feedback may negatively influence individual B's performance. This simultaneous positive and negative influence may lead to negative or ineffective teamwork between A and B, which may lead to poor team performance.

It is important to note that there are forces that may influence a team's performance that is hard to minimize. For example, researchers argue that the attributes of team members (i.e., team composition) can influence team performance and outcomes (Bell, Brown, Colaneri, & Outland, 2018). This conclusion could imply that in an ideal situation, the teams within a team study should have similar team compositions, since having similar teams would increase confidence that any change in behavior will be a result of the intervention. However, it can be difficult to ensure similar team compositions across a study, which may

introduce noise into the study from artificially matched team members, or simply from the difficulty of finding each team members to create appropriate teams. Perhaps teams can be chosen carefully to reduce the influence of team composition. More work is needed to discover how team composition influences team performance and team outcomes. In summary, teams add complexity to a research study at multiple levels.

Research Question

This research attempts to answer the following question: How will teams' performance change when given feedback that displays indicators based on individual performance, team performance, or both?

Organization of This Dissertation

Chapter 2 offers a more detailed literature review of Intelligent Team Tutoring Systems and feedback design. Chapter 3 describes the methods for this research, which include both a description of the team-based software platform created to study team dynamics for this research, the Team Multiple Errands Task (TMET), and the experimental design used. Chapters 4-6 describe the results of the study by examining different dependent measures. Chapter 4 focuses on the basic performance measures for the TMET task and whether they were affected by the feedback intervention. Chapter 5 focuses on how strategies used by teams may have affected team performance. Chapter 6 focuses on qualitative and self-report data from surveys of team members, exploring teams' perception of individual and team performance. Chapter 7 summarizes the findings and describes future work.

Contributions of This Research

This dissertation will add several contributions to the team research field. This study will address the question of whether feedback presented to a team should be based more on the assessment of individual team members or the whole team. In particular, it will present

data on how the assessment basis of feedback influenced team performance (Chapter 4), how the feedback intervention influenced team performance based on the strategies implemented (Chapter 5), and how the feedback intervention influenced participants' perception of their performance and team performance (Chapter 6). Lastly, this research reaffirms the value of the Team Multiple Errands Test (TMET) as a platform for studying team behaviors.

CHAPTER 2. LITERATURE REVIEW

This chapter describes previous research that informs the current study from multiple perspectives. In particular, this literature review focuses on feedback design, the process of team adaptation to a task and how to measure it, and previous team training methods.

Feedback

A challenge that exists in many collaborative systems is that they are made up of complex instructional elements (Suh & Lee, 2006). One complex instructional element is the distribution of feedback to teams. In general, feedback has been defined as a special case of communication that occurs between a source and a recipient where the source conveys a message to the recipient. A recipient's response and perception of the information provided depends on the recipient's personal characteristics, the source's characteristics, and the message's characteristics (Ilgen et al., 1979). This current study focuses on the characteristics of the message. Specifically, this study focuses on the impact of feedback based on assessment at the individual level, the team level, or the individual and team level (I&T). The current literature review explores researchers' current understanding of feedback at these different levels.

There is little agreement among researchers as to which level of feedback (Individual or Team) produces the highest performance. Some researchers argue that one level of feedback (individual or team level) produces the highest performance (DeShon et al., 2004). Other researchers have not found strong differences in performance as a result of individual or team level feedback (Walton, Gilbert, Winer, Dorneich, & Bonner, 2015). These results suggest that more work is needed to gain more insight into how feedback at the individual level and team level influences performance.

Generally, researchers categorize feedback presented to a team as either individual or team feedback (Tindale, 1989). Previous research aimed to produce evidence that determines if the individual, team, or a combination of individual and team (I&T) feedback produce optimal team performance. The results of these studies are inconsistent or conflicting (Table 1). For example, Archer-Kath, Johnson, and Johnson (1994), Smith-Jentsch, Salas, and Baker (1996), and Moreland & Myaskovsky (2000) support the claim that individual feedback in a group setting produces high team performance. However, Scott-Young & Samson (2006) and Walter & Van Der Vegt (2009) provided evidence that suggests team feedback produces higher team performance.

Table 1 - Studies that support different claims of optimal feedback content to present to teams

Optimal Feedback Level	**Supporting Research**
Individual Feedback	(Archer-Kath et al., 1994) (Smith-Jentsch et al., 1996) (Moreland & Myaskovsky, 2000)
Team Feedback	(Scott-Young & Samson, 2006) (Walter & Van Der Vegt, 2009)
Individual and Team Feedback (I&T)	(Austin, Kessler, Riccobono, & Bailey, 1996) (Sivunen, 2006)
Individual or Team Feedback	(Emmert, 1978)
Individual or Team Feedback, Not both	(DeShon et al., 2004)

A possible conclusion to these results is that both individual and team feedback are needed to produce the highest team performance. Emmert (1978) concluded that performance would increase with individual or team feedback. If both individual and team feedback improve group performance, then one could conclude that both individual and team feedback are needed during team training. This conclusion is not difficult to understand because it is generally believed that the more elaborate the feedback, the more effective the feedback (Van

der Kleij, Feskens, & Eggen, 2015). Based on this conclusion, training should provide I&T feedback. Austin, Kessler, Riccobono, and Bailey (1996) and Sivunen (2006) offered support to the conclusion that I&T feedback should be given to teams to improve performance. It makes logical sense that a team will perform better when given more information because they will have a better understanding of the team's performance and the behavioral changes needed to improve performance in the future. However, Deshon et al. (2004) concluded that team performance was highest when given individual or team feedback but not when given I&T feedback. The authors argue that team members should only be given either individual or team feedback because they are not able to utilize the information provided with I&T feedback. This current work aims to provide more insight into which form of feedback produces increased team performance.

A possible explanation of the conflicting results attempting to determine feedback effectiveness is that feedback is a complex element in training that has several variables to consider (Table 2) (Gabelica et al., 2012; Walton et al., 2014). Gabelica et al. (2012) discovered a pattern within literature focusing on feedback within teams after conducting a meta-analysis. These results suggested that studies focused less on the question of whether feedback improves team effectiveness and more on the question of which feedback condition is effective in teams. This distinction is important because it suggests a shift from attempting to prove the effectiveness of feedback in all conditions to demonstrating the effectiveness of feedback in certain situations.

Table 2 - Elements of feedback that could be considered, adapted from Gabelica et al. (2012)

Characteristics	
	Performance
	Process
Type of dependent variables the feedback interventions targeted	
	Team or individual outcomes
	Emergent states (cognitive or affective)
	Team processes
The study design	
	Laboratory study
	Field study
The team type	
	Knowledge work teams
	Physical work teams
Individual or team situation and characteristics	
Perception of feedback	
Processing of feedback	

Another possible explanation as to why there are conflicting results when studying the feedback process is that the feedback process contains several failure modes that have not been considered consistently in previous research (Figure 1). The recipient must perceive, accept, be inclined to respond to, and respond intentionally to feedback to modify behavior (Ilgen et al., 1979). The team member may not receive the feedback accurately because of noise in the communication process, leading to missed feedback. Or, the team member may consciously reject the feedback, based on mistrust of the feedback source or understanding the current context more accurately than the feedback source ("Good suggestion at other times, but that does not apply right now…"). The team member may accept the validity of the feedback, but decide not to respond, perhaps based on current priorities or available cognitive resources. Finally, the team member may take no action because the feedback is not relevant to the person's current goal. Shute (2008) summarizes these conditions with the note that feedback, specifically formative feedback, requires motive, means, and opportunity

to change the recipient's behavior or take action. In the conflicting studies reviewed above, when feedback of a certain kind did not work, it would have been useful of previous researchers had described the failure mode per a process such as Figure 1, because then results could be categorized more systematically, e.g., in a context in which team members have the ability and motivation to respond, one type of feedback was effective, but in a context of heavy cognitive load with lower ability to respond, another type of feedback might be effective.

A component of feedback that influences its effectiveness is the timing of feedback (Walton et al., 2014). The timing of the feedback, such as delayed, immediate, or concurrent, is an important step in the feedback process that can positively or negatively influence the recipient's performance (Corbett & Anderson, 2001).

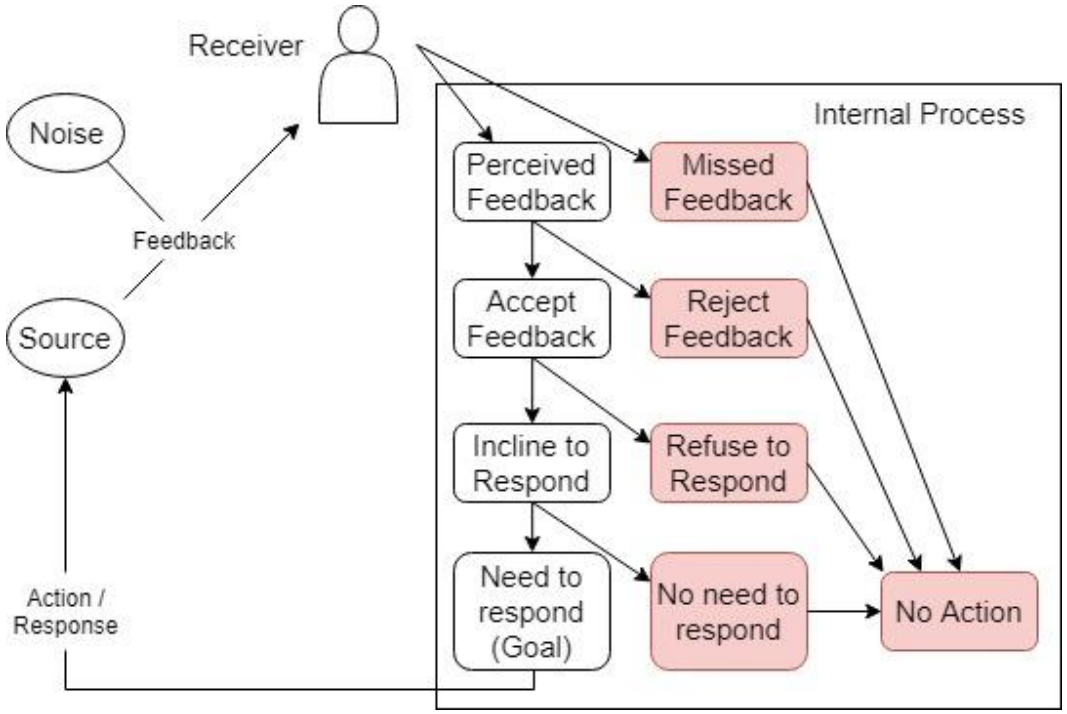


Figure 1 – Feedback process for an individual. Adapted from Ilgen et al. (1979) and Shannon (1948).

Several studies focus on feedback timing, but there is limited agreement among researchers on the most effective approach (Butler, Karpicke, & Roediger, 2007; Corbett & Anderson, 2001; Gabelica et al., 2012; Walsh, Ling, Wang, & Carnahan, 2009). Researchers have developed guidelines on the timing of feedback, despite the limited agreement among researchers. For example, guidelines presented by Salas, Cannon-Bowers, and Johnston (1997) suggest that it is important that the team leader knows how to provide periodic updates to the team regarding important aspects of team performance. Researchers have presented conclusions that generally support this guideline. In one example, Marks, Zaccaro, and Mathieu (2000) provided evidence that supports the guideline. The author's findings suggest that having leaders regularly provide teams with an update, or report, on the current goals will allow teams to effectively and accurately modify their mental model of a task. In another example, Entin and Serfaty (1999) demonstrated that periodically providing situational updates (every three minutes) can help improve team performance, though the experimental design was not intended to study the effect of the periodic situational update provided by the leader. This periodic feedback, delivered by the team leader, was given in combination with instructions on how to identify stress within the environment, team, and the team's members. As mentioned before, Salas et al. (1997) suggest that team leaders should provide performance updates to the team members. In contrast to a human team leader, an ITTS can provide feedback to a team at a near-continuous rate. Can an ITTS that provides continuous team, individual, or I&T performance updates produce similar positive effects on team performance?

Performance updates can be automated if an ITTS has enough information about the current state of the team. Providing enough information to the ITTS is not difficult if the

team conducts their task within a virtual environment, and the tutor has access to the environment's state (Devasani, Gilbert, Shetty, Ramaswamy, & Blessing, 2011; Gilbert, Devasani, Kodavali, & Blessing, 2011). Access to multiple data sources about team members' behavior allows the ITTS to provide feedback in a way that is difficult, or impossible, for a human tutor to accomplish in real-time manually. For example, imagine there is a team of three conducting a task that requires some degree of collaboration. While conducting this task, the team must follow a set of rules. During that task, an ITTS may count an error every time a team member breaks a rule and update the team on how many errors they committed during that task. The ITTS developed in this current study will provide performance feedback regarding important components of performance to each team. The feedback will display metrics based on individual, team, or I&T performance. Similar feedback methods were in other fields, such as the medical field (Fralick & Kesselheim, 2017).

As previously mentioned, the common purpose of delivering feedback is to provoke a change in behavior. The performance feedback delivered by the ITTS in this current work displays metrics about individual, team, or I&T performance. The researcher anticipated that teams would adapt their behavior to improve team performance. The next section briefly discusses the complexities of team adaptation.

Team Adaptation

Team adaptability is a trait that is relevant to many teams in various situations across industries (Salas, Sims, & Burke, 2005). Researchers define team adaptation as changes that occur within a team's procedure or method as a result of an unexpected intervention (Maynard, Kennedy, & Sommer, 2015). Feedback is a mechanism that is a part of the adaptation process and has the potential to increase team adaptation (Burke, Stagl, Salas,

Pierce, & Kendall, 2006). The author of this current work argues that it is important to review the complexities of team adaptation because it further illustrates the complexities of understanding the influence feedback has on teams.

Providing information to a team that relates to that team's performance in real-time may support a team's adaptation ability. Team adaptation is an active area of research opportunities (Gilson, Maynard, Jones Young, Vartiainen, & Hakonen, 2015). The increased activity may be because teams have a high potential for adaptation, which is at the core of team effectiveness (Burke et al., 2006). Team adaptation is a complex process (Sander, van Doorn, van der Pal, & Zijlstra, 2015), and the definition within literature has not been consistent (Maynard et al., 2015). Though there is a lack of agreement on the exact definition of team adaptation, researchers generally agree on the components of team adaptation models. Models of team adaptation generally have three components (Figure 2): Input, team process, and output (Ilgen, Hollenbeck, Johnson, & Jundt, 2005). Definitions of adaptation have focused on the Input (Randall, Resick, & DeChurch, 2011), process (LePine, 2003), or the outcome (Shoss, Witt, & Vera, 2012) component of team adaptation. These components are discussed below in more detail in the following sections.

Team Adaptation: Input

The input component of team adaptation refers to the diverse characteristics that describe a team. These characteristics are commonly used to choose the members assigned to a team. Though there are several characteristics that can describe a team (Bonner et al., 2014), the characteristics used by researchers to describe teams are generally categorized as Team, Individual, and Work Design (Burke et al., 2006; Entin & Serfaty, 1999; Grant & Parker, 2009; Maynard et al., 2015; Mickan & Rodger, 2000). Work Design has been defined as characteristics that describe how a job is structured and how this job can impact the

individual, group, and organization outcomes (Grant & Parker, 2009). It is important to note that this is different from Job Design because Job Design focuses more on the Job itself, whereas Work Design focuses on the job and outside elements that will influence the job. Work Design, by definition, includes many different characteristics that can describe the job and the elements that have an impact on the job. Out of the many characteristics that teams possess, it is important to understand which team characteristics connect to a team's effectiveness.

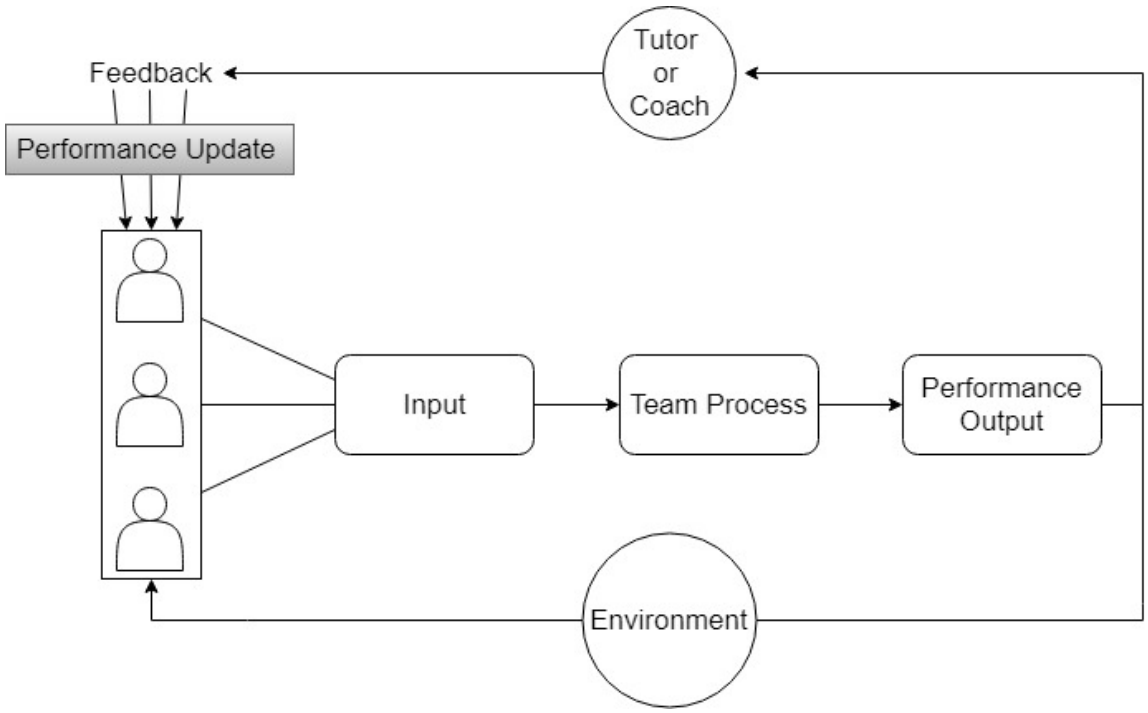


Figure 2 –Feedback within the adaptation process. Based on models from Burke et al. (2006), Christian et al. (2017), Entin and Serfaty (1999), and Maynard et al. (2015)

Multiple studies have sought to identify characteristics associated with effective teams. Generally, researchers have different names for characteristics that influence team effectiveness, and there is at least one characteristic that can be categorized as Team, Individual, or Work Design. Campion, Medsker, and Higgs (1993) concluded that there are five categories, or themes, that are related to team effectiveness: Job Design,

Interdependence, Composition, Context, and Process. These characteristics were also shown to be generalizable (Campion, Papper, & Medsker, 1996). Job design and context are related to work design characteristics, whereas interdependence, composition, and process are related to team characteristics. Mickan and Rodger (2000), with a focus on healthcare teams, conducted a literature review of effective teams and found that the categories for each characteristic, connected to effective teams, were Organizational Structure, Team Processes, and Individual Contribution. All three of these elements are related to Work Design, team, and individual characteristics, respectively. Bannister, Wickenheiser, and Keegan (2014) argue that the key elements of highly effective teams are Purpose, Openness, and Roles and Skills. Purpose can be categorized as a Work Design characteristic, Openness as a team characteristic, and Roles and Skills and individual characteristics. The characteristics described in each study previously mentioned can be categorized as Team, Individual, and Work Design category. This suggests that when studying adaptation, it is important to gather data related to the team and the task at hand. Gathering this data could provide insight into how the input of a team influences the effectiveness.

As previously mentioned, researchers are focusing on verifying the feedback condition that is most effective (Gabelica et al., 2012). The effectiveness of feedback can be dependent on the input component of a team, but it is unclear if configuration characteristics, such as diversity, influence the effectiveness of team feedback. For example, Devlin, Flynn, and Riggs (2018) presented evidence that highly adaptable teams had low levels of non-job related diversity (i.e., similar personalities) but it is unclear how low levels of diversity will influence how the team perceives or receives various types of feedback. The author recorded information about the participants that conducted this study since it is important to collecting

information about the team. The next section will briefly discuss the process component of team adaptation.

Team Adaptation: Process

The process, or throughput, of adaptation refers to the way in which a team works together to complete the task at hand. A team's process will vary from task to task because each task will potentially require different skill sets to complete it successfully. The variation in a team's process from task to task has resulted in various models of how the team process occurs. For example, Entin and Serfaty (1999) present a team adaptation model that contains a team processes component that only includes taskwork and teamwork. Maynard et al. (2015) present a model based on an examination of team adaptation literature over fifteen years (1998-2013). The process component in the model describes an iterative relationship between team adaptation process and team mediators (e.g., communication or coordination). Burke et al. (2006) presented a more in-depth model of team adaptation. The process component of this model consists of an adaptive cycle and emergent states that influence this adaptive process. Overall, several studies have focused on the process component within team adaptation, but more work is needed to develop a consistent model.

While the process component is important in team adaptation, this work focuses on the output metrics. The researcher attempted to provide insight into how feedback influences the process component by analyzing qualitative and quantitative metrics, such as team strategies. The author of this current work hopes that the data collected can be used to guide future studies to extend the understanding of the process component of team adaptation. The next section will briefly discuss the output component of team adaptation.

Team Adaptation: Output

The output component of the team adaptation process refers to the outputs that follow the process component. The output component has been defined as the results, or constructs, that are produced as a by-product of a team's adaptation process (Maynard et al., 2015, p. 654). As suggested by definition, the output is connected to the process of a team. In other words, if the team process is positive, then it will positively influence the team output, and if the team process is negative, then it will negatively influence the team output. This logic also suggests that positive changes to the team process will result in positive changes to team output.

The goal of displaying performance feedback to participants is to invoke a change in team behavior or team process to improve team outcomes. An important component of feedback is the focus level (Walton et al., 2014). According to Hattie and Timperley (2007), there are four levels of feedback. First, it can focus on whether the task at hand was done correctly or not. Second, it can focus on the process utilized to accomplish a task. Third, it can focus on an individual's, or team's, compacity to self-evaluate or self-regulate. Lastly, it can personally focus on the entity (an individual or a team) of interest. For example, the feedback content may contain information that says, "your team did well" or "you are a good team member." Determining the focus level of the feedback given to a team is different from determining the focus level of the feedback given to an individual because each focus level can contain important information about an individual or the team. No matter what the focus level may be, the goal of the performance feedback displayed to a team is to encourage the team to improve its team process in a way that results in an improved outcome. This goal is especially prominent during training.

Training is commonly used to help teams improve overall team process, which in turn improves the team's overall outcome. The feedback given during training is important in any domain. What should the feedback focus on during training? The four levels of feedback (i.e., Task, Process, Self-regulation, and Self) can benefit learning if done effectivity. For example, feedback focused on the task is particularly effective when the entity (individual or team) receiving the feedback is learning a new task or skill (Hattie & Timperley, 2007). However, the benefit gained from the same task-focused feedback can be reduced if the information displayed by the feedback is misperceived by the entity receiving the feedback (Howie, Sy, Ford, & Vicente, 2000). Schmutz and Manser (2013) conducted a literature review with a focus on clinical performance and concluded that implementing training that focuses on team process behaviors will influence outcomes. However, the authors also concluded that since the primary method used to measure team process is observation, they could only assume that interventions did influence the process, which in turn influenced the team outcome. It is important to understand how team performance is measured. The next section briefly discusses measuring performance and performance adaptation.

Measuring Performance and Performance Adaptation

Measuring team performance is not a trivial task. How many metrics are needed to measure team performance? One metric can be used to measure a team's performance, but it is unlikely that that metric will provide a deep insight into the basis of team performance because team performance is multi-layered and requires different metrics to study it (Salas, Rosen, Burke, Nicholson, & Howse, 2007). Several considerations need to be addressed when deciding which metrics to use when measuring team performance. Rosen et al. (2008) provided eleven suggestions on how to develop and choose metrics to measure team

performance in simulation-based training. This current work will use the guidelines provided by Rosen et al. (2008) to develop metrics for team performance.

It is important to note that discovering measurements that accurately represent team adaptation is not always straightforward (Burke et al., 2006). LePine (2003) concluded that the measurements used to predict team performance in an unexpected situation are different from those used to measure team performance in a routine situation. This conclusion helps explain why some teams are ideal on paper, but they do not perform well in practice. This disconnect between the anticipated team outcome and the correct measurements for team adaptation may lead to unexpected team performance. LePine (2003) also suggests that cognitive ability, dependability, achievement, and openness are all critical elements of adaptation. This suggestion implies that multiple metrics can be used to potentially measure team adaptation despite the disconnect between expected team performance and team performance in unexpected and expected situations.

Since multiple metrics are needed to measure adaptation, it is important to state the current author's perspective on team adaptation clearly. The two main perspectives of adaptation are domain-specific and domain-general. The domain-specific perspective focuses on key elements, such as skills or proficiencies, that relate to adaptation inside a domain. The domain-general perspective views adaptation as a general concept that can occur in different situations (Baard, Rench, & Kozlowski, 2014). Neither perspective of performance adaptation is preferred over the other, but the conclusions made will differ depending on the perspective. The author of this current work used the domain-general perspective when analyzing the data because the author seeks to present a conclusion that holds across different domains. The author believes that having conclusions that hold across domains will help

guide the development of the feedback component of any ITTS in any domain. There are many training methods that an ITTS can implement to help a team achieve the desired outcome. The next section will briefly discuss various team training methods.

Team Training Methods

Training is a common method used to improve teamwork and team performance and has been shown to increase both across various domains (McEwan, Ruissen, Eys, Zumbo, & Beauchamp, 2017). Also, Entin and Serfaty (1999) demonstrated that teams could be trained to improve adaptation. Authors have employed a variety of team training methods, such as cross-training (Blickensderfer, Cannon-Bowers, & Salas, 1998; Volpe, Cannon-Bowers, Salas, & Spector, 1996), guided team self-correction (Smith-Jentsch, Cannon-Bowers, Tannenbaum, Tannenbaum, & Salas, 2008), team coordination and adaptation training (Cannon-Bowers & Salas, 1998; Entin & Serfaty, 1999), assertiveness training (Smith-Jentsch et al., 1996), team-leader training (Cannon-Bowers & Salas, 1998), team-interaction training (Marks et al., 2000), procedural training (Hockey, Sauer, & Wastell, 2007), and perturbation training (Gorman, Amazeen, & Cooke, 2010). All the training method mentioned were previously implemented and have the potential to increase team performance. What training method will improve team adaptation and what should be the focus of training when developing it to improve a team's adaption? Sander et al. (2015) concluded that an adaptive team needs to be trained in updating one another with new information regarding any change that has occurred for the team to be successful. In other words, a team needs to learn how to correct their behavior when needed. This objective is like the goal of self-correction training.

The idea behind the guided team self-correction training method is to allow a team to identify and correct their problems (Smith-Jentsch et al., 2008). In this training method, there

is a facilitator, or a leader, that asks the team to describe their performance and describe areas that need improvement. Similar to the team self-correct training method, the performance feedback method described in this current work relies on the idea that teams can self-correct. Instead of assigning one member as a facilitator, or a leader, that guides the team in the self-correction process, the author of this current work believe that a team will use the performance feedback provided by an ITS to self-correct in real-time.

Developing a virtual training environment that utilizes the guided team self-correct method and implements an ITTS to guide a team has the potential to be the next step in the continuing evolution of team training. A critical question regarding ITTSs focuses on the characteristics of an effective tutor. Having an accurate model of the team will support the benefits provided by an ITTS. There are ongoing challenges, such as the methods used, to modeling individuals and teams (Sottilare & Boyce, 2016). The current state of team models are complex and, as a result, hinders researchers ability to develop important research questions that can be tested (Schmutz & Manser, 2013), making it difficult to develop effective ITTSs. However, the author of this current work believes that it is important to examine the characteristics of an effective human tutor to determine if researchers can apply these characteristics to an ITTS. The next section will briefly discuss the characteristics of effective tutors.

Characteristics of Effective Tutors

It is difficult to develop effective ITTSs due to the minimal agreement among researchers on the model for teams. Perhaps insights into characteristics of an effective ITTS can be gained from characteristics of effective human tutors. When comparing human tutors to intelligent tutoring systems, it is important to question if an intelligent tutoring system should mimic the characteristics of a human tutor (du Boulay & Luckin, 2001). A previous

study discovered characteristics of highly effective human tutors for individuals. These characteristics include Intelligent, Nurturant, Socratic, Progressive, Indirect, Reflective, and Encouraging (INSPIRE) (Lepper, Drake, & O'Donnell-Johnson, 1997; Lepper & Woolverton, 2002). The characteristic Indirect is most applicable to the present research.

Indirect – A Characteristic of Effective Human Tutors

Effective tutors are indirect with learners when they provide feedback without identifying the student's error (Lepper & Woolverton, 2002). The goal is to encourage students to identify their errors. This approach of having students exert effort to identify their errors and recall previous information to correct their actions has positive benefits to students' learning (Storm, Bjork, & Storm, 2010). However, the effort needed to correct their actions should not far exceed the student's capability. The balance of effort and difficulty is known as "desirable difficulty" (Dobson, 2011). The author of this current work suggests that this principle will apply to teams as well. The question that needs to be answered is what information will prompt this recall effort in a way that increases team and individual performance. The author suggests that providing a team with indirect information about incorrect choices made will improve its team performance.

Also, effective tutors use positive feedback sparingly and in a way that helps reinforce their learning (Lepper & Woolverton, 2002). The spare use of positive feedback suggests that there is a need for some amount of positive feedback, even if it is a small amount. Therefore, the author of this current work suggests that offering teams information about the correct actions taken will increase performance. Combining the above two suggestions about negative and positive feedback yields this hypothesis:

H_1: A team that receives feedback that provides information about both team and individual errors and correct items collected will perform better than teams that receive

error and correct items collected feedback that provides only team or only individual information.

The somewhat general hypothesis must be justified within the specific context of the TMET team task and its feedback, which the researcher described in Chapter 3.

In summary, this brief exploration of the literature reveals a wide range of complex challenges that arise when developing an ITTS. Feedback is one of the top five attributes that are important to students, or tutored (McAndrew, Mucciolo, & Jahangiri, 2016) and is critical to learning. As a result, the implementation of feedback via an ITTS must be considered carefully. Lepper and Woolverton (2002) argue that an influential tutor should provide constant feedback, or information, to the person receiving instruction. This research will explore whether this idea applies to an ITTS.

CHAPTER 3. METHODS

This chapter describes the methods used to collect performance data from each team. The next section briefly describes the background of the task, called the Team Multiple Errands Test, the teams conducted during the study. The following sections describe the experimental design, the independent and dependent variables used in the experiment, how the individual and team score are calculated, the surveys distributed to the participants, and the experimental procedure.

Overview of the Team Multiple Errands Test (TMET)

The current work uses a task presented by Walton, Bonner et al. (2015) called the Team Multiple Errands Test (TMET). The TMET is based on an ecologically valid test (Alderman, Burgess, Knight, & Henman, 2003) presented originally by Shallice and Burgess (1991), called the Multiple Errands Test (MET). The purpose of the original MET was to examine the capability of patients with injuries to their prefrontal structures by having them carry out different cognitive tasks.

The original MET was a shopping mall task in which the participants were asked to complete eight tasks within an unfamiliar area. The first six tasks were simple (e.g., purchase a packet of throat lozenges). The seventh task was for participants to be in a location 15 minutes after the beginning of the task. The eighth (and most complex) task required participants to obtain four pieces of information during the task and write each down on a postcard. The information included 1) the name of the store that might have the most expensive item, 2) the cost of one pound of tomatoes, 3) the name of the coldest location in Britain the previous day, and 4) the exchange rate for a French franc the previous day.

Shallice and Burgess (1991) added further instructions to the task to increase the complexity; the participants had to adhere to a set of rules while conducting the task (Table 3).

Table 3 - Rules that participants had to follow during the original Multiple Errands Test.

1. Spend as little money as possible while completing the task as quickly as possible.
2. Do not enter a store unless you plan to buy something.
3. Tell a researcher what you bought when you leave a store.
4. You can only use items bought on the street to assist you.
5. You may complete the tasks in any order.

Recently, a growing number of researchers have modified the MET to fit various contexts. Examples include the MET – Hospital Version (MET-HV) (Knight, Alderman, & Burgess, 2002), MET – Simplified Version (MET-SV) (Alderman et al., 2003), Virtual MET (VMET) (Cipresso et al., 2014; Rand, Basha-Abu Rukan, Weiss, & Katz, 2009), MET-Home (Burns et al., 2019), MET Contextualized Version (MET-CV) (Valls-Serrano, Verdejo-García, Noël, & Caracuel, 2018), modified MET for intellectual disabilities (mMET-ID) (Steverson, Adlam, & Langdon, 2017), and the youth MET (yMET) (Hanberg, MacKenzie, & Versnel, 2019). The MET-SV and VMET researchers focused specifically on validating their versions of the MET as equally applicable cognitive assessment tools. The MET-Home researchers adapted the task to test older adults within their own home after they have had a stroke. The MET-CV, mMET-ID, and yMET were adapted for specific populations: people with substance dependence, people with intellectual disabilities, and youth. Other researchers have adapted the MET as a basis task for other unrelated research, e.g., to study whether people would trust robotic peacekeepers, James Bliss and colleagues created a virtual village in which participants completed the MET while being periodically interrupted by robots who demanded that the participants give up a personal item for safety's sake (Long, Karpinsky, & Bliss, 2017).

This current work uses a modification of the MET for teams called the TMET. The TMET was first implemented in a study described in Walton, Gilbert, et al. (2015). The current TMET implementation includes scores (individual and team), deception, and a new data collection method. It is worth noting that these TMET implementations were not designed with the goal of producing a validated clinical diagnostic instrument of executive function, like some other modifications of the MET, but rather with the goal of providing a platform for teamwork task that would pose a reasonable challenge for teams that could be scaled in difficulty if needed for assessment purposes. The author suspected that by keeping rules similar to the original MET within the TMET, individual teammates would face cognitive challenges similarly to participants in the original MET, and by requiring team coordination and communication to succeed at the TMET, participants would be challenged at a team level as well. In this specific research, the purpose of using the TMET was to provide a platform that allowed the researcher to evaluate the impact of feedback interventions on team performance.

It is important to note that the development and implementation of TMET was completed in previous work (Walton, 2015) and is not a major contribution of this current work. The study presented by Walton (2015) primarily focused on how the privacy and audience dimensions of feedback influenced team performance. This current work focuses on how the assessment dimension of feedback influences performance, i.e., displaying feedback that is based on an assessment of individual metrics, team metrics, or both (I&T metrics). A brief overview of the TMET is described below.

Description of the TMET

The objective of the task in the TMET is for teams to collect all the items on their shopping lists as quickly as possible. The current research implemented TMET for teams of

three participants. There are two list types for each team, three different individual lists with six items each and a team list with 18 items (APPENDIX F). The items on the team list are not assigned to a specific player. In other words, any team member can collect any item on the team list. The items on an individual list (e.g., Player 2's list) are assigned to a specific player, and that player is the only one who should buy that item. For example, imagine that Player 2 has Party Hat on her individual list. This means that Player 1 and Player 3 are not responsible for buying the Party Hat. The items on the team list do not appear on any individual list. The items on the individual lists do not appear on the team list or the other individual lists. For example, if Player 2 has Party Hat on her list, then Party Hat will not appear on Player 1's list, Player 3's list, or the team list. Each player had a physical hardcopy of the team list and his or her own individual list during each session, and the lists differed across different sessions within the same team. The players were not allowed to cross off items on the physical copy of the list because they could view the items they had bought in the virtual mall software they used. Furthermore, the inability to cross-off items during the task increased the task difficulty. Shallice and Burgess (1991) did not indicate if similar restrictions were placed on their participants in the original MET.

Like the original MET, the participants had several rules they had to follow while completing the task (Table 4). Rules 1 and 2 are like rules found in the original MET (Table 3), and Rule 6 was created as a timing constraint similar the location-based timing task of the original MET that also added an element of teamwork. The rest of the TMET rules were added to increase complexity. For example, imagine Player 1 enters a store and finds multiple team items. In order to follow the rules, Player 1 must buy an item (Rule 2), and she can only buy one team item from the store (Rule 3). The software environment counted an

error each time a player broke a rule. The task ended when each team member signaled via the environment that each had completed the task or when the game timer reached zero seconds. The players could signal that they had completed the session even if some of the items were not collected. The following section describes the study procedure.

Table 4 - Rules that participants had to follow during each session

1. Do not spend over your allotted amount of money.
2. If you enter a store, you must buy something.
3. You must buy only 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 individual or team list.
6. Meet up with your teammates at the fountain when the timer is at 0:30 (30 seconds remaining) or earlier, and before the game has ended.
7. Signal when you are finished or before time has run out.

Deception to Promote Participant Engagement

The researcher deceived each team into motivating participants to engage seriously with the research task. Each team was told that they needed to perform better than at least 50% of the teams that came before them to receive full compensation. Otherwise, they would only receive half of their compensation. In reality, each team received full compensation. The Institutional Review Board approved this deception and included in the study because of feedback received from pilot studies (APPENDIX I). Pilot participants indicated they did not have a significant motivation to do well in the study. There was no indication that any of the participants were aware of this deception.

Experimental Design

This study had one independent variable (feedback content) with four levels. All participants in the study experienced each feedback condition. The four levels were: No Feedback, Individual Feedback, Team Feedback, and I&T Feedback (Table 5). The feedback

conditions always displayed information regarding the correct items collected. The feedback conditions displayed information regarding rules broken if at least one error was committed. No feedback was displayed to teams in the No Feedback condition. Feedback displayed in the Individual Feedback condition only contained performance metrics at the individual level (Figure 3 and Figure 6). Feedback displayed in the Team Feedback condition only contained performance metrics at the team level (Figure 4 and Figure 7). Feedback displayed in the I&T condition contained performance metrics for both the individual and team level (Figure 5 and Figure 8). The individual and team scores were displayed in the top corner of each participant's screen in every feedback condition (Figure 9). The calculation for the individual and team score is described below. It is important to note that each member of the team has the same performance metrics feedback.

Table 5 - Description of the different feedback conditions. In each condition, all feedback information was displayed to every member of the team (i.e., every member could see information regarding their own performance and their teammate's performance).

Feedback Condition	Description
A. No Feedback	No feedback was given to the participants during the session.
B. Individual Feedback	Feedback displayed in this condition only contained performance information at the individual level (Figure 3 and Figure 6).
C. Team Feedback	Feedback displayed in this condition only contained performance information at the team level (Figure 4 and Figure 7).
D. Individual and Team Feedback (I&T)	Feedback displayed in this condition contained performance information at the team and individual level (Figure 5 and Figure 8).

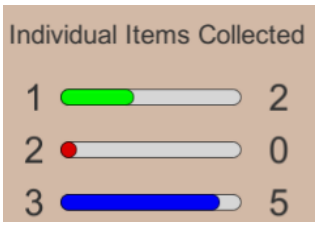


Figure 3 - Correct item report (Individual feedback condition), which shows the items collected by each of the three players from their respective shopping lists. Each individual list contained six items.

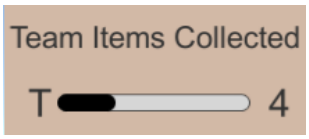


Figure 4 - Correct item report (Team feedback condition), which shows the number of items from the team shopping list collected by the entire team. This count does not include items from the individual shopping lists. The team list contained 18 items.

Figure 5 - Correct item report (I&T feedback condition), which shows the team list items collected (T) out of 18 and the individual list items collected by each player (out of six).

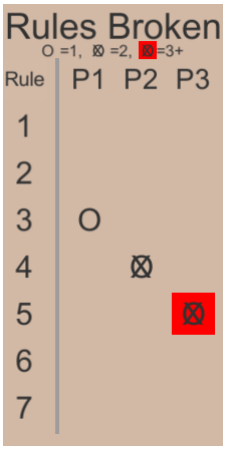


Figure 6 - Error report (Individual feedback condition). In this case, P1 has broken Rule 3 once, P2 has broken Rule 4 two times, and P3 has broken Rule 5 three or more times.

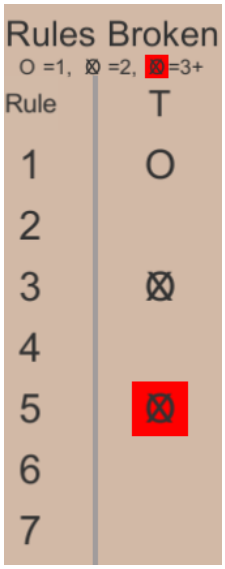


Figure 7 - Error report (Team feedback condition). In this case, the team overall broke Rule 1 one time, Rule 3 two times, and Rule 5 three or more times. This team error count is the sum of individual players' errors.

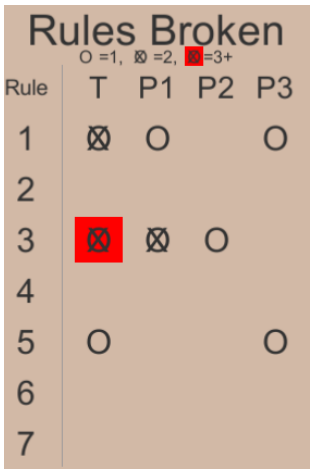


Figure 8 - Error report (I&T condition). In this case, the team broke Rule 1 two times, Rule 3 three or more times, and Rule 5 one time. These errors are also shown in the form of P1 and P3 breaking Rule 1 once, P1 and P2 breaking Rule 3 twice and once, respectively, and P3 breaking Rule 5 once. The team counts are the sum of the players' counts.

Figure 9 - Individual and team score display. Both the individual and team scores were visible for each feedback condition.

Design Justification for Feedback Displays

The user interfaces displayed to users (Figure 3 to Figure 8) were guided by display design principles presented by Wickens, Lee, Liu, and Becker (2003). The researcher utilized the following design principles: placing knowledge in the world, proximity compatibility principle, and minimizing the effort needed to access information. See Figure 14 for an example of the feedback displayed to the participants.

The knowledge in the world principle suggests that information be placed in the environment to reduce the extent to which users depend on their working memory or long

term memory (Norman, 1988). During each session, each team was given an individual score and team score, which are discussed in greater detail later in this chapter. The scores were based on multiple components, including errors committed and correct items collected. Using the knowledge in the world principle, the researcher designed an interface that displayed information regarding the errors committed (i.e., rules broken) and the correct items collected, to reduce the amount of information that the members needed to store in memory.

The proximity compatibility principle suggests that information from multiple sources that relate to the same task (e.g., collecting all the items on the shopping list) should be displayed close to one another (Wickens & Carswell, 1995). Using this principle, the researcher grouped the metrics into two panels of information, the correct items collected, and the errors committed (i.e., the rules broken). Depending on the condition, the display showed information related to Individual, Team, or I&T information performance metrics. The information related to the correct items collected was placed near the center of the screen so that participants could easily access that information while inside a store. The information related to errors was placed on the far left side of the screen to increase the distance between the information related to the correct items collected, thus maintaining the proximity principle.

The minimizing effort to access information principle suggests that a display should minimize the time and effort needed to find the "correct" or relevant information. Using this principle, the researcher displayed the feedback information continuously or "persistently" to reduce the amount of time and effort needed to locate the information. Also, the team measures were labeled "T," while the player measures were numbered (1, 2, 3), to minimize the effort required to visually distinguish the team measures and individual measures by

mixing the letter T with the numerals. Note that the minimizing effort to access information principle cannot be used to justify adding information on a display repeatedly; it must be balanced with overloading the user with information, as measured by the Hick-Hyman law (Hick, 1952; Hyman, 1953), which notes that increasing the user's choices of information sources can increase decision making time logarithmically. In the context of TMET, the addition of the single team metric (T) to the three player metrics (1, 2, 3) was thought to be a justifiable increase in useful information without a dramatic increase in total information content.

Persistent Feedback

As described in Chapter 2, feedback has three content dimensions, as well as many other characteristics, such as timing, format, and style of presentation. In the current research, because the goal is to explore the assessment dimension of feedback, feedback is provided continuously using the displays of metrics shown in Figure 3 through Figure 8. While the previous TMET study (Walton, 2015; Walton, Bonner, et al., 2015) and other ITSs (Graesser, Wiemer-Hastings, Wiemer-Hastings, & Kreuz, 1999; Koedinger, Anderson, Hadley, & Mark, 1997) have triggered the presentation of feedback based on specific actions taken by the learner and then removed the feedback after a specific time duration or new event trigger, this persistent feedback approach was taken in order to ensure that the task state information was available to team members at any time. The persistent feedback approach echoes the principles of ubiquitous computing (Weiser, 1993), offering continuous unobtrusive feedback to be used at will rather than intrusive feedback that might distract team members from their task at hand.

Expected Influence of Feedback Conditions on Performance Metrics

Researchers have identified five core components (i.e., the "Big Five") that contribute to successful teamwork: team leadership, mutual performance monitoring, backup behavior, adaptability, and team orientation (Salas et al., 2005). Mutual performance monitoring is relevant to this current work and was defined in that research as the team members' ability to track their own performance and team performance to ensure the team is proceeding appropriately. Studies have shown that teams who engage in performance monitoring, as well as adaptation and leadership, can improve performance (Serfaty & Entin, 1997).

The researcher anticipated that providing performance information (i.e., correct items collected, and errors committed) would support the mutual performance monitoring component of effective teamwork and improve performance. Specifically, the researcher anticipated that the condition with the most information (i.e., the I&T condition) would yield the highest performance because team members would have a more complete understanding of their own performance and the team's performance.

Additionally, the researcher predicted that the information provided in the feedback conditions would give teams, as noted by Shute (2008), the motive, means, and opportunity needed to take the necessary steps to change their behavior. For example, the feedback regarding the rules broken was actionable information that allowed participants to know what rules the team, or an individual, had broken so they could focus on not breaking those rules in the future. It was anticipated that the I&T condition, with the most information, would best reduce the chance of participants taking no action because of a failure in the internal process (Figure 1). Consequently, it was expected that teams in the I&T condition would yield the highest performance, which led to the development of H_1.

Dependent Variables

There were several dependent variables collected during each session and calculated post hoc. Information regarding each variable (e.g., metrics, units, collection frequency) is displayed in Table 6. The main dependent variables were Performance, Items Collected, Errors, and Time Remaining. A higher value of Performance, Items Collected, and Time Remaining indicates a higher performing team, while a lower number of Errors indicates a higher-performing team.

Determining the Dominant Task Strategy

Walton (2015) observed two strategies implemented by participants: Go Together and Go Alone. The author of this current work used the position proximity of the avatars within the virtual environment to assign the dominant strategy used by each team (see Figure 10 and Figure 11). The author calculated the centroid, or center of mass, and drew a circle with the centroid at the center of that circle. The border of the circle is called the Strategy Border. A pilot study was used to determine the radius (R) of the circle for the Strategy Boarder. The researcher used the data collected from the pilot study to determine the radius that correctly identifies teams that used the different strategies. Generally, teams that used Go Together typically stayed close or within the line of sight of each other, while teams that used Go Alone moved independently throughout the task. The position of each avatar within the virtual environment determined if the teams were using the Go Together or Go Alone strategy (or neither). If all three avatars were within the Strategy Border (i.e., inside the circle) in an instant, then the group was labeled as using the Go Together strategy. If all three avatars were outside the Strategy Border (i.e., outside the circle) in an instant, then the group was labeled as using the Go Alone strategy. If some of the members were outside the Strategy Border and the other members were inside the Strategy Border, then the group was

labeled as using a Mix Strategy. The author summed each strategy instance (i.e., Go Together, Go Alone, and Mix) and the strategy with the highest instance count was assigned to the team as the dominant strategy during a specific session. It is important to note that no team had a dominant strategy of Mix.

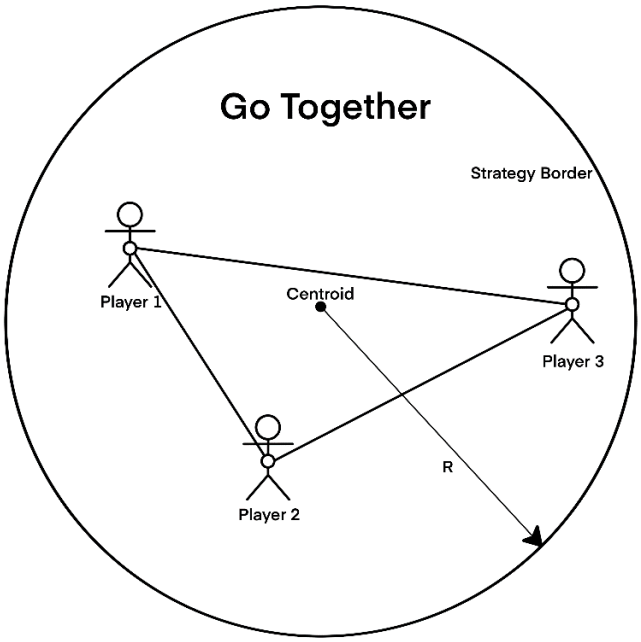


Figure 10 - Go together strategy calculation by position.

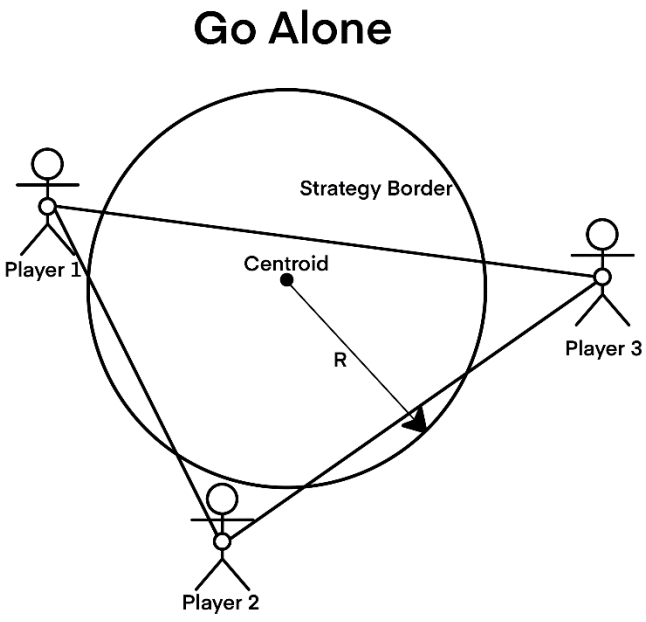


Figure 11 - Go alone strategy calculation by position.

Table 6 - Description of the dependent variables collected and how each was collected. The word experiment in the "Collection / Calculation Frequency" column refers to the time after the experiment for a team concluded. The word session refers to the time teams are conducting the task (i.e., the TMET).

Dependent Variable	**Metric**	**Units**	**Collection / Calculation Frequency**
Main Metrics			
Performance	Team Score	Weighted Sum	Collected during each session (4 times per second).
	Individual Score	Weighted Sum	Collected during each session (4 times per second).
Items Collected	Correct Items Collected	Count	Checked every time an item is collected during each session. Increased by 1 if correct item was collected.
	Incorrect Items Collected	Count	Checked every time an item is collected during each session. Increased by 1 if an incorrect item was collected.
Errors	Unique Error Count	Count	Behaviors (e.g., entering a store) were checked throughout the session. Increased by 1 every time a new rule was broken.
	Total Error Count	Count	Behaviors (e.g., entering a store) were checked throughout the session. Increased by 1 every time a rule was broken.
Time	Time Remaining	Seconds	Collected once the participants signaled they completed the session or the timer reached zero.
Other Metrics			
Position	Unity Position Coordinate	Meters	Collected during each session (4 times per second).
	Unity Rotation Coordinate	Meters	Collected during each session (4 times per second).
Dominant Task Strategy: Go together (Figure 10) and Go alone (Figure 11)	Position Proximity	Meters	Calculated at the end of an experiment. The position of each avatar was collected approximately four times a second.
Distance Traveled	Position coordinates	Meters	Calculated at the end of an experiment.
Collection Time Per Item	Correct Items and Duration	Duration divided by Count of correct items collected.	Calculated at the end of an experiment.
	Incorrect Items, Correct Items, and Duration	Duration divided by Count of correct items collected.	Calculated at the end of an experiment.

Team and Individual Score Calculations

The author of the current work used the studies presented by von Ahn & Dabbish (2008) and Zapata-Rivera & Katz (2014) to guide score development. Four components were used to calculate the scores: correct items collected, incorrect items collected, time remaining, and the unique errors broken. The scores were calculated using a weighted sum where each of the four components was weighted (W) equally (i.e., 0.25). Table 7 and Table 9 presents how each component of the individual score and team score, respectively, were calculated. Table 8 and Table 10 provide clarification of how values were calculated in the individual score and team score, respectively. Each participant saw their individual score and team score as a multiple of 100 (see Table 7 and Table 9). The participants did not know exactly how the score was calculated, but they did know that their scores depended on correct items collected, incorrect items collected, errors committed (i.e., rules broken), and time remaining. They could only see their score change in the top right corner of their screen (Figure 9).

It is important to note the different impact each component has on the individual score and the team score. Even though each component was weighted equally, they impacted the score differently because of different denominators. For example, the denominator for CI_{ind} (six items) is much smaller than the denominator for II_{ind} (77 items). Based on denominators, a participant needed to collect only three correct individual items to give CI_{ind} a value of fifty percent and a participant needed to collect 39 incorrect individual items to give II_{ind} a value of fifty percent. In future studies, the score weights could be adjusted to address the impact of each component in a more balanced fashion. For this current work, the

researcher analyzed the scores calculated as described in Table 7 and Table 9 and each component for the individual score and team score.

Table 7 - The components of the individual score. Each component is a percentage, with II_{team} counting the number of incorrect items not collected (larger is better) and E_{team} counting the number of rules not broken (larger is better). The maximum score is 100 percent; the minimum is 0 percent. The weight (W) for each component was 0.25.

Individual Score Components
Correct Items (CI_{ind}) = $\frac{Total\ \#\ of\ Correct\ Individual\ Items\ collected}{Total\ \#\ of\ Correct\ Individual\ Items} \times 100$
Incorrect Items (II_{ind}) = $\frac{Total\ \#\ of\ Incorrect\ Items - Total\ \#\ of\ Incorrect\ Items\ collected}{Total\ \#\ of\ Incorrect\ Items} \times 100$
Time (T_{ind}) = $\frac{Total\ remaining\ time}{Total\ Time\ to\ Complete\ Task} \times 100$
Error (E_{ind}) = $\frac{Total\ \#\ of\ rules - Total\ \#\ of\ rules\ broken}{Total\ \#\ of\ rules} \times 100$
Score = $(\mathrm{CI}_{ind} + \mathrm{II}_{ind} + T_{ind} + \mathrm{E}_{ind}) \times (W)$

Table 8 - Clarification on how components for the individual score were calculated.

Individual Score Calculation Clarification			
Dependent Variable	Calculation	Value	Definition
CI_{ind}	Total # of Correct Individual Items	6 Items	This is the number of correct individual items.
II_{ind}	Total # of Incorrect Items	77 Items	This number includes items that are on another player's list but did not include team items since they were not assigned to a specific player. Participants were not penalized for collecting team items.
T_{ind}	Total remaining time	600 Seconds (10 Minutes)	The time remaining for the individual score was determined when the participant had signaled the completion of the session or the session ends.
E_{ind}	Total # of rules broken	7 Rules	The number of unique rules broken by an individual.

Table 9 - The components of the team score. Each component is a percentage, with II_{team} counting the number of incorrect items not collected (larger is better) and E_{team} counting the number of rules not broken (larger is better). The maximum score is 100 percent; the minimum is 0 percent. The weight (W) for each component was 0.25.

Team Score Components
Correct Items (CI_{team}) = $\frac{Total\ \#\ of\ Correct\ Team\ Items\ collected}{Total\ \#\ of\ Correct\ Team\ Items} \times 100$
Incorrect Items (II_{team}) = $\frac{Total\ \#\ of\ Incorrect\ Items - Total\ \#\ of\ Incorrect\ Items\ collected}{Total\ \#\ of\ Incorrect\ Items} \times 100$
Time (T_{team}) = $\frac{Total\ remaining\ time}{Total\ Time\ to\ Complete\ Task} \times 100$
Error (E_{team}) = $\frac{Total\ \#\ of\ rules - Total\ \#\ of\ rules\ broken}{Total\ \#\ of\ rules} \times 100$
Score (Team) = $(\mathrm{CI}_{team} + \mathrm{II}_{team} + T_{team} + E_{team}) \times (W)$

Table 10 - Clarification on how components for the team score were calculated

Team Score Calculation Clarification			
Dependent Variable	Calculation	Value	Definition
CI_{team}	$Total\ \#\ of\ Correct\ Team$	18 Items	This is the number of correct individual items.
II_{team}	Total # of Incorrect Items	83 Items	This number included the items that were not assigned to a team or individual and the items that were assigned to an individual. The researcher included individual items because each individual item had the potential to be collected incorrectly. This number excluded the correct team items collected.
T_{team}	Total remaining time	600 Seconds (10 Minutes)	The time remaining for the team score was determined once all the players had signaled they completed or the session had ended.
E_{team}	Total # of rules broken	7 Rules	The number of unique rules broken by any team member.

Surveys

Each participant completed four surveys: demographics/pre-survey (APPENDIX B), post-session survey (APPENDIX C), NASA TLX survey (APPENDIX A), and overall post-session survey (APPENDIX E). Participants completed the demographics/pre-survey before they arrived began their session. They completed the post-trail and NASA TLX survey after each session. They completed the overall post-session survey after they completed all four sessions.

Experimental Procedure

A general procedure overview is displayed in Figure 12. Participants were asked to complete a consent survey, demographic survey, and pre-survey before signing up to participate in the study. Each team that conducted the study consisted of three members. When participants arrived at the lab, they were introduced to the experiment and told that the purpose of the study was to understand better how different components of training influences team performance.

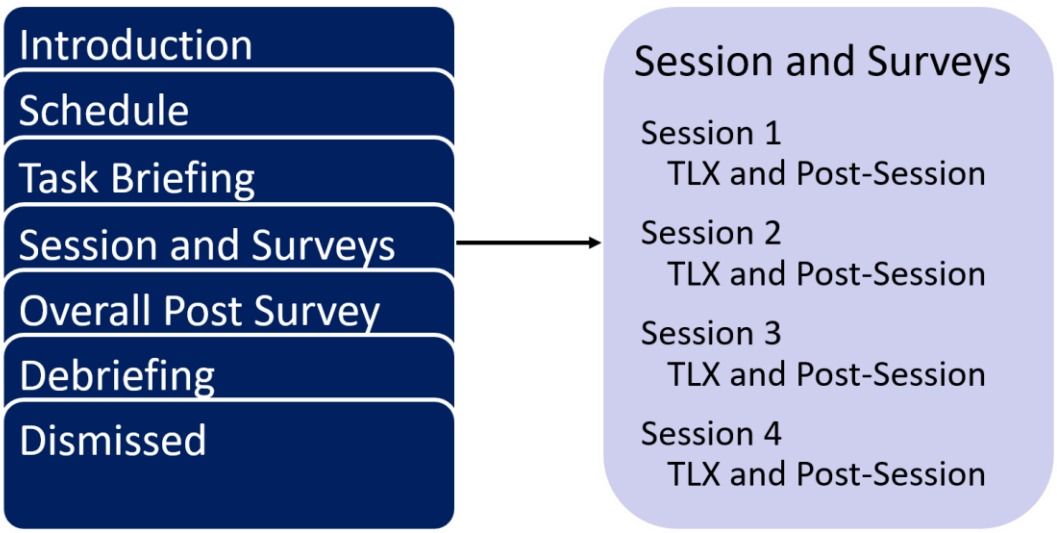


Figure 12 - General procedure overview.

Once the introduction was complete, the researcher described the task and the virtual environment to the team. Participants were told they would be completing a shopping mall task. They were told that the objective of the task was to buy all the items on the shopping lists (i.e., Team and Individual List) as quickly as possible. They were then told that they must work together to collect all the items on the team list, but they are only responsible for the items on their own individual lists. The researcher read the rules (Table 4) to the articipants and placed a copy of the rule at each team member's station during the entire experiment, so the members did not have to remember the rules. The researcher told the participants that their score was based on correct items collected, incorrect items collected, errors committed (i.e., rules broken), and time remaining. The participants knew the four main components that contributed to their score, but they were unaware of how the scores were exactly calculated (Table 7 and Table 9).

After the team understood the task objectives and rules, the researcher explained how to navigate within the environment. The researcher gave the participants a printed screenshot image (Figure 13) of the interface to the participants while explaining how to navigate through the environment. The teams began their sessions once every member understood the rules and task objective. Each team completed four sessions overall. Every session had different feedback interventions (Table 5). The stores in each session had different position configuration and different items in the stores. Each configuration was like the store configuration displayed in Figure 15. Each first session was always the No Feedback session but ordered in a counterbalanced fashion. After each session, the participants completed two surveys: NASA Task Load Index (NASA TLX) survey (Hart & Staveland, 1988)(APPENDIX A) and post-session survey (APPENDIX C). After the last session (i.e.,

the fourth session), the participants completed the NASA TLX, post-session, and overall post-session surveys. Participants participated in a debriefing session, or semi-structured group interview, where the researcher asked them questions (APPENDIX D), and the participants could answer freely. The researcher told the participants that they successfully completed the session (i.e., regardless of their scores). Participants were dismissed shortly after the debriefing session and given their compensation. Participants typically completed the entire session in 90 – 120 minutes.

Figure 13 - User interface example for each participant. This participant has "entered" a store by approaching it and is examining the goods for sale (center gray panel). The participant's current shopping cart contains one item (lower left panel). The participant's game information is displayed in the top right corner, including the amount of money available to buy items, time remaining in the session, player name and team name, team score (percent), and individual score (percent). This is an example of the No Feedback condition (see Table 5).

Figure 14 - User interface example for each participant. This participant has "entered" a store by approaching it and is examining the goods for sale (center gray panel). The participant's current shopping cart contains one item (lower left panel). The participant's game information is displayed in the top right corner, including the amount of money available to buy items, time remaining in the session, player name and team name, team score (percent), and individual score (percent). This is an example of the I&T Feedback condition, in which the panels of rules broken and items collected are also present (see Table 5).

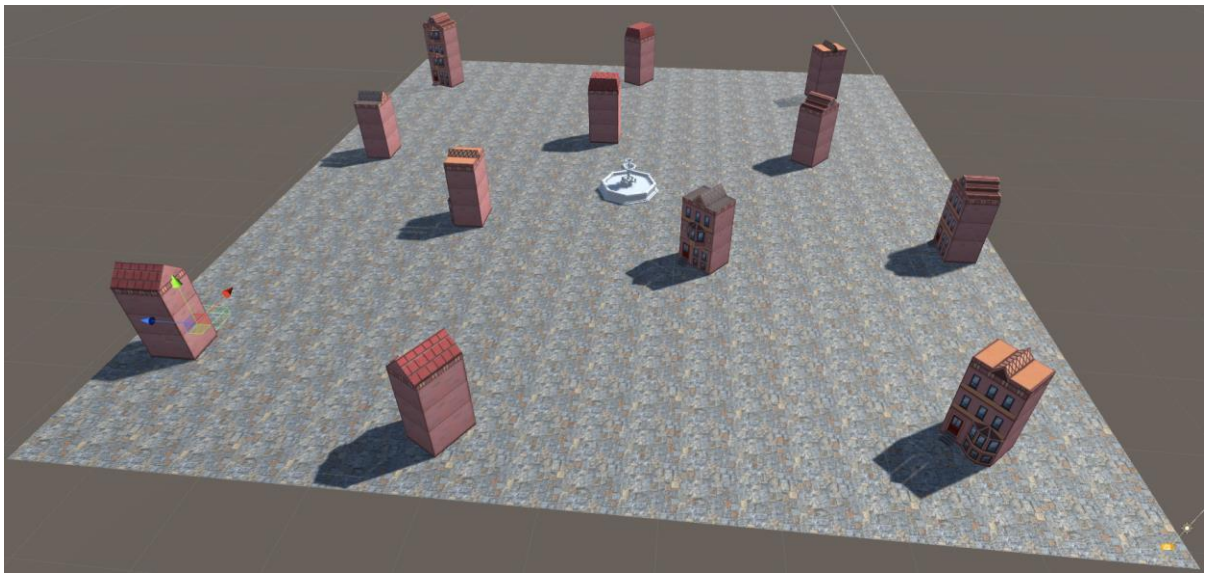


Figure 15 – Top-level view of one configuration of the virtual shopping mall (of four possible configurations).

CHAPTER 4. PERFORMANCE METRICS ANALYSIS

This chapter offers the results and discussion of how the feedback condition (Individual, Team, or I&T) affected the main dependent variables related to performance: scores, correct items collected, incorrect items collected, time remaining, and errors. Each analysis was done at both the team level (how the team performed) and at the individual level (how each team member performed). The researcher used Restricted Maximum Likelihood (REML) estimation instead of ANOVA because Maximum Likelihood Estimations (MLE) utilize methods that use all the data. Furthermore, MLE produces better-predicted values for unbalanced data (McCulloch, 2005). Thus, a regression modeling approach was used instead and explained in detail below. Later chapters explore team strategies for completing the task and how team members perceived their individual and team performance.

Focus of Analysis

The focus of the chapter analysis is to understand how the feedback interventions influenced team and individual performance. In particular, the question is how the feedback assessment conditions (Individual, Team, and I&T) affected the dependent performance variables: score, correct items collected, incorrect items collected, time remaining, and errors. The researcher expected the analysis to show a significant increase in performance over time because previous studies have shown that practice improves performance (Ericsson, 2008; Pusic, Boutis, Hatala, & Cook, 2015). The first part of the analysis focused on team-level metrics, and the second analysis focused on individual-level metrics. A full analysis of the Team and Individual score was presented while a shorten analysis was presented for the other metrics. Along the way, a short discussion about each result and its implication is provided.

Participant Demographics

There were 117 participants (39 teams) that participated in this study. Demographic breakdowns are provided by gender (Table 11), age (Table 12), academic standing (Table 13), academic major (Table 14), and highest degree acquired (Table 15). It is important to note that over half of the participants were engineering students (Table 14) and that most of the participants were between the ages of 18 and 30 (Table 12). The higher proportion of engineering students could be a concern if it led to a ceiling effect in performance, e.g., if engineering participants were better problem solvers than other participants since an engineering curriculum focuses on problem solving, but that did not seem to be the case. Some percentages do not add to 100% due to rounding.

Table 11 - Participant gender demographics

	Count	Percent
Female	42	35.9%
Male	74	63.2%
Other	1	0.9%

Table 12 - Participant age demographics

Age	Count	Percent
18-21	85	72.6%
22-30	23	19.6%
31-40	7	6.0%
41-50	1	0.9%
51-60	1	0.9%

Table 13 – Participant academic standing

Year in school	Count	Percent
Freshman	19	16.2%
Sophomore	27	23.1%
Junior	21	17.9%
Senior	28	23.9%
Not an undergraduate	21	17.9%
Prefer not to answer	1	0.9%

Table 14 - Academic major

Major	Count	Percent
Business	6	5.1%
Design	5	4.3%
Education & Human Sciences	9	7.7%
Engineering	71	60.7%
Liberal Arts	8	6.8%
Sciences	9	7.7%
Other	9	7.7%

Table 15 - Highest degree acquired

Highest Degree	Count	Percent
Associate's	8	6.8%
Bachelor's	18	15.4%
High School	81	69.2%
Master's	10	8.5%

Linear Mixed Model (LMM) Regression Modeling

The researcher used an LMM regression approach, instead of a linear model, to model team performance and analyze the influence of the feedback condition because LMMs can account for order effects due to a within-subject (repeated measures) experimental design (since each team performed four sessions) as well as effects due to a team's ability. When working together as a team, certain individuals perform better than other individuals due to uncontrollable effects (e.g., individual talents and interactions with other participants). For example, imagine that Player A is a high performing individual. Player A may perform better when partnered with Player D than when partnered with Player C. An LMM can account for this random effect. Other effects are designated fixed effects because they are based on the study design. In this case, the session order and the feedback condition are fixed effects. An LMM can include these effects as well.

The model developed by the researcher must also take experience (i.e., session order) into account because teams will improve their performance over time. The models developed accounted for the experience of each team. The following full model was developed for performance at the team level:

$$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt} \quad (1)$$

where y_{ijt} is the response variable (e.g., team score), μ is the baseline (or intercept), α_i is the fixed effect for the ith feedback category, β_j is the fixed effect for the jth session order, $\alpha_i\beta_j$ is the fixed effect of the interaction between feedback category and session order, γ_t is the random effect for the tth team, and ϵ_{ijt} is the residual for the model. The following full model, with an additional term for the effect of each team member, was developed for performance at the individual level:

$$y_{ijtp} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \theta_p + \epsilon_{ijtp} \quad (2)$$

where y_{ijtp} is the response variable (e.g., team score), μ is the baseline (or intercept), α_i is the fixed effect for the ith feedback category, β_j is the fixed effect for the jth session order, $\alpha_i\beta_j$ is the fixed effect of the interaction between feedback category and session order, γ_t is the random effect for the tth team, θ_p is the random effect for the pth individual, and ϵ_{ijtp} is the residual for the model.

For each dependent variable below, residual assumptions for LMMs are tested, the restricted maximum likelihood estimation is calculated, the effect size (R^2) is calculated, figures of the raw data are shown to display possible patterns, and implications are discussed. It is important to note that dummy variables were used to conduct the restricted maximum likelihood (REML) estimation. It is important to note that the REML estimation uses a

coding technique called dummy variable coding (Suits, 1957). Using this method drops a level in a categorical parameter. Dropping a level in the categorical data avoids breaking a model due to multicollinearity. The missing category is called the reference category. All interpretation is in reference to that reference category. The analysis in this chapter will use the individual feedback level and Session 2 as the reference category.

Model Selection Process

The full model for the team level (1) and individual level (2) assume that each independent variable improves the model for each dependent variable. However, each dependent variable may not significantly improve the model. The researcher systematically chose a model that "best" described the data using the Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) (Gałecki & Burzykowski, 2013). Generally speaking, the lower the AIC and BIC values, the better the model. If the AIC and BIC values gave conflicting recommendations between two models that were not significantly different, the simpler model was chosen. The steps used to select a model for each dependent variable at the team and individual level are shown in Table 16. The sub-models compared to the full model are based on the full model. The models used in the selection process at the team level are presented in Table 17, and the models used in the selection process at the individual level are presented in Table 18. In each sub-model, the researcher removed one or two terms from the full model. The following are the names of each sub-model: Null, All, No Interaction, No Interaction No Feedback, and No Interaction No Session Order. The Null model removes the terms for the fixed effects of the feedback condition and session order. The All model contains the full model. The other sub-models leave out the terms for feedback condition, session order, and/or the term for the interaction between feedback condition and session order.

Table 16 - Model selection process

1. Fit the multiple sub-models using the data provided.
2. Compare those models to the null model.
3. Pick the models that are significantly different from the null model.
4. Among those models, determine which model best describes the data.
5. Evaluate the effect size (R^2) of the model using the method presented by Nakagawa and Schielzeth (2013).

Table 17 - Models used in the model selection process at the team level

Model Name	Equation
Null	$y_{ijt} = \mu + \gamma_t + \epsilon_{ijt}$
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$
No Interaction No Feedback	$y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}$
No Interaction No Session Order	$y_{ijt} = \mu + \alpha_i + \gamma_t + \epsilon_{ijt}$

Table 18 - Models used in the model selection process at the individual level

Model Name	Equation
Null	$y_{ijtp} = \mu + \gamma_t + \theta_p + \epsilon_{ijtp}$
All	$y_{ijtp} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$
No Interaction	$y_{ijtp} = \mu + \alpha_i + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$
No Interaction No Feedback	$y_{ijtp} = \mu + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$
No Interaction No Session Order	$y_{ijtp} = \mu + \alpha_i + \gamma_t + \theta_p + \epsilon_{ijtp}$

Assumption Tests

The residuals of an LMM must fulfill certain assumption to be analyzed as a linear mixed-effect model. The assumptions for the residuals are shown in Table 19.

Table 19 - LMM assumptions for residuals (Galwey, 2014)

1. The residuals should have an approximately normal distribution (i.e., a bell curve) when plotted on a histogram.

2. A fitted-value plot should show an almost constant width when viewed from left to right.
3. The points on a normal plot should lie on an almost straight line from the bottom left to the top right.

For some dependent variables, the selected models violated these assumptions. According to Field and Wilcox (2017), the best way to examine the influence of assumption violations is to compare a robust model to a classic model. In this context, the classic model is a model based on non-robust estimators, while a robust model is a model that is based on robust estimators. The researcher compared a robust model to a classic model when assumptions were violated. If there was a noticeable difference between the two models (i.e., if the estimated coefficient of a fixed effect was noticeably different), then the researcher used a step on the power ladder to transform (Tukey, 1977), or re-express, the data to address the assumption violation. The procedure used by the researcher to resolve assumption violations is summarized in Figure 16.

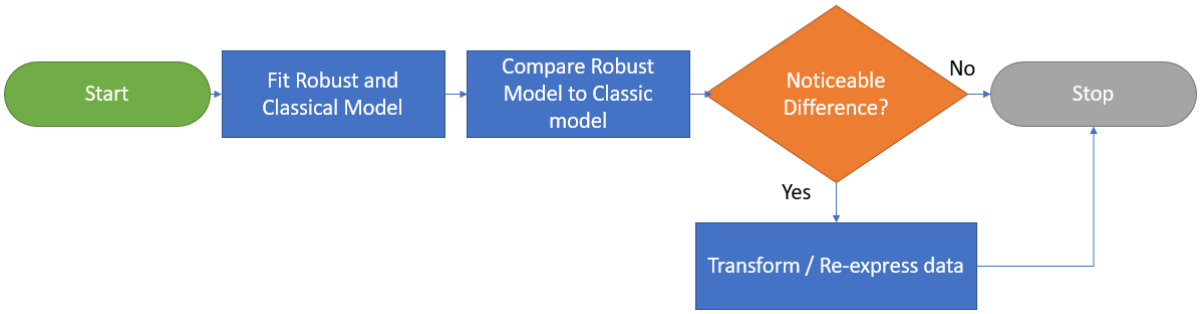


Figure 16 - Flow chart for assumption violations

Analysis

Each of the dependent variables analyzed and described above is displayed in Table 20. It is important to note that in the analysis, Session 1 was dropped from the analysis

because this session was always the condition with No Feedback intervention. The No Feedback condition was used as training for the teams. In the analysis, the No Feedback condition was dropped. Because the analysis of each variable requires several steps, the full process is documented in this chapter only for Team Score and Individual Score. The complete analysis for each variable is available in APPENDIX G, however.

Table 20 - Dependent variables analyzed in Chapter 4.

Dependent Variables Analyzed
Team Score
Correct Team Items
Incorrect Items (Team)
Time Remaining (Team)
Unique Errors Committed (Team)
Individual Score
Correct Individual Items
Incorrect Items (Individual)
Time Remaining (Individual)
Unique Errors Committed (Individual)

Table 21 – Overall Analysis process

1. Find models that are significantly different from the null model.
2. Pick the model the describes the data best.
3. Test assumptions for linear mixed models.
4. Use REML to estimate the fixed effect values for independent variables.
5. Evaluate the marginal R^2 (R_m^2) and conditional R^2 (R_c^2).
6. Generate pairwise comparison of estimated marginal means.

Team Score

Team score was a dependent variable used to measure performance. This variable is a weighted sum of the correct items collected from the team list by any player, incorrect items not collected, time remaining, and errors not committed (see Table 9). The max score is 100

percent, and the minimum score is 0 percent. The analysis below explores whether the feedback condition influenced team score.

Distribution Overview (Team Score)

The overall distribution of the scores at the team level shows a slight skew to the left with a center around 57 (Figure 17). The distribution of the scores, when grouped by Feedback, is similar to the overall distribution (Figure 18). The distribution of the scores in the Individual condition is similar to the distribution of the scores in the Team condition. Over sessions, the distribution of the scores flattens and spreads in each Feedback condition (Figure 19).

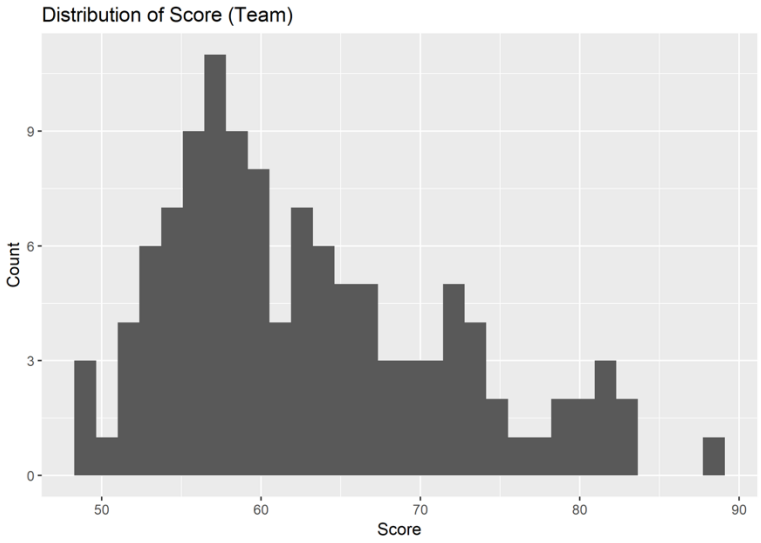


Figure 17 - Distribution overview of score (Team)

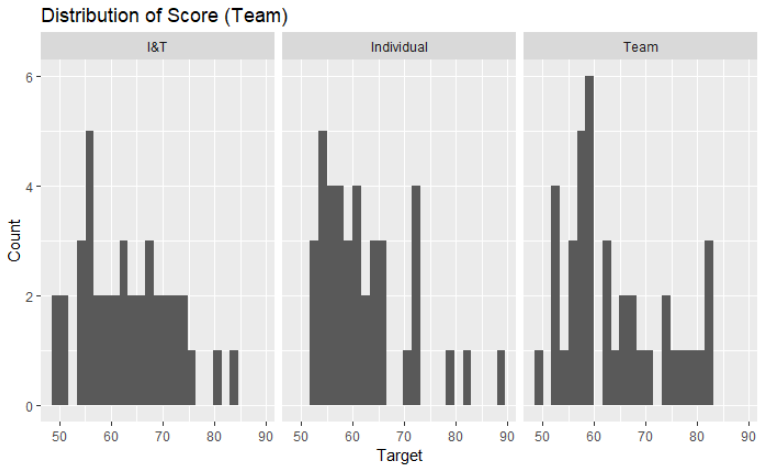


Figure 18 - Distribution of score grouped by Feedback (Team)

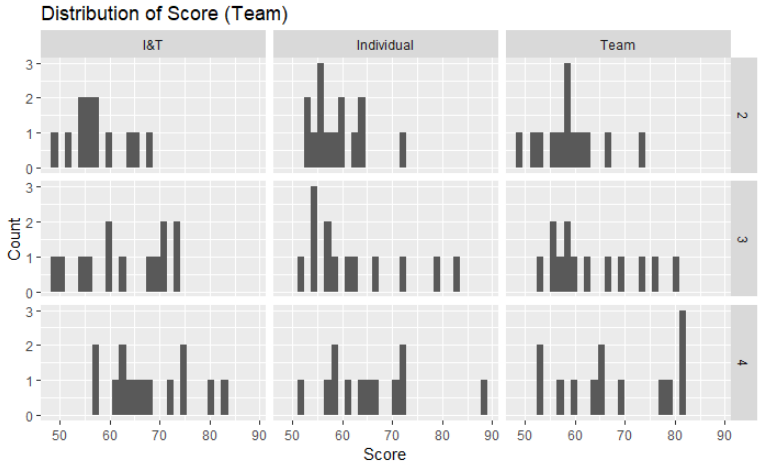


Figure 19 - Distribution of score grouped by Feedback and Session Order (Team)

Model Selection (Team Score)

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 22. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect was negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible, and the simplest model that described the data best was the No Interaction No Feedback model.

Table 22 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	790.20	820.59
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	786.67	806.01
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**784.28**	**798.09**

Testing Assumptions: Residuals (Team Score)

The histogram of the residuals (Figure 20) appeared to be approximately distributed. The residual fitted-value (Figure 21) showed a relatively constant variance, despite the slight increase in variance around the fitted value of 70. The points on the residual normal Q-Q plot (Figure 22) lay in a roughly straight line, but the tail end of the data seem to trail away from the diagonal, which suggest a violation of Assumption 3 (Table 19). The classic LMM and the robust LMM produced similar coefficient values (Table 23), so the researcher did not transform the data.

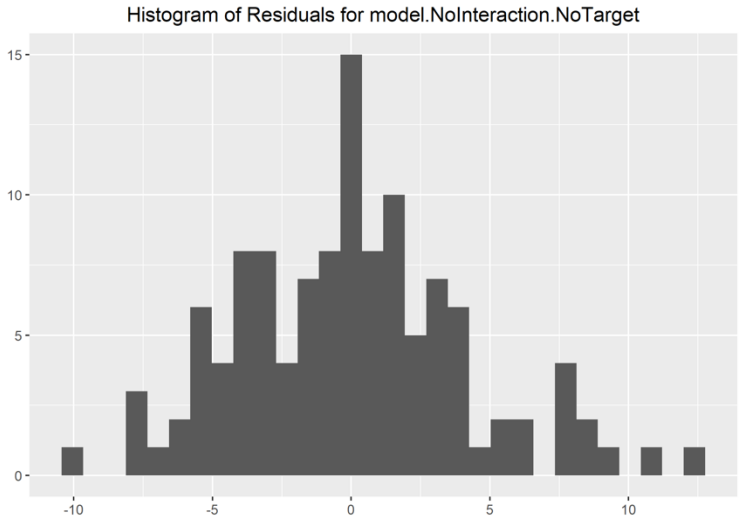


Figure 20 - Histogram of Residuals (Team Score)

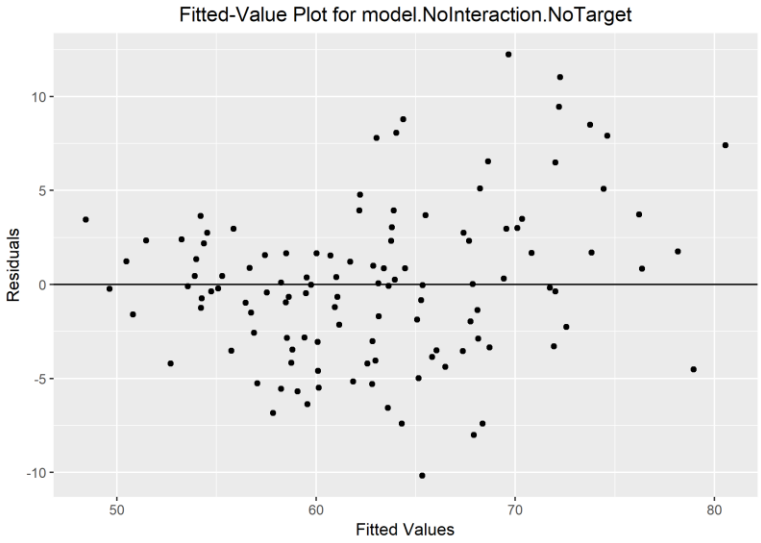


Figure 21 - Residual fitted plot (Team Score)

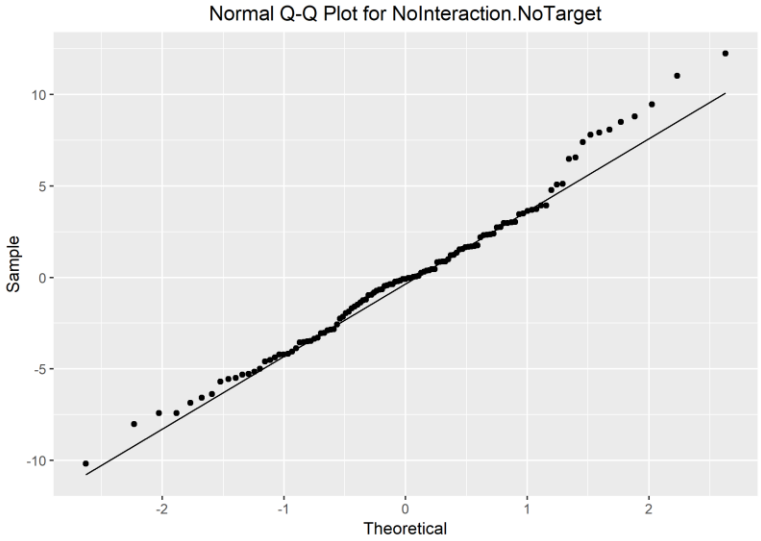


Figure 22 - Normal Q-Q plot for residuals (Team Score)

Table 23 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	58.7	58.3
SessionOrder: 3	4.27	4.18
SessionOrder: 4	8.61	7.97

Restricted Maximum Likelihood Estimation (Team Score)

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 24. The results showed that the effects of Session 3 and 4 are significant in reference to Session 2, the reference category.

Table 24 - Fixed effect result of REML Estimation (Team Score)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	58.7	1.31	64	45.0	< .05*
Session: 3	4.27	1.12	76	3.80	< .05*
Session: 4	8.61	1.12	76	7.66	< .05*

Evaluating Effect size (Team Score)

The coefficient of determination, R-squared (R^2), value is used to report the goodness-of-fit of a regression line (Nakagawa & Schielzeth, 2013). There are traditional properties of R^2, such as the value should not rely on a particular unit of measure (Orelien & Edwards, 2008). However, there are problems that arise when seeking to determine the R^2 for linear mixed-models. An issue that contributes to the problem of calculating R^2 is that researchers have not agreed on a single definition of R^2 (Nakagawa & Schielzeth, 2013). Johnson (2014) offers an extension of a simple method, developed by Nakagawa and Schielzeth (2013), to generate R^2 for LMMs. This method generates two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables (session order) while the R_c^2 describes how the

variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 are presented in Table 25. The researcher used R^2 values .01, .09, and .25 to be small, medium, and large effect sizes, respectively (J. Cohen, 1992). The results suggest that the fixed variables have a medium effect size (i.e., explain 16% of the variance) and the fixed and random variables have a large effect size (i.e., explain 69% of the variance).

Table 25 - Effect size for LMM (Team Score)

R^2 Type	Value
R_m^2	0.158
R_c^2	0.688

Estimated Marginal Means (Team Score)

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 26) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 27). The pairwise comparison showed a statistical difference between each session. The statistical difference suggests that the team scores improved over time.

Table 26 - Estimated marginal means (Team Score)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	58.7	1.30	64	56.1	61.3
3	62.9	1.30	64	60.3	65.5
4	67.3	1.30	64	64.7	69.9
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 27 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-4.27	1.12	76	-3.80	< .05*
2 – 4	-8.61	1.12	76	-7.66	< .05*
3 – 4	-4.34	1.12	76	-3.86	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (Team Score)

The best model, No Interaction No Feedback, excluded the effect of feedback condition (Table 22). None of the assumptions for LMM modeling of team score were violated. REML estimation (Table 24) indicated the effect of the session order was not zero, though the R_m^2 value indicated that the session order explained 16% of the variance. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the team scores in Sessions 2, 3, and 4 were statistically significantly different.

The feedback did not have a significant effect on the Team Score, or the effect is so small (or the variance so large) that more data are needed. The next sections present the analysis of the impact of feedback condition on each of the four components that are used to calculate the team score. These analyses are presented in a shorter form, but the full analyses are available in APPENDIX G.

Correct Items Collected (Team)

Correct Items Collected (Team), or CI_{team}, was a dependent variable used to measure performance and was one of the four components of Team Score. This variable measures the total count of team list items collected by all of the team members. The analysis below explores whether CI_{team} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (CI_{team})

Estimated marginal means (EMMs) (Table 28) were generated to analyze the means of each session. The researcher used a pairwise difference analysis to compare the means differences across each session (Table 29). The pairwise comparison showed no statistical difference between Sessions 2 and 3. It also showed no statistical difference between Sessions 3 and 4. The comparison does show a statistical difference between Session 2 and 4. The comparison suggests the number of correct items collected increased over overall (i.e., from Session 2 to Session 4).

Table 28 - Estimated marginal means (Correct Item Collected)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	15.1	0.313	96	14.4	15.7
3	15.7	0.313	96	15.0	16.3
4	16.2	0.313	96	15.6	16.8
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 29 - Pairwise difference of estimated marginal means (EMMs)(Correct Item Collected)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-0.615	0.368	76	-1.67	.222
2 – 4	-1.13	0.368	76	-3.07	< .05*
3 – 4	-0.513	0.368	76	-1.39	.349
Results are averaged over the levels of Session Order					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (CI_{team})

The best model, No Interaction No Feedback, excluded the effect of feedback condition (Table 99). Some of the assumptions for LMM modeling of CI_{team} were violated. Comparing the classic LMM to the robust LMM indicated the violation did not drastically influence the estimated results. REML estimation (Table 101) indicated that the effect of session 4 was not zero, though the R_m^2 value indicated that the session order explained only 5% of the variance. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the CI_{team} in Sessions 2 and 4 were statistically significantly different (Table 28 and Table 29).

Overall, the results suggest that the feedback did not have a significant effect on the CI_{team} or the effect is so small that more data are needed to detect it with the amount of variance present in the population. The next section will present the analysis of the incorrect items collected metric to explore how feedback influenced the metric.

Incorrect Items Collected (Team)

Incorrect Items Collected (Team), or II_{team}, was a dependent variable used to measure performance and was one of the four components of Team Score. This variable measures the total count of incorrect items collected by any of the team (see Table 10 for clarification of incorrect items collected for teams). The analysis below explores whether II_{team} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (II_{team})

Estimated marginal means (EMMs) (Table 30) were generated to analyze the means of each level (i.e., Individual, team, and I&T). A pairwise difference analysis was used to

compare the means differences across each condition (Table 31). The pairwise comparison showed no statistical difference between the groups. EMMs for the means for each session order was calculated (Table 32). A pairwise difference analysis was used to compare the means differences across each session (Table 33). The pairwise comparison showed no statistical difference among the session order.

Table 30 - Estimated marginal means (Incorrect Item Collected)

Estimated Marginal Means					
Feedback	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
Individual	0.988	0.123	92	0.743	1.23
Individual and Team	0.819	0.123	92	0.574	1.06
Team	0.930	0.123	92	0.685	1.18
Results are averaged over the levels of Session Order					
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 31 - Pairwise difference of estimated marginal means (EMMs)(Incorrect Item Collected)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
Individual – Individual and Team	0.169	0.141	74	1.20	.459
Individual – Team	0.058	0.141	74	0.411	.911
Individual and Team – Team	-0.111	0.141	74	-0.788	.712
Results are averaged over the levels of Session Order					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Table 32 - Estimated marginal means for session order (Incorrect Item Collected)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	0.991	0.123	92	0.746	1.24
3	0.966	0.123	92	0.721	1.21
4	0.780	0.123	92	0.535	1.03
Results are averaged over the levels of Feedback					
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 33 - Pairwise difference of estimated marginal means for session order (EMMs)(Incorrect Item Collected)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	0.025	0.141	74	0.178	.983
2 – 4	0.211	0.141	74	1.49	.302
3 – 4	0.185	0.141	74	1.31	.393
Results are averaged over the levels of Feedback					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (II_{team})

The best model, Null, excluded the effect of the feedback condition. No model described the data better than the Null model. The researcher used the No Interaction model to examine the effect of session order and the feedback variable. All of the assumptions for LMM modeling of II_{team} were violated. The researcher transformed the data because the classic LMM and the robust LMM produce concernedly different coefficient values (Table 103). REML estimation (Table 105) indicated the effect of the session order and feedback was not significantly different from zero, the R_m^2 value indicated that session order and feedback explained only 0.8% of the variance. Other variation must be due to factors other than session order or feedback condition, since none of the models described the data better

than the Null model. A comparison of the EMMs indicated no significant difference among the session order and the feedback variable (Table 31 and Table 33).

Overall, the results suggest that the feedback did not have a significant effect on the $II_{team,}$ or the effect is so small that more data are needed. The next section will present the analysis of the time remaining metric to explore how feedback influenced the metric.

Time remaining (Team)

The Time Remaining (Team), or T_{team}, was a dependent variable used to measure performance and was one of the four components of Team Score. This variable measures the total amount of time remaining at the end of the session by the team. The analysis below explores whether T_{team} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (T_{team})

Estimated marginal means (EMMs) (Table 34) were generated to analyze the means of each session. A pairwise difference analysis was used to compare the means differences across each session (Table 35). The pairwise comparison showed a statistical difference among each session, indicating time remaining increased over time.

Table 34 - Estimated marginal means (Time remaining - Team)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	5.13	0.605	55	3.92	6.34
3	6.71	0.605	55	5.50	7.92
4	8.21	0.605	55	7.00	9.42
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 35 - Pairwise difference of estimated marginal means (EMMs) (Time remaining - Team)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-1.58	0.440	76	-3.58	< .05*
2 – 4	-3.08	0.440	76	-7.00	< .05*
3 – 4	-1.50	0.440	76	-3.41	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (T_{team})

The best model, No Interaction No Feedback, excluded the effect of feedback condition (Table 107). Some of the assumptions for LMM modeling of T_{team} were violated. The researcher transformed the data because the classic LMM and the robust LMM produce concernedly different coefficient values (Table 108). REML estimation (Table 109) indicated the effect of the session order was not zero, though the R_m^2 value indicated that the session order explained only 8.2% of the variance. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the T_{team} in Sessions 2, 3 and 4 were all statistically significantly different (Table 35).

Overall, the results suggest that the feedback did not have a significant effect on the $T_{team,}$ or the effect is so small that more data are needed. The next section will present the analysis of the unique errors metric to explore how feedback influences the metric.

Unique Errors Committed (Team)

Unique Errors Committed (Team), or E_{team}, was a dependent variable used to measure performance and was one of the four components of Team Score. This variable measures the total count of unique errors committed by any member of the team. The analysis below explores whether E_{team} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (E_{team})

Estimated marginal means (EMMs) (Table 36) were generated to analyze the means of each session. A pairwise difference analysis was used to compare the means differences across each session (Table 37). The pairwise comparison showed a statistical difference among each session. Specifically, E_{team} decreased from session 2 to session 4.

Table 36 - Estimated marginal means (Unique Errors Committed - Team)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	3.74	0.245	70	3.25	4.23
3	3.03	0.245	70	2.54	3.51
4	2.33	0.245	70	1.84	2.82
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 37 - Pairwise difference of estimated marginal means (EMMs) (Unique Errors Committed - Team)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	0.718	0.229	76	3.13	< .05*
2 – 4	1.41	0.229	76	6.15	< .05*
3 – 4	0.692	0.229	76	3.02	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (E_{team})

The best model, No Interaction No Feedback, excluded the effect of feedback condition (Table 111). Some of the assumptions for LMM modeling of E_{team} were violated. The researcher transformed the data because the classic LMM and the robust LMM produce concernedly different coefficient values (Table 108). The researcher did not transform the data because the classic LMM and the robust LMM produce similar coefficient values (Table 112). REML estimation (Table 109) indicated the effect of the session order was not zero,

though the R_m^2 value indicated that the session order explained only 8%of the variance. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the E_{team} in Sessions 2, 3 and 4 were all statistically significantly different (Table 37). Overall, the results suggest that the feedback did not have a significant effect on the E_{team} or the effect is so small that more data are needed.

Analysis Summary of the Four Team Score Components

The researcher used a weighted sum to determine the team score. The four components had an equal-weighted value. However, there is no evidential basis to indicate that giving equal weight to the components is an optimal representation of team performance for this task. Each component was calculated as a percentage (Table 9), and it was assumed that each component gave an accurate representation of how teams perform. It is possible that other dependent variables may more accurately represent performance.

The model selected in for the Team Score, T_{team}, CI_{team}, and E_{team}, did not include the feedback condition and only included the Session Order. No sub-model for II_{team} was significantly different from the null model, suggesting that neither the feedback condition or the session order had a significant influence on II_{team}. There was a significant difference among each Session Order for Team Score, T_{team}, and E_{team}, indicating that Team Score and T_{team} increased over time and that E_{team} decreased over time. For CI_{team}, only session 4 and session 2 were significantly different, indicating that teams improved CI_{team} overall but did not make a significant improvement until after the fourth session.

Overall, the results suggest that the feedback condition had little influence on the Team Score and its components. The results indicate that a model only including the session order accurately described the data better than models including the feedback condition. This

result could indicate that the feedback condition did not influence the team performance or that the effect is weak, and more teams are needed to detect significance.

Individual Score

The individual score was a dependent variable used to measure the performance of individual team members. This variable is a weighted sum of the correct items collected from the individual's shopping list, incorrect items collected, time remaining, and errors committed (see Table 7). The analysis below explores whether the feedback condition influenced the individual score. As with Team Score, this analysis is presented in full detail to illustrate the process of analyzing individual team member variables, but subsequent analyses of individual variables are abbreviated in this chapter and presented in detail in APPENDIX G.

Distribution Overview (Individual Score)

The overall distribution of the scores at the individual level shows a normal distribution with a center around 67 - 68 (Figure 23). The distribution of the scores, when grouped by Feedback, is similar to the overall distribution (Figure 27). Over time, the distribution of the scores flattens in each Feedback condition (Figure 19). It is interesting to note that the distribution in the Team condition and Session 4 seems to skew to the right.

Model Selection (Individual Score)

The author used a model selection process outlined in Table 16. The models that are significantly different from the null model are displayed in Table 38. The All model was not significantly different from the models with No Interaction model or the model with No Interaction No Feedback, indicating that the interaction effect is negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible and the simplest model that describes the data best is the No Interaction No Feedback model. It is important to note that the

difference between the No Interaction model and the No Interaction No Feedback model was approaching significance (i.e., $p < 0.1$), indicating more data is needed to reach significance. However, the researcher did not analyze the No Interaction model because it was not significantly different from the No Interaction No Feedback model.

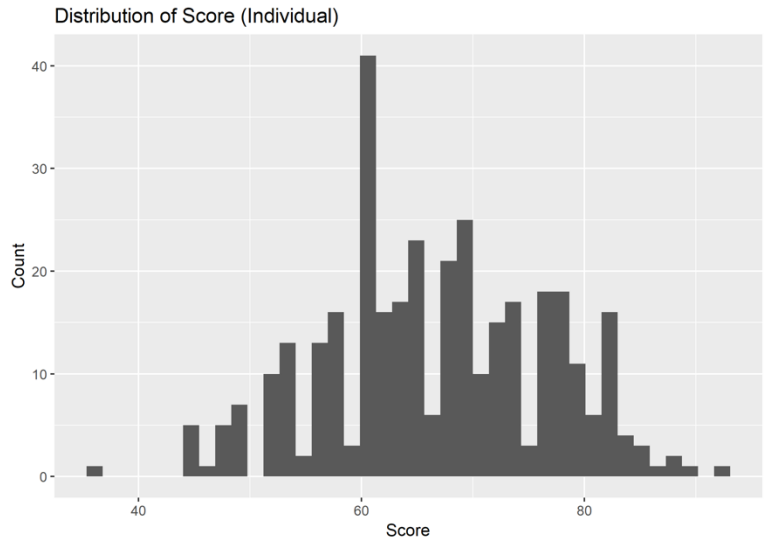


Figure 23 - Distribution overview of score (Individual)

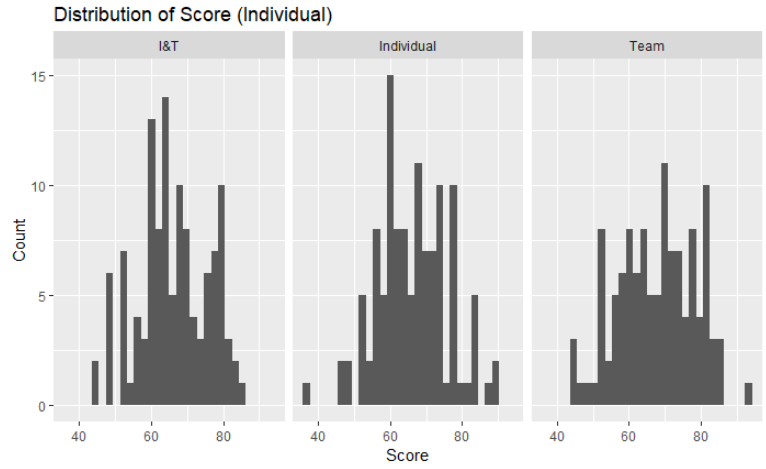


Figure 24 - Distribution of score grouped by Feedback (Individual)

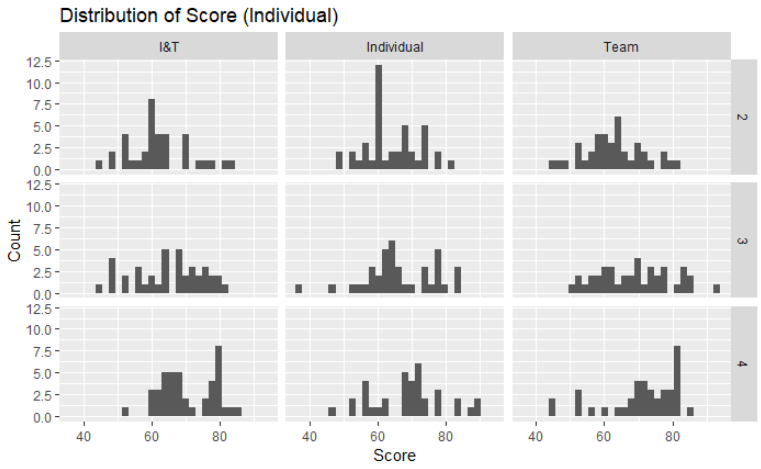


Figure 25 - Distribution of score grouped by Feedback and Session Order (Individual)

Table 38 – Models that are significantly different from the null model

Model Name	Equation	AIC	BIC
All	$y_{ijtp} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	2439.02	2485.35
No Interaction	$y_{ijtp} = \mu + \alpha_i + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	2433.73	2464.61
No Interaction No Feedback	$\boldsymbol{y_{ijtp} = \mu + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}}$	**2434.76**	**2457.92**

Testing Assumptions: Residuals (Individual Score)

The histogram of the residuals (Figure 26) appears to be approximately distributed. The residual fitted-value (Figure 27) shows a relatively constant variance. The points on the residual normal Q-Q plot (Figure 28) lie in a roughly straight line, which suggests the residuals are normally distributed. The researcher did not transform the data because all assumptions were met.

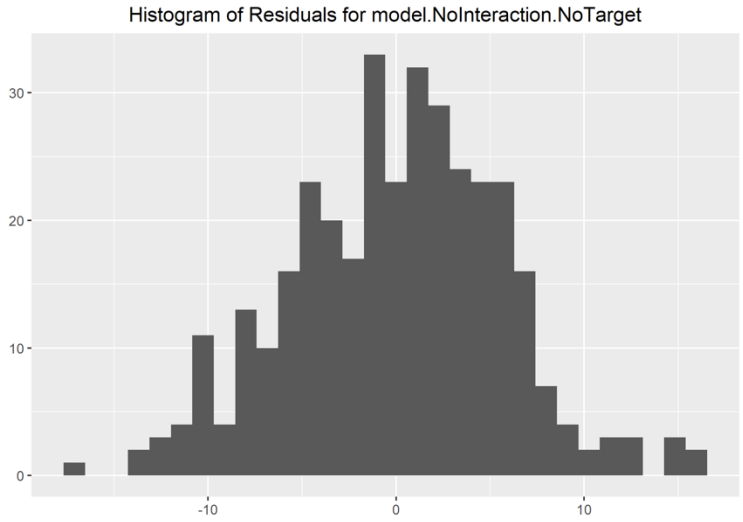


Figure 26 - Histogram of Residuals (Individual Score)

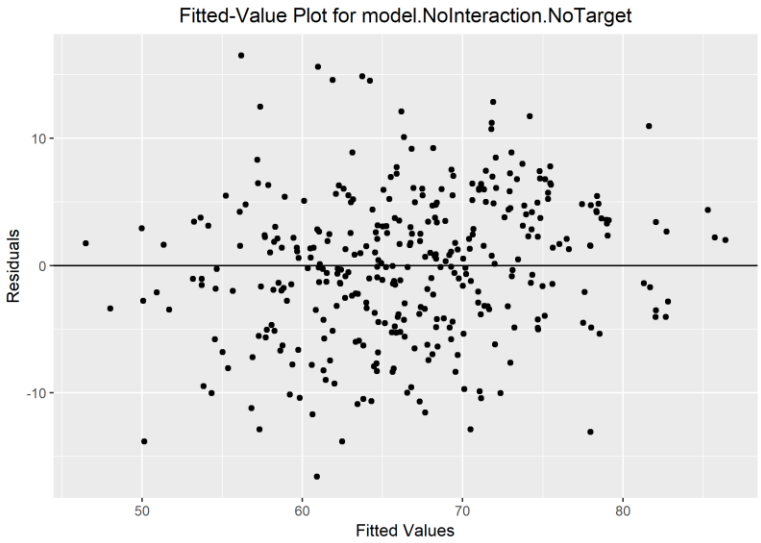


Figure 27 - Residual fitted plot (Individual Score)

Figure 28 - Normal Q-Q plot for residuals (Individual Score)

Restricted Maximum Likelihood Estimation (Individual Score)

The results for the fixed effect variables are displayed in Table 39.

Table 39 - Fixed effect result of REML Estimation (Individual Score)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	63.0	1.23	52	51.1	< .05*
Session: 3	3.66	0.826	232	4.43	< .05*
Session: 4	7.35	0.826	232	8.89	< .05*

The results showed that the effect of session 3 and 4 was significant in reference to session 2.

Evaluating Effect size (Individual Score)

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the

fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 40. The results suggest that the fixed variables have a small effect size (i.e., explains 9% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 61% of the variance).

Table 40 - Effect size for LMM (Individual Score)

R^2 Type	Value
R_m^2	0.089
R_c^2	0.605

Estimated Marginal Means (EMMs) (Individual Score)

Estimated marginal means were calculated to analyze the differences among groups in different variables (i.e., feedback condition and session order). The EMMs were generated (Table 41) to analyze the means of each level (i.e., session 2, 3, and 4). A pairwise difference analysis was used to compare the mean differences across each session (Table 42). The pairwise comparison showed a statistical difference among each session. The statistical difference suggests that the Individual Score.

Table 41 - Estimated marginal means (Individual Score)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	63.0	1.23	52	60.5	65.5
3	66.7	1.23	52	64.2	69.1
4	70.4	1.23	52	67.9	72.8
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 42 - Pairwise difference of estimated marginal means (EMMs)(Individual Score)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value

2 – 3	-3.67	0.821	230	-4.47	< .05*
2 – 4	-7.37	0.822	230	-8.97	< .05*
3 – 4	-3.70	0.821	230	-4.50	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (Individual Score)

The best model, No Interaction No Feedback, excluded the effect of feedback condition (Table 38). None of the assumptions for LMM modeling of Individual Score were violated. REML estimation (Table 39) indicated the effect of the session order was not zero, though the R_m^2 value indicated that the session order explained only 9% of the variance. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the Individual Score in Sessions 2, 3 and 4 were statistically significantly different (Table 42).

Overall, the results suggest that the feedback did not have a significant effect on the individual score or the effect is so small that more data are needed. The next sections present the analysis of the impact of feedback condition on each of the four components that are used to calculate the individual score.

Correct Items Collected (Individual)

Correct Items Collected (Individual), or CI_{ind}, was a dependent variable used to measure performance and was one of the four components of Individual Score. This variable measures the total count of individual items collected by a player. The analysis below explores whether CI_{ind} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (CI_{ind})

Estimated marginal means (EMMs) (Table 43) were generated to analyze the means of each session. A pairwise difference analysis was used to compare the means differences across each session (

Table 44). The pairwise comparison showed a statistical difference among each session, indicating CI_{ind} increased over time.

Table 43 - Estimated marginal means (Correct Items Collected – Individual)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	4.74	0.120	80	4.51	4.98
3	4.86	0.120	80	4.62	5.10
4	5.10	0.120	80	4.86	5.34
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 44 - Pairwise difference of estimated marginal means (EMMs)(Correct Items Collected – Individual)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-0.120	0.118	232	-1.02	.566
2 – 4	-0.359	0.118	232	-3.05	< .05*
3 – 4	-0.239	0.118	232	-2.04	.106
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (CI_{ind})

The best model, No Interaction No Feedback, excluded the effect of feedback condition (Table 115). All of the assumptions for LMM modeling of CI_{ind} were violated. The researcher did not transform the data because the classic LMM and the robust LMM produced similar coefficient values. REML estimation (Table 117) indicated that the effect of session 4 was not zero, though the R_m^2 value indicated that the session order explained only 2% of the variance. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the CI_{ind} in Sessions 2 and 4 were statistically significantly different (Table 44).

Overall, the results suggest that the feedback did not have a significant effect on CI_{ind} or the effect is so small that more data are needed. The next section will present the analysis of the incorrect items collected metric to explore how feedback influences the metric.

Incorrect Items Collected (Individual)

Incorrect Items Collected (Individual), or II_{ind}, was a dependent variable used to measure performance and was one of the four components of Individual Score. This variable measures the total count of individual items incorrectly collected by a player. The analysis below explores whether II_{ind} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (II_{ind})

Estimated marginal means (EMMs) (Table 45 and Table 122) were generated to analyze the means of each level (i.e., Individual, team, and I&T) and each session. A pairwise difference analysis was used to compare the means differences across each session (

Table 46 and Table 48). The pairwise comparison showed no statistical difference between the Feedback levels.

Table 45 - Estimated marginal means (Incorrect Items Collected – Individual)

Estimated Marginal Means					
Feedback	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
Individual	0.439	0.062	91	0.317	0.562
Individual and Team	0.349	0.062	91	0.227	0.472
Team	0.401	0.062	91	0.279	0.524
Results are averaged over the levels of Session Order					
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 46 - Pairwise difference of estimated marginal means (EMMs)(Incorrect Items Collected – Individual)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
Individual – Individual and Team	0.090	0.065	230	1.38	.352
Individual – Team	0.038	0.065	230	0.584	.829
Individual and Team – Team	-0.052	0.065	230	-0.799	.704
Results are averaged over the levels of Session Order					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Table 47 - Estimated marginal means (Incorrect Items Collected – Individual)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	0.435	0.062	91	0.313	0.558
3	0.415	0.062	91	0.292	0.537
4	0.340	0.062	91	0.217	0.462
Results are averaged over the levels of Feedback					
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 48 - Pairwise difference of estimated marginal means (EMMs)(Incorrect Items Collected – Individual)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	0.021	0.065	230	0.321	.945
2 – 4	0.096	0.065	230	1.47	.307
3 – 4	0.075	0.065	230	1.15	.484
Results are averaged over the levels of Feedback					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (II_{ind})

The best model, Null, excluded the effect of the feedback condition. No model described the data better than the Null model. The researcher used the No Interaction model to examine the effect of session order and the feedback variable. All of the assumptions for LMM modeling of II_{ind} were violated. The researcher used square root to transform the data to reduce the heterogeneity because the classic LMM and the robust LMM produce concernedly different coefficient values (Table 119). REML estimation (Table 120) indicated the effect of the session order and feedback was not significantly different from zero, the R_m^2 value indicated that session order and feedback explained only 1% of the variance. Other variation must be due to factors other than session order or feedback condition, since none of the models described the data better than the Null model. A comparison of the EMMs indicated no significant difference among the session order and the feedback variable (

Table 46 and Table 48).

Overall, the results suggest that the feedback did not have a significant effect on the $II_{ind,}$ or the effect is so small that more data are needed. The next section will present the analysis of the time remaining metric to explore how feedback influences the metric

Time remaining (Individual)

Time remaining (Individual), or T_{ind}, was a dependent variable used to measure performance and was one of the four components of Individual Score. This variable measures the total amount of time remaining and the end of the session by an individual. The analysis below explores whether T_{ind} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (T_{ind})

Estimated marginal means (EMMs) (

Table 49 and Table 51) were generated to analyze the means of each level (i.e., Individual, team, and I&T) and each session. A pairwise difference analysis was used to compare the means differences across each session (Table 50 and Table 128). The pairwise comparison showed a statistical difference between the Team level and the I&T level, suggesting that T_{ind} is higher in the Team condition when compared to the I&T condition.

Table 49 - Estimated marginal means (Time remaining – Individual)

Estimated Marginal Means					
Feedback	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
Individual	7.30	0.561	45	6.17	8.43
Individual and Team	6.87	0.561	45	5.74	8.00
Team	7.78	0.561	45	6.65	8.91
Results are averaged over the levels of Session Order					
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 50 - Pairwise difference of estimated marginal means (EMMs) (Time remaining – Individual)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
Individual – Individual and Team	0.436	0.269	308	1.62	.238
Individual – Team	-0.475	0.269	308	-1.77	.182
Individual and Team – Team	-0.911	0.269	308	-3.39	< .05*
Results are averaged over the levels of Session Order					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Table 51 - Estimated marginal means (Time remaining – Individual)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	5.60	0.561	45	4.47	6.73
3	7.48	0.561	45	6.35	8.60

4	8.87	0.561	45	7.74	10.00
Results are averaged over the levels of Feedback					
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 52 - Pairwise difference of estimated marginal means (EMMs) (Time remaining – Individual)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-1.88	0.269	308	-7.00	< .05*
2 - 4	-3.28	0.269	308	-12.12	< .05*
3 - 4	-1.40	0.269	308	-5.20	< .05*
Results are averaged over the levels of Feedback					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (T_{ind})

The best model, No Interaction, included the effect of feedback condition (Table 122). Some of the assumptions for LMM modeling of T_{ind} were violated. The researcher used square root to transform the data to reduce the heterogeneity because the classic LMM and the robust LMM did not produce similar results coefficient values (Table 123). REML estimation (Table 124) indicated the effect of the session order was not zero, though the R_m^2 value indicated that session order and feedback condition explained only 8% of the variance. Other variation must be due to factors other than session order or feedback condition. A comparison of the EMMs indicated that the T_{ind} in Team and I&T condition were statistically significantly different (Table 50) and Sessions 2, 3, and 4 were statistically significantly different (Table 52).

Unexpectedly, the results suggested that the T_{ind} in the Team condition is higher than in the I&T condition. The results contradicted H_3, which predicted that the I&T level would produce the highest performance. The next section will present the analysis of the errors committed metric to explore how feedback influences the metric.

Unique Errors Committed (Individual)

Unique Errors Committed (Team), or E_{ind}, was a dependent variable used to measure performance and was one of the four components of Team Score. This variable measures the total count of unique errors committed by a team member. The analysis below explores whether E_{ind} was influenced by feedback condition.

Estimated Marginal Means (EMMs) (E_{ind})

Estimated marginal means (EMMs) (Table 53) were generated to analyze the means of each session. A pairwise difference analysis was used to compare the means differences across each session (Table 54). The pairwise comparison showed among each session-level, indicating E_{ind} decreased over time.

Table 53 - Estimated marginal means (Unique Errors Committed – Individual)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	2.38	0.179	57	2.026	2.74
3	1.83	0.179	57	1.470	2.19
4	1.35	0.179	57	0.992	1.71
Degrees-of-freedom method: Satterthwaite					
Confidence level used: 0.95					

Table 54 - Pairwise difference of estimated marginal means (EMMs)(Unique Errors Committed – Individual)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	0.556	0.135	232	4.12	< .05*

2 – 4	1.03	0.135	232	7.66	< .05*
3 – 4	0.479	0.135	232	3.55	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (E_{ind})

The best model, No Interaction No Feedback, excluded the effect of feedback condition (Table 126). Some of the assumptions for LMM modeling of E_{ind} were violated. REML estimation (Table 128) indicated the effect of the session order was not zero, though the R_m^2 value indicated that the session order explained only 8% of the variance. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the E_{ind} in Sessions 2, 3, and 4 were statistically significantly different (Table 54). Overall, the results suggest that the feedback did not have a significant effect on the E_{ind} or the effect is so small that more data are needed.

Analysis Summary of the Four Individual Score Components

The researcher used a weighted sum to determine the individual score. The four components had an equal-weighted value. However, there is no evidential basis to indicate that giving equal weight to the components is an optimal representation of individual performance in this task. Each component was calculated as a percentage (Table 8), and it was assumed that each component gave an accurate representation of how individuals performed. It is possible that other dependent variables more accurately represent performance.

The model selected in for the Individual Score, CI_{ind}, and E_{ind} did not include the feedback condition and only included the Session Order. The model selected for T_{ind} included the Session Order and feedback condition. No sub-model for II_{ind} was significantly different

from the null model, suggesting that neither the feedback condition nor the session order had a significant influence on II_{ind}. There was a significant difference among each Session Order for Individual Score and E_{ind}, indicating Individual Score increase over time and E_{ind} decreased over time. There was a significant difference between session 2 and 4 for CI_{ind}, indicating that CI_{ind} increased overall but only increase significantly from session 2 to session 4. Regarding T_{ind}, there was a significant difference between Team and I&T, indicating that the T_{ind} in the Team condition is greater than T_{ind} in the I&T condition.

Overall, the results suggest that T_{ind} in the Team condition was higher than T_{ind} in the I&T condition. This result contradicted H_3, which predicted that the I&T level would produce the highest performance. Regarding the remaining variables (i.e., Individual Score, CI_{ind}, II_{ind}, and E_{ind}), the results indicate that the feedback condition did not have an influence, or the effect is weak and more teams are needed to detect significance.

Summary Discussion

This chapter analyzed the main performance metrics: Individual score and Team score. This chapter also analyzed the components of the Individual and Team scores: Correct Items Collected, Incorrect Items Collected, Time Remaining, and Unique Errors Committed. The results are summarized in Table 55. The metrics analyzed in this chapter tested hypothesis three (H_1). H_1 predicted that teams that received information about I&T errors and correct items collected would perform better than teams that received information only about team or individual errors or correct items collected.

The researcher cannot make a definitive conclusion about H_1 based on the individual and team score, because the respective null hypotheses could not be rejected. However, the data did indicate that the mean (estimated marginal mean) of the time remaining (T_{ind}) with Team level feedback was statistically significantly greater than the time remaining (T_{ind}) with

I&T level feedback. This result contradicted H_1 and was unexpected since Team level feedback contained less overall information than I&T level feedback. This result suggests that feedback containing only Team content should be given to the teams while conducting a team task. Based on initial impressions of the semi-structured group interviews conducted after each study, participants often felt the Team level feedback was best because it gave them a better picture of how the team was doing and allowed each member to focus on his or her own performance. As a result, this method could lead team members to have increased confidence when collecting items and increased the rate at which items were collected, resulting in the increased time remaining.

Table 55 – Results of Chapter 4 Analysis. There were no interaction effects.

Dependent Variables Analyzed	Feedback Effect	Session Effect	R_m^2
Team Score	No	Yes	.158
Correct Team Items	No	Yes	.053
Incorrect Items (Team)	No	No	.008
Time Remaining (Team)	No	Yes	.082
Unique Errors Committed (Team)	No	Yes	.081
Individual Score	No	Yes	.089
Correct Individual Items	No	Yes	.017
Incorrect Items (Individual)	No	No	.008
Time Remaining (Individual)	Yes	Yes	.080
Unique Errors Committed (Individual)	No	Yes	.081

Did Participants Notice the Feedback?

The results of this chapter introduce some uncertainty as to whether the feedback had any impact on performance since there were no statistically significant differences in all three feedback conditions for most dependent variables. This section addresses the question of whether participants paid attention to the feedback. After each condition, participants were asked if they noticed any feedback during the task. Overall, most participants, approximately

89% in each condition, noticed the feedback during the session (Figure 29 and Table 130), suggesting that most participants noticed the feedback during the sessions. Over time, in any given condition, at least 84% of the participants noticed the feedback during the session (Figure 30 and Table 131), suggesting that most participants consistently noticed the feedback over time in each condition. Overall, in response to the questions, "Did you find the feedback helpful," at least 67% of the participants found the feedback at least somewhat helpful (Figure 31 and Table 132), suggesting that participants may have used the feedback provided to improve their performance. Over time, in any given condition, at least 56% of the participants found the feedback at least somewhat helpful (Figure 32 and Table 133), suggesting that participants may have used the feedback provided to improve their performance no matter how many times they had completed the task. While the data provided in Figure 29 to Figure 32 and Table 130 to Table 133 does not conclusively prove that the feedback had an impact on performance, it does support the idea that the participants perceived the feedback and a majority found it at least somewhat helpful.

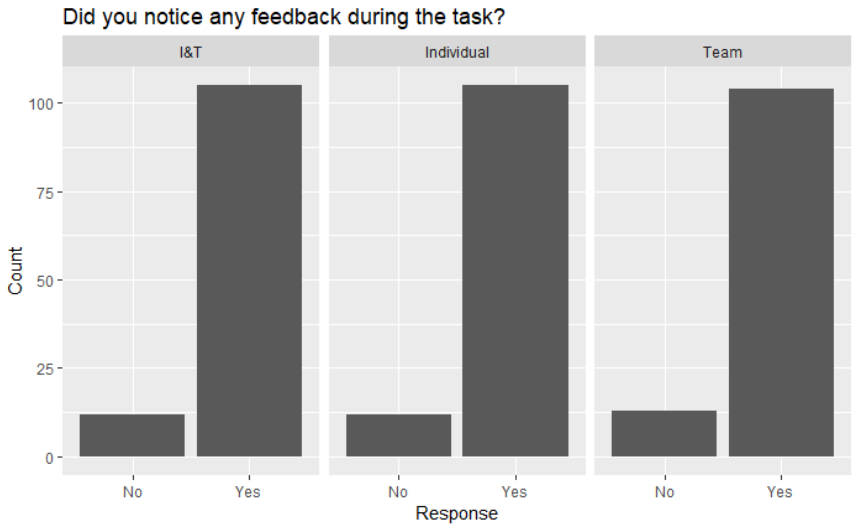


Figure 29 - Bar plot for participant response to a question after each session, “Did you notice any feedback during the task?” The data is grouped by feedback condition (*n* = 117 per plot).

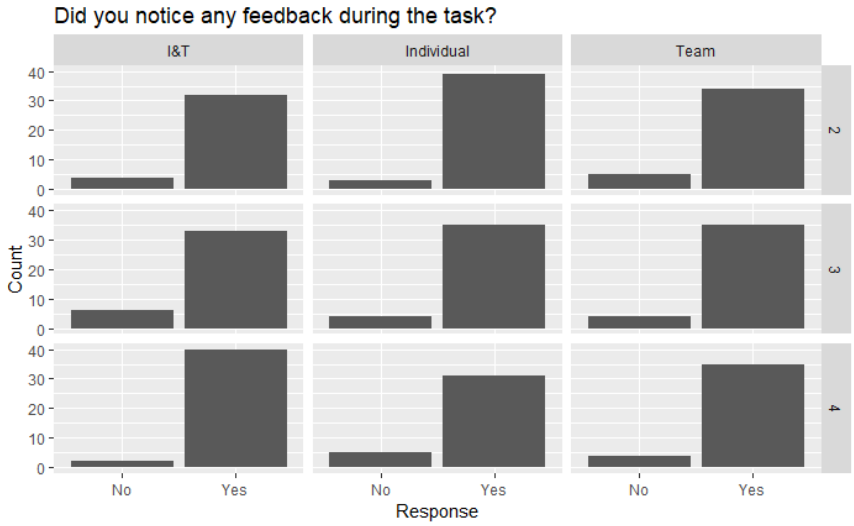


Figure 30 - Bar plot for participant response to a question, "Did you notice any feedback during the task?" The data is grouped by the feedback condition and session order (see Table 131 for *n* per plot).

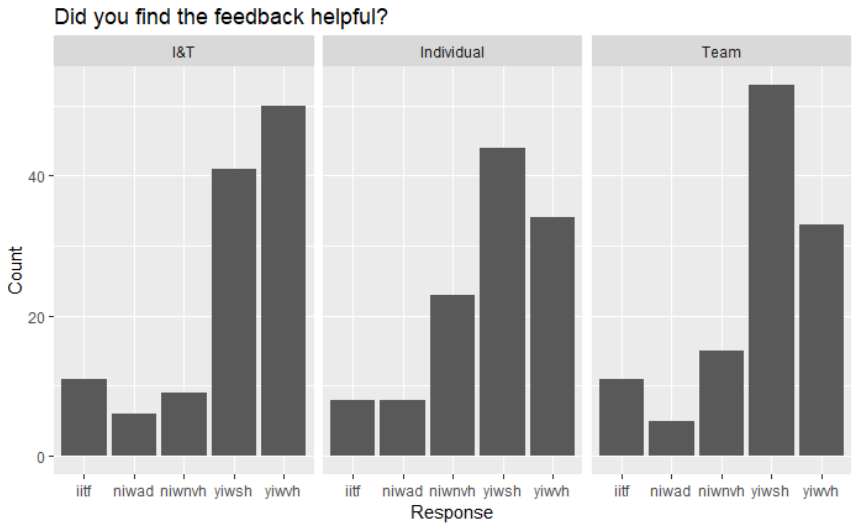


Figure 31 - Bar plot for participant responses to a question, "Did you find the feedback helpful?" The data is grouped by the feedback condition. The response abbreviations on the horizontal axis are as follows: I ignored the feedback (*iitf*), No, it was actually distracting (*niwad*), No, it was not very helpful (*miwnvh*), Yes, it was somewhat helpful (*yiwsh*), and Yes, it was very helpful (*yiwvh*) (*n* = 117 per plot).

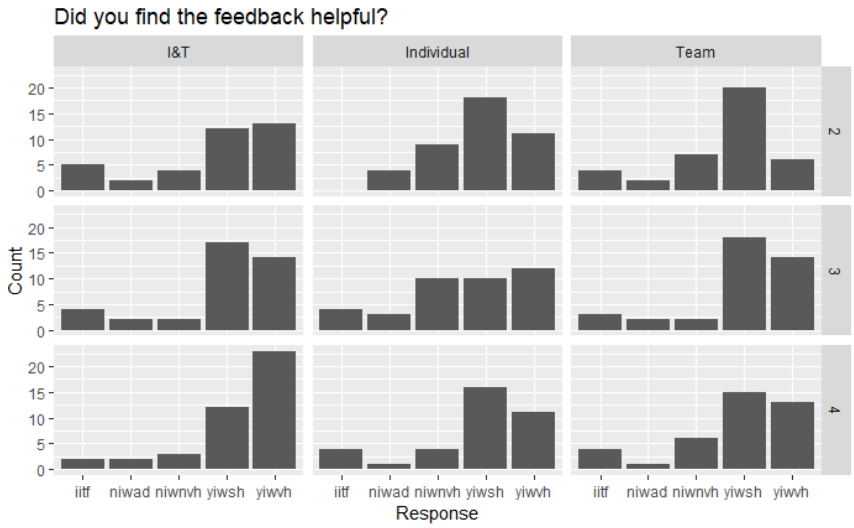


Figure 32 - Bar plot for participant response to a question, "Did you find the feedback helpful?" The data is grouped by feedback condition and session order. The response abbreviations on the horizontal axis are as follows: I ignored the feedback (*iitf*), No, it was actually distracting (*niwad*), No, it was not very helpful (*miwnvh*), Yes, it was somewhat helpful (*yiwsh*), and Yes, it was very helpful (*yiwvh*) (see Table 133 for *n* per plot).

CHAPTER 5. BEHAVIORAL AND TEAM CHARACTERISTIC METRICS ANALYSIS

This chapter explores two topics: how the feedback condition (Individual, Team, or I&T) related to the strategy used by the team, and why the time remaining for individuals (T_{ind}) in the team condition is significantly higher than in the I&T condition (a finding in Chapter 4). The author suspected that dependent measures such as time remaining might be affected by the strategy chosen by the team.

Strategy

The researcher observed two main strategies developed by each team: Go Together and Go Alone (Figure 10 and Figure 11). In general, the members of teams that used Go Alone tended to collect items and visit different stores at their own pace and complete the session on their own. The members of teams that used Go Together stayed close to their teammates throughout the session and visited each store together as a team. Some teams began with a Go Alone strategy and migrated to a Go Together strategy in later sessions. The researcher suspected that the feedback conditions might have had different influences on performance depending on the strategy used during the session because each strategy required a different type of cognitive workload based on coordination with teammates. The analysis below explores the impact that strategy had on how feedback condition affected the dependent measures.

Time Remaining

Regarding the time remaining for individuals, the above results suggested that T_{ind} is higher in the team condition when compared to the I&T condition. There are a few questions that are worth exploring. For example, T_{ind} does not tell us if the feedback condition influences T_{ind} differently depending on the strategy used by the teams. T_{ind} tells how much

time was remaining at the end of the session, but it does not tell how well the teams were able to collect correct items. T_{ind} does not tell us if the feedback condition, depending on the strategy used by the team, influences the task workload of coordination (e.g., frustration). Similarly to Chapter 4, the researcher expected the analysis to show a significant increase in performance over time because previous studies have shown that practice improves performance (Ericsson, 2008; Pusic et al., 2015). The following section examines whether the feedback condition had an effect on time remaining (T_{ind} and T_{team}), the time it took for a team and individual to collect a correct item, and the self-reported frustration levels, and the strategy used by the team. First, a new dependent measure is introduced.

Description of Collection Time Per Item

The average Collection Time per Item (CTI) was generated by the researcher to examine further how well teams and individuals collect correct items. Specifically, CTI describes the average amount of time a team or individual needed to collect a correct item. A lower value reflects better performance. The following is the equation of the CTI for teams:

$$CTI_{Team} = \frac{Total\ Duration\ Used_{Team}}{Total\ Correct\ Items\ Collected_{Team}} \quad (3)$$

Where $Duration_{Team}$ is the amount of time (in seconds) the team spent in a session, $Total\ Correct\ Items\ Collected_{Team}$ is the total number of correct items (team or individual) collected during that session by the entire team. If a team collected 0 correct items, the total duration was used (e.g., the Collection Time per Item for a team that had a duration of 462 seconds and collected no correct item is 462). The following is the equation for CTI for individuals:

$$CTI_{ind} = \frac{Total\ Duration\ Used_{ind}}{Total\ Correct\ Items\ Collected_{ind}} \quad (4)$$

Where $Duration_{ind}$ is the amount of time (in seconds) an individual spent in a session, $Total\ Correct\ Items\ Collected_{ind}$ is the total number of correct items (team or individual) collected during that session by the an individual. If an individual collected 0 correct items, the total duration was used (e.g., the Collection Time per Item for an individual that had a duration of 462 seconds and collected no correct item is 462). The units for CTI at the team and individual level is seconds per item. It is important to note that a lower CTI value corresponds to a high performing team. The researcher expected high performing teams to need less time (i.e., a lower duration value) to collect correct items (i.e., maximize the total correct items collected).

Time Remaining (Team) by Strategy

The Time Remaining (Team), or T_{team}, was a dependent variable used to measure performance and was one of the four components of Team Score. T_{team} measured the total amount of time remaining at the end of the session by the team. The analysis below explores whether T_{team} was influenced by feedback condition differently based on the strategy used by the team. Because the analysis of each variable requires several steps, the full process is documented in this chapter only for T_{team}. The complete analysis for each variable is available in APPENDIX H, however.

Distribution Overview (T_{team})

The overall distribution of T_{team} shows a skew to the left with a center a little lower than 50 for teams using Go Alone and a center around 50 for teams using Go Together (Figure 33). The distribution of the T_{team}, when grouped by Feedback, is similar to the overall

distribution for using either Go Alone or Go Together (Figure 34). Over session, the researcher noticed no obvious pattern in the distribution for teams using Go Alone (Figure 35). Over time, the researcher noticed that the distribution for T_{team} has a wider spread in the Team condition when compared to the Individual condition by session 4 for teams using Go Together (Figure 36).

Table 56 – Data for distribution overview of T_{team} divided by strategy

Strategy	N (Observations)
Go Alone	71
Go Together	46

Table 57 - Data for the distribution of T_{team} grouped by Feedback and divided by strategy

Feedback Condition	Strategy	N (Observations)
Individual	Go Alone	24
Individual	Go Together	15
I&T	Go Alone	23
I&T	Go Together	16
Team	Go Alone	24
Team	Go Together	15

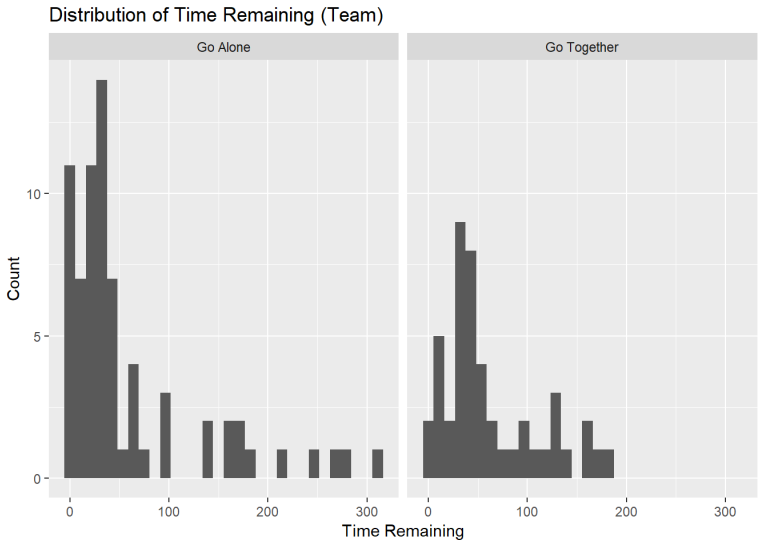


Figure 33 - Distribution overview of T_{team} divided by strategy

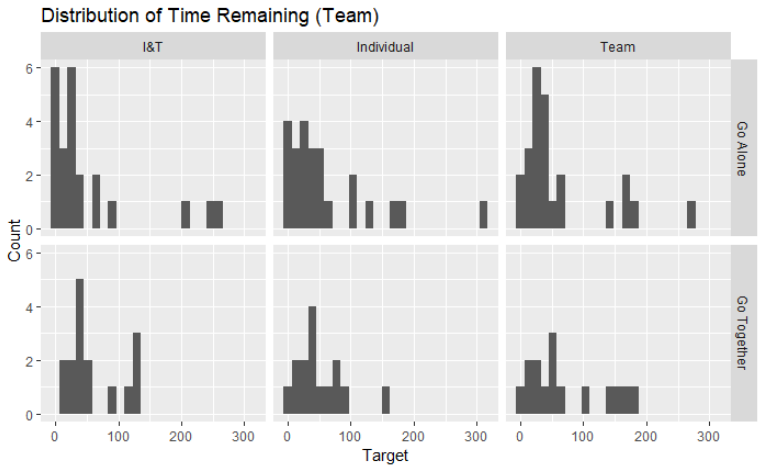


Figure 34 - Distribution of T_{team} grouped by Feedback and divided by strategy

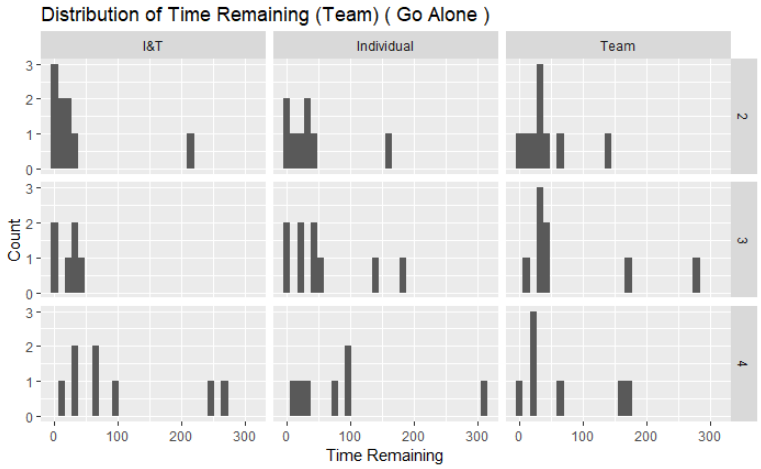


Figure 35 - Distribution of T_{team} grouped by Feedback and Session Order for teams using Go Alone

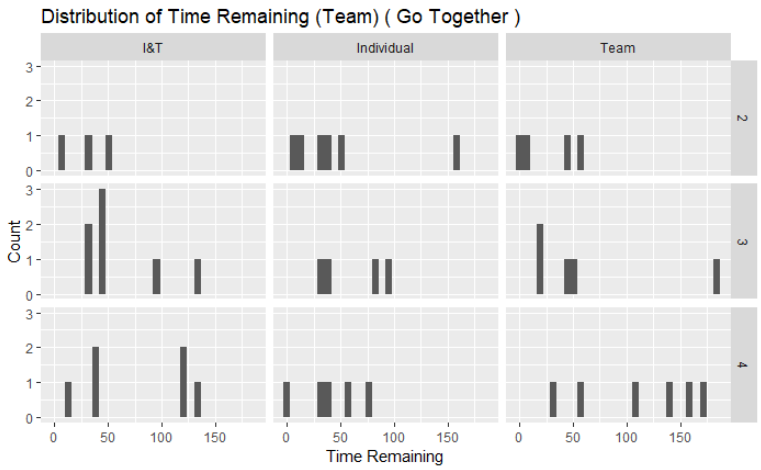


Figure 36 – Distribution of T_{team} grouped by Feedback and Session Order for teams using Go Together

Model Selection (T_{team})

Go Alone

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 58. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect was negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible, and the simplest model that described the data best was the No Interaction No Feedback model.

Table 58 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	768.57	793.46
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	762.53	778.37
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**759.48**	**770.79**

Go Together

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 59. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect was negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible, and the simplest model that described the data best was the No Interaction No Feedback model.

Table 59 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	484.83	504.95
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	483.21	496.01
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**481.46**	**490.61**

Testing Assumptions: Residuals (T_{team})

Go Alone

The histogram of the residuals (Figure 37) appeared to be approximately distributed. The residual fitted-value (Figure 38) showed a violation of constant variance. The points on the residual normal Q-Q plot (Figure 45) do not lay in a roughly straight line. The classic LMM and the robust LMM produced different coefficient values (Table 23), so the researcher used square root to transform the data.

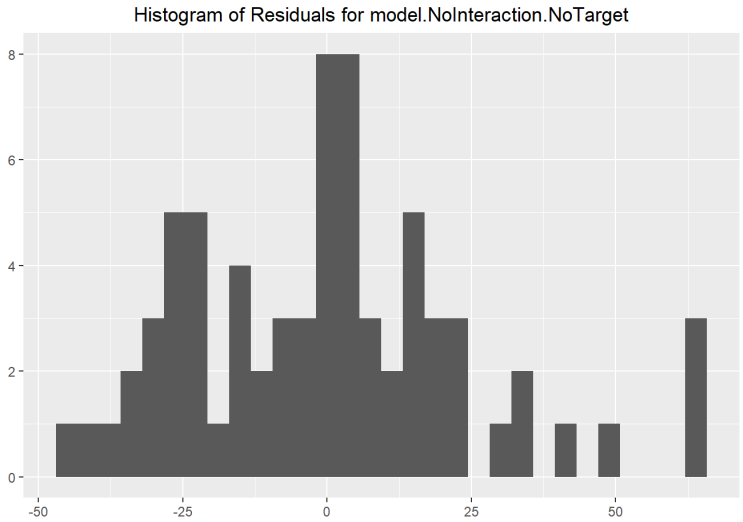


Figure 37 - Histogram of Residuals (T_{team})

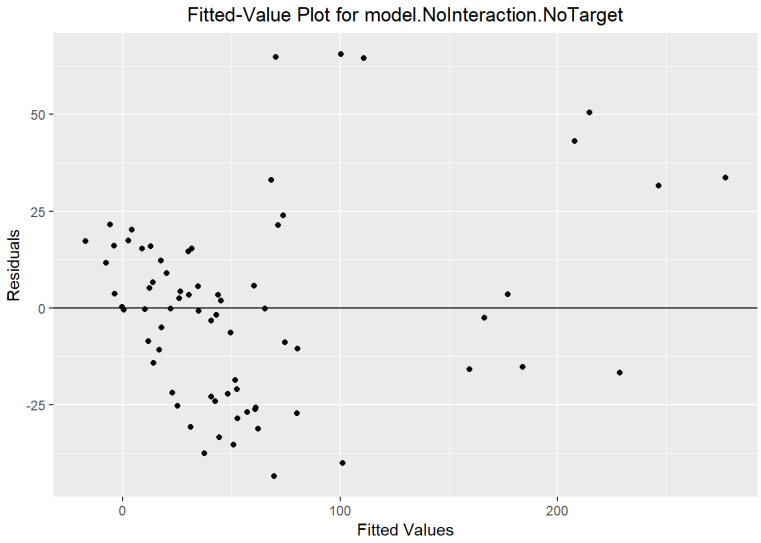


Figure 38 - Residual fitted plot (T_{team})

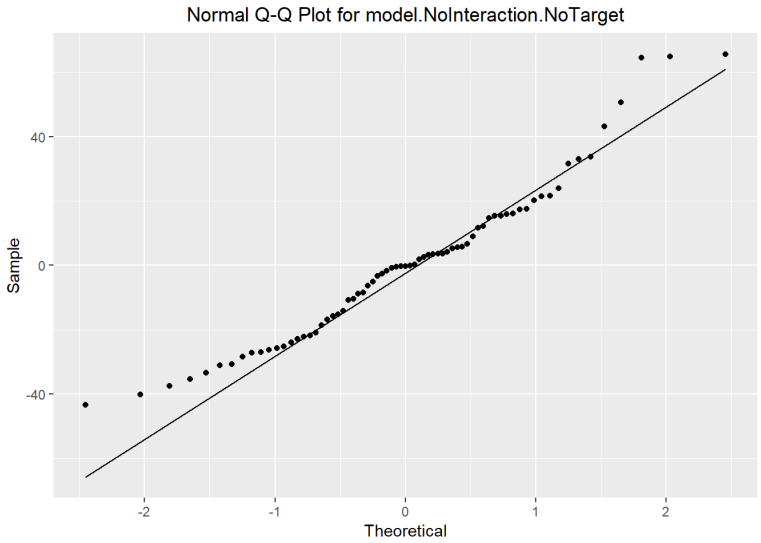


Figure 39 - Normal Q-Q plot for residuals (T_{team})

Table 60 - Comparing classic model to a robust model

Coefficient	**No Interaction No Feedback**	**Robust No Interaction No Feedback**
Intercept	37.9	24.0
SessionOrder: 3	17.8	14.4
SessionOrder: 4	48.5	38.8

Go Together

The histogram of the residuals (Figure 46) appeared to be approximately distributed. The residual fitted-value (Figure 47) showed a violation of constant variance. The points on the residual normal Q-Q plot (Figure 42) do not lay in a roughly straight line. The classic LMM and the robust LMM produced similar coefficient values (Table 61), so the researcher did not transform the data.

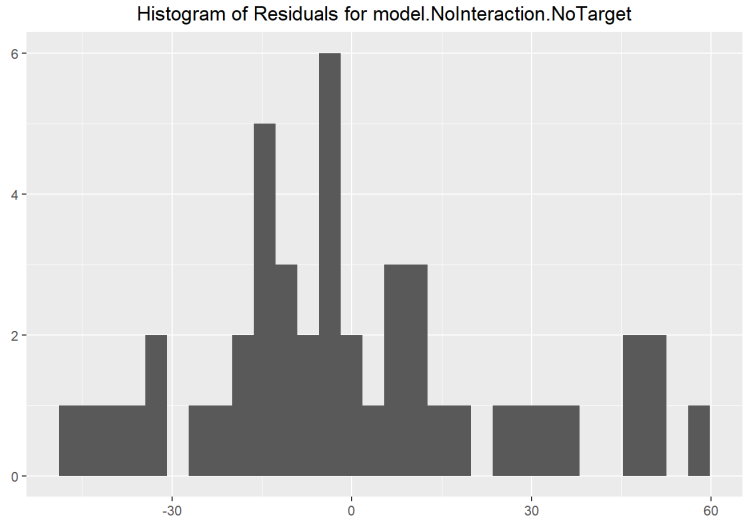


Figure 40 - Histogram of Residuals (T_{team})

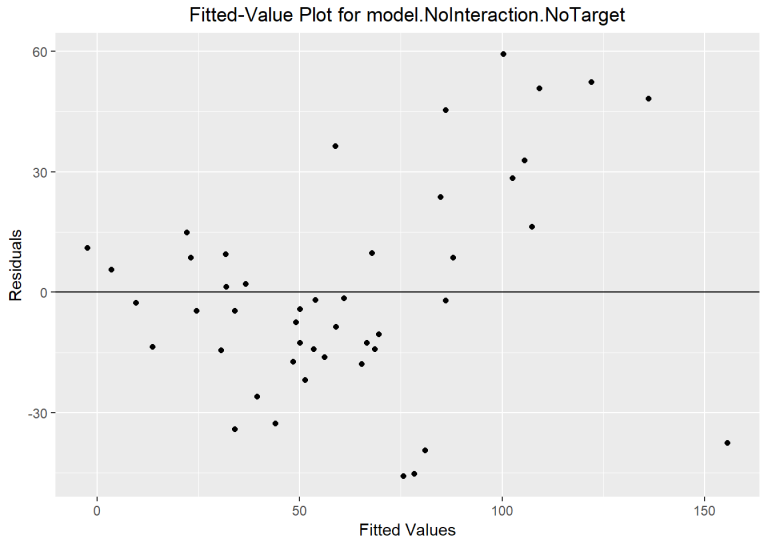


Figure 41 - Residual fitted plot (T_{team})

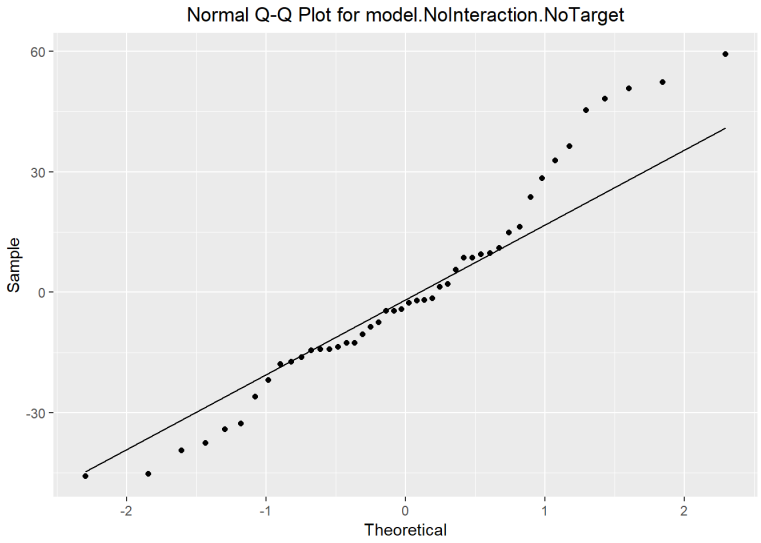


Figure 42 - Normal Q-Q plot for residuals (T_{team})

Table 61 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	31.1	29.2
SessionOrder: 3	27.0	26.8
SessionOrder: 4	46.4	43.0

Restricted Maximum Likelihood Estimation (T_{team})

Go Alone

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 62. The results showed that the effects of Session 3 and 4 are significant in reference to Session 2, the reference category.

Table 62 - Fixed effect result of REML Estimation (T_{team})

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	4.91	0.792	37	6.20	< .05*
Session: 3	1.44	0.612	45	2.35	< .05*
Session: 4	3.29	0.622	45	5.29	< .05*

Go Together

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 100. The results showed that the effects of Session 3 and 4 are significant in reference to Session 2, the reference category.

Table 63 - Fixed effect result of REML Estimation (T_{team})

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	31.1	12.6	35	2.47	< .05*
Session: 3	27.0	12.3	28	2.19	< .05*
Session: 4	46.4	12.3	28	3.79	< .05*

Evaluating Effect size (T_{team})

Go Alone

The coefficient of determination, R-squared (R^2), value is used to report the goodness-of-fit of a regression line (Nakagawa & Schielzeth, 2013). There are traditional properties of R^2, such as the value should not rely on a particular unit of measure (Orelien & Edwards, 2008). However, there are problems that arise when seeking to determine the R^2 for linear mixed-models. An issue that contributes to the problem of calculating R^2 is that

researchers have not agreed on a single definition of R^2 (Nakagawa & Schielzeth, 2013). Johnson (2014) offers an extension of a simple method, developed by Nakagawa and Schielzeth (2013), to generate R^2 for LMMs. This method generates two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables (session order) while the R_c^2 describes how the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 are presented in Table 64. The results suggest that the fixed variables have a small effect size (i.e., explain 8% of the variance) and the fixed and random variables have a large effect size (i.e., explain 81% of the variance).

Table 64 - Effect size for LMM (T_{team})

R^2 Type	Value
R_m^2	0.079
R_c^2	0.814

Go Together

The values for R_m^2 and R_c^2 are presented in Table 102. The results suggest that the fixed variables have a medium effect size (i.e., explain 13% of the variance) and the fixed and random variables have a large effect size (i.e., explain 62% of the variance).

Table 65 - Effect size for LMM (T_{team})

R^2 Type	Value
R_m^2	0.133
R_c^2	0.618

Estimated Marginal Means (T_{team})

Go Alone

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 66) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 104). The pairwise comparison showed a statistical difference between each session. The statistical difference suggests that the T_{team} improved over time.

Table 66 - Estimated marginal means (T_{team})

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	4.91	0.792	36	3.30	6.52
3	6.35	0.817	39	4.69	8.00
4	8.20	0.825	40	6.53	9.87
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 67 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-1.44	0.613	44	-2.35	< .05*
2 – 4	-3.29	0.623	44	-5.28	< .05*
3 – 4	-1.85	0.618	42	-2.99	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Go Together

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 105) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 106). The pairwise comparison showed a statistical difference between each session 2 and 4. The statistical difference suggests that the T_{team} improved overall but not during session 3.

Table 68 - Estimated marginal means (T_{team})

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	31.1	12.6	35	5.40	56.7
3	58.1	11.8	31	34.0	82.2
4	77.5	11.6	30	53.9	101.1
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 69 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-27.0	12.4	28	-2.18	.092
2 – 4	-46.4	12.3	29	-3.76	< .05*
3 – 4	-19.4	11.2	27	-1.73	.212
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (T_{team})

The best model both strategies, No Interaction No Feedback, excluded the effect of feedback condition (Table 58 and Table 59). Some of the assumptions for LMM modeling of T_{team} were violated for Go Alone and Go Together. REML estimation (Table 62 and Table 63) indicated that the effect of the session order for Go Alone and Go Together was not zero. The R_m^2 value for Go Alone and Go Together indicated that session order explained 8% and 13% of the variance, respectively. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the T_{team} for Go Alone in Sessions 2, 3, and 4 were statistically significantly different and T_{team} for Go Together in Sessions 2 and 4 were statistically significantly different.

The feedback did not have a significant effect on the $T_{team,}$ or the effect is so small (or the variance so large) that more data are needed. The next sections present the analysis of the Collection Time per Item. These analyses are presented in a shorter form, but the full analysis is available in APPENDIX H.

Collection Time per Item (Team) by Strategy

The Collection Time per Item (Team), or CTI_{team}, was a dependent variable used to measure performance.

Estimated Marginal Means (CTI_{team})

Go Alone

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 70) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not

directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 71). The pairwise comparison showed a statistical difference between session 2 and 4 and between session 3 and 4. The statistical difference suggests that the CTI_{team} generally improved over time.

Table 70 - Estimated marginal means (CTI_{team})

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	18.9	0.619	37	17.7	20.2
3	18.2	0.641	41	16.9	19.5
4	16.6	0.647	42	15.3	17.9
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 71 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	0.726	0.509	44	1.43	.336
2 – 4	2.31	0.517	44	4.47	< .05*
3 – 4	1.59	0.514	42	3.09	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Go Together

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 72) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis

was used to compare the differences in the means across session (Table 73). The pairwise comparison showed a statistical difference between session 2 and 3 and between session 2 and 4. The statistical difference suggests that the CTI_{team} generally improved over time.

Table 72 - Estimated marginal means (CTI_{team})

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	21.6	1.05	27	19.5	23.8
3	18.6	1.01	24	16.6	20.7
4	16.9	0.991	23	14.9	19.0
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 73 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	2.97	0.805	27	3.69	< .05*
2 – 4	4.68	0.803	27	5.83	< .05*
3 – 4	1.71	0.724	26	2.36	.065
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (CTI_{team})

The best model for both strategies, No Interaction No Feedback, excluded the effect of feedback condition (Table 134 and Table 135). Some of the assumptions for LMM modeling of CTI_{team} were violated for Go Alone and Go Together. REML estimation for Go Alone (Table 138) indicated the effect of session 4 was not zero, and the REML estimation for Go Together (Table 139) indicated the effect of the session order was not zero. The R_m^2 value for Go Alone and Go Together indicated that session order explained 8% and 17% of the variance, respectively. Other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison

of the EMMs indicated that the CTI_{team} for Go Alone in Sessions 2 significantly different from session 4, and session 3 was significantly different from 4, and CTI_{team} for Go Together in Sessions 2 was significantly different from session 3, and session 2 was significantly different from session 4.

The feedback did not have a significant effect on the $CTI_{team,}$ or the effect is so small (or the variance so large) that more data are needed. The next sections present the analysis of the Time Remaining.

Analysis Summary of Feedback on Variables Based on Strategy at the Team Level

The researcher examined two dependent variables based on the strategy used by teams: T_{team} and CTI_{team}. The researcher focused on T_{team} and CTI_{team} to further explore the significant results discovered in the analysis of the main variables. The model selected for the T_{team} and CTI_{team} for teams using either Go Alone or Go Together did not include the feedback condition and only included the Session Order. For T_{team}, there was a significant difference among each session for teams using Go Alone, and there was a significant difference between sessions 2 and 4 for teams using Go Together, indicating that T_{team} increased over time for teams in both strategy groups.

Overall, the results suggest that the feedback condition had little influence on the T_{team} and CTI_{team}. The results indicate that a model only including the session order accurately described the data better than models including the feedback condition. This result could indicate that the feedback condition did not influence T_{team} or CTI_{team} or that the effect is weak, and more teams are needed to detect significance.

Time Remaining (Individual) by Strategy

The Time Remaining (Individual), or T_{ind}, was a dependent variable used to measure performance. It is possible that the results for T_{ind} will differ from the above results for $T_{team.}$

Estimated Marginal Means (T_{ind})

Go Alone

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 74) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 75). The pairwise comparison showed a statistical difference between each session. The statistical difference suggests that the T_{ind} improved over time.

Table 74 - Estimated marginal means (T_{ind})

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	5.47	0.747	30	3.95	7.00
3	7.54	0.759	32	5.99	9.08
4	9.23	0.762	33	7.67	10.78
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 75 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-2.07	0.400	136	-5.17	< .05*
2 – 4	-3.75	0.407	137	-9.23	< .05*
3 – 4	-1.69	0.402	134	-4.21	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Go Together

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 74) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session and feedback condition (Table 77 and Table 79). However, the results may be misleading due to the presence of an interaction term. The presence of an interaction term means that the researcher should always consider the effect of both terms (i.e., session order and feedback condition). A pairwise difference analysis was used to compare the differences in the means across the feedback condition based on the session order (Table 81). The pairwise comparison showed a statistical difference between in session 4 between Team and Individual condition and between Team and I&T condition. The statistical difference suggests that the T_{ind} (square root) in the Team condition in session 4 is significantly greater than T_{ind} (square root) in the Individual or I&T condition.

Table 76 - Estimated marginal means (T_{ind}). Note: result may be misleading, due to the presence of the interaction term.

Estimated Marginal Means					
Feedback	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
Individual	6.38	0.693	19	4.93	7.82
Individual and Team	6.59	0.695	20	5.14	8.04
Team	7.01	0.692	19	5.57	8.46
Results are averaged over the levels of Session Order					
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 77 - Pairwise difference of estimated marginal means (EMMs). Note: result may be misleading, due to the presence of the interaction term.

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
Individual – Individual and Team	-0.217	0.315	80	-0.690	.770
Individual – Team	-0.638	0.326	81	-1.95	.131
Individual and Team – Team	-0.420	0.323	80	-1.30	.398
Results are averaged over the levels of Session Order					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Table 78 - Estimated marginal means (T_{ind}). Note: result may be misleading, due to the presence of the interaction term.

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	5.00	0.704	21	3.53	6.46
3	6.97	0.691	20	5.52	8.41
4	8.02	0.685	19	6.58	9.45
Results are averaged over the levels of Feedback					
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 79 - Pairwise difference of estimated marginal means (EMMs). Note: result may be misleading, due to the presence of the interaction term.

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	-1.97	0.333	80	-5.92	< .05*
2 – 4	-3.02	0.336	81	-8.99	< .05*
3 – 4	-1.05	0.297	80	-3.52	< .05*
Results are averaged over the levels of Feedback					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Table 80 - Estimated marginal means (T_{ind})

Estimated Marginal Means					
Session 2					
Feedback	*EMM*	*Standard Error*	*Degree of Freedom*	*Lower CL*	*Upper CL*
Individual	5.14	0.768	28	3.56	6.71
I&T	5.17	0.872	43	3.41	6.92
Team	4.69	0.851	38	2.97	6.41
Session 3					
Feedback	*EMM*	*Standard Error*	*Degree of Freedom*	*Lower CL*	*Upper CL*
Individual	7.00	0.838	37	5.30	8.70
I&T	7.22	0.764	27	5.65	8.78
Team	6.69	0.793	32	5.07	8.31
Session 4					
Feedback	*EMM*	*Standard Error*	*Degree of Freedom*	*Lower CL*	*Upper CL*
Individual	6.99	0.810	33	5.34	8.64
I&T	7.40	0.780	29	5.80	8.99
Team	9.66	0.783	30	8.06	11.26
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 81 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Session 2					
Contrast	*Estimate*	*Standard Error*	*Degree of Freedom*	*T ratio*	*P-value*
Individual – Individual and Team	-0.028	0.731	82	-0.038	.999
Individual – Team	0.450	0.732	86	0.614	.813
Individual and Team – Team	0.477	0.835	85	0.572	0.836
Session 3					
Contrast	*Estimate*	*Standard Error*	*Degree of Freedom*	*T ratio*	*P-value*
Individual – Individual and Team	-0.215	0.722	86	-0.298	.952
Individual – Team	0.313	0.710	83	0.440	.899
Individual and Team – Team	0.528	0.648	84	0.814	.695
Session 4					
Contrast	*Estimate*	*Standard Error*	*Degree of Freedom*	*T ratio*	*P-value*
Individual – Individual and Team	-0.409	0.700	86	-0.584	.829
Individual – Team	-2.68	0.691	86	-3.87	< .05*
Individual and Team – Team	-2.27	0.664	86	-3.42	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (T_{ind})

The best model for Go Alone teams, No Interaction No Feedback, excluded the effect of feedback condition (Table 142) and the best model for Go Together teams, All, included the interaction effect (Table 143). Some of the assumptions for LMM modeling of T_{ind} were violated for Go Alone and Go Together teams. The researcher used square root to transform data for Go Alone and Go Together. REML estimation for Go Alone (Table 146) indicated the effect of the session order was not zero, and the REML estimation for Go Together (Table 147) indicated the effect of the session order was not zero and the effect of Team feedback in session 4 was not zero. The R_m^2 value for Go Alone indicated that the fixed effects explained 8% of the variance and the value for Go Together indicated that the fixed effects explained 24% of the variance. For Go Alone, other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the T_{ind} for Go Alone in all session were statistically significantly different for each other. A comparison of the EMMs indicated that the T_{ind} for Go Together in the Team feedback condition in session four was significantly greater than T_{ind} in the Individual or I&T condition.

For teams that used Go Alone, the feedback did not have a significant effect on the T_{ind} or the effect is so small (or the variance so large) that more data are needed. For teams that used Go Together, the results suggest that by session 4, T_{ind} is significantly greater in the Team condition than the Individual or I&T condition. This result contradicted H_1, which predicted that the I&T level would produce the highest performance. The next sections present the analysis of the Collection Time per Item.

Collection Time per Item (Individual) by Strategy

The Collection Time per Item (Individual), or CTI_{ind}, was a dependent variable used to measure performance.

Estimated Marginal Means (CTI_{ind})

Go Alone

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 82) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 83). The pairwise comparison showed a statistical difference between session 2 and 4 and between session 3 and 4. The statistical difference suggests that the CTI_{ind} generally improved over time.

Table 82 - Estimated marginal means (CTI_{ind})

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	59.1	2.35	40	54.4	63.9
3	55.2	2.44	45	50.2	60.1
4	49.5	2.47	46	44.5	54.5
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 83 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	3.98	2.05	137	1.94	.131
2 – 4	9.62	2.08	137	4.62	< .05*
3 – 4	5.64	2.07	132	2.73	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Go Together

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 84 and Table 86) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Team, Individual, and I&T). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 85 and Table 87). The pairwise comparison showed a statistical difference between the Team and Individual feedback condition. The statistical difference suggests that CTI_{ind} is greater in the Team condition when compared to the Individual feedback condition. The pairwise comparison also showed a statistically significant difference between session 2, 3, and 4. The statistical difference suggests that the CTI_{ind} improved over time.

Table 84 - Estimated marginal means (CTI_{ind})

Estimated Marginal Means					
Feedback	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
Individual	60.0	2.91	22	54.0	66.0
Individual and Team	58.5	2.89	22	52.5	64.5
Team	54.6	2.91	22	48.6	60.6
Results are averaged over the levels of SessionOrder					
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 85 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
Individual – Individual and Team	1.49	1.79	84	0.828	.687
Individual – Team	5.41	1.88	86	2.87	< .05*
Individual and Team – Team	3.92	1.76	83	2.23	.072
Results are averaged over the levels o SessionOrder					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Table 86 - Estimated marginal means (CTI_{ind})

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	65.5	2.97	24	59.3	71.6
3	56.4	2.89	22	50.5	62.4
4	51.2	2.86	21	45.2	57.1
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 87 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	9.04	1.90	84	4.75	< .05*
2 – 4	14.3	1.90	85	7.53	< .05*
3 – 4	5.26	1.67	83	3.15	< .05*
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (CTI_{ind})

The best model for Go Alone teams, No Interaction No Feedback, excluded the effect of feedback condition (Table 150) and the best mode for Go Together teams, No Interaction, excluded the interaction effect (Table 151). Some of the assumptions for LMM modeling of CTI_{ind} were violated for Go Alone and Go Together. The researcher did not transform data for Go Alone or Go Together because the classic model and the robust model produced similar coefficient values for fixed effects. REML estimation for Go Alone (Table 154) indicated the effect of session 4 was not zero, and the REML estimation for Go Together (Table 155) indicated the effect of the session order was not zero and the effect of Team feedback condition was not zero. The R_m^2 value for Go Alone indicated that the fixed effects explained 4% of the variance and the value for Go Together indicated that the fixed effects explained 17% of the variance. For Go Alone, other variation must be due to factors other than session order or feedback condition, since the best model did not include feedback condition. A comparison of the EMMs indicated that the CTI_{ind} for Go Alone in session 2 and 4 were statistically different and session 3 and 4 were statistically significantly different for each other. A comparison of the EMMs indicated that the CTI_{ind} for Go Together in the Team feedback condition was significantly greater than CTI_{ind} in the Individual condition and marginally significantly different (p = .07) from I&T condition.

For teams that used Go Alone, the feedback did not have a significant effect on the CTI_{ind} or the effect is so small (or the variance so large) that more data are needed. For teams that used Go Together, the results suggest that CTI_{ind} is significantly greater in the Team condition than the Individual. This result contradicted H_1, which predicted that the I&T level

would produce the highest performance. The next sections present the analysis of self-reported Frustration.

Frustration by Strategy

The Frustration was a dependent variable used to report the level of frustration experienced by each participant. It was one construct of the NASA-TLX survey that was given to participants after each session and ranged from 0 – 100 (Where 0 means very low frustration and 100 means very high frustration).

Estimated Marginal Means (Frustration)

Go Alone

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 88) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Session 2, 3, and 4). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across session (Table 89). The pairwise comparison showed no statistical difference among any of the sessions.

Table 88 - Estimated marginal means (Frustration)

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	47.4	3.33	45	40.7	54.1
3	44.9	3.47	50	37.9	51.8
4	40.4	3.52	52	33.3	47.4
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 89 - Pairwise difference of estimated marginal means (EMMs)

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	2.57	3.17	138	0.810	.697
2 – 4	7.05	3.22	138	2.19	< .05*
3 – 4	4.48	3.21	132	1.39	.347
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Go Together

Estimated marginal means (EMMs) were calculated to analyze the differences among groups for the variables feedback condition and session order. The EMMs were generated (Table 90 and Table 92) using the emmean package in R (Lenth, 2016) to analyze the means of each level (i.e., Individual, Team, and I&T). It is important to note that the EMMs are based on the model, not directly on the data. Thus, the standard error of those EMMs is the same for each one, since an assumption of the model is that errors are homogeneous. A pairwise difference analysis was used to compare the differences in the means across feedback conditions (Table 91). The pairwise comparison showed a significant difference between Individual and I&T feedback condition and between Individual and Team condition, indicating Frustration was higher in the Individual condition. A pairwise difference analysis was used to compare the differences in the means across session (Table 93). The pairwise comparison showed a statistical difference between session 2 and 4, suggesting that Frustration was reduced over time.

Table 90 - Estimated marginal means (Frustration).

Estimated Marginal Means					
Feedback	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
Individual	36.3	3.55	30	29.0	43.6
Individual and Team	27.6	3.51	29	20.4	34.8
Team	25.3	3.57	30	18.0	32.5
Results are averaged over the levels of Session Order					
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 91 - Pairwise difference of estimated marginal means (EMMs).

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
Individual – Individual and Team	8.72	3.09	83	2.82	< .05*
Individual – Team	11.0	3.22	87	3.42	< .05*
Individual and Team – Team	2.31	3.03	83	0.760	.728
Results are averaged over the levels of Session Order					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Table 92 - Estimated marginal means (Frustration).

Estimated Marginal Means					
Session	EMM	Standard Error	Degree of Freedom	Lower CL	Upper CL
2	33.3	3.71	34	25.8	40.8
3	30.4	3.50	28	23.2	37.6
4	25.4	3.44	27	18.4	32.5
Results are averaged over the levels of Feedback					
Degrees-of-freedom method: Satterthwaite					
Confidence level (CL) used: 0.95					

Table 93 - Pairwise difference of estimated marginal means (EMMs).

Pairwise Difference					
Contrast	Estimate	Standard Error	Degree of Freedom	T ratio	P-value
2 – 3	2.91	3.27	85	0.889	.649
2 – 4	7.86	3.25	87	2.42	< .05*
3 – 4	4.95	2.89	82	1.71	.207
Results are averaged over the levels of Feedback					
P-value adjustment: Tukey method for comparing a family of 3 estimates					

Analysis Summary (Frustration)

There were no models significantly better than the Null model for Go Alone, but the researcher analyzed the No Interaction No Feedback model because it was the simplest model and was marginally significantly better than the Null model (*p = .088*) (Table 158). The best model for Go Together, No Interaction, excluded the interaction effect (Table 159). Some of the assumptions for LMM modeling of Frustration were violated for Go Alone and Go Together. The researcher did not transform the data because the classic model and the robust model produced similar coefficient values for Go Alone and Go Together. The R_m^2 value for Go Alone indicated that the fixed effects explained 1% of the variance and the value for Go Together indicated that the fixed effects explained 7% of the variance. A comparison of the EMMs indicated that the Frustration for Go Alone in session 2 and 4 were statistically significantly different for each other. A comparison of the EMMs indicated that the Frustration for Go Together in the Team and I&T feedback condition was significantly smaller than Frustration in the Individual condition.

For teams that used Go Alone, the results indicate that feedback and session order did not have a significant effect on the Frustration or the effect is so small (or the variance so large) that more data are needed. For teams that used Go Together, the results suggest that Frustration is significantly greater in the Individual condition when compared to the Team or

I&T condition. This result partially supports H_1, which predicted that the I&T level would produce the highest performance. The result only partially supports H_1 because lower Frustration does not necessarily mean higher performance.

Analysis Summary of Feedback on Variables Based on Strategy at the Individual Level

The researcher examined T_{ind} and CTI_{ind} based on the strategy used by teams and participants self-reported frustration towards the task. The researcher focused on T_{ind}, CTI_{ind}, and Frustration to further explore the significant results discovered in the analysis of the main variables.

The model selected for T_{ind} for teams that used Go Alone did not include the feedback condition and only included the Session Order. The model selected for T_{ind} for teams that used Go Together included feedback condition, session order, and the interaction effect. For teams that used Go Alone, there was a significant difference among each session, indicating that T_{ind} increased over time. For teams that used Go Together, in session 4, T_{ind} was significantly greater than the Individual and I&T condition.

The model selected for CTI_{ind} for teams that used Go Alone did not include the feedback condition and only included the session order. The mode selected for CTI_{ind} for teams that used Go Together included feedback condition and session order. For teams that used Go Alone, there was a significant difference between session 2 and 4, and session 3 and 4, indicating that CTI_{ind} decreased over time. For teams that used Go Together, there was a significant difference between the Individual and Team condition and marginally significant difference between I&T and Team condition (p = .07), indicating that CTI_{ind} in the Team condition is significantly smaller than in the Individual condition and marginally significantly (p = .07) smaller than in the I&T condition.

Overall, the results suggest that the feedback condition had little influence on T_{ind} or CTI_{ind} when teams used Go Alone, but it did have an influence on T_{ind} or CTI_{ind} when teams used Go Together. Specifically, T_{ind} in session 4 when teams used Go Together was significantly higher in the Team condition when compared to Individual and I&T condition. CTI_{ind} when teams used Go Together was significantly smaller in the Team condition when compared to Individual condition and marginally significantly (p = .07) smaller in the Team condition when compared to I&T condition.

Summary Discussion

This chapter analyzed two performance metrics, *T* and *CTI*, and self-reported Frustration. This chapter analyzed these metrics grouped by the Go Alone and Go Together strategy to see whether the strategy differentially affected the impact of feedback condition on these metrics. The metrics analyzed in this chapter test hypothesis three (H_1). H_1 predicted that teams that received I&T information about errors and correct items collected would perform better than teams that received information only about team or individual errors or correct items collected. Table 94 displays the effects included in the model that described the data for a given dependent variable.

The researcher cannot make a definitive conclusion about H_1 based on T_{team} or CTI_{team} for teams that used Go Alone or Go Together because the respective null hypotheses could not be rejected. The researcher also could not make a definitive conclusion about H_1 based on T_{ind}, CTI_{ind}, or Frustration for teams that used Go Alone because the respective null hypotheses could not be rejected.

However, the data indicated that the mean (estimated marginal mean) of T_{ind} for teams that used Go Together in session 4 was significantly greater in the Team condition than in the Individual and I&T conditions. This result contradicted H_1 and was unexpected

since Team level feedback contained less overall information than I&T level feedback. The result suggests that feedback containing Team information could be more effective than the other conditions for teams that have experience with the task.

Table 94 – Results of Chapter 5 Analysis.

Dependent Variables	Strategy	Feedback Effect	Session Effect	Interaction Effect	R_m^2
Time Remaining (Team)	Go Alone	No	Yes	No	.079
	Go Together	No	Yes	No	.133
Collection Time per Item (Team)	Go Alone	No	Yes	No	.084
	Go Together	No	Yes	No	.173
Time Remaining (Individual)	Go Alone	No	Yes	No	.078
	Go Together	Yes	Yes	Yes	.236
Collection Time per Item (Individual)	Go Alone	No	Yes	No	.044
	Go Together	Yes	Yes	No	.173
Frustration (Individual)	Go Alone	No	Yes	No	.011
	Go Together	Yes	Yes	No	.069

The data also indicated that the mean (estimated marginal mean) of CTI_{ind} for teams that used Go Together was significantly lower in the Team condition than the Individual condition and marginally significantly (p = .07) (Table 85) lower than the I&T condition. This result contradicted H_1 and was unexpected since Team level feedback contained less overall information than I&T level feedback. The result suggests that feedback containing Team information should be given to teams while conducting a team task.

Finally, the data indicated that the mean (estimated marginal mean) of Frustration for teams that used Go Together was significantly higher in the Individual condition than in the Team or I&T conditions. This result indirectly supports H_1 because the researcher expected

Frustration to be lower in the I&T condition because more information is given to the participants. The result suggests that feedback containing Team or I&T information should be given to teams while conducting a team task to reduce frustration.

CHAPTER 6. PERCEPTION OF INDIVIDUAL AND TEAM PERFORMANCE ANALYSIS

This chapter explores team's perception of performance to analyze the influence of feedback. Data were collected from an overall post-survey given to participants after all the sessions were completed, as well as from a post-session survey given after each session (APPENDIX E and APPENDIX C respectively). These data are used to explore how participants perceived their own performance and team performance and analyze if their perception correlated with their actual performance.

Focus of Analysis

The focus of this analysis is to examine participants' perception of their own performance and their team's performance. The metrics used to represent individual and team performance were time remaining and collection time per item, introduced in Chapter 5. The next sections examine the correlations between those metrics and self-reported perception of performance at the individual level (NASA-TLX) and team level (post-session survey).

Individual Perception Versus Individual Performance

This section is focused on answering the question of how well did a participant's perception of his or her own performance correlate with actual performance. The analysis plotted the self-reported performance from the TLX survey against the time remaining (T_{ind}) and collection item per item (CTI_{ind}). Spearman's rank-order correlation was used to assess the relationship between performance metrics (T_{ind} and CTI_{ind}) and self-reported performance. The self-reported TLX values for performance were reversed coded to reduce confusion since the original TLX uses 0 for success and 100 for failure. As a result, a value of 100

indicated the participant felt their performance was perfect, and 0 indicated the participant felt they failed.

For T_{ind}, a positive correlation with self-reported performance indicated a correct perception of individual performance (i.e., as T_{ind} increase, so did the self-reported performance value), whereas a negative correlation indicated an incorrect perception of individual performance (i.e., as T_{ind} increased, the self-reported performance value decreased). For CTI_{ind}, a negative correlation with self-reported performance indicated a correct perception of individual performance (i.e., as CTI_{ind} decreased, the self-reported performance value increased), whereas a positive correlation with self-reported performance indicated an incorrect perception of individual performance (i.e., as CTI_{ind}, decreased, the self-reported performance value decreased). The next sections present the results.

Individual Perception of Performance: T_{ind}

The data were analyzed in several groupings based on what metrics might be correlated with the NASA-TLX performance rating: feedback condition, session order, and participants' own self-response on the post-session survey (Very poor to Excellent). Findings are summarized in Table 95. When the data were grouped by feedback condition, the results showed a significant positive correlation at the Individual ($r_s(115)= .258$, $p < .05$) and Team ($r_s(115)= .235$, $p < .05$) condition, indicating a correct perception of individual performance. When the data were grouped by session order, the results showed a significant positive correlation in session 2 ($r_s(115) = .252$, $p < .05$) and session 4 ($r_s(115) = .193$, $p < .05$), indicating a correct perception of individual performance. The results also showed a marginally significant positive correlation in session 3 ($r_s(115) = .157$, $p = .091$). When the data were grouped by survey response performance (i.e., Very poor to Excellent), the results showed a significant positive correlation for participants who rated their own performance as

Good (r_s(173)=.159, p < .05). The results also showed a marginally significant positive correlation for participants who rated their own performance as Excellent (r_s(75)=.216, p = .060). When the data was grouped by feedback condition and session order, the results showed a significant positive correlation for participants in the Team condition in session 2 (r_s(37)=.431, p < .05) and for participants in the Individual condition in session 3 (r_s(37)= .435, p < .05). When the data were grouped by self-reported individual performance and session order, the results showed a significant positive correlation in session 2 for participants who rated their performance as excellent (r_s(12)= .590, p < .05). When the data were grouped by feedback condition, session order, and self-reported performance, the results showed a marginally significant negative correlation in the I&T condition, session 3, and who rated their performance as Excellent (r_s(3)= -.9, p = .083). The results also showed a marginally significant positive correlation in the Team condition, session 2, and who rated their performance as Good (r_s(16)= .403, p = .098). The results of the correlation results grouped by feedback condition, session order, and/or survey response are summarized in Table 94.

Individual Perception of Performance: *CTI_{ind}*

The data was analyzed by seven groupings of feedback condition, session order, and survey response (post-session survey). When the data were grouped by feedback condition, the results showed a significant negative correlation in Team condition (r_s(115) = -.230, p < .05). The results also showed a marginally significant negative correlation in the I&T condition (r_s(115) = -.176, p = .058). When the data were grouped by session order, the results showed a significant negative correlation in session 4 (r_s(115)= -.239, p < .05). The results also showed a marginally significant negative correlation in session 2 (r_s(115) = -.162, p =.081). When the data was grouped by feedback condition and session order, the results

showed a significant negative correlation in the Team condition in session 2 ($r_s(37) = -.355$, $p < .05$) and the I&T condition in session 4 ($r_s(40) = -.359$, $p < .05$).

Table 95 - Correlation results of participant perception of individual performance (i.e., TLX) versus time remaining. Only results that were significant ($p < .05$) or marginally significant ($p < .1$) are displayed in the table.

	Category	DF	r_s	p
Feedback				
	Individual	115	.258	< .05
	Team	115	.235	< .05
Session				
	2	115	.252	< .05
	3	115	.157	0.091
	4	115	.193	< .05
Survey Response				
	Excellent	75	.216	.060
	Good	173	.159	< .05
Feedback & Session				
	Individual – 3	37	.435	< .05
	Team – 2	37	.431	< .05
Survey Response & Session				
	Excellent – 2	12	.590	< .05
Feedback & Survey Response & Session				
	Team – Good – 2	16	.403	.098
	I&T – Excellent – 3	3	-.900	.083

When the data were grouped by self-reported performance, the results showed a significant negative correlation for participants who rated their performance as Excellent ($r_s(75) = -.270$, $p < .05$). The results also showed a significant positive correlation for participants who rated their performance as Poor ($r_s(21) = .701$, $p < .05$), indicating an incorrect perception of their performance. When the data were grouped by self-reported performance and feedback condition, the results showed a marginally significant positive correlation in for participants in the Team condition that labeled their performance as Poor ($r_s(5) = .739$, $p = .058$) and for participants in the I&T condition that labeled performance as

Poor ($r_s(4) = .829$, $p = .058$). When the data were grouped by session order and self-reported performance, the results showed a significant negative correlation for session 2 for participants who labeled their performance Excellent ($r_s(12) = -.731$, $p < .05$). The results also showed a significant positive correlation for session 2 for participants who labeled their performance Poor ($r_s(13) = .603$, $p < .05$). When the data were grouped by feedback condition, session order, and self-reported performance, the results showed a significant negative correlation for participants in the Individual condition, session 3, and rated their performance as Excellent ($r_s(8) = -.719$, $p < .05$). The results also showed a significant negative correlation in I&T condition, session 4, and rated their performance as Excellent ($r_s(14) = -.623$, $p < .05$). The results of the correlation results grouped by feedback condition, session order, and/or survey response are summarized in Table 96.

Individual Perception verse Team Performance

This section is focused on answering the question of whether a participant's perception of team performance correlated with the team's actual performance. The analysis compared how participants rated their team performance (Very poor, Poor, Average, Good, or Excellent) against time remaining (T_{team}) and collection item per item (CTI_{team}). Kendall's tau-b was used to assess the relationship between performance metrics (T_{team} and CTI_{team}) because it is suited to handle data with ties (Siegel & Castellan Jr., 1988).

Table 96 - Correlation results of participant perception of individual performance (i.e., TLX) verses collection time per item. Only results that were significant ($p < .05$) or marginally significant ($p < .1$) are displayed in the table.

	Category	DF	r_s	p
	Feedback			
	Team	115	-.230	< .05
	I&T	115	-.176	.058
	Session			
	2	115	-.162	.081
	4	115	-.239	< .05
	Survey Response			
	Excellent	75	-.270	< .05
	Poor	21	.701	< .05
	Feedback & Session			
	Team – 2	37	-.355	< .05
	I&T – 4	40	-.359	< .05
	Feedback & Survey Response			
	Team – Poor	5	.739	.058
	I&T – Poor	4	.829	.058
	Session & Survey Response			
	Excellent – 2	12	-.731	< .05
	Poor – 2	13	.603	< .05
	Feedback & Survey Response & Session			
	Individual – Excellent – 3	8	-.719	< .05
	I&T – Excellent – 4	14	-.623	< .05

For T_{team}, a positive correlation with how participants rated their team performance indicated a correct perception of team performance (i.e., as T_{team} increase, so did participant's team performance rating), whereas a negative correlation indicated an incorrect perception of team performance (i.e., as T_{team} increased, participant's team performance rating decreased). For CTI_{team}, a negative correlation with self-reported performance indicated a correct perception of team performance (i.e., as CTI_{team} decreased, participant's team performance rating increased), whereas a positive correlation with self-reported performance indicated an incorrect perception of team performance (i.e., as CTI_{team}, decreased, participant's team performance rating decreased). The next sections present the results.

Individual Perception verse Team Performance: T_{team}

The data were analyzed by three groupings of feedback condition and session order. When the data were grouped by feedback condition, the results showed a significant positive correlation in the Individual condition ($\tau_b = .286$, $p < .05$), Team condition ($\tau_b = .246$, $p < .05$), and I&T condition ($\tau_b = .204$, $p < .05$). When the data were grouped by session order, the results showed a significant positive correlation in session 3 ($\tau_b = .181$, $p < .05$) and 4 ($\tau_b = .363$, $p < .05$). When the data were grouped by feedback condition and session order, the results showed a significant positive correlation in the Individual condition in session 3 ($\tau_b = .292$, $p < .05$), in the I&T condition in session 3 ($\tau_b = .273$, $p < .05$), in the Individual condition in session 4 ($\tau_b = .556$, $p < .05$), and in the Team condition in session 4 ($\tau_b = .389$, $p < .05$). The results of the correlation results grouped by feedback condition, session order, and/or survey response are summarized in Table 97.

Table 97 - Correlation results of participant perception of team performance (i.e., post-session survey) versus time remaining. Only results that were significant ($p < .05$) or marginally significant ($p < .1$) are displayed in the table.

	Category	τ_b	p
Feedback			
	Individual	.286	< .05
	Team	.246	< .05
	I&T	.204	< .05
Session			
	3	.181	< .05
	4	.363	< .05
Feedback & Session			
	Individual – 3	.292	< .05
	Individual – 4	.556	< .05
	Team – 4	.389	< .05
	I&T – 3	.273	< .05

Individual Perception verse Team Performance: CTI_{team}

The data was analyzed by three groupings of feedback condition and session order. When the data was grouped by feedback condition, the results showed a significant negative correlation in the Individual (τ_b = -.279, p < .05), Team (τ_b = -.328, p < .05), and I&T (τ_b = -.204, p < .05) condition. When the data was grouped by session order, the results showed a significant negative correlation in session 3 (τ_b = -.266, p < .05) and 4 (τ_b = -.330, p < .05). When the data was grouped by feedback condition and session order, the results showed a significant negative correlation in Individual in session 3 (τ_b = -.527, p < .05), Individual in session 4 (τ_b = -.386, p < .05), and Team in session 4 (τ_b = -.450, p < .05). The results also showed a marginally significant negative correlation in Team in session 2 (τ_b = -.222, p = .090). The results of the correlation results grouped by feedback condition, session order, and/or survey response are summarized in Table 98.

Table 98 - Correlation results of participant perception of team performance (i.e., post-session survey) verses collection time per item. Only results that were significant ($p < .05$) or marginally significant ($p < .1$) are displayed in the table.

	Category	τ_b	p
Feedback			
	Individual	-0.279	< .05
	Team	-0.328	< .05
	I&T	-0.204	< .05
Session			
	3	-0.266	< .05
	4	-0.330	< .05
Feedback & Session			
	Individual – 3	-0.527	< .05
	Individual – 4	-0.386	< .05
	Team – 2	-0.222	0.090
	Team – 4	-0.450	< .05

Summary Discussion

This chapter analyzed participants' perception of their individual and team performance by comparing participants' self-reported performance, collected by NASA-TLX survey and a post-session survey, to performance metrics (T_{ind}, CTI_{ind}, T_{team}, and CTI_{team}). The focus of this analysis was to answer two questions: how the participant's perception of his or her own performance correlated with actual performance, and how the participant's perception of team performance correlated with the team's actual performance. This analysis was accomplished by comparing survey responses to actual performance metrics.

The analysis that focused on individual-level metrics compared participants NASA-TLX response to their T_{ind} and CTI_{ind} values. For T_{ind}, the researcher expected a positive correlation between the participant's self-reported performance, indicating a correct perception of their actual performance. In other words, as their self-report performance increased, T_{ind} should also increase. Overall, there was a positive significant correlation at the Individual ($r_s(115)= .258$, $p < .05$) and Team ($r_s(115)= .235$, $p < .05$) condition, indicating a correct perception of their individual performance in the Individual and Team conditions. This result suggests that giving Individual or Team information to participants will give them a correct perception of their individual performance. Furthermore, the result also showed a significant positive correlation at the Team level in session 2 ($r_s(37)=.431$, $p < .05$) and at the Individual level in session 3 ($r_s(37)= .435$, $p < .05$). This result suggests that Team information may be more useful early on, while Individual information may be more useful later. See Table 95 for a summary of the results.

For CTI_{ind}, the researcher expected a negative correlation between the participant's self-reported performance, indicating a correct perception of their actual performance. In other words, as their self-reported performance increased, CTI_{ind} should decrease. Overall,

there was a significant negative correlation in the Team condition ($r_s(115) = -.230$, $p < .05$), indicating that participants had a correct perception of their individual performance. This result suggests that giving team information can give participants a correct perception of their individual performance. Furthermore, the result also showed a significant negative correlation at the Team level in session 2 ($r_s(37) = -.355$, $p < .05$), indicating that giving team information early on may give the participants a correct perception of their team performance. A result that is worth noting that participants who labeled their performance as Excellent ($r_s(75) = -.270$, $p < .05$) had a correct perception of their performance, while participants who labeled their performance as Poor ($r_s(21) = .701$, $p < .05$) had an incorrect perception of their performance. See Table 96 for a summary of the results.

The analysis also focused on team-level metrics (T_{team} and CTI_{team}) to compare participants' perception of team performance. For T_{team}, the researcher expected a positive correlation between the participant's perception of team performance and actual team performance, indicating a correct perception of actual team performance. In other words, as their self-report of term performance increased, T_{team} should also increase. The result showed a significant positive correlation for all feedback conditions, indicating that the participants had a correct perception of their team's performance. The results suggest that giving any individual or team feedback will give participants a correct perception of their team performance. The result also showed a positive correlation for participants in session 3 and 4, indicating that participants had a correct perception of their team performance after the second session. This result, unsurprisingly, indicates that over time, participants began to have a correct perception of their team's performance. Specifically, the results in session 3 for participants in the Individual and I&T conditions showed a significant positive

correlation, indicating that at some point in time (i.e., after session 2) Individual or I&T information should be given to participants to allow participants to have a correct perception of their team performance. The results also showed, in session 4, a significant positive correlation for participants in the Team or Individual condition, indicating that participants with task experience should be given Individual or Team feedback to give them a correct perception of their team performance. See Table 97 for a summary of the results.

For CTI_{team}, the researcher expected a negative correlation between the participant's perception of team performance and actual team performance, indicating a correct perception of actual team performance. In other words, as their self-report of team performance increased, CTI_{team} should decrease. The results showed a significant negative correlation in all conditions, indicating that giving participants individual, team, or I&T information generally gave them a correct perception of their team's performance. The results also showed a significant negative correlation for sessions 3 and 4, indicating a correct perception of team performance over time. The results also showed a significant negative correlation in session 3 for participants in the individual condition. This result indicates that after some time, participants could be given Individual information to support a correct perception of team performance. The results also showed a significant negative correlation in session 4 for participants in the Team and Individual condition. This result indicates that participants with task experience might benefit from Individual or Team information. See Table 98 for a summary of the results.

Overall (i.e., across session order), the results indicate that providing information regarding team performance consistently gives participants a correct perception of individual and team performance. However, taking the session order into consideration, the results

suggest the effectiveness of the feedback conditions to give participants a correct perception of their own performance and team performance changes over time. Both implications are consistent with the results presented in previous chapters.

CHAPTER 7. CONCLUSIONS & FUTURE WORK

As discussed in the introduction chapter, the purpose of this study was to answer the following research question: How will teams' performance change when given feedback that displays indicators based on an assessment of individual performance, team performance, or both? This chapter explores the conclusions of the data presented and future work.

Discussion

Chapter 4 – Performance Metrics Analysis

The focus of Chapter 4's analysis was to understand how the feedback interventions influenced team and individual performance. Specifically, the question is how the feedback assessment conditions (Individual, Team, and I&T) affected the dependent performance variables: score, correct items collected, incorrect items collected, time remaining, and errors. Results suggested that individuals moved through the task more quickly in the Team condition than in the I&T condition. This result contradicted H_1, which predicted that teams that received information about I&T errors and correct items collected would perform better than teams that received information only about team or individual metrics.

The researcher predicted that teams would move quicker in the task when given I&T feedback because they would have access to a more complete performance information while pursuing a task that had a relatively low complexity level. This prediction was not supported. One conclusion, based on these results, is that giving too much information to teams can overwhelm its members and reduce the speed with which they complete a task. For example, in post-task interviews, one participant said, "I liked when there wasn't too much information on the screen." Another conclusion could be that individuals were able to process the information without feeling overwhelmed, but that increased information on-screen required

them to spend more time assessing their own performance and their team's performance, causing them to move more slowly through the task.

If this were true, then why did individuals move more quickly through the task when given team information? Having seen the results and interviewed participants after the last session, the researcher suggests that the I&T condition posed unanticipated additional cognitive load for participants in potentially three ways. First, participants were able to keep track of their individual performance using their own memory (i.e., how many of the six individual items they have collected), but offering them that same information on-screen posed additional cognitive load due to the need to confirm their mental calculations with the information on screen periodically. In contrast, players were not able to easily keep track of the team's overall performance in their own memory (i.e., how many of the 18 correct team items collected). For example, one participant noted in the post-interview, "I liked the ability to know how many team items we collected. I had an easier time keeping track of my individual items, but there were too many team items for me to keep track in my head." In addition, a different participant supported this idea by saying, "I did not like knowing how many of my individual items I collected, because I could do that on my own...."

A second potential cognitive load presented by the I&T condition (8 metrics displayed for items and errors) was simply that it offered essentially four times as much information as the Team condition (2 metrics displayed for items and errors), requiring players to consider more potential goals for their performance metrics rather than just two goals. Thirdly, with the full I&T feedback, players might have experienced additional cognitive load due to focusing on individual teammates and their performance levels, e.g., Player 1 might think something like, "Oh, Player 2 is not doing well. I need to help him out.

And Player 3 is doing better than I am. I wonder how she is doing that." The I&T and Individual conditions afforded this detailed player-by-player analysis, which could exert an additional mental load, while the Team condition did not. These two potential additional forms of mental load are only conjectured at this point, based on the idea that additional cognitive load will arise from processing additional information and considering interactions with others, but could be explored in the future.

Note that this result, just discussed, of individuals completing the task more quickly with Team feedback, was revealed with the T_{ind} metric, but not with T_{team}. This difference is likely due to the design of the T_{team} metric; by measuring the team completion time when the last team member completed the task, this metric essentially penalized a team with two faster members and one slower member, or, less so, teams with one faster member and two slower members. By equating all of these teams within the T_{team} metric, the variance was reduced, which likely made the impact of the feedback conditions less notable during analysis.

As noted in Chapter 2, there is little agreement among researchers as to which level of feedback produces optimal performance (see Table 1). In short, some researchers support the claim that Individual feedback in a group setting improves team performance (Archer-Kath et al., 1994; Moreland & Myaskovsky, 2000; Smith-Jentsch et al., 1996), some researchers support Team feedback (Scott-Young & Samson, 2006; Walter & Van Der Vegt, 2009), and some researchers support I&T feedback (Austin et al., 1996; Sivunen, 2006). Overall, the results presented in this Chapter 4 suggest that Team performance feedback should be given in the context of ITTSs providing real-time feedback in a group setting, which is in keeping with researchers who also conclude that Team feedback should be given in a group setting. Why did the results presented in Chapter 4 contradict the researchers that suggested

Individual or I&T Feedback? A difference between a human tutor and an ITTS is that an ITTS can provide real-time feedback during a task because it can process and present feedback almost instantaneously, while it is near impossible for a human tutor to process and present feedback instantaneously. The feedback implemented in previously discussed studies presented feedback before or after a session or task, while the feedback implemented in this study was during the task in real-time. These results suggest that the effect of assessment feedback delivered before or after a task is different from the effect of assessment feedback that is delivered during a task in real-time, contributing a new insight to this body of work. More studies are needed to explore this possibility since few studies have explored the effects of real-time assessment feedback delivered ITTSs.

The results suggested that giving participants Team level performance feedback encouraged members to move more quickly through the task. Specifically, participants in the Team condition had more time remaining at the end of the sessions than participants in the I&T condition. This result suggested that user interface (UI) designers should include only feedback information based on team assessment, as opposed to I&T assessment, in persistent feedback displayed on the UI during a virtual team task. However, it is important to note that the results indicated that displaying team assessment information increased the speed with which the team completed the task, but those faster teams did not necessarily complete the task well, e.g., with the most correct items and fewest errors. (The data showed no particular relationship between completion time and the number of correct items or errors.) The results also did not indicate if the feedback condition had a similar or different effect on team performance depending on the strategy, or process, teams used to complete the task. The

analysis in Chapter 5 examined if the strategy used by teams had any influence on the effect that the feedback conditions may have had on performance.

Chapter 5 – Behavioral and Team Characteristic Metrics Analysis

The focus of Chapter 5 was to explore two topics: how the feedback condition (Individual, Team, or I&T) related to strategy used by the team, and why the time remaining for individuals (T_{ind}) in the Team condition was significantly higher than in the I&T condition (a finding in Chapter 4). This analysis also introduced and examined the average collection time per item metric to understand how well the teams collected correct items. It was expected that high performing teams would need less time to collect a correct item. Therefore, high performing teams would have a lower collection time per item value, and low performing teams would have a high collection time per item value.

Results suggested that once participants had experience with a task (i.e., in session 4), T_{ind} was greater in the Team condition than in the Individual and I&T conditions for teams that used Go Together strategy. This result contradicted H_1. Results also suggested that the collection time per item at the individual level (CTI_{ind}) for teams that used the Go Together strategy was significantly lower in the Team condition than teams in the Individual condition. It is important to note that the results also showed that the CTI_{ind} for teams that used Go Together strategy was marginally significantly lower (p = .072, see Table 85) in the Team condition than for teams in the I&T condition. Generally, the results suggested that the influence of the feedback condition is different depending on the strategy used to complete the task.

This difference in influence could be a result of the ideas behind the different strategies. The idea behind teams using the Go Alone strategy is to complete the task by "divide and conquer." In other words, the members completed the task at their own pace and

communicated only information they felt was relevant to the whole team, even if the information was not relevant. In short, the team generally did not work closely with one another. Perhaps the feedback intervention did not affect teams using Go Alone strategy because they only needed information regarding their own performance, which they could track on their own. The members in groups using the Go Alone strategy may have an increased cognitive workload because they were individually responsible for completing the task on their own at their own speed. The feedback conditions may have had no significant influence for teams using Go Alone because they were generally focused on their own individual performance and could track that information on their own, even though the cognitive load for each team member was high.

The idea behind teams using the Go Together strategy is to complete the task by coordinating closely with one another. When teams worked together closely, it was important that they communicated accurate team information. The members in groups using the Go Together strategy may have had an increased cognitive workload based on the communication required by this strategy. Perhaps the team information provided in the Team condition supported better teamwork by 1) not creating as much competition among team members by highlighting individual performance, and 2) providing less information to communicate about. This increased teamwork, in turn, may have led to collecting correct items more quickly and, for experienced teams, moving through the task more quickly.

These ideas are supported by post-task interviews. For example, one team member who had the team feedback condition in session 4 noted, “The last session was the one I enjoyed the most. The characteristics of this session were that all of the team members had achieved proficiency at our tasks, there was no personal quota, and there was a team quota to

see how we were doing overall." Another team member noted, "There was no feedback (beige box) for us to focus on our individual objective, which made achieving our team objective that much easier." Both statements support the idea that individual feedback could distract from the team camaraderie.

As noted in Chapter 2, the Process component in team adaptation describes the way in which teams work together to complete a task (e.g., Go Together vs. Go Alone). Researchers have presented multiple models of the Process component in team adaptation (Burke et al., 2006; Entin & Serfaty, 1999; Maynard et al., 2015). This component was not a major focus of the study. However, the results presented in Chapter 5 suggest that the strategy with which teams chose to complete the task influenced the effect the feedback had on performance at the individual level (i.e., T_{ind} and CTI_{ind}). This result highlights the importance of researchers developing a consistent model of the Process component of team adaptation.

This result also suggests that the degree of collaboration implemented in the strategy used by a team is important to consider when UI designers develop displays that include assessment feedback that is persistent during a virtual team task. Specifically, the results suggest that if designers know, or anticipate, that teams will implement a strategy that is high in collaboration, then designers should include only feedback information based on Team assessment, as opposed to I&T (supported by results from T_{ind}) or Individual (supported by results from T_{ind} and CTI_{ind}). This conclusion is consistent with researchers who suggest giving Team feedback in a group setting (Scott-Young & Samson, 2006; Walter & Van Der Vegt, 2009).

A potential explanation for the results that suggested providing Team feedback helped teams complete the task more quickly and spend less time collecting correct items is that providing Team feedback helped improved camaraderie by allowing group members to focus on working together as a team and less on competing with their teammates. However, this idea is only a supposition about this study that is supported by the results of DeShon et al. (2004), who concluded that group members focused more on team performance when they received only Team feedback. However, DeShon et al. (2004) also concluded that their group members focused more on individual performance when they received only Individual feedback, while their group members that received I&T feedback were not able to take advantage of the information provided in the I&T feedback. The results from Chapter 4 and 5, specifically the results for T_{ind} and CTI_{ind}, support those findings, implying that the feedback given in a group setting should be focused on one level, but the results more specifically support the idea that the source should focus on the Team level, especially if the strategy implemented is highly collaborative.

One issue presented by the strategy analysis is whether the strategy chosen by the teams should be a discrete or continuous variable. Rather than Go Alone and Go Together, for example, as defined by whether the team members stayed inside or outside the Strategy Border on average, a continuous measure of "togetherness" could have been used. This approach has both pros and cons. The advantage would be that a continuous metric would likely simplify statistical analysis and afford a more detailed comparison between teams, e.g., offering the ability to distinguish between teams that were tightly together vs. only loosely together. The con, however, is that a continuous measure would be a less realistic representation of the very different team behaviors and communication patterns present in the

two strategies. An additional study would likely be required to fully explore the full dynamic range of Go Together vs. Go Alone strategies, but the proximity of team members was actually an emergent property of very different attitudes among the team members. The Go Together team members appear to approach the task as an actual team, while the Go Alone team members did not. These disparate attitudes about teaming would not likely be reflected well in a continuous "togetherness" variable.

The results in Chapter 5 indicated that the feedback condition influenced performance if the strategy implemented by the team had a high degree of collaboration. The performance metrics and strategies used by teams does not indicate how the participants perceived their own performance. The analysis in Chapter 6 examined whether the feedback conditions influenced the participants' perception of their own performance and team performance.

Chapter 6 – Perception of Individual and Team Performance Analysis

The metrics used to represent individual and team performance were time remaining and collection time per item introduced in Chapter 5. The metric used to represent individuals' perception of their own performance was the NASA-TLX survey, and the metric used to represent individuals' perception of team performance was the post-session survey. The researcher expected teams to have a correct perception of their own performance and team performance when given I&T feedback because the increased performance information would give team members a more accurate perception of their team performance. This expectation was not supported.

Results indicated by T_{ind} that participants had a correct perception of their own performance in the Individual and Team condition. Results also indicated that in session 2, participants had a correct perception of their own performance in the Team condition, and in session 3 participants had a correct perception of their own performance in the Individual

condition. A possible conclusion is that participants were better able to assess their individual performance when given feedback that focused on either Individual or Team level information. Perhaps giving participants both levels of information at the same time clouded their perception of how well they were doing. Though providing Individual or Team level information to participants correlated with participants having a correct perception of their own performance, it is important to consider the team's task experience (i.e., session 2, 3, and 4).

Specifically, based on these results, Team information should be provided to teams early on, and then Individual information should be provided after teams have gained some experience in the task. No conclusions can be drawn regarding I&T feedback because there were no statistically significant correlations in the I&T condition. It is possible that the I&T condition is essentially the same as the Individual and Team condition per these metrics, but a conclusion cannot be made with the current data. If the I&T condition is no different from the Individual and Team condition, then it is possible that including extra information, independent of task experience, will not significantly influence team members' perception of their performance. The researcher recommends that UI designers consider giving at least Team feedback information to teams with less experience in a task and then giving at least Individual information when that team gains a little more experience in the task to support a correct perception of individual performance.

Results for CTI_{ind} indicated that participants in the Team condition had a correct perception of their individual performance. Results also indicated that in session 2, participants had a correct perception of individual performance in the Team condition. Finally, results also indicated that in session 4, participants had a correct perception of

individual performance in the I&T condition. Similar to conclusions of previous chapters, it is possible that participants had a correct perception of their own performance in only the Team condition because they could privately manage their own performance and use the Team feedback information displayed to them to help improve the overall team performance. Perhaps when their own performance was displayed to the team, like in the Individual or I&T conditions, their perception of their own performance may have been skewed if they felt their performance was especially low. This idea is supported by the fact that participants who labeled their performance as Poor had an overall incorrect perception of their performance, while participants who labeled their performance as Excellent had a correct perception of their performance.

Results for T_{team} showed an overall trend that participants in the Individual, Team, and I&T condition had a correct perception of team performance. A significant correlation indicated that participants in session 3 had a correct perception of team performance in the Individual and I&T conditions. Also, significant correlations indicated that participants in session 4 had a correct perception of team performance in the Team and Individual condition. A conclusion could be that no matter what level of information one gives to teams, they will have a correct perception of the team's performance. However, the task experience of the teams is important to consider when giving feedback information. Specifically, these results suggest that when a team has limited experience with a task, Individual or I&T information should be provided to support the correct perception of team performance, while teams with a greater amount of task experience should be given Individual or Team feedback, but not both. This conclusion of providing one source of information (Individual or Team) is in alignment with other researchers (DeShon et al., 2004).

Overall, results for CTI_{team} showed the trend that participants had a correct perception of team performance in the Individual, Team, and I&T conditions. A significant correlation showed that in session 3, participants had a correct perception of team performance in the Individual condition. Another significant correlation showed that in session 4, participants had a correct perception of team performance in the Individual and Team condition. A possible conclusion is that giving teams any level of feedback information will support a correct perception of team performance. Specifically, Individual feedback information should be given to teams with some experience so they can improve their own performance, while Individual or Team feedback should be given to experienced teams, because they may be able to correctly assess the team information to generate a correct perception of team performance.

Overall, at the individual level, the results suggest that participants seemed to have a correct perception of their own performance when given either Individual (supported by results from T_{ind}) or Team feedback (supported by results from T_{ind} and CTI_{ind}). This implies that one level of information (Individual or Team) should be given to teams but not both, which is similar to conclusions presented by DeShon et al. (2004). Participants may not have been able to build a correct perception of their own performance in the I&T condition because having both sources of information clouded their perception of their own performance.

The results indicated that the influence of the feedback conditions is more complex when task experience is considered. Specifically, if participants have little task experience (i.e., session 2), then Team feedback (supported by results from T_{ind} and CTI_{ind}) should be given to participants to support a correct perception of their own performance. If participants

have some task experience (i.e., session 3), then Individual feedback (supported by results from T_{ind}) should be given to participants to support a correct perception of their own performance. If participants have more task experience (i.e., session 4), then I&T feedback (supported by results from CTI_{ind}) should be given to participants to support a correct perception of their own performance. These results suggest that UI designers should consider a team's experience with a task when developing a display that consists of persistent assessment feedback in order to maximize the positive effect it may have on users' perception of their own performance. However, if it is not possible, or not practical, to consider task experience, then the results suggest that designers should display to participants persistent assessment feedback that focuses on Team performance metrics to support a correct perception of their own performance.

Overall, at the team level, the results suggest that participants seemed to have a correct perception of their team's performance when given Individual, Team, or I&T feedback (supported by results from T_{team} and CTI_{team}), suggesting that any level of feedback supported a correct perception of the team's performance. This result is different from the individual level analysis which suggested that Individual or Team feedback should be given but not both, suggesting the feedback condition had an influence on the Team level that is different from its influence on the individual level.

Similar to the above analysis of a participant's perception on their own performance, the result indicated that the influence of the feedback conditions on participant's perception of their team's performance is different depending on a team's task experience. If participants had little task experience (i.e., session 2), no specific feedback condition showed evidence of supporting a correct perception of team performance. If participants had some

task experience (i.e., session 3), then Individual (supported by results from T_{team} and CTI_{team}) or I&T (supported by results from T_{team}) feedback should be given to support a correct perception of team performance. If participants had more task experience (i.e., session 4), then Team or Individual (supported by results from T_{team} and CTI_{team}), feedback should be given to support a correct perception of team performance. Similar to the individual level analysis, these results suggest that UI designers should consider a team's experience with a task when developing a display that consists of persistent assessment feedback in order to maximize the positive effect it may have on users' perception of their team's performance. However, if it is not possible, or not practical, to consider task experience then the results suggest that displaying any level of feedback (i.e., Individual, Team, or I&T) on a UI that includes persistent assessment feedback will support a correct perception of their team's performance.

Implications of Chapter Findings

Overall, the implication of the results suggests that the Team feedback condition improved performance at the individual level and correlated with participants having the correct perception of individual or team performance. This implication contradicted H_1, which predicted that teams that received information about I&T errors and correct items collected would perform better than teams that received information only about team or individual errors or correct items collected. The idea behind H_1 was that providing group members with as much information as possible (i.e., I&T feedback) would support performance monitoring (Salas et al., 2005) by allowing them to better track their own performance and their team performance, which would result in improved performance. The researcher considered the possibility that displaying I&T feedback could be overwhelming but believed that that would not be an issue in this current study because the task was

relatively simple (i.e., shop for items as quickly as possible). A possible flaw of H_1 was that it did not consider which condition (i.e., Individual, Team, or I&T) provided information that was more useful to participants in this particular task.

The researcher argues that the Team feedback condition generally provided more useful information to participants than the Individual and I&T condition. Team feedback information is available in the I&T condition, but the researcher believes that including Individual feedback information provided redundant information that induced an increased cognitive workload that led to reduced team performance. Generally, receiving Team feedback information allowed team members to keep track of their own performance while using the Team information to contribute to the team. As a result, group members moved more quickly through the task in the Team condition as opposed to the I&T condition (see Chapter 4).

Another potential flaw in H_1 was that it did not consider how the degree of collaboration in strategies implemented by teams may have influenced the effect of the feedback condition. As noted in Chapter 2, the Process component is one of three main elements of team adaptation and it is the element that a has a wide variation, of how it occurs, in models presented by researchers (Burke et al., 2006; Entin & Serfaty, 1999; Maynard et al., 2015). The process component of team adaption was not a major focus of this current study, but the results indicate that the influence of persistent assessment is affected by the degree of collaboration in the strategy implemented by groups. Specifically, the results indicated that teams with more experience (i.e., session 4) that used a collaborative strategy (i.e., Go Together) moved more quickly through the task in the Team condition than in the Individual or I&T conditions. The results also suggested that group members spent less time

collecting items in the Team condition than in Individual condition (see Chapter 5). The implications of these finds suggest that designers should strongly consider giving Team information level feedback to teams during a collaborative virtual task.

Finally, an additional potential flaw in H_1 is that it does not consider the team's task experience. An interesting implication from the results is that the influence of the feedback conditions on a team's performance and perception appeared to change depending on the team's task experience (i.e., session 2, 3, and 4). For example, in Chapter 5, the results suggested that participants with task experience (i.e., in session 4) moved more quickly through the task in the Team condition than in the Individual and I&T condition. This implies that the positive influence of teams receiving Team feedback may not be realized until the team has more experience with the task. This makes sense because teams that are unfamiliar with the task may not know how to effectively use the information that is given to them. The influence that feedback condition had on the perception of performance appeared to also change depending on the session as well. In Chapter 6, the results suggested that in session 2 participants had a correct perception of their performance (T_{ind} compared to TLX response) in the Team condition, but in session 3 participants had a correct perception of their performance in the Individual condition. These results imply that a team's task experience should be considered when deciding the level of performance feedback to give to teams. Overall, the results indicate that if UI designers want to optimize the positive effect of persistent assessment feedback, then the team's task experience must be considered.

Limitations

This experiment does have some limitations. Team familiarity was not controlled in the experiment. Studies have shown that familiarity can have a positive influence on team performance (Espinosa, Slaughter, Kraut, & Herbsleb, 2007). As a result, participants who

were familiar with one another may have had higher performance compared to teams that were strangers.

Also, the results showed a statistically significant finding in only one out of the ten performance metrics examined in Chapter 4. While we know that most participants noticed the feedback and that a majority found it somewhat helpful (see Chapter 4), the lack of significance regarding feedback condition introduces some uncertainty as to whether the feedback conditions had any impact on performance since most performance metrics showed no statistical difference in performance. The participants did experience a session with no feedback (Table 5), but all teams experienced this condition first because the researcher used this condition as training to minimize any learning effect. The absence of a true control group increases the uncertainty that the feedback implemented in this had any impact on performance.

Future Directions

There are numerous opportunities for future research. First, future studies should further explore the significant results presented in previous chapters. Specifically, studies should implement qualitative analysis on other performance metrics (e.g., communication frequencies and durations) and survey data (e.g., overall post-session survey) to better understand the significant results presented. For example, these studies could explore potential influences the feedback condition and the strategies implemented by teams may have had on team communication.

Second, future studies should also explore different modalities for team feedback. For example, a study could explore the influence of audio feedback on team performance. Also, the results suggest that the influence of the feedback conditions changed over time, studies

could implement an adaptive instruction architecture, similar to the one presented by Sottilare et al. (2018), and explore the influence it may have on team performance.

Third, future studies could explore different improvements for the TMET. For example, the researcher gave equal weight to the four components in the scores (Table 7 and Table 9), but there is no evidential basis to indicate that giving equal weight to the components is an optimal representation of team performance for this task. A study would explore how well the components of the individual and team score map onto the overall scores. Also, another study could develop a performance metric that incorporates multiple sources and is a better representation of performance. A study could explore different methods of calculating team metrics. For example, T_{team} was recorded after all members of a team had completed the task, which removed any measure of the variation that occurred between teams with one, two, or three slower team members. Instead of recording T_{team} only when all members of a team had completed the task, T_{team} could have been an average of each member's T_{ind}. A study could explore whether either approach better represented the overall team performance. Furthermore, another study could explore the effectiveness of the feedback implemented in this current work by including a control group.

Fourth, future studies should explore the effect that deception has on the participants and their performance during the task. The researcher implemented deception to promote participant engagement (see Chapter 3). Studies showed that financial incentive techniques could effectively increase group motivation (Clark, 2003). The researcher believes that lab studies that focus on teams should include some form of motivation to ensure that participants are engaged in the experiment and that the data collected are reliable. However, in a group setting, there is a social aspect that must be considered when implementing

motivational techniques. For example, Clark (2003) describe a phenomenon in teams called social loafing. In short, this occurs when a member of a group is not contributing a proportionate share of the work. A method to reduce social loafing suggested by Clark (2003) is to inform the team that they will be assessed by their individual contribution, not just team performance. Future studies could explore how different motivational techniques influence participant performance and engagement. Finally, a future study should explore the extent to which the implications of this study are generalizable and can be implemented in real-work team training.

REFERENCES

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. https://doi.org/10.1.1.121.9138

Archer-Kath, J., Johnson, D. W., & Johnson, R. T. (1994). Individual versus Group Feedback in Cooperative Groups. *The Journal of Social Psychology*, *134*(5), 681–694. https://doi.org/10.1080/00224545.1994.9922999

Austin, J., Kessler, M. L., Riccobono, J. E., & Bailey, J. S. (1996). Using Feedback and Reinforcement to Improve the Performance and Safety of a Roofing Crew. *Journal of Organizational Behavior Management*, *16*(2), 49–75. https://doi.org/10.1300/J075v16n02_04

Azevedo, R., & Bernard, R. M. (1995). A Meta-Analysis of the Effects of Feedback in Computer-Based Instruction. *Journal of Educational Computing Research*, *13*(2), 111–127. https://doi.org/10.2190/9LMD-3U28-3A0G-FTQT

Baard, S. K., Rench, T. A., & Kozlowski, S. W. J. (2014). Performance Adaptation. *Journal of Management*, *40*(1), 48–99. https://doi.org/10.1177/0149206313488210

Bannister, S. L., Wickenheiser, H. M., & Keegan, D. A. (2014). Key Elements of Highly Effective Teams. *PEDIATRICS*, *133*(2), 184–186. https://doi.org/10.1542/peds.2013-3734

Bell, S. T., Brown, S. G., Colaneri, A., & Outland, N. (2018). Team composition and the ABCs of teamwork. *American Psychologist*, *73*(4), 349–362. https://doi.org/10.1037/amp0000305

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1998). Cross-training and team performance. In *Making decisions under stress: Implications for individual and team training.* (pp. 299–311). https://doi.org/10.1037/10278-011

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

Bonner, D., Gilbert, S. B., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., … Holub, J. (2016). The Challenges of Building Intelligent Tutoring Systems for Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 1981–1985. https://doi.org/10.1177/1541931213601451

Burke, C. S., Georganta, E., & Hernandez, C. (2017). The Importance of Time in Team Leadership Research. In E. Salas, W. B. Vessey, & L. B. Landon (Eds.), *Team Dynamics Over Time* (pp. 95–122). Bingley, UK: Emerald Publishing Limited.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology*, *91*(6), 1189–1207. https://doi.org/10.1037/0021-9010.91.6.1189

Burns, S. P., Dawson, D. R., Perea, J. D., Vas, A., Pickens, N. D., & Neville, M. (2019). Development, Reliability, and Validity of the Multiple Errands Test Home Version (MET–Home) in Adults With Stroke. *American Journal of Occupational Therapy*, *73*(3). https://doi.org/10.5014/ajot.2019.027755

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology. Applied*, *13*(4), 273–281. https://doi.org/10.1037/1076-898X.13.4.273

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations Between Work Group Characteristics and Effectiveness: Implications for Designing Effective Work Groups. *Personnel Psychology*, *46*(4), 823–847. https://doi.org/10.1111/j.1744-6570.1993.tb01571.x

Campion, M. A., Papper, E. M., & Medsker, G. J. (1996). Relations Between Work Group Characteristics and Effectiveness: A replication and Extension. *Personnel Psychology*, *49*(2), 429–452. https://doi.org/10.1111/j.1744-6570.1996.tb01806.x

Cannon-Bowers, J. a., & Salas, E. (1998). Team Performance and Training in Complex Environments: Recent Findings From Applied Research. *Current Directions in Psychological Science*, *7*(3), 83–87. https://doi.org/10.1111/1467-8721.ep10773005

Christian, J. S., Christian, M. S., Pearsall, M. J., & Long, E. C. (2017). Team adaptation in context: An integrated conceptual model and meta-analytic review. *Organizational Behavior and Human Decision Processes*, *140*, 62–89. https://doi.org/10.1016/j.obhdp.2017.01.003

Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., & Riva, G. (2014). Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinsonâ€™s disease. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00405

Clark, R. E. (2003). Fostering the work motivation of individuals and teams. *Performance Improvement*, *42*(3), 21–29. https://doi.org/10.1002/pfi.4930420305

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, P. A., & Dacanay, L. S. (1992). Computer-Based Instruction and Health Professions Education. *Evaluation & the Health Professions*, *15*(3), 259–281. https://doi.org/10.1177/016327879201500301

Corbett, A. T., & Anderson, J. R. (2001). Locus of Feedback Control in Computer-Based Tutoring: Impact on Learning Rate, Achievement and Attitudes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 245–252.

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), 1035–1056. https://doi.org/10.1037/0021-9010.89.6.1035

Devasani, S., Gilbert, S. B., Shetty, S., Ramaswamy, N., & Blessing, S. (2011). Authoring Intelligent Tutoring Systems for 3D Game Environments 2 Previous Work. *Aied*.

Devlin, S. P., Flynn, J. R., & Riggs, S. L. (2018). Connecting the Big Five Taxonomies: Understanding how Individual Traits Contribute to Team Adaptability under Workload Transitions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 119–123. https://doi.org/10.1177/1541931218621027

Dobson, J. L. (2011). Effect of selected "desirable difficulty" learning strategies on the retention of physiology information. *Advances in Physiology Education*, *35*(4), 378–383. https://doi.org/10.1152/advan.00039.2011

du Boulay, B., & Luckin, R. (2001). Modelling human teaching tactics and strategies for tutoring systems. *International Journal of Artificial Intelligence in Education*, *12*(3), 235–256.

Emmert, G. D. (1978). Measuring the Impact of Group Performance Feedback Versus Individual Performance Feedback in an Industrial Setting. *Journal of Organizational Behavior Management*, *1*(2), 134–141. https://doi.org/10.1300/J075v01n02_04

Entin, E. E., & Serfaty, D. (1999). Adaptive Team Coordination. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*(2), 312–325. https://doi.org/10.1518/001872099779591196

Ericsson, K. A. (2008). Deliberate Practice and Acquisition of Expert Performance: A General Overview. *Academic Emergency Medicine*, *15*(11), 988–994. https://doi.org/10.1111/j.1553-2712.2008.00227.x

Espinosa, J. A., Slaughter, S. A., Kraut, R. E., & Herbsleb, J. D. (2007). Familiarity, Complexity, and Team Performance in Geographically Distributed Software Development. *Organization Science*, *18*(4), 613–630. https://doi.org/10.1287/orsc.1070.0297

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19–38. https://doi.org/10.1016/j.brat.2017.05.013

Fralick, M., & Kesselheim, A. S. (2017). Periodic benefit-risk evaluation reports have substantial promise to guide patient care and should be made publicly available. *Pharmacoepidemiology and Drug Safety*. https://doi.org/10.1002/pds.4177

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. https://doi.org/10.1016/j.edurev.2011.11.003

Gałecki, A., & Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R*. https://doi.org/10.1007/978-1-4614-3900-4

Galwey, N. W. (2014). Testing the assumptions of the analyses: Inspection of the residual values. In *Introduction to Mixed Modelling: Beyond Regression and Analysis of Variance* (Second). Chichester, United Kingdom: John Wiley & Sons, Ltd.

Gilbert, S. B., Devasani, S., Kodavali, S. K., & Blessing, S. B. (2011). Easy authoring of intelligent tutoring systems for synthetic environments. *Twentieth Conference on Behavior Representation in Modeling and Simulation*. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84865398675&partnerID=40&md5=39567eac1121150c640f6182db2525e5

Gilbert, S. B., Dorneich, M., Walton, J., & Winer, E. (2018). Five Lenses on Team Tutor Challenges: A Multidisciplinary Approach. In J. Johnston, R. Sottilare, A. M. Sinatra, & C. S. Burke (Eds.), *Building Intelligent Tutoring Systems for Teams* (pp. 247–277). Bingley, UK: Emerald Publishing.

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., … Winer, E. (2018). Creating a Team Tutor Using GIFT. *Int'l Journal of Artificial Intelligence in Education*, *28*(2), 286–313. https://doi.org/10.1007/s40593-017-0151-2

Gilson, L. L., Maynard, M. T., Jones Young, N. C., Vartiainen, M., & Hakonen, M. (2015). Virtual Teams Research. *Journal of Management*, *41*(5), 1313–1337. https://doi.org/10.1177/0149206314559946

Gonzalez-Mulé, E., Courtright, S. H., DeGeest, D., Seong, J.-Y., & Hong, D.-S. (2016). Channeled Autonomy. *Journal of Management*, *42*(7), 2018–2033. https://doi.org/10.1177/0149206314535443

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team Coordination Dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences*, *14*(3), 265–289.

Graesser, A. C., Hu, X., & Sottilare, R. A. (2018). Intelligent tutoring systems. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 246–255). New York: Routledge.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, *1*(1), 35–51. https://doi.org/10.1016/S1389-0417(99)00005-4

Grant, A. M., & Parker, S. K. (2009). 7 Redesigning Work Design Theories: The Rise of Relational and Proactive Perspectives. *The Academy of Management Annals*, *3*(1), 317–375. https://doi.org/10.1080/19416520903047327

Hanberg, V. L., MacKenzie, D. E., & Versnel, J. (2019). Face validity of the youth Multiple Errands Test (yMET) in the community: A focus group and pilot study. *British Journal of Occupational Therapy*, *82*(4), 248–258. https://doi.org/10.1177/0308022618813098

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*, 139–183.

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. B., & Blessing, S. (2008). Effectiveness of using an intelligent tutoring system to train users on off-the-shelf software. *Society for Information Technology & Teacher Education International Conference*, 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. https://doi.org/10.3102/003465430298487

Hick, W. E. (1952). On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. https://doi.org/10.1080/17470215208416600

Hockey, G. R. J., Sauer, J., & Wastell, D. G. (2007). Adaptability of Training in Simulated Process Control: Knowledge- Versus Rule-Based Guidance Under Task Changes and Environmental Stress. *Human Factors*, *49*(1), 158–174. https://doi.org/10.1518/001872007779598000

Howie, E., Sy, S., Ford, L., & Vicente, K. J. (2000). Human - computer interface design can reduce misperceptions of feedback. *System Dynamics Review*, *16*(3), 151–171. https://doi.org/10.1002/1099-1727(200023)16:3<151::AID-SDR191>3.0.CO;2-0

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188–196. https://doi.org/10.1037/h0056940

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. https://doi.org/10.1037//0021-9010.64.4.349

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in Organizations: From Input-Process-Output Models to IMOI Models. *Annual Review of Psychology*, *56*, 517–543. https://doi.org/10.1146/annurev.psych.56.091103.070250

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R 2 GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944–946. https://doi.org/10.1111/2041-210X.12225

Johnston, J. H., Burke, C. S., Milham, L. A., Ross, W. M., & Salas, E. (2018). Challenges and Propositions for Developing Effective Team Training with Adaptive Tutors. In Joan Johnston, R. Sottilare, A. M. Sinatra, & C. S. Burke (Eds.), *Building Intelligent Tutoring Systems for Teams* (pp. 75–97). https://doi.org/10.1108/S1534-085620180000019008

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. https://doi.org/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. *Proceedings of the Seventh International Conference of Intelligent Tutoring Systems*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-30139-4_16

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artifical Intelligence in Education*, *8*, 30–43.

Larreamendy-Joerns, J., Leinhardt, G., & Corredor, J. (2005). Six Online Statistics Courses: Examination and Review. *The American Statistician*, *59*(3), 240–251. https://doi.org/10.1198/000313005X54162

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1). https://doi.org/10.18637/jss.v069.i01

LePine, J. A. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology*, *88*(1), 27–39. https://doi.org/10.1037/0021-9010.88.1.27

Lepper, M. R., Drake, M. F., & O'Donnell-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 108–144). Brookline Books.

Lepper, M. R., & Woolverton, M. (2002). The Wisdom of Practice: Lessons Learned from the Study of Highly Effective Tutors. In *Improving Academic Achievement* (pp. 135–158). https://doi.org/10.1016/B978-012064455-1/50010-5

Long, S. K., Karpinsky, N. D., & Bliss, J. P. (2017). Trust of Simulated Robotic Peacekeepers among Resident and Expatriate Americans. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 2091–2095. https://doi.org/10.1177/1541931213602005

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. https://doi.org/10.1037/a0037123

Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance Implications of Leader Briefings and Team-Interaction Training for Team Adaptation to Novel Environments. *Journal of Applied Psychology*, *85*(6), 971–986. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11125660

Martin, F., Klein, J. D., & Sullivan, H. (2007). The impact of instructional elements in computer-based instruction. *British Journal of Educational Technology*, *38*(4), 623–636. https://doi.org/10.1111/j.1467-8535.2006.00670.x

Maynard, M. T., Kennedy, D. M., & Sommer, S. A. (2015). Team adaptation: A fifteen-year synthesis (1998–2013) and framework for how this literature needs to "adapt" going forward. *European Journal of Work and Organizational Psychology*, *24*(5), 652–677. https://doi.org/10.1080/1359432X.2014.1001376

McAndrew, M., Mucciolo, T. W., & Jahangiri, L. (2016). Characteristics of Effective Simulation (Preclinical) Teachers as Identified by Dental Students: A Qualitative Study. *Journal of Dental Education*, *80*(11). https://doi.org/10.1111/eje.12012

McCulloch, C. E. (2005). Repeated Measures ANOVA, R.I.P.? *CHANCE*, *18*(3), 29–33. https://doi.org/10.1080/09332480.2005.10722732

McEwan, D., Ruissen, G. R., Eys, M. A., Zumbo, B. D., & Beauchamp, M. R. (2017). The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions. *PLOS ONE*, *12*(1). https://doi.org/10.1371/journal.pone.0169604

Mickan, S., & Rodger, S. (2000). Characteristics of effective teams: a literature review. *Australian Health Review*, *23*(3), 201–208. https://doi.org/http://dx.doi.org/10.1071/AH000201

Moreland, R. L., & Myaskovsky, L. (2000). Exploring the Performance Benefits of Group Training: Transactive Memory or Improved Communication? *Organizational Behavior and Human Decision Processes*, *82*(1), 117–133. https://doi.org/10.1006/obhd.2000.2891

Morris, E. (2001). The design and evaluation of Link: A computer-based learning system for correlation. *British Journal of Educational Technology*, *32*(1), 39–52. https://doi.org/10.1111/1467-8535.00175

Mory, E. (2004). Feedback research revisited. In *Handbook of research on educational communications and technology* (pp. 745–784). https://doi.org/10.1007/s00127-009-0052-2

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R 2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Norman, D. A. (1988). The Psychopathology of Everyday Things. In *The Psychology of Everyday Things* (pp. 1–33). New York: Basic Books.

Orelien, J. G., & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using statistics. *Computational Statistics & Data Analysis*, *52*(4), 1896–1907. https://doi.org/10.1016/j.csda.2007.06.006

Pusic, M. V., Boutis, K., Hatala, R., & Cook, D. A. (2015). Learning Curves in Health Professions Education. *Academic Medicine*. https://doi.org/10.1097/ACM.0000000000000681

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. https://doi.org/10.1080/09602010802469074

Randall, K. R., Resick, C. J., & DeChurch, L. A. (2011). Building team adaptive capacity: The roles of sensegiving and team composition. *Journal of Applied Psychology*, *96*(3), 525–540. https://doi.org/10.1037/a0022622

Rosen, M. a, Salas, E., Wilson, K. a, King, H. B., Salisbury, M., Augenstein, J. S., … Birnbach, D. J. (2008). Measuring Team Performance in Simulation-Based Training: Adopting Best Practices for Healthcare. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *3*(1), 33–41. https://doi.org/10.1097/SIH.0b013e3181626276

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How Can you Turn a Team of Experts Into an Expert Team?: Emerging Training Strategies. In *Naturalistic Decision Making* (pp. 359–370).

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for Enhancing Team Cognition in Complex Environments: The Power of Team Performance Diagnosis. *Aviation, Space, and Environmental Medicine*, *78*(5), B77–B85.

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

Sander, P. C., van Doorn, R. R. A., van der Pal, J., & Zijlstra, F. R. H. (2015). Team adaptation to an unforeseen system failure: Limits of the potential aids of shared knowledge and standardized communication. *European Journal of Work and Organizational Psychology*, *24*(5), 796–811. https://doi.org/10.1080/1359432X.2015.1006199

Schmutz, J., & Manser, T. (2013). Do team processes really have an effect on clinical performance? A systematic literature review. *British Journal of Anaesthesia*, *110*(4), 529–544. https://doi.org/10.1093/bja/aes513

Scott-Young, C. M., & Samson, D. (2006). Modeling Team Efficacy in Project Management: an explorartory study. *Academy of Management Proceedings*, *2006*(1), D1–D6. https://doi.org/10.5465/ambpp.2006.27175183

Serfaty, D., & Entin, E. E. (1997). Team adaptation and coordination training. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision Making Under Stress: Emerging Themes and Application* (Report No., pp. 170–184).

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

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shoss, M. K., Witt, L. A., & Vera, D. (2012). When does adaptive performance lead to higher task performance? *Journal of Organizational Behavior*, *33*(7), 910–924. https://doi.org/10.1002/job.780

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

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*. https://doi.org/10.1126/science.228.4698.456

Siegel, S., & Castellan Jr., N. J. (1988). Nonparametric statistics for the behavioral sciences, 2nd ed. In *Nonparametric statistics for the behavioral sciences, 2nd ed.* New York, NY, England: Mcgraw-Hill Book Company.

Sivunen, A. (2006). Strengthening Identification with the Team in Virtual Teams: The Leaders' Perspective. *Group Decision and Negotiation*, *15*(4), 345–366. https://doi.org/10.1007/s10726-006-9046-6

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., Tannenbaum, S. I., & Salas, E. (2008). Guided Team Self-Correction: Impacts on Team Mental Models, Processes, and Effectiveness. *Small Group Research*, *39*(3), 303–327. https://doi.org/10.1177/1046496408317794

Smith-Jentsch, K. A., Salas, E., & Baker, D. p. (1996). Training Team Performance-Related Assertiveness. *Personnel Psychology*, *49*(4), 909–936. https://doi.org/10.1111/j.1744-6570.1996.tb02454.x

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of Computer-Assisted Instruction in Statistics. *Review of Educational Research*, *81*(1), 97–128. https://doi.org/10.3102/0034654310378174

Sottilare, R. A., & Boyce, M. W. (2016). *Elements of Adaptive Instruction for Training and Education*. https://doi.org/10.1007/978-3-319-39952-2_9

Sottilare, R. A., 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*.

Sottilare, R. A., Shawn Burke, C., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing Adaptive Instruction for Teams: a Meta-Analysis. *International Journal of Artificial Intelligence in Education*, *28*(2), 225–264. https://doi.org/10.1007/s40593-017-0146-z

Steverson, T., Adlam, A.-L. R., & Langdon, P. E. (2017). Development and Validation of a Modified Multiple Errands Test for Adults with Intellectual Disabilities. *Journal of Applied Research in Intellectual Disabilities*, *30*(2), 255–268. https://doi.org/10.1111/jar.12236

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*(2), 244–253. https://doi.org/10.3758/MC.38.2.244

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

Suits, D. B. (1957). Use of Dummy Variables in Regression Equations. *Journal of the American Statistical Association*, *52*(280), 548–551.

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.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass.: Addison-Wesley Pub.

Valls-Serrano, C., Verdejo-García, A., Noël, X., & Caracuel, A. (2018). Development of a Contextualized Version of the Multiple Errands Test for People with Substance Dependence. *Journal of the International Neuropsychological Society*, *24*(4), 347–359. https://doi.org/10.1017/S1355617717001023

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes. *Review of Educational Research*, *85*(4), 475–511. https://doi.org/10.3102/0034654314564881

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Volpe, C., Cannon-Bowers, J., Salas, E., & Spector, P. (1996). The Impact of Cross-Training on Team Functioning: An Empirical Investigation. *Human Factors*, *38*(1), 87–100. https://doi.org/10.1037/0021-9010.87.1.3

von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM*, *51*(8), 57. https://doi.org/10.1145/1378704.1378719

Walsh, C. M., Ling, S. C., Wang, C. S., & Carnahan, H. (2009). Concurrent versus terminal feedback: it may be better to wait. *Academic Medicine : Journal of the Association of American Medical Colleges*, *84*(10), S54-7. https://doi.org/10.1097/ACM.0b013e3181b38daf

Walter, F., & Van Der Vegt, G. S. (2009). Harnessing Positive Mood for Team Learning Facilitation: the role of perceived team feedback. *Academy of Management Proceedings*, *2009*(1), 1–6. https://doi.org/10.5465/ambpp.2009.44243035

Walton, J. (2015). *Comparing team feedback in a Team Multiple Errands Test* (Iowa State University, Digital Repository). https://doi.org/10.31274/etd-180810-4264

Walton, J., Bonner, D., Walker, K., Mater, S., Dorneich, M., Gilbert, S. B., & West, R. (2015). The Team Multiple Errands Test: A Platform to Evaluate Distributed Teams. *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing - CSCW'15 Companion*, 247–250. https://doi.org/10.1145/2685553.2699018

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

Walton, J., Gilbert, S. B., Winer, E., Dorneich, M., & Bonner, D. (2015). Evaluating Distributed Teams with the Team Multiple Errands Test. *Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Weiser, M. (1993). Ubiquitous Computing. *Computer*, *26*(10), 71–72. https://doi.org/10.1109/2.237456

Wickens, C. D., & Carswell, C. M. (1995). The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(3), 473–494. https://doi.org/10.1518/001872095779049408

Wickens, C. D., Lee, J. D., Liu, Y., & Becker, S. E. G. (2003). Displays. In *An Introduction to Human Factors Engineering* (2nd ed., pp. 184–217). Pearson Prentice Hall.

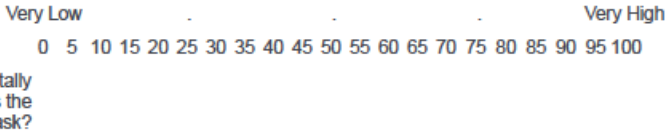
Yang, E., & Dorneich, M. C. (2016). Evaluation of Etiquette Strategies to Adapt Feedback In Affect-Aware Tutoring. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 393–397. https://doi.org/10.1177/1541931213601089

Zapata-Rivera, J. D., & Katz, I. R. (2014). Keeping your audience in mind: applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 442–463. https://doi.org/10.1080/0969594X.2014.936357

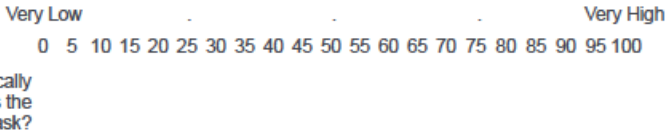
APPENDIX A NASA TLX SURVEY

NASA TLX

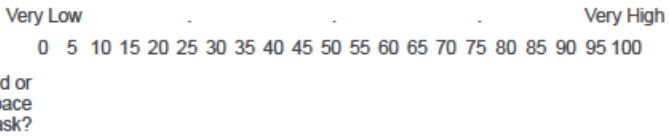
Mental Demand



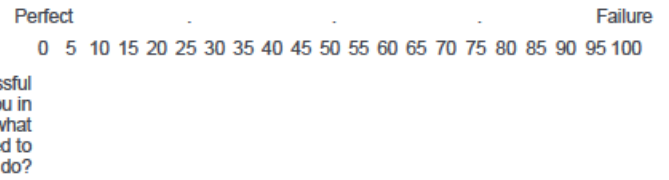
Physical Demand



Temporal Demand



Performance



Effort

APPENDIX B DEMOGRAPHICS / PRE SURVEY

Please select the gender with which you identify.

- ○ Male
- ○ Female
- ○ Other
- ○ Prefer not to answer

Age

- ○ 18-21
- ○ 22-30
- ○ 31-40
- ○ 41-50
- ○ 51-60
- ○ 61+
- ○ Prefer not to answer

What is your highest degree?

- ○ High School
- ○ Associate's
- ○ Bachelor's
- ○ Master's
- ○ PhD
- ○ Prefer not to answer

What year are you?

- ○ Freshman
- ○ Sophomore
- ○ Junior
- ○ Senior
- ○ I am not an undergraduate
- ○ Prefer not to answer

What is your major of study?

- Agricultural
- Business
- Design
- Education & Human Sciences
- Engineering
- Liberal Arts
- Sciences
- Veterinary Medicine
- Other _____
- Prefer not to answer

Do you have near 20/20 vision either naturally or with corrective lenses?

- Yes
- No
- Prefer not to answer

Please drag the bar to the number that best represents how true the statements are for you.

Strongly Agree		Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree		Strongly Disagree
1	2	3	4	5	6	7

I am confident in my ability to communicate well in a team.	
I hold my team back in teamwork situations.	
Team members slow me down	
Teams perform better together than individuals.	
Feedback is useful for teamwork.	
I learn better in teams than on my own.	
I am confident in my ability to navigate a virtual environment.	

How often do you play video games?

- ○ 0 hours per week (I don't play video games)
- ○ Less than 1 hour per week
- ○ 1 to less than 2 hours per week
- ○ 2 to less than 5 hours per week
- ○ 5 to less than 10 hours per week
- ○ 10 to less than 15 hours per week
- ○ 15 to less than 20 hours per week
- ○ More than 20 hours per week

How often do you work in teams?

- ○ I rarely work in teams
- ○ Daily
- ○ Once or twice a week
- ○ Once or twice every two weeks
- ○ Once a month
- ○ Once or twice every year

What kind of teams do you work on? (The emphasis in this survey is *work*, not sports teams or club teams, etc.)

_____

_____

_____

_____

_____

APPENDIX C POST-SESSION SURVEY

What player name was assigned to you (e.g., player 2)?

- ○ Player 1
- ○ Player 2
- ○ Player 3

Did you notice any feedback during the task?

- ○ Yes
- ○ No

Did you find the feedback helpful?

- ○ Yes, it was very helpful
- ○ Yes, it was somewhat helpful
- ○ No, it was not very helpful
- ○ No, it was actually distracting
- ○ I ignored the feedback

Please rate the performance for this task:

	Please select the answer that best corresponds to how you rate performance for the task you just completed				
	Very poor	Poor	Average	Good	Excellent

My individual performance	○	○	○	○	○
My team's performance	○	○	○	○	○
Our communication	○	○	○	○	○

Please rate player 1's performance:

	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
I am confident in the communication ability of my team member	○	○	○	○	○	○
My team member performed poorly.	○	○	○	○	○	○
I would work with my team member again.	○	○	○	○	○	○

Please rate player 2's performance:

	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
I am confident in the communication ability of my team member	○	○	○	○	○	○
My team member performed poorly.	○	○	○	○	○	○
I would work with my team member again.	○	○	○	○	○	○

Please rate player 3's performance:

	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
I am confident in the communication ability of my team member	○	○	○	○	○	○
My team member performed poorly.	○	○	○	○	○	○
I would work with my team member again.	○	○	○	○	○	○

Please indicate the degree to which you believe the following statements to be true.

I am satisfied with my team's performance on the task we just completed.

- ○ Very Inaccurate
- ○ Inaccurate
- ○ Neither Inaccurate nor Accurate
- ○ Accurate
- ○ Very Accurate

Do you think that your beliefs about the task are incompatible or inconsistent with your team members' beliefs about the task?

- ○ Yes
- ○ No

Do you think one of your team member's beliefs about the task is incompatible or inconsistent with another team member's beliefs about the task?

- ○ Yes
- ○ No

Below is a list of different tasks important to teamwork. Rate how confident you are that your team can do them as of now on the five-point Likert scale (1 = not at all confident to 5 =

extremely confident):

	1	2	3	4	5
My team can communicate important details in a timely manner	○	○	○	○	○
My team can satisfactorily communicate about important events	○	○	○	○	○
My team can accurately assess how to handle information we receive	○	○	○	○	○
My team can quickly assess how to handle information we receive	○	○	○	○	○
My team can accurately transfer information to one another	○	○	○	○	○

APPENDIX D SEMI-STRUCTURED GROUP INTERVIEW QUESTIONS

1. How noticeable was the feedback (i.e., did you notice it?)? Why or why not?
2. Were you motivated to do well during the trials (i.e., were you motived to achieve the highest score possible)?
3. How did you feel about the frequency of the feedback (i.e., did the feedback display often enough)?
    a. How could it be better?
4. Did you notice the content of the feedback?
5. How helpful was the content of the periodic feedback? Why was it helpful or why was it not helpful?
6. Did the information being displayed by the feedback change your behavior?
    a. If so, why did you change your behavior and in what way did you change your behavior?
    b. If not, what information would have prompted you to change your behavior?

Do you have any general comments regarding any aspect of the feedback?

APPENDIX E OVERALL POST SESSION SURVEY

Mark the extent to which you agree with these statements:

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
	0	1	2	3	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 I had information regarding my own performance.					
It was better when I had information regarding the team's performance.					

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 the session(s) that you enjoyed. What made a session a better experience?

_____

_____

_____

_____

_____

Please list 3 or more characteristics of the 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?

_____

_____

_____

_____

_____

APPENDIX F SHOPPING LISTS

Conditions A – No Feedback

Player 1	Player 2	Player 3
Party Hat	Chocolate Cake	Dryer
Armageddon	Key Lime Pie	Peanut Butter
Cake	Grand Piano	Tomato Soup
Cake Pan	Paper Towels	Stereo
Crowbar	Blue Present	The Quest Into the Sixth Age
Dress Shirt	Pepsi	Green Invitations

Team	Team	Team
Printer	Printer	Printer
Invisible Queen	Invisible Queen	Invisible Queen
Tape Measurer	Tape Measurer	Tape Measurer
Red Candle	Red Candle	Red Candle
Tissues	Tissues	Tissues
Water colors	Water colors	Water colors
Batteries	Batteries	Batteries
White Invitations	White Invitations	White Invitations
Pocket Knife	Pocket Knife	Pocket Knife
Newspaper	Newspaper	Newspaper
Ice Cream	Ice Cream	Ice Cream
Cupcake	Cupcake	Cupcake
First Aid Kit	First Aid Kit	First Aid Kit
Jasmine Tea	Jasmine Tea	Jasmine Tea
Cowboy Boots	Cowboy Boots	Cowboy Boots
Milk	Milk	Milk
Safety Goggles	Safety Goggles	Safety Goggles
Spam	Spam	Spam

Conditions B – Individual Feedback

Player 1	Player 2	Player 3
Pepsi	Table Saw	Carmel Apple
T-shirt	Ice Cream Cone	Grand Piano
Lighter Fluid	Paper Towels	Spam
Hard Hat	Toaster	Pound Cake
Condiments	Peanut Butter	Streamers
Drill	Coffee	Book of Matches
Team	**Team**	**Team**
The Quest Into the Sixth Age	The Quest Into the Sixth Age	The Quest Into the Sixth Age
Red Candle	Red Candle	Red Candle
Harmonica	Harmonica	Harmonica
Dryer	Dryer	Dryer
The Comb of Closing Bloodline	The Comb of Closing Bloodline	The Comb of Closing Bloodline
Red Envelopes	Red Envelopes	Red Envelopes
Cardboard Box	Cardboard Box	Cardboard Box
Table	Table	Table
Green Candle	Green Candle	Green Candle
Microwave	Microwave	Microwave
Yellow Cake	Yellow Cake	Yellow Cake
Jasmine Tea	Jasmine Tea	Jasmine Tea
Tape	Tape	Tape
Knife	Knife	Knife
Water colors	Water colors	Water colors
Glue	Glue	Glue
Ice Cream	Ice Cream	Ice Cream
Milk	Milk	Milk

Conditions C – Team Feedback

Player 1	Player 2	Player 3

Condiments	Crayons	Hard Hat
Printer	Dress Shirt	Fruit Tart
First Aid Kit	Ice Cream Cone	Pencil
Mint Tea	Tape Measurer	Colored Pencils
Eraser	Knife	Balloons
Watch	Yellow Cake	Newspaper

Team	Team	Team
Batteries	Batteries	Batteries
Armageddon	Armageddon	Armageddon
Cake Pan	Cake Pan	Cake Pan
Anvil	Anvil	Anvil
Box of Matches	Box of Matches	Box of Matches
The Quest Into the Sixth Age	The Quest Into the Sixth Age	The Quest Into the Sixth Age
The Hieroglyph Bridge	The Hieroglyph Bridge	The Hieroglyph Bridge
Drill	Drill	Drill
Cupcake	Cupcake	Cupcake
Hot Dog	Hot Dog	Hot Dog
Tool Set	Tool Set	Tool Set
Diet Coke	Diet Coke	Diet Coke
Blue Present	Blue Present	Blue Present
Jasmine Tea	Jasmine Tea	Jasmine Tea
Tomato Soup	Tomato Soup	Tomato Soup
White Invitations	White Invitations	White Invitations
Pencil Sharpener	Pencil Sharpener	Pencil Sharpener
Spam	Spam	Spam

Conditions D – Individual and Team (I&T) Feedback

Player 1	Player 2	Player 3
Coca-Cola	Safety Goggles	The Comb of Closing Bloodline
Party Hat	Milk	Balloons
Pound Cake	Watch	Superdog

Hammer	Dress Shirt	T-shirt
Diet Coke	Jasmine Tea	Batteries
Cardboard Box	Pocket Knife	Red Candle

Team	Team	Team
Cowboy Boots	Cowboy Boots	Cowboy Boots
Book of Matches	Book of Matches	Book of Matches
Tape Measurer	Tape Measurer	Tape Measurer
Tool Set	Tool Set	Tool Set
Knife	Knife	Knife
Hard Hat	Hard Hat	Hard Hat
Paper Towels	Paper Towels	Paper Towels
Table Saw	Table Saw	Table Saw
Newspaper	Newspaper	Newspaper
Lighter Fluid	Lighter Fluid	Lighter Fluid
Etch-a-sketch	Etch-a-sketch	Etch-a-sketch
Colored Pencils	Colored Pencils	Colored Pencils
Red Envelopes	Red Envelopes	Red Envelopes
Cake	Cake	Cake
Condiments	Condiments	Condiments
Anvil	Anvil	Anvil
Tomato Soup	Tomato Soup	Tomato Soup
Nails	Nails	Nails

APPENDIX G SUPPLEMENTARY ANALYSIS FOR CHAPTER 4

Correct Items Collected (Team)

The following analysis performed for CI_{team}.

Distribution Overview (CI_{team})

The overall distribution of the correct items collected at the team level shows a skew to the right (Figure 43). The distribution of the scores, when grouped by Feedback, is similar to the overall distribution (Figure 44). Over time, the distribution of each feedback group skews to the right (Figure 45).

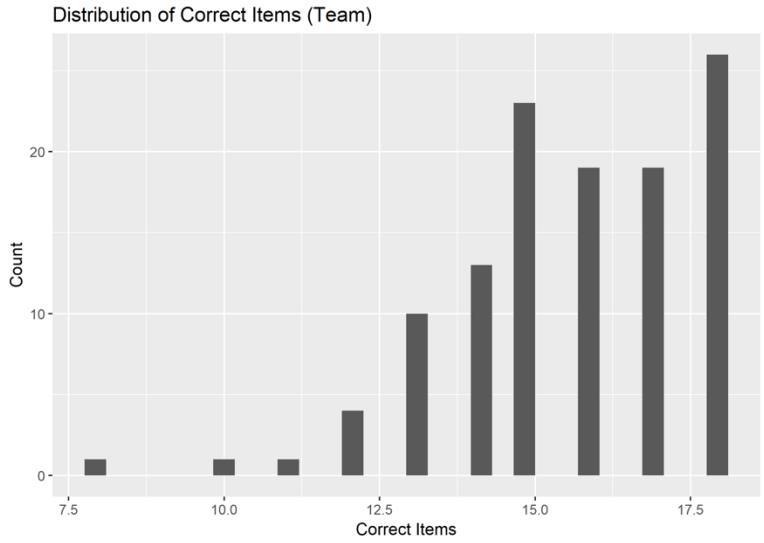


Figure 43 - Distribution overview of correct items collected (Team)

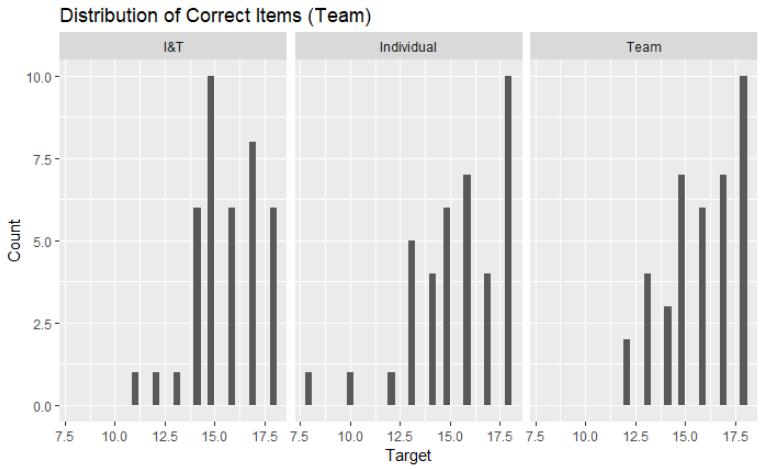


Figure 44 - Distribution of correct items collected grouped by Feedback (Team)

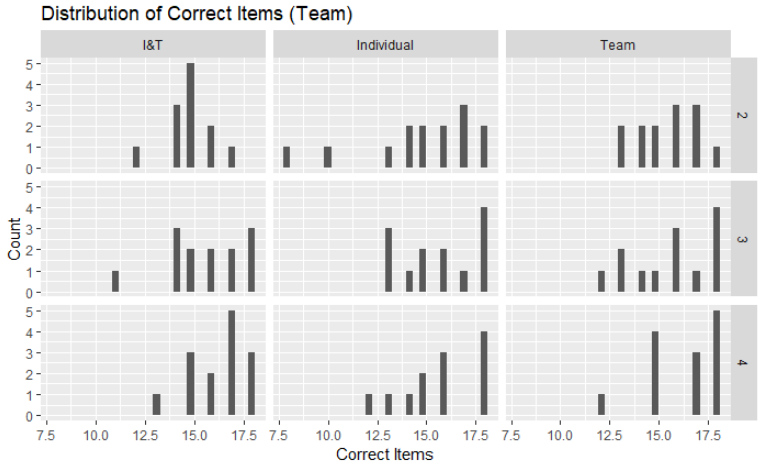


Figure 45 - Distribution of correct items collected grouped by Feedback and Session Order (Team)

Model Selection (CI_{team})

The author used a model selection process outlined in Table 16. The models that are significantly different from the null model are displayed in Table 99. None of the models that included the interaction effect were significantly different from the null model, indicating that the interaction effect is negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is

negligible and the simplest model that describes the data best in the No Interaction No Feedback model.

Table 99 – Models that are significantly different from the null model

Model Name	Equation	AIC	BIC
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	488.56	507.90
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**485.84**	**499.65**

Testing Assumptions: Residuals (CI_{team})

The histogram of the residuals (Figure 46) appeared to be approximately distributed. The residual fitted-value (Figure 47) showed a violation of constant variance. The points on the residual normal Q-Q plot (Figure 48) lay in a roughly straight line but trails from a straight line at both ends, which suggest the residuals violate normal distribution assumption. Overall, there is a visual violation of assumption two and three (Table 19). The classic LMM and the robust LMM produce similar results coefficient values (Table 100), so the researcher did not transform the data.

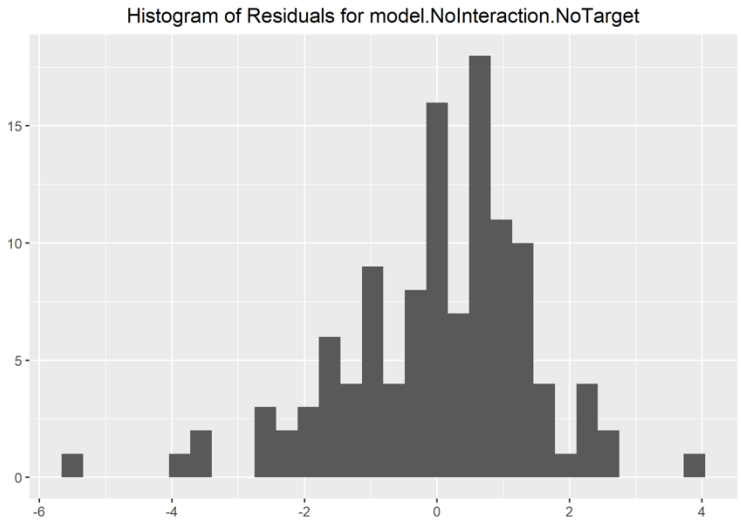


Figure 46 - Histogram of Residuals (Correct Items Collected - Team)

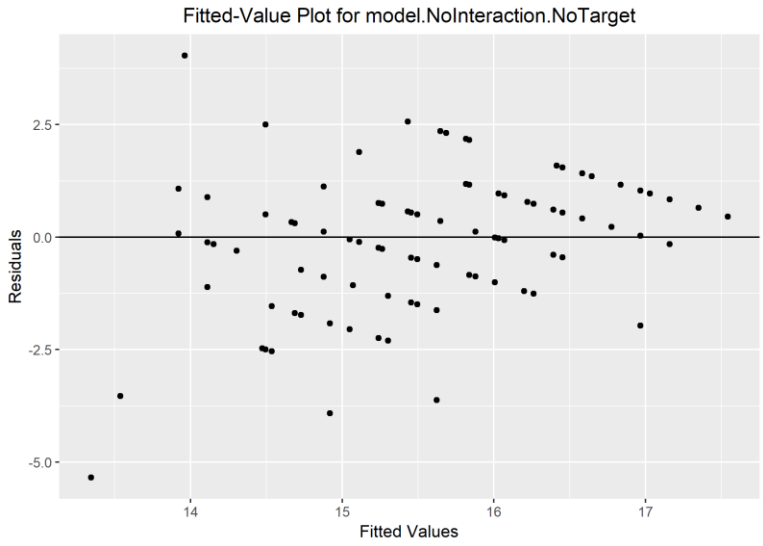


Figure 47 - Residual fitted plot (Correct Items Collected - Team)

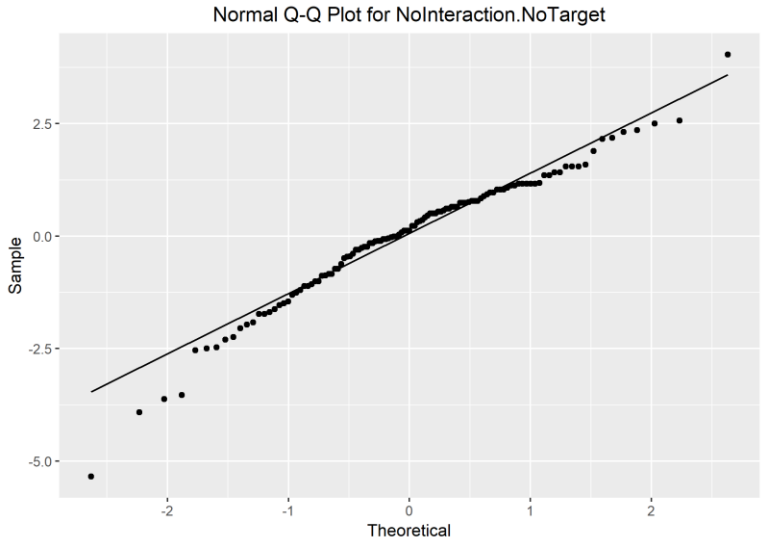


Figure 48 - Normal Q-Q plot for residuals (Correct Items Collected - Team)

Table 100 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	15.1	15.2
SessionOrder: 3	0.615	0.479
SessionOrder: 4	1.13	1.05

Restricted Maximum Likelihood Estimation (CI_{team})

The results for the fixed effect variables are displayed in and Table 101.

Table 101 - Fixed effect result of REML Estimation (Correct Items Collected)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	15.1	0.313	96	48.0	< .05*
Session: 3	0.615	0.368	76	1.67	.098
Session: 4	1.13	0.368	76	3.07	< .05*

The results show that the effect of session 3 was not significantly different from session 2, though it is approaching significance (p = 0.098). The results also indicate that the effect of session 4 is significantly different from session 2.

Evaluating Effect size (CI_{team})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 102. The results suggest that the fixed variables have a small effect size (i.e., explains 5% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 35% of the variance).

Table 102 - Effect size for LMM (Correct Items Collected - Team)

R^2 Type	Value
R_m^2	0.053
R_c^2	0.347

Incorrect Items Collected (Team)

Distribution Overview (II_{team})

The overall distribution of the incorrect items collected teams at the team level shows a skew to the left (Figure 49). The distribution of the incorrect items collected, when grouped by Feedback, is similar to the overall distribution (Figure 50). Over time, the distribution of the incorrect items collected skews to the left (Figure 51).

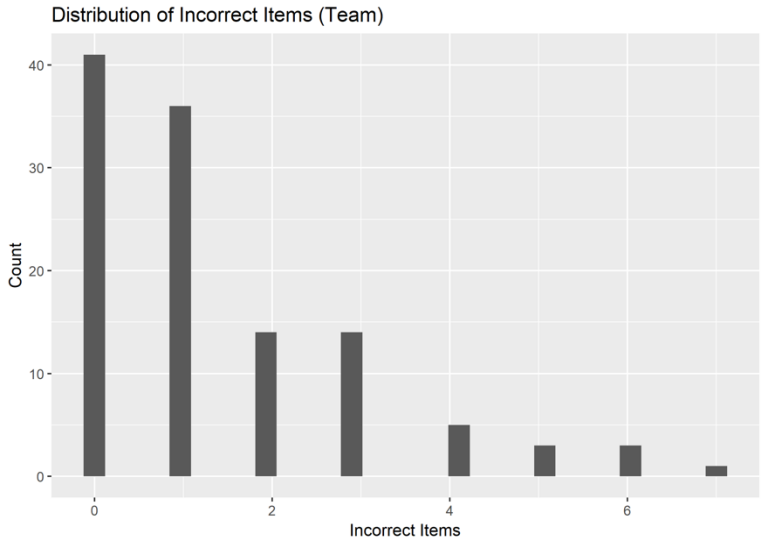


Figure 49 - Distribution overview of incorrect items collected (Team)

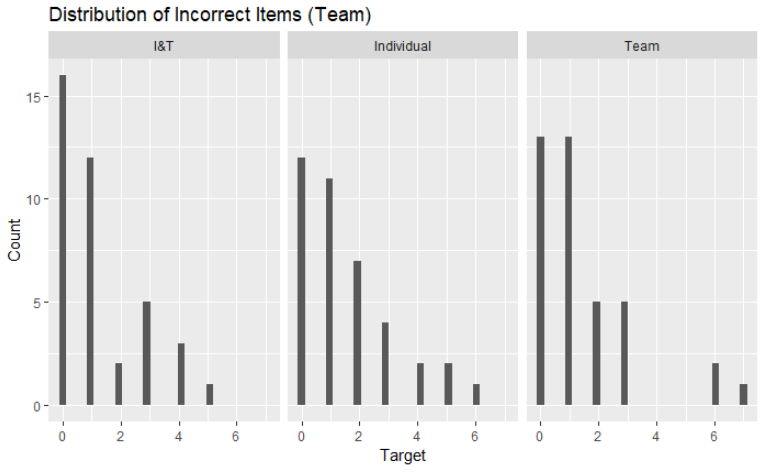


Figure 50 - Distribution of incorrect items collected grouped by Feedback (Team)

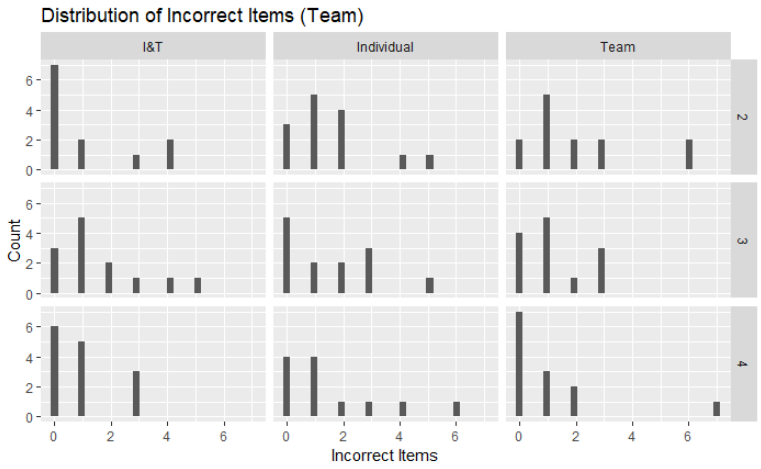


Figure 51 - Distribution of incorrect items collected grouped by Feedback and Session Order (Team)

Model Selection (II_{team})

The author used a model selection process outlined in Table 16. No models were significantly different from the null model. Indicating Feedback and Session Order did not significantly influence the incorrect items collected. The model used in the rest of the analysis is the No Interaction model since it includes both the Session Order and the Feedback variable.

Testing Assumptions: Residuals (II_{team})

The histogram of the residuals (Figure 52) appear to skew slightly to the left. The residual fitted-value (Figure 53) shows a violation of constant variance. The points on the residual normal Q-Q plot (Figure 54) do not lie in a roughly straight line, which suggests the residuals are not normally distributed. The classic LMM and the robust LMM produce concernedly different coefficient values (Table 103), so the researcher transformed the data. The classic LMM and the robust LMM for the transformed data similar coefficient values (Table 104).

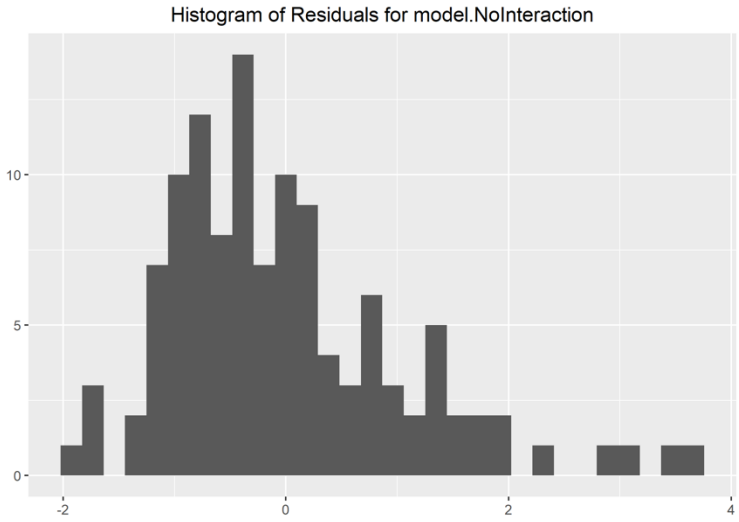


Figure 52 - Histogram of Residuals (Incorrect Items Collected - Team)

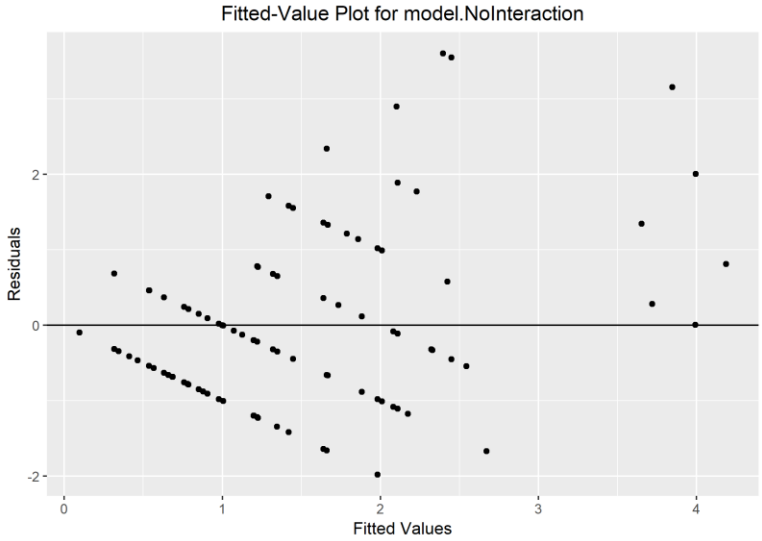


Figure 53 - Residual fitted plot (Incorrect Items Collected - Team)

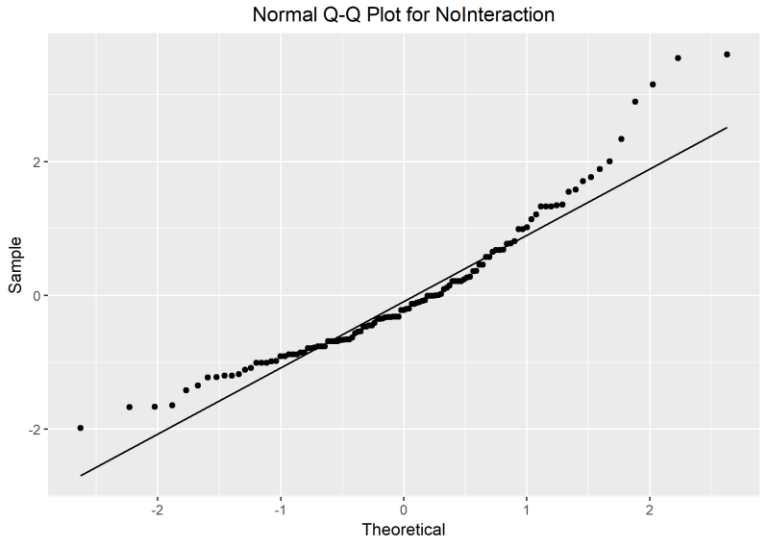


Figure 54 - Normal Q-Q plot for residuals (Incorrect Items Collected - Team)

Table 103 - Comparing classic model to a robust model

Coefficient	No Interaction	Robust No Interaction
Intercept	1.72	1.38
Feedback: Ind_Team	-0.314	-0.184
Feedback: Team	-0.093	-0.054
SessionOrder: 3	-0.120	0.061
SessionOrder: 4	-0.369	-0.329

Table 104 - Comparing classic model to a robust model for transformed data

Coefficient	No Interaction	Robust No Interaction
Intercept	1.07	1.04
Feedback: Ind_Team	-0.169	-0.162
Feedback: Team	-0.058	-0.059
SessionOrder: 3	-0.025	-0.005
SessionOrder: 4	-0.211	-0.210

Restricted Maximum Likelihood Estimation (II_{team})

The results for the fixed effect variables are displayed in Table 105.

Table 105 - Fixed effect result of REML estimation (Incorrect Items Collected)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	1.07	0.146	110	7.29	< .05*
Feedback: Individual and Team	-0.169	0.141	74	-1.20	.235
Feedback: Team	-0.058	0.141	74	-0.411	.683
Session: 3	-0.025	0.141	74	-0.178	.859
Session: 4	-0.211	0.141	74	-1.49	.141

The results showed that the effect of the I&T and the Team condition was not significantly different from zero in reference to the Ind level. The results also show that session 3 and session 4 did not have a significant effect in reference to session 2.

Evaluating Effect size (II_{team})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 106. The results suggest that the fixed variables have a unmeaningful effect size (i.e., explains 0.8% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 23.3% of the variance).

Table 106 - Effect size for LMM (Incorrect Items Collected - Team)

R^2 Type	Value
R_m^2	0.008
R_c^2	0.233

Time remaining (Team)

Distribution Overview (T_{team})

The overall distribution of the T_{team} shows a skew to the left (Figure 55). The distribution of T_{team}, when grouped by Feedback, are similar to the overall distribution (Figure 56). Over time, the distribution of each feedback group flattens, indicating teams were completing the session faster with each passing session (Figure 57).

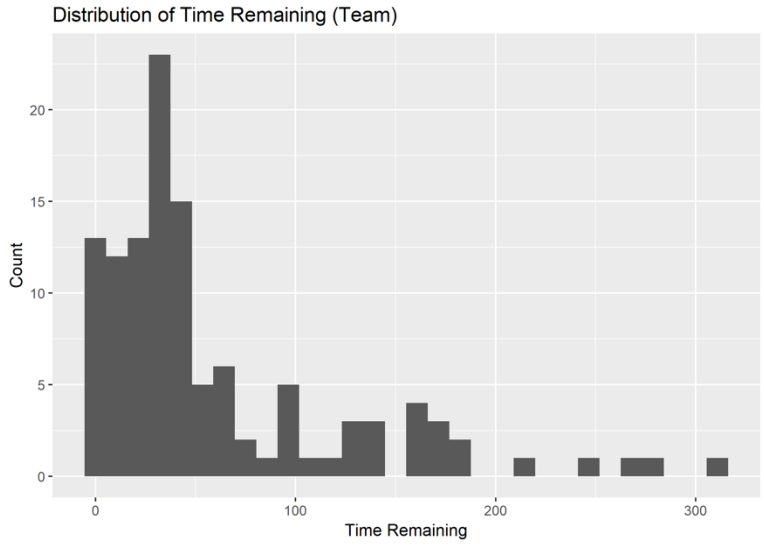


Figure 55 - Distribution overview of time remaining (Team)

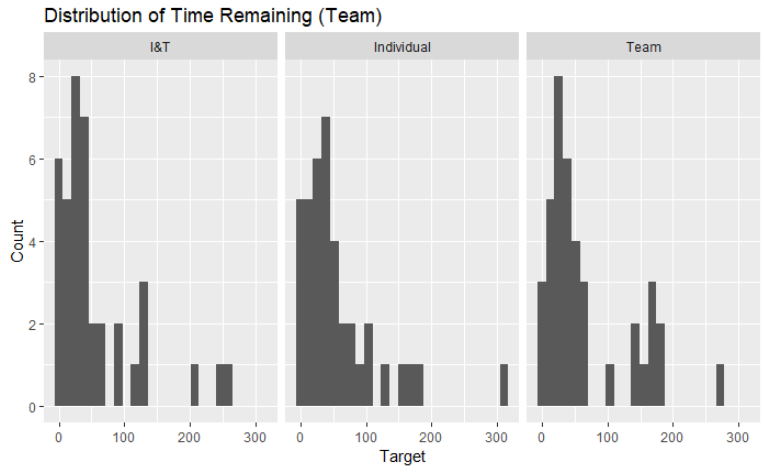


Figure 56 - Distribution of time remaining grouped by Feedback (Team)

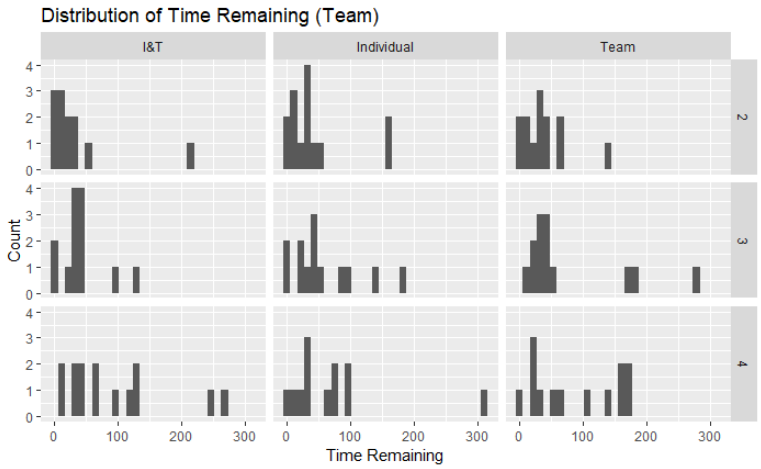


Figure 57 - Distribution of time remaining grouped by Feedback and Session Order (Team)

Model Selection (T_{team})

The author used a model selection process outlined in Table 16. The models that are significantly different from the null model are displayed in Table 107. The All model was not significantly different from the No Interaction or the No Interaction No Feedback model, indicating the interaction effect can be excluded from the model. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback variable can be excluded from the model. The model to best represent the data is the No Interaction No Feedback model.

Table 107 – Models that are significantly different from the null model

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	1238.67	1269.06
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	1232.76	1252.09
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**1231.50**	**1245.31**

Testing Assumptions: Residuals (T_{team})

The histogram of the residuals (Figure 58) appears to be approximately distributed. The residual fitted-value (Figure 59) shows a violation of constant variance. The points on

the residual normal Q-Q plot (Figure 60) lie in a roughly straight line but trail away from a straight line when moving from left to right, violating the assumption that the residuals are normal. The classic LMM and the robust LMM produce noticeably different results (Table 108), so the researcher transformed the data.

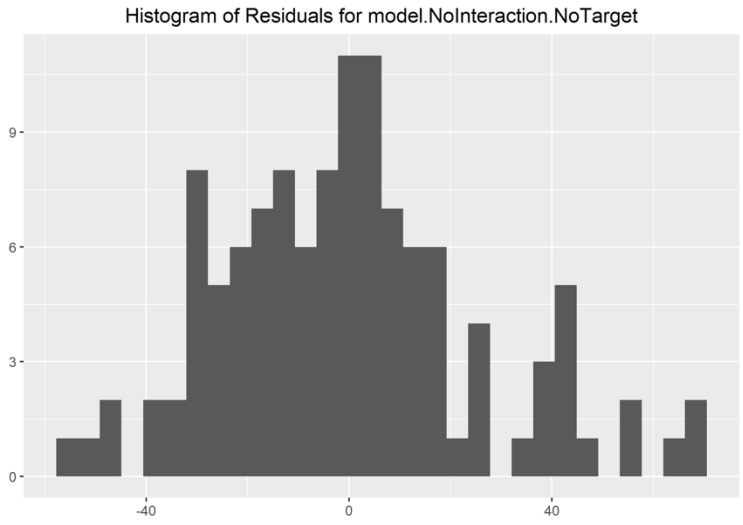


Figure 58 - Histogram of Residuals (T_{team})

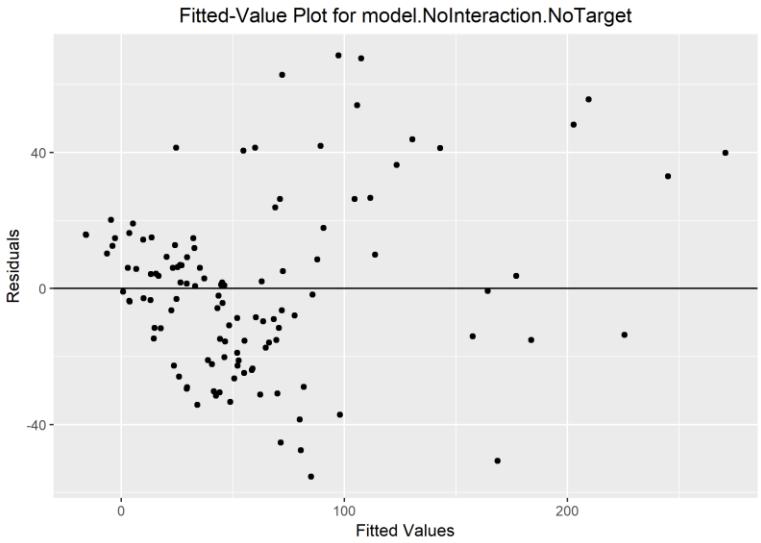


Figure 59 - Residual fitted plot (T_{team})

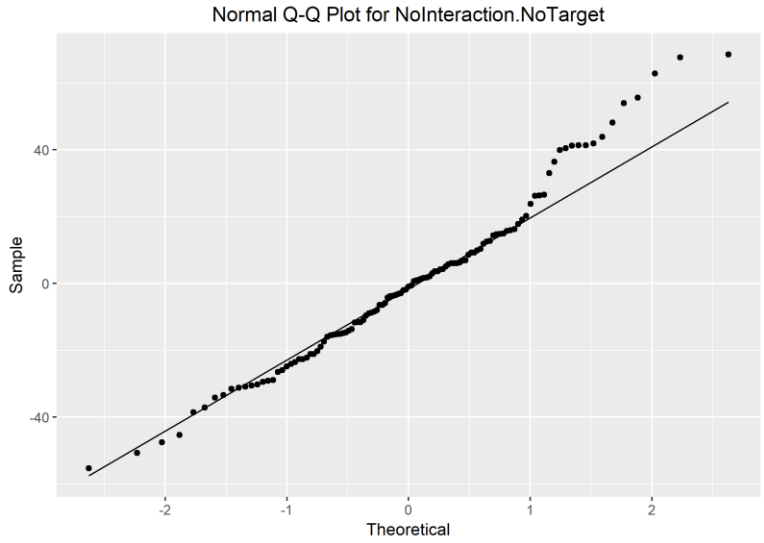


Figure 60 - Normal Q-Q plot for residuals (T_{team})

Table 108 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	38.5	28.2
SessionOrder: 3	19.5	17.9
SessionOrder: 4	45.3	39.2

Restricted Maximum Likelihood Estimation (T_{team})

The results for the fixed effect variables are displayed in Table 109, respectively.

Table 109 - Fixed effect result of REML Estimation (Time remaining - Team)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	5.13	0.605	55	8.48	< .05*
Session: 3	1.58	0.440	76	3.58	< .05*
Session: 4	3.08	0.440	76	7.00	< .05*

The results showed that session 3 and session 4 are significant in reference to session 2.

Evaluating Effect size (T_{team})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 110. The results suggest that the fixed variables have a small effect size (i.e., explains 8% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 77.9% of the variance).

Table 110 - Effect size for LMM (Time remaining - Team)

R^2 Type	Value
R_m^2	0.082
R_c^2	0.779

Unique Errors Committed (Team)

Distribution Overview (E_{team})

The overall distribution of E_{team} skews to the left (Figure 61). The distribution of E_{team}, when grouped by Feedback, are similar to the overall distribution (Figure 62). The distribution of E_{team} flattens over time (Figure 63).

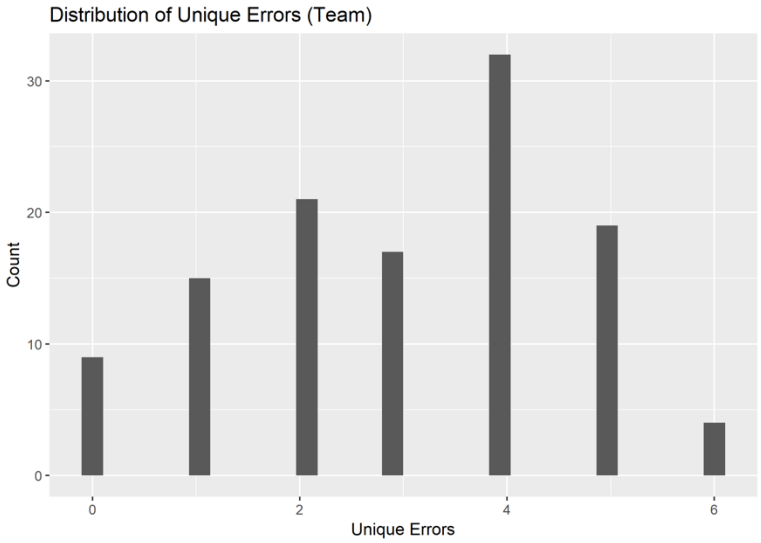


Figure 61 - Distribution overview of unique errors (Team)

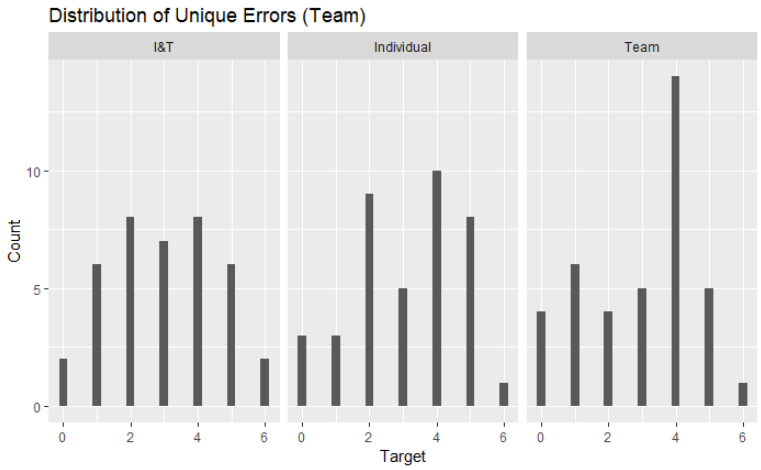


Figure 62 - Distribution of unique errors grouped by Feedback (Team)

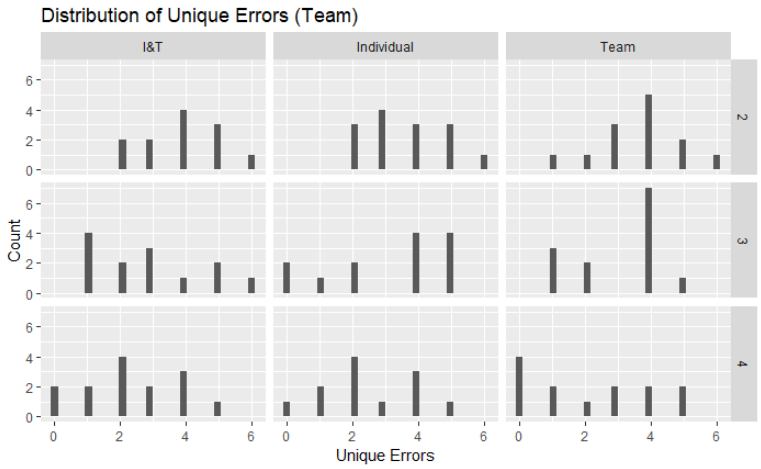


Figure 63 - Distribution of unique errors grouped by Feedback and Session Order (Team)

Model Selection (E_{team})

The author used a model selection process outlined in Table 16. The models that are significantly different from the null model are displayed in Table 111. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible and the simplest model that describes the data best in the No Interaction No Feedback model.

Table 111 – Models that are significantly different from the null model

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	410.71	441.09
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	407.23	426.57
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**403.50**	**417.31**

Testing Assumptions: Residuals (E_{team})

The histogram of the residuals (Figure 64) appears to be approximately distributed. The residual fitted-value (Figure 65) shows a violation of constant variance. The points on the residual normal Q-Q plot (Figure 66) lie in a roughly straight line but appears to trail off at the end of the line when moving from left to right, which suggests a possible violation of assumption three (Table 19). The classic LMM and robust LMM produce similar coefficient values (Table 112). The researcher did not transform the data.

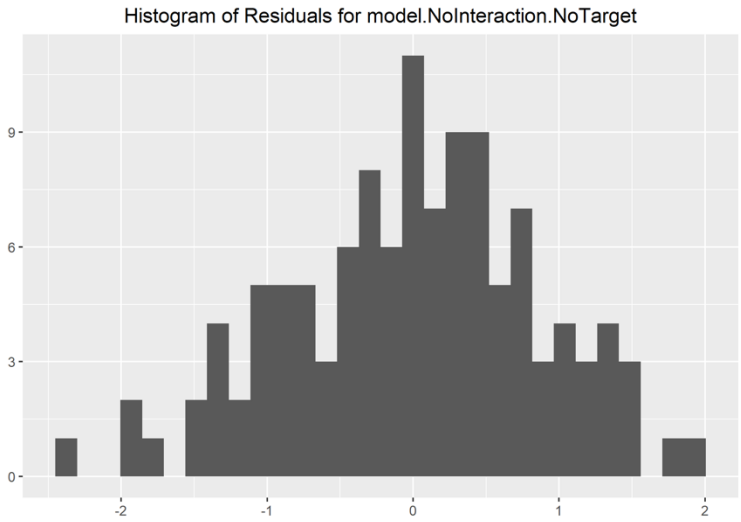


Figure 64 - Histogram of Residuals (Unique Errors Committed - Team)

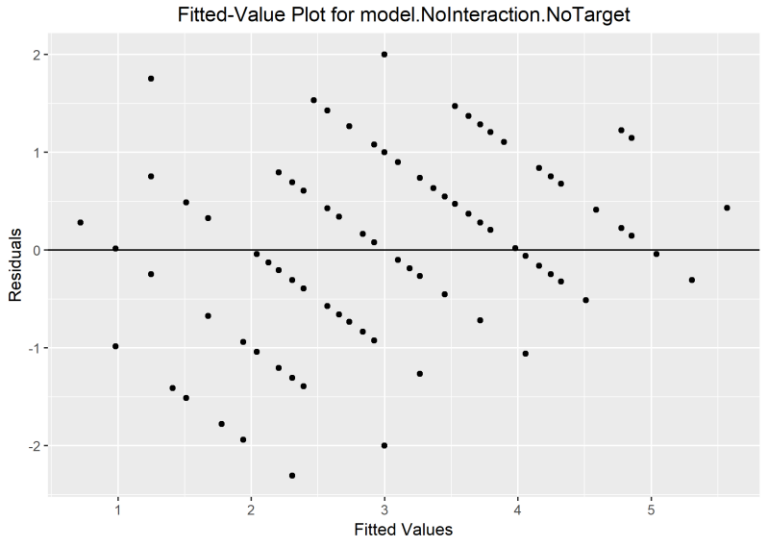


Figure 65 - Residual fitted plot (Unique Errors Committed - Team)

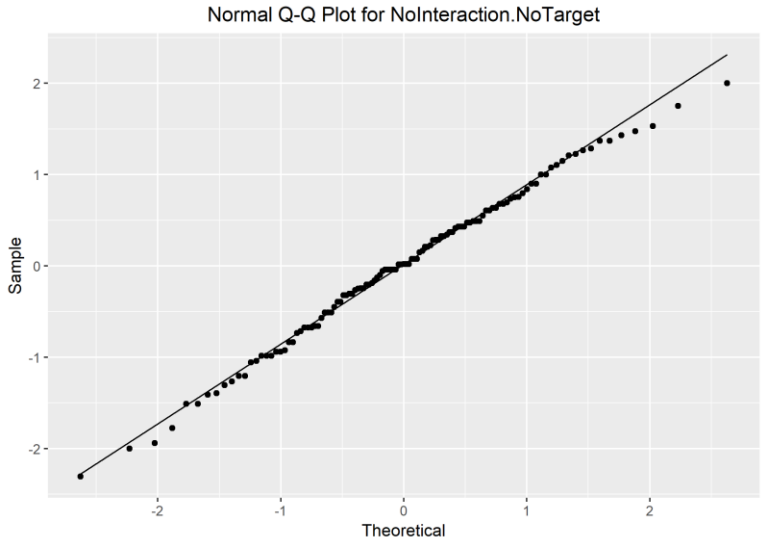


Figure 66 - Normal Q-Q plot for residuals (Unique Errors Committed - Team)

Table 112 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	3.74	3.74
SessionOrder: 3	-0.718	-0.698
SessionOrder: 4	-1.41	-1.37

Restricted Maximum Likelihood Estimation (E_{team})

The results for the fixed effect variables are displayed in Table 113.

Table 113 - Fixed effect result of REML Estimation (Unique Errors Committed - Team)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	3.74	0.245	70	15.3	< .05*
Session: 3	-0.718	0.229	76	-3.13	< .05*
Session: 4	-1.41	0.229	76	-6.15	< .05*

The results showed that the effect of session 3 and session 4 are significant in reference to session 2.

Evaluating Effect size (E_{team})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 114. The results suggest that the fixed variables have a small effect size (i.e., explains 8% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 52% of the variance).

Table 114 - Effect size for LMM (Unique Errors Committed - Team)

R^2 Type	Value
R_m^2	0.081
R_c^2	0.517

Correct Items Collected (Individual)

Distribution Overview (CI_{ind})

The overall distribution of the CI_{ind} shows a skew to the right (Figure 67). The distribution of the CI_{ind}, when grouped by Feedback, is similar to the overall distribution (Figure 68). Over time, the distribution of CI_{ind} skews to the right (Figure 69). It is interesting to note that the distribution in Team condition and Session 4 appears to have a higher skew than the Individual and I&T condition in session 4.

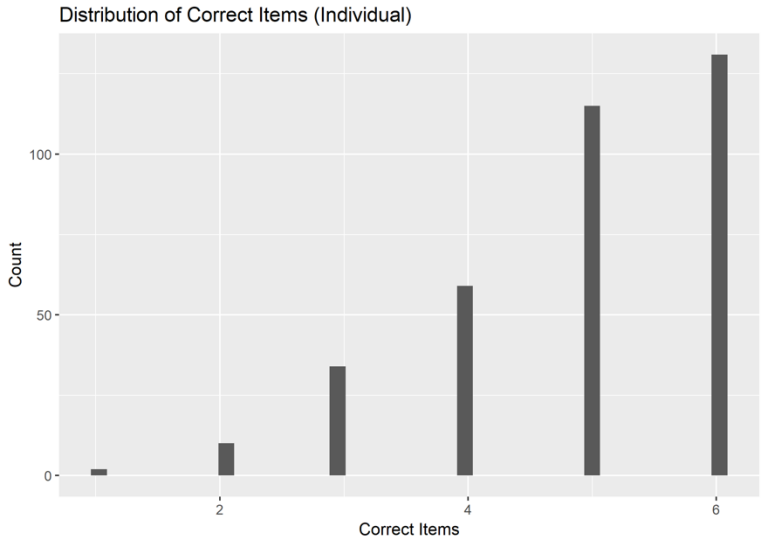


Figure 67 - Distribution overview of CI_{ind} (Individual)

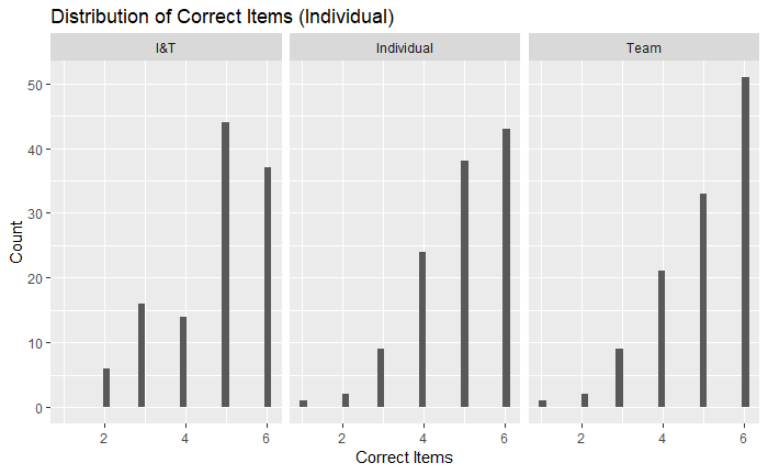


Figure 68 - Distribution of CI_{ind} grouped by Feedback (Individual)

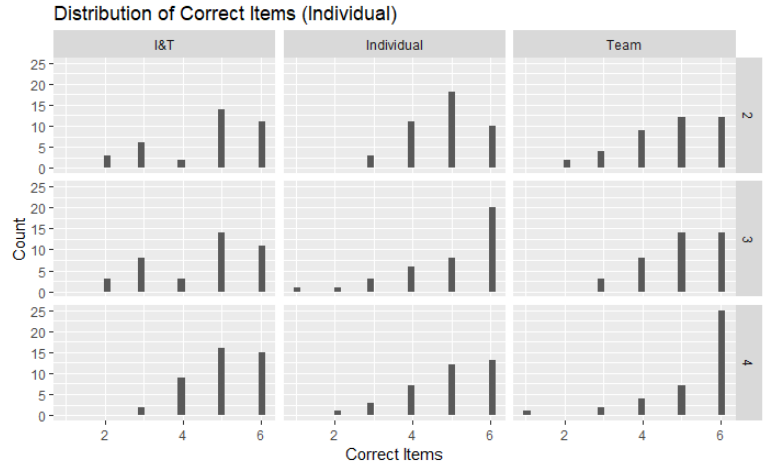


Figure 69 - Distribution of CI_{ind} grouped by Feedback and Session Order (Individual)

Model Selection (CI_{ind})

The author used a model selection process outlined in Table 16. The models that are significantly different from the null model are displayed in Table 115. The All model was not significantly different from the models with No Interaction model or the model with No Interaction No Feedback, indicating that the interaction effect is negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible and the simplest model that describes the data best is the No Interaction No Feedback model (Table 115). It is important to note that the difference between the No Interaction model and the No Interaction No Feedback model was approaching significance (i.e., $p < 0.1$), indicating more data is needed to reach significance. However, the researcher did not analyze the No Interaction model because it was not significantly different from the No Interaction No Feedback model.

Table 115 – Models that are significantly different from the null model

Model Name	Equation	AIC	BIC
All	$y_{ijtp} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	1039.16	1085.49
No Interaction	$y_{ijtp} = \mu + \alpha_i + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	1035.41	1066.30
No Interaction No Feedback	$\boldsymbol{y_{ijtp} = \mu + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}}$	**1036.49**	**1059.65**

Testing Assumptions: Residuals (CI_{ind})

The histogram of the residuals (Figure 70) appears to be slightly skewed the right. The residual fitted-value (Figure 71) shows a violation of constant variance. The points on the residual normal Q-Q plot (Figure 72) appear to not lie on a roughly straight line, which suggests a violation of the Q-Q plot assumption. The researcher did not transform the data because the classic LMM and the robust LMM produced similar coefficient values (Table 116).

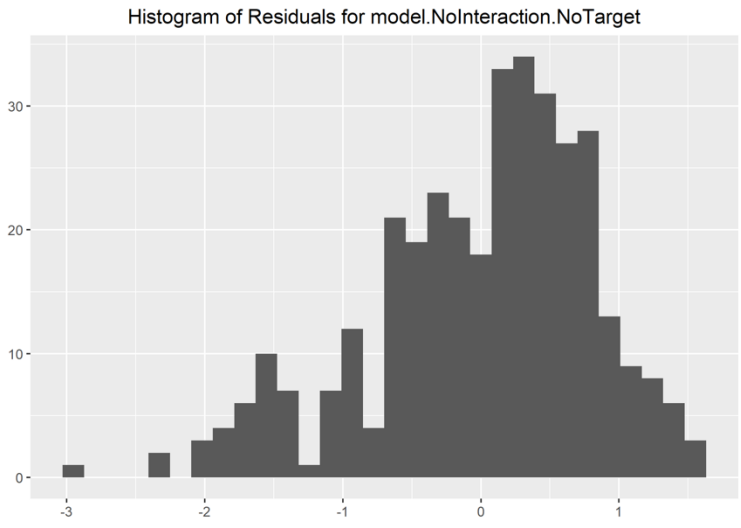


Figure 70 - Histogram of Residuals (Correct Items Collected – Individual)

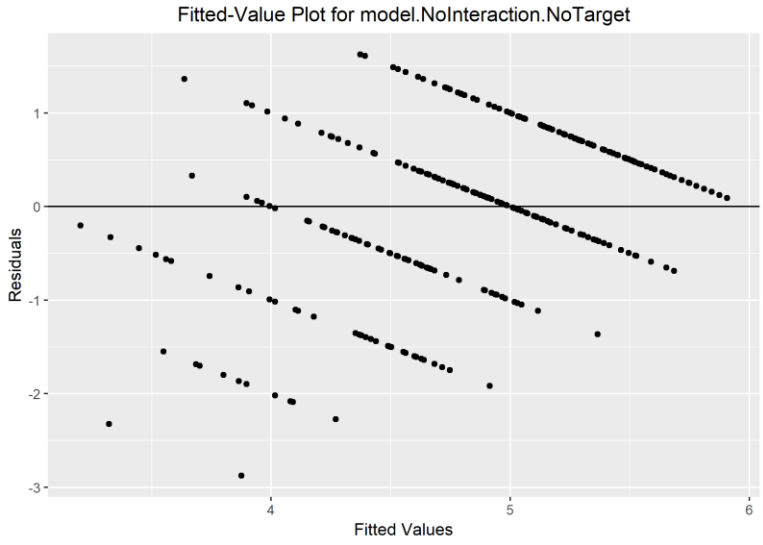


Figure 71 - Residual fitted plot (Correct Items Collected – Individual)

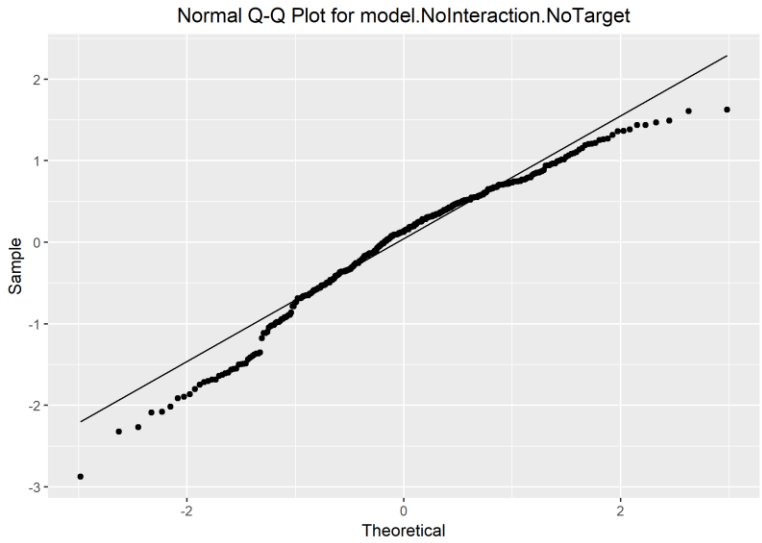


Figure 72 - Normal Q-Q plot for residuals (Correct Items Collected – Individual)

Table 116 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	4.74	4.82
SessionOrder: 3	0.120	0.140
SessionOrder: 4	0.359	0.346

Restricted Maximum Likelihood Estimation (CI_{ind})

The results for the fixed effect variables are displayed in Table 117.

Table 117 - Fixed effect result of REML Estimation (Correct Items Collected – Individual)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	4.74	0.120	80	39.6	< .05*
Session: 3	0.120	0.118	232	1.02	.310
Session: 4	0.359	0.118	232	3.05	< .05*

The results showed that the effect of session 3 and 4 are significant in reference to session 2.

Evaluating Effect size (CI_{ind})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 118. The results suggest that the fixed variables have a small effect size (i.e., explains 2% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 37% of the variance).

Table 118 - Effect size for LMM (Correct Items Collected – Individual)

R^2 Type	Value
R_m^2	0.017
R_c^2	0.369

Incorrect Items Collected (Individual)

Distribution Overview (II_{ind})

The overall distribution of the II_{ind} shows a skew to the left (Figure 73). The distribution of the II_{ind}, when grouped by Feedback, is similar to the overall distribution (Figure 74). Over time, the distribution of II_{ind} begins with a skew to the left and end with a skew to the left (Figure 75).

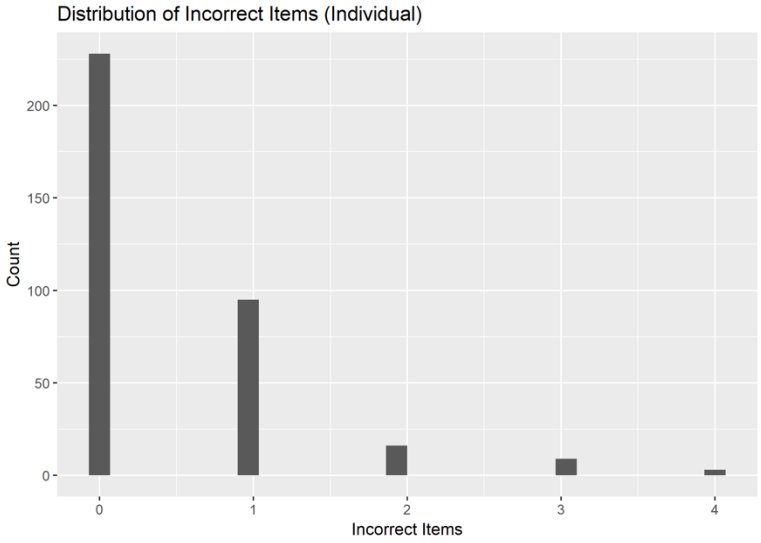


Figure 73 - Distribution overview of II_{ind} (Individual)

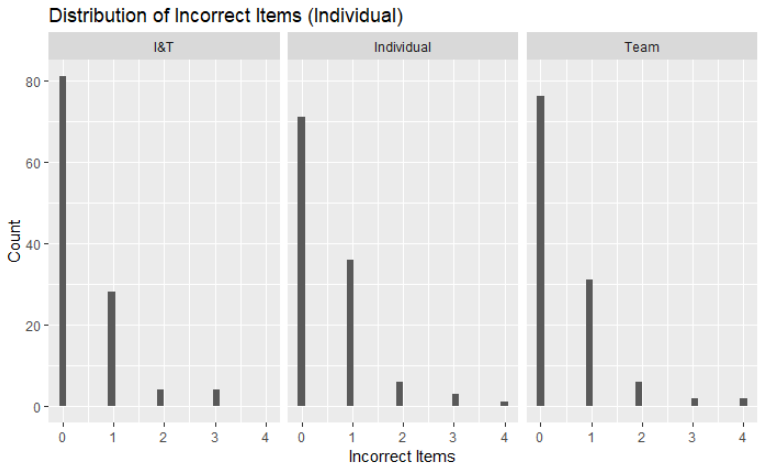


Figure 74 - Distribution of II_{ind} grouped by Feedback (Individual)

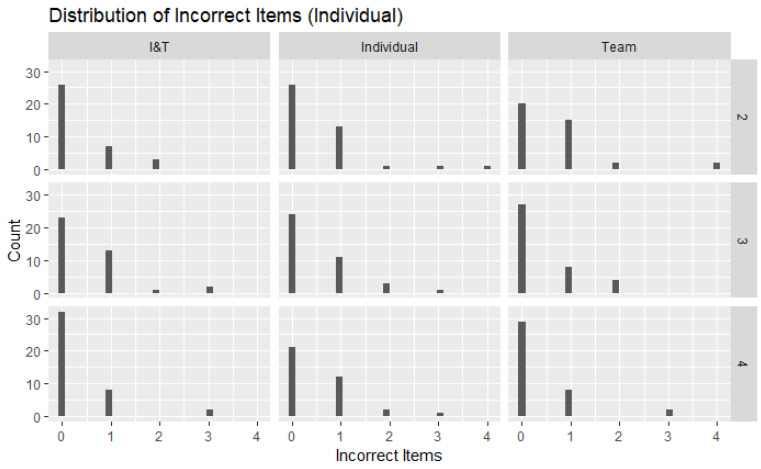


Figure 75 - Distribution of II_{ind} grouped by Feedback and Session Order (Individual)

Model Selection (II_{ind})

The author used a model selection process outlined in Table 16. No models were significantly different from the null model. Indicating Feedback and Session Order did not significantly influence II_{ind}. The model used in the rest of the analysis is the No Interaction model since it includes both the Session Order and the Feedback variable.

Testing Assumptions: Residuals (II_{ind})

The histogram of the residuals (Figure 76) appears to be skewed to the left, indicating a violation in the normal distribution assumption. The residual fitted-value (Figure 77) shows a violation of constant variance. The points on the residual normal Q-Q plot (Figure 78) does not lie in a roughly straight line, which suggests the residuals are not normally distributed. The researcher transformed the data because the classic LMM and the robust LMM did not produce similar results coefficient values (Table 125).

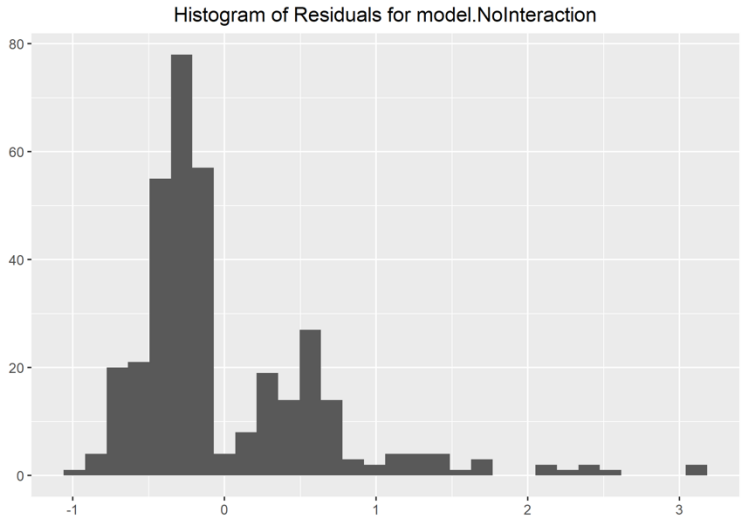


Figure 76 - Histogram of Residuals (Incorrect Items Collected – Individual)

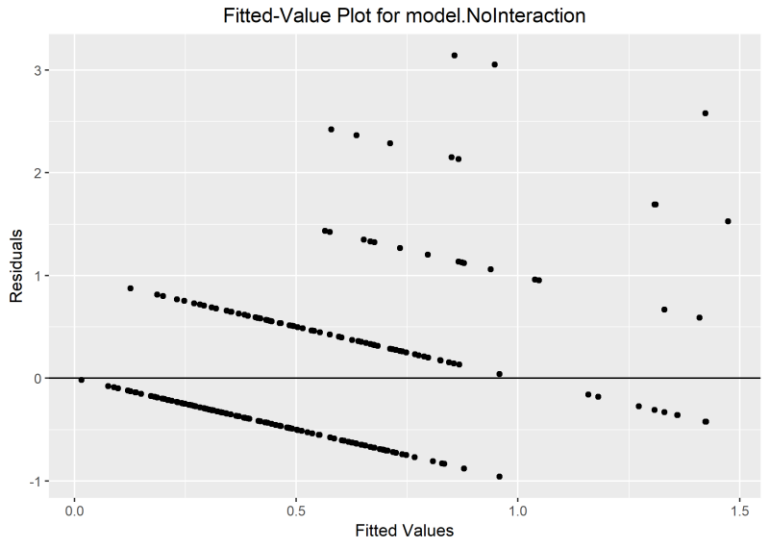


Figure 77 - Residual fitted plot (Incorrect Items Collected – Individual)

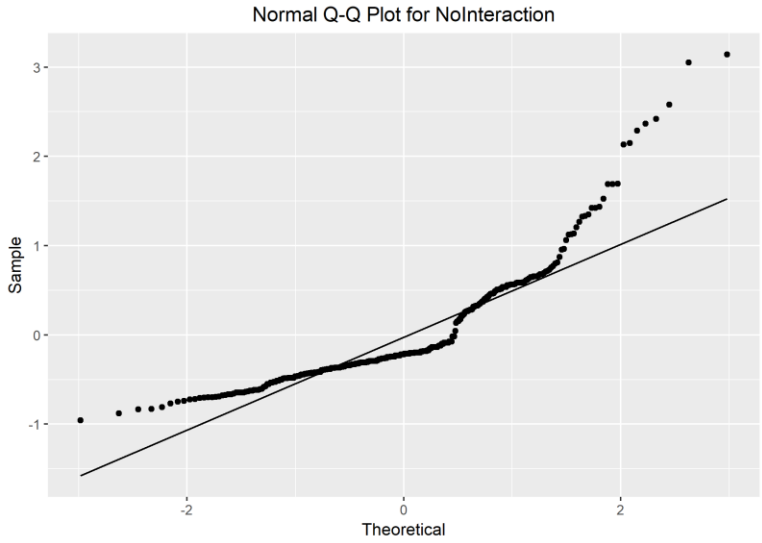


Figure 78 - Normal Q-Q plot for residuals (Incorrect Items Collected – Individual)

Table 119 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	0.573	0.431
Feedback: I&T	-0.105	-0.094
Feedback: Team	-0.031	-0.040
SessionOrder: 3	-0.040	-0.014
SessionOrder: 4	-0.123	-0.088

Restricted Maximum Likelihood Estimation (II_{ind})

The results for the fixed effect variables are displayed in Table 120.

Table 120 - Fixed effect result of REML Estimation (Incorrect Items Collected – Individual)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	0.478	0.071	143	6.69	< .05*
Feedback: Individual and Team	-0.090	0.065	230	-1.38	.168
Feedback: Team	-0.038	0.065	230	-0.584	.559
Session: 3	-0.021	0.065	230	-0.321	.749
Session: 4	-0.096	0.065	230	-1.47	.143

The results showed that the effect of the Feedback variable and the session order variable was not significantly different from zero.

Evaluating Effect size (II_{ind})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 121. The results suggest that the fixed variables have a unmeaningful effect size (i.e., explains 1% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 23% of the variance).

Table 121 - Effect size for LMM (Incorrect Items Collected – Individual)

R^2 Type	Value
R_m^2	0.008
R_c^2	0.233

Time remaining (Individual)

Distribution Overview (T_{ind})

The overall distribution of the T_{ind} shows a skew to the left (Figure 79). The distribution of the T_{ind}, when grouped by Feedback, are similar to the overall distribution (Figure 80). Over time, the distribution of T_{ind} flattens slightly (Figure 81).

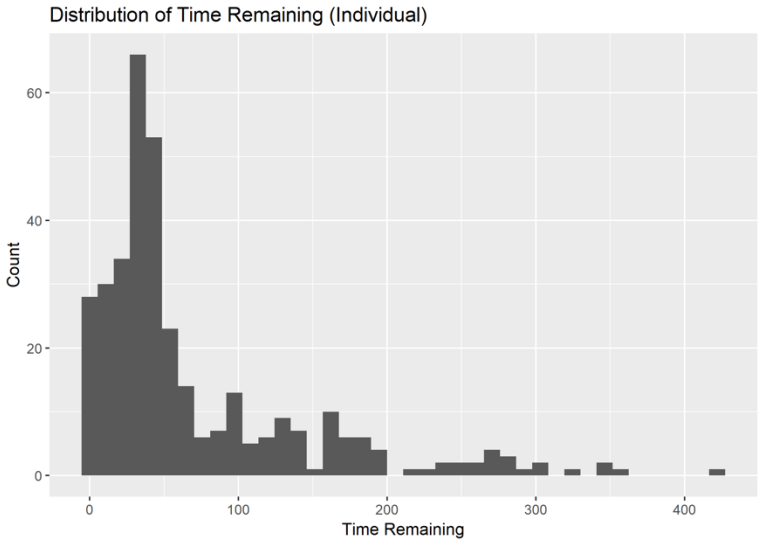


Figure 79 - Distribution overview of T_{ind} (Individual)

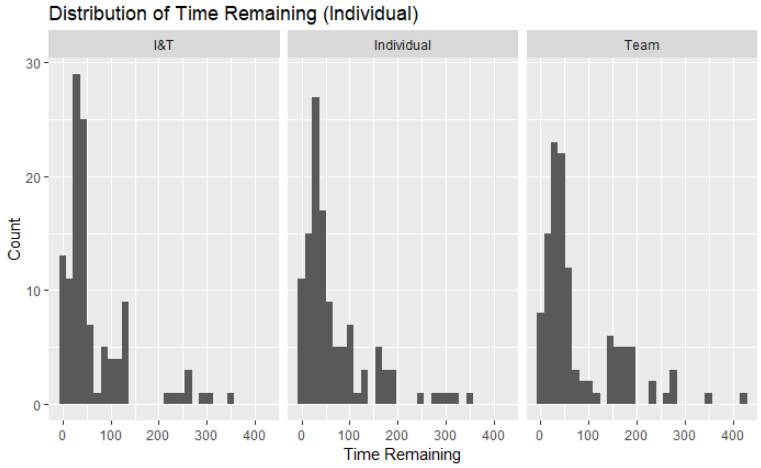


Figure 80 - Distribution of T_{ind} grouped by Feedback (Individual)

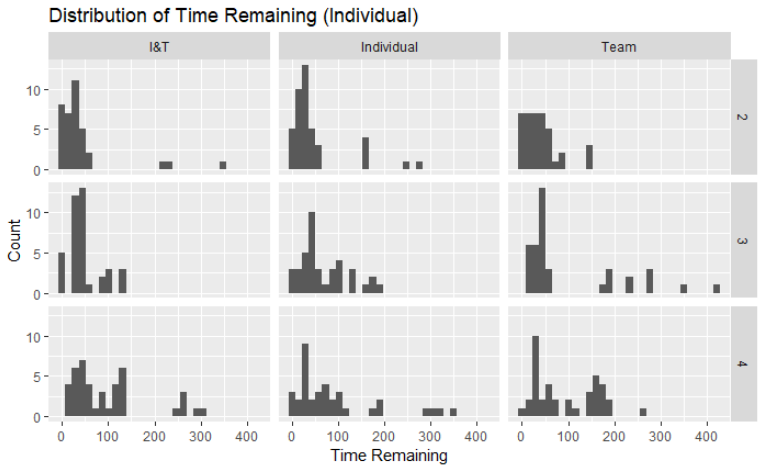


Figure 81 - Distribution of T_{ind} grouped by Feedback and Session Order (Individual)

Model Selection (T_{ind})

The author used a model selection process outlined in Table 16. The models that are significantly different from the null model are displayed in Table 122. The All model was significantly different from the No Interaction No Feedback model and the No Interaction No Session Order model, indicating that neither the Session Order nor the Feedback condition alone is best at describing the data. However, the All model was no significantly different from the No Interaction model, indicating that the interaction effect is negligible. The No Interaction model is used in this analysis.

Table 122 – Models that are significantly different from the null model

Model Name	Equation	AIC	BIC
All	$y_{ijtp} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	3658.26	3704.59
No Interaction	$\boldsymbol{y_{ijtp} = \mu + \alpha_i + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}}$	**3652.92**	**3683.81**
No Interaction No Feedback	$y_{ijtp} = \mu + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	3659.46	3682.63
No Interaction No Session Order	$y_{ijtp} = \mu + \alpha_i + \gamma_t + \theta_p + \epsilon_{ijtp}$	3747.24	3770.40

Testing Assumptions: Residuals (T_{ind})

The histogram of the residuals (Figure 82) appears to be approximately distributed. The residual fitted-value (Figure 83) shows a concentration of errors, indicating an assumption violation. The points on the residual normal Q-Q plot (Figure 84) lie in a roughly straight line but trails off on the tail end, moving from left to right, which suggests violation in residual assumption. The researcher transformed the data because the classic LMM and the robust LMM did not produce similar results coefficient values (Table 123).

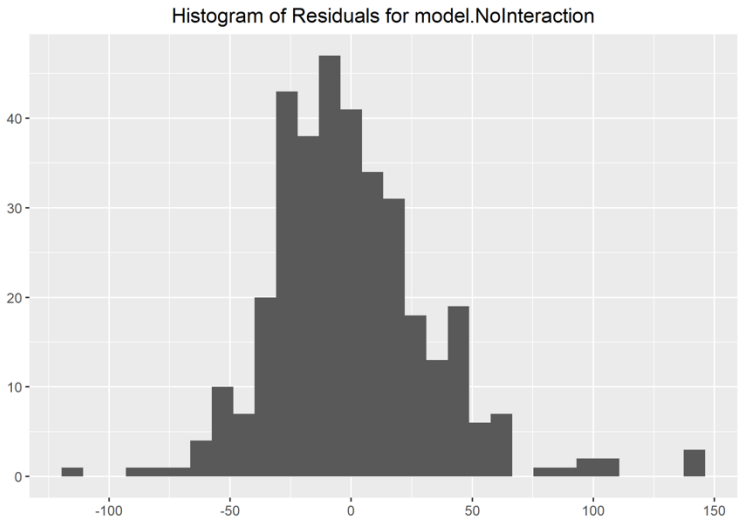


Figure 82 - Histogram of Residuals (Time remaining – Individual)

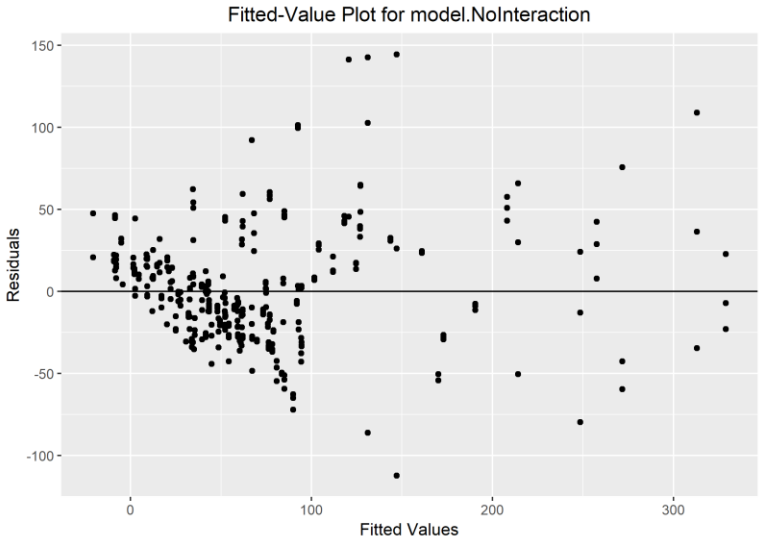


Figure 83 - Residual fitted plot (Time remaining – Individual)

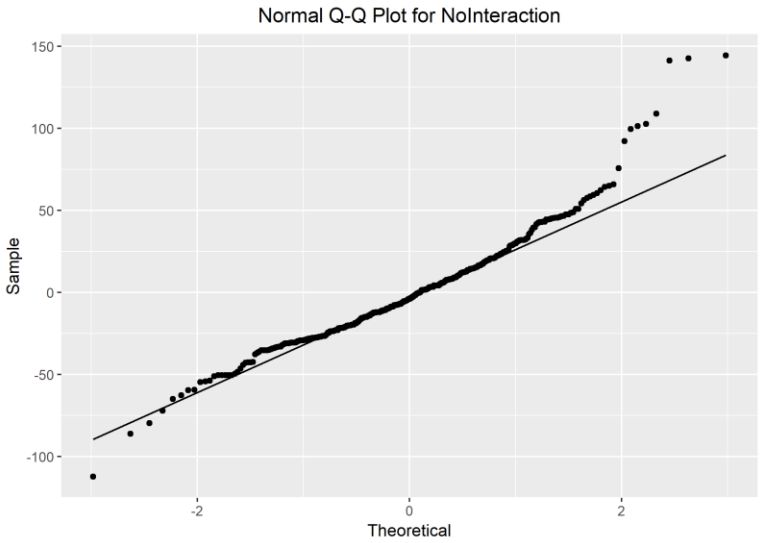


Figure 84 - Normal Q-Q plot for residuals (Time remaining – Individual)

Table 123 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	44.3	30.9
Feedback: I&T	-6.96	-3.11
Feedback: Team	8.33	6.07
SessionOrder: 3	25.9	20.5
SessionOrder: 4	50.3	41.9

Restricted Maximum Likelihood Estimation (T_{ind})

The results for the fixed effect variables are displayed in Table 124, respectively.

Table 124 - Fixed effect result of REML Estimation (Time remaining – Individual)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	5.58	0.580	51	9.62	< .05*
Feedback: Individual and Team	-0.436	0.269	308	-1.62	.106
Feedback: Team	0.475	0.269	308	1.77	.078
Session: 3	1.88	0.269	308	7.00	< .05*
Session: 4	3.28	0.269	308	12.2	< .05*

The results showed that the effect of the Feedback variable was not significantly different from zero. However, it is important to note that the effect of the Feedback variable at the Team level ($p = .08$) is trending to significance. This suggests that more participants are needed to reject the null hypothesis. The results also show that session 3 and session 4 have a significant effect in reference to session 2.

Evaluating Effect size (T_{ind})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 118. The

results suggest that the fixed variables have a small effect size (i.e., explains 8% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 77% of the variance).

Table 125 - Effect size for LMM (Time remaining – Individual)

R^2 Type	Value
R_m^2	0.080
R_c^2	0.773

Unique Errors Committed (Individual)

Distribution Overview (E_{ind})

The overall distribution of the E_{ind} shows a skew to the left (Figure 85). The distribution of the E_{ind}, when grouped by Feedback, is similar to the overall distribution (Figure 86). Over time, the distribution of E_{ind} skews to the left (Figure 87).

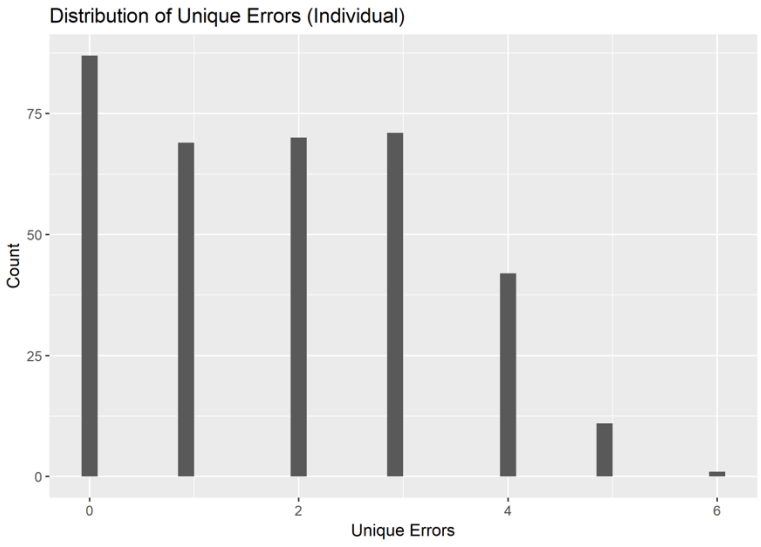


Figure 85 - Distribution overview of T_{ind} (Individual)

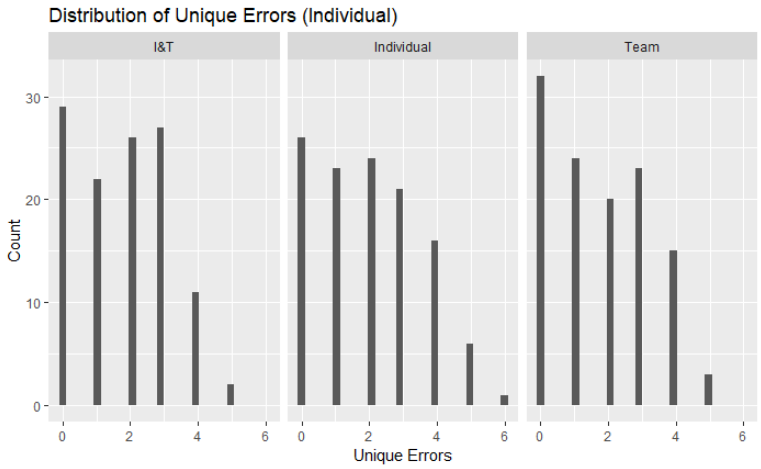


Figure 86 - Distribution of T_{ind} grouped by Feedback (Individual)

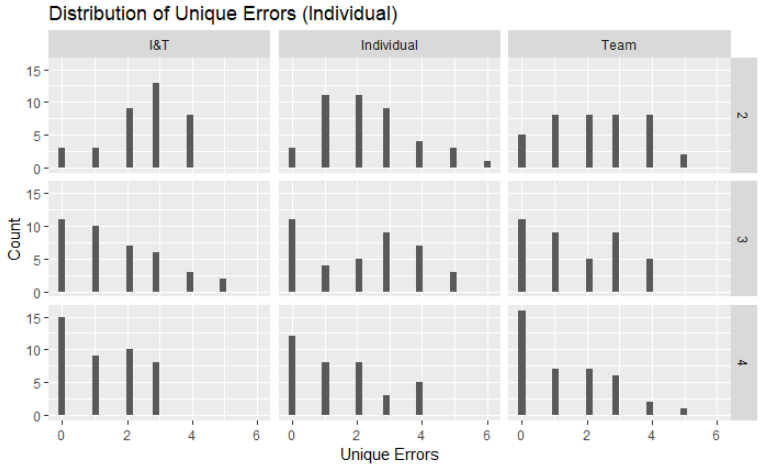


Figure 87 - Distribution of T_{ind} grouped by Feedback and Session Order (Individual)

Model Selection (E_{ind})

The author used a model selection process outlined in Table 16. The models that are significantly different from the null model are displayed in Table 126. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect is negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating the simplest model that describes the data best is the No Interaction No Feedback model.

Table 126 – Models that are significantly different from the null model

Model Name	Equation	AIC	BIC
All	$y_{ijtp} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	1139.84	1186.17
No Interaction	$y_{ijtp} = \mu + \alpha_i + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}$	1132.27	1163.16
No Interaction No Feedback	$\boldsymbol{y_{ijtp} = \mu + \beta_j + \gamma_t + \theta_p + \epsilon_{ijtp}}$	**1130.70**	**1153.86**

Testing Assumptions: Residuals (E_{ind})

The histogram of the residuals (Figure 88) appears to be approximately distributed. The residual fitted-value (Figure 89) shows a violation in the assumption of constant residuals. The points on the residual normal Q-Q plot (Figure 90) lie in a roughly straight line, which suggests the residuals are normally distributed. The researcher did not transform the data because the classic LMM and the robust LMM produce similar coefficient values (Table 123).

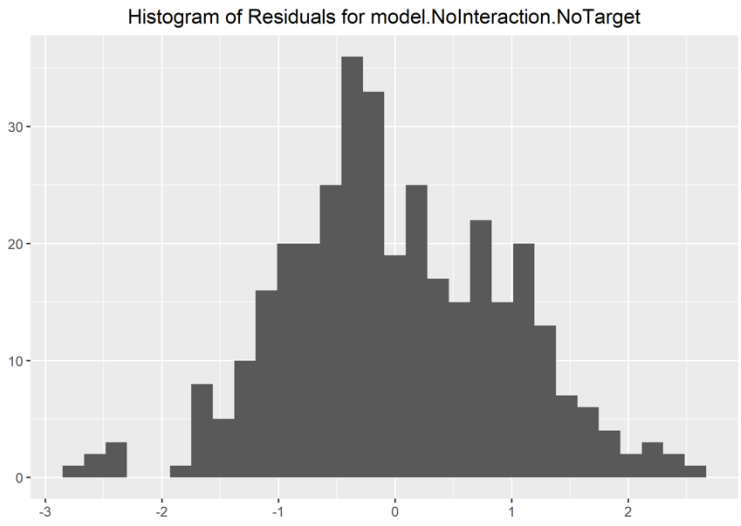


Figure 88 - Histogram of Residuals (Unique Errors Committed – Individual)

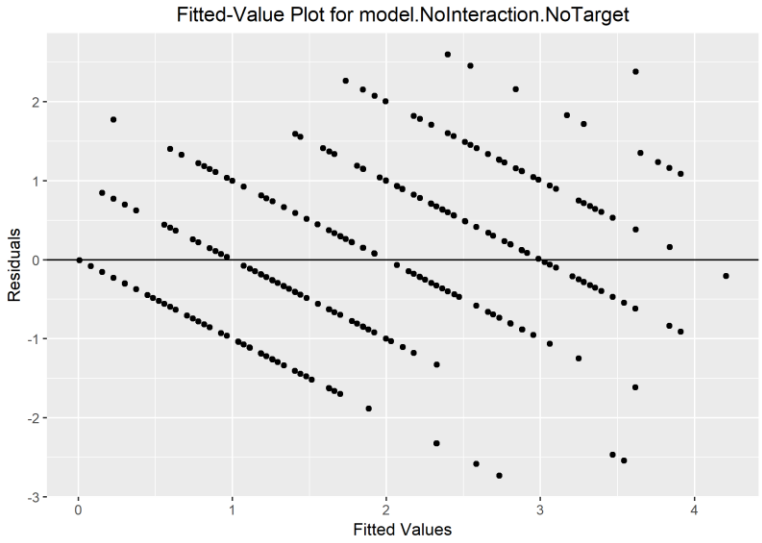


Figure 89 - Residual fitted plot (Unique Errors Committed – Individual)

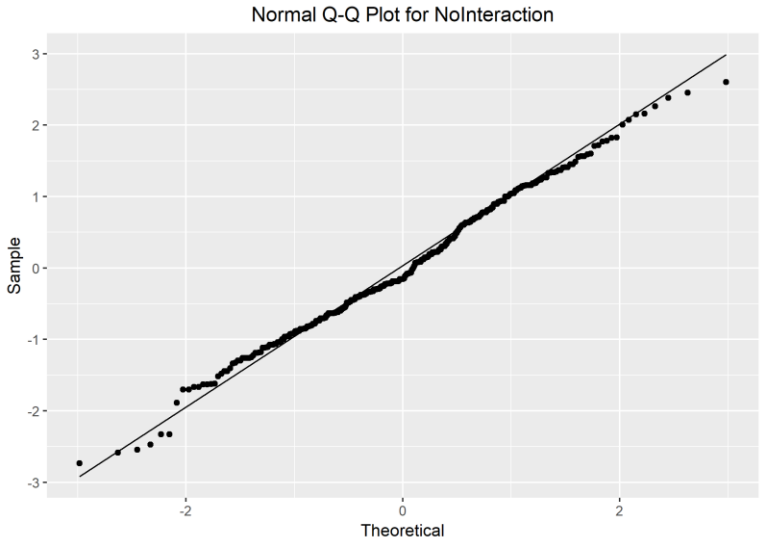


Figure 90 - Normal Q-Q plot for residuals (Unique Errors Committed – Individual)

Table 127 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	2.39	2.40
SessionOrder: 3	-0.556	-0.597
SessionOrder: 4	-1.03	-1.05

Restricted Maximum Likelihood Estimation (E_{ind})

The results for the fixed effect variables are displayed in Table 128.

Table 128 - Fixed effect result of REML estimation (Unique Errors Committed – Individual)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	2.38	0.179	57	13.3	< .05*
Session: 3	-0.556	0.135	232	-4.12	< .05*
Session: 4	-1.03	0.135	232	-7.66	< .05*

The results showed that the effect of the session order variable was not significantly different from zero.

Evaluating Effect size (E_{ind})

Johnson (2014) method was used to generate two types of R^2 values called marginal R^2 (R_m^2) and conditional R^2 (R_c^2). The R_m^2 describes how the variance is described by the fixed effect variables while the R_c^2 describes how both the variance is described by both the fixed and random effect variables. The values for R_m^2 and R_c^2 and presented in Table 129. The results suggest that the fixed variables have a small effect size (i.e., explains 8% of the variance) and the fixed and rand variables have a large effect size (i.e., explains 52% of the variance).

Table 129 - Effect size for LMM (Unique Errors Committed – Individual)

R^2 Type	Value
R_m^2	0.081
R_c^2	0.517

Summary Discussion Plots

Table 130 - Data for participant response to a question, "Did you notice any feedback during the task?" The data is grouped by the feedback condition.

Feedback	Response	N	Percent
Individual	No	12	10.3
Individual	Yes	105	89.7
I&T	No	12	10.3
I&T	Yes	105	89.7
Team	No	13	11.1
Team	Yes	104	88.9

Table 131 - Data for participant response to a question, "Did you notice any feedback during the task?" The data is grouped by feedback condition and session order.

Feedback	Session	Response	N	Percent
Individual	2	No	3	7.1
Individual	2	Yes	39	92.9
Individual	3	No	4	10.3
Individual	3	Yes	35	89.7
Individual	4	No	5	13.9
Individual	4	Yes	31	86.1
I&T	2	No	4	11.1
I&T	2	Yes	32	88. 9
I&T	3	No	6	15.4
I&T	3	Yes	33	84.6
I&T	4	No	2	4.8
I&T	4	Yes	40	95.2
Team	2	No	5	12.8
Team	2	Yes	34	87.2
Team	3	No	4	10.3
Team	3	Yes	35	89.7
Team	4	No	4	10.3
Team	4	Yes	35	89.7

Table 132 - Data for participant response to a question, " Did you find the feedback helpful?" The data is grouped by the feedback condition.

Feedback	Response	N	Percent
Individual	I ignored the feedback	8	6.8
Individual	No, it was actually distracting	8	6.8
Individual	No, it was not very helpful	23	19.7
Individual	Yes, it was somewhat helpful	44	37.6
Individual	Yes, it was very helpful	34	29.1
I&T	I ignored the feedback	11	9.4
I&T	No, it was actually distracting	6	5.1
I&T	No, it was not very helpful	9	7.7
I&T	Yes, it was somewhat helpful	41	35.0
I&T	Yes, it was very helpful	50	42.7
Team	I ignored the feedback	11	9.4
Team	No, it was actually distracting	5	4.3
Team	No, it was not very helpful	15	12.8
Team	Yes, it was somewhat helpful	53	45.3
Team	Yes, it was very helpful	33	28.2

Table 133 - Data for participant response to a question, " Did you find the feedback helpful?" The data is grouped by feedback condition and session order.

Feedback	Session	Response	N	Percent
Individual	2	I ignored the feedback	0	0.0
Individual	2	No, it was actually distracting	4	9.5
Individual	2	No, it was not very helpful	9	21.4
Individual	2	Yes, it was somewhat helpful	18	42.9
Individual	2	Yes, it was very helpful	11	26.2
Individual	3	I ignored the feedback	4	10.3
Individual	3	No, it was actually distracting	3	7.7
Individual	3	No, it was not very helpful	10	25.6
Individual	3	Yes, it was somewhat helpful	10	25.6
Individual	3	Yes, it was very helpful	12	30.8
Individual	4	I ignored the feedback	4	11.1
Individual	4	No, it was actually distracting	1	2. 8
Individual	4	No, it was not very helpful	4	11.1
Individual	4	Yes, it was somewhat helpful	16	44.4
Individual	4	Yes, it was very helpful	11	30. 6
I&T	2	I ignored the feedback	5	13. 9
I&T	2	No, it was actually distracting	2	5. 6
I&T	2	No, it was not very helpful	4	11.1
I&T	2	Yes, it was somewhat helpful	12	33.3
I&T	2	Yes, it was very helpful	13	36.1

Table 133 (continued)

Feedback	Session	Response	N	Percent
I&T	3	I ignored the feedback	4	10.3
I&T	3	No, it was actually distracting	2	5.1
I&T	3	No, it was not very helpful	2	5.1
I&T	3	Yes, it was somewhat helpful	17	43.6
I&T	3	Yes, it was very helpful	14	35.9
I&T	4	I ignored the feedback	2	4.8
I&T	4	No, it was actually distracting	2	4.8
I&T	4	No, it was not very helpful	3	7.1
I&T	4	Yes, it was somewhat helpful	12	28.6
I&T	4	Yes, it was very helpful	23	54.8
Team	2	I ignored the feedback	4	10.3
Team	2	No, it was actually distracting	2	5.1
Team	2	No, it was not very helpful	7	17.9
Team	2	Yes, it was somewhat helpful	20	51.3
Team	2	Yes, it was very helpful	6	15.4
Team	3	I ignored the feedback	3	7.7
Team	3	No, it was actually distracting	2	5.1
Team	3	No, it was not very helpful	2	5.1
Team	3	Yes, it was somewhat helpful	18	46.2
Team	3	Yes, it was very helpful	14	35.9
Team	4	I ignored the feedback	4	10.3
Team	4	No, it was actually distracting	1	2.6
Team	4	No, it was not very helpful	6	15.4
Team	4	Yes, it was somewhat helpful	15	38.5
Team	4	Yes, it was very helpful	13	33.3

APPENDIX H SUPPLEMENTARY ANALYSIS FOR CHAPTER 5

Collection Time per Item (Team) by Strategy

The Collection Time per Item (Team), or CTI_{team}, was a dependent variable used to measure performance.

Distribution Overview (CTI_{team})

The overall distribution of CTI_{team} shows a relatively normal distribution for teams using Go Alone or Go Together (Figure 91). The distribution of the CTI_{team}, when grouped by Feedback, is similar to the overall distribution for using either Go Alone or Go Together (Figure 44). Over time, the researcher noticed no obvious pattern in the distribution for teams using Go Alone (Figure 93). Over time, the researcher noticed that the distribution for CTI_{team} skews to the left more noticeably in the Team condition than the other conditions (Figure 94).

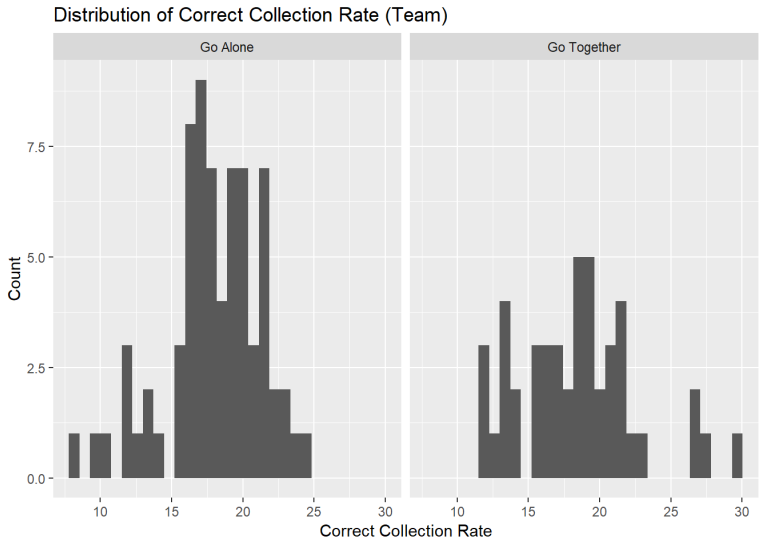


Figure 91 - Distribution overview of CTI_{team} divided by strategy

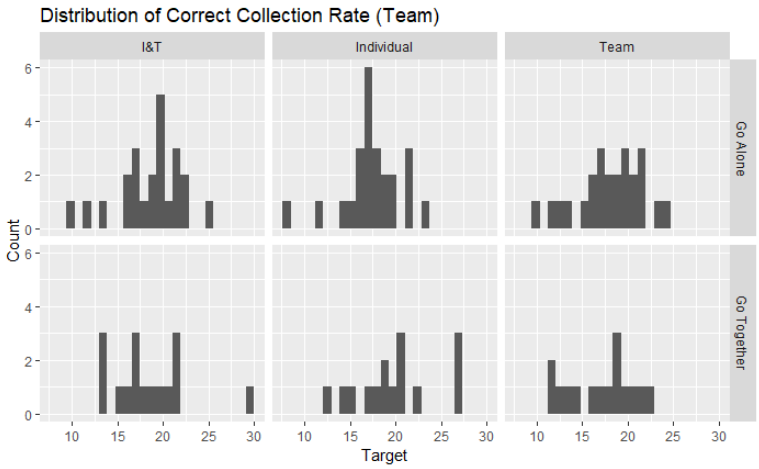


Figure 92 - Distribution of CTI_{team} grouped by Feedback and divided by strategy

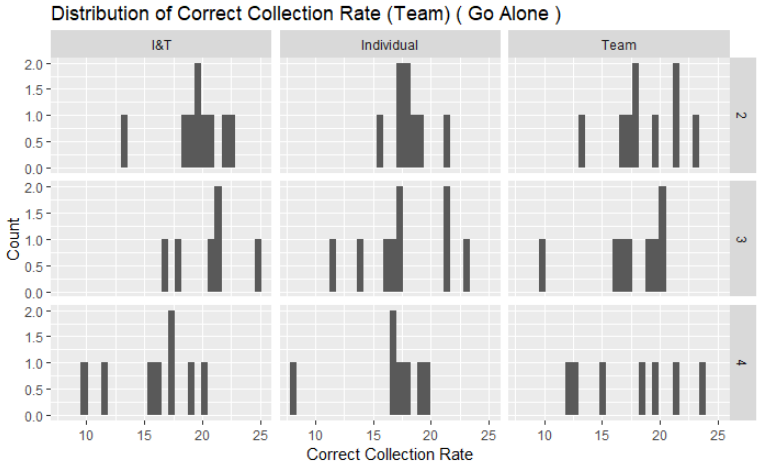


Figure 93 - Distribution of CTI_{team} grouped by Feedback and Session Order for teams using Go Alone

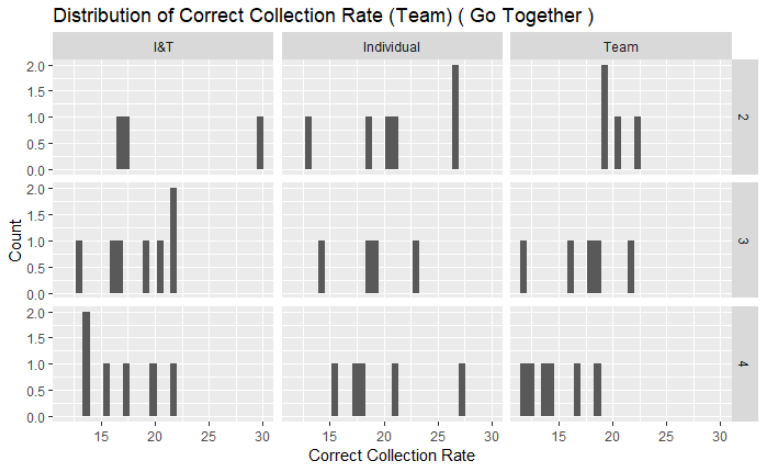


Figure 94 – Distribution of CTI_{team} grouped by Feedback and Session Order for teams using Go Together

Model Selection (CTI_{team})

Go Alone

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 134. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect was negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible, and the simplest model that described the data best was the No Interaction No Feedback model.

Table 134 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	339.36	364.25
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	337.78	353.61
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**338.20**	**349.52**

Go Together

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 135. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect was negligible. The No Interaction model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible, and the simplest model that described the data best was the No Interaction No Feedback model.

Table 135 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	248.27	268.39
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	240.61	253.41
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**242.23**	**251.38**

Testing Assumptions: Residuals (CTI_{team})

Go Alone

The histogram of the residuals (Figure 95) appeared to be approximately distributed. The residual fitted-value (Figure 96) showed a slight violation of constant variance. The points on the residual normal Q-Q plot (Figure 97) do not lay in a roughly straight line. The classic LMM and the robust LMM produced similar coefficient values (Table 136), so the researcher did not transform the data.

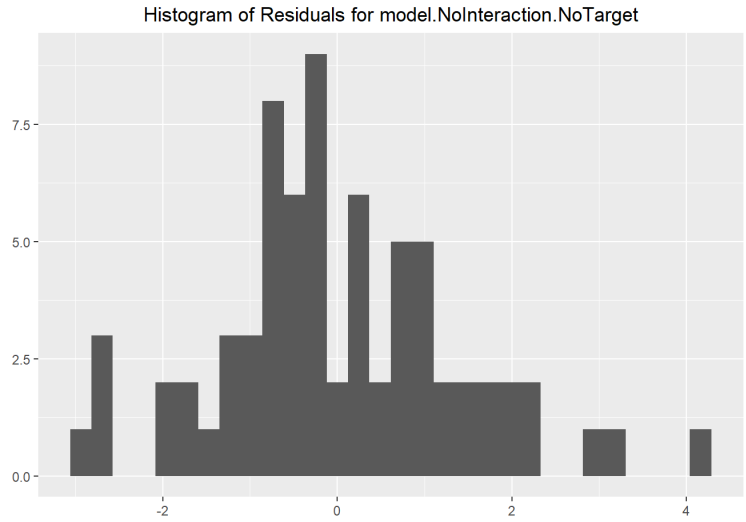


Figure 95 - Histogram of Residuals (CTI_{team})

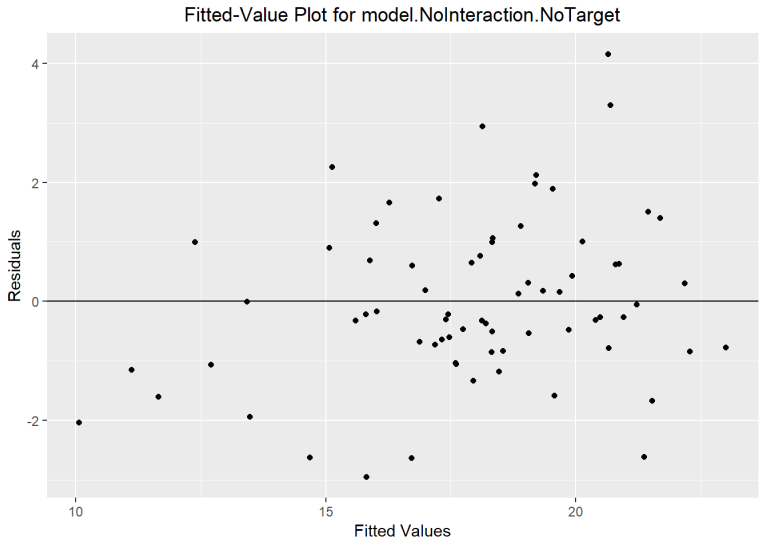


Figure 96 - Residual fitted plot (CTI_{team})

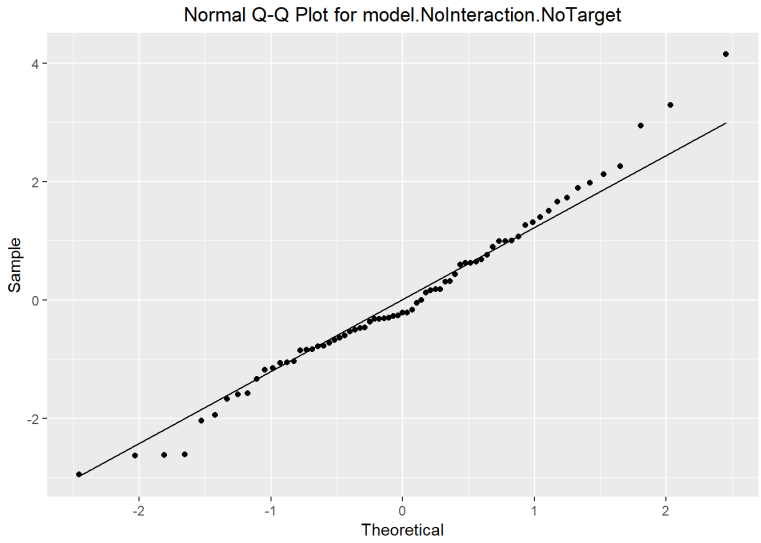


Figure 97 - Normal Q-Q plot for residuals (CTI_{team})

Table 136 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	18.9	19.1
SessionOrder: 3	-0.726	-0.780
SessionOrder: 4	-2.31	-2.25

Go Together

The histogram of the residuals (Figure 98) appeared to be approximately distributed. The residual fitted-value (Figure 99) showed a relatively constant variance. The points on the residual normal Q-Q plot (Figure 100) do not lay in a roughly straight line but seems to move away from the straight line on the tail ends. The classic LMM and the robust LMM produced similar coefficient values (Table 137), so the researcher did not transform the data.

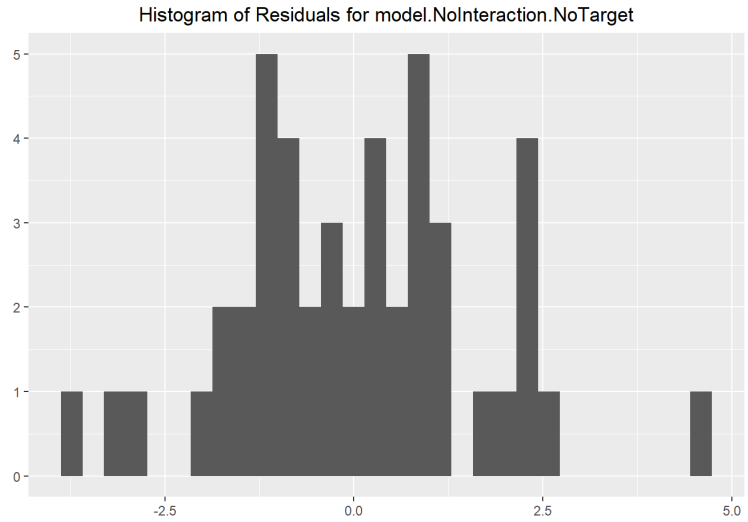


Figure 98 - Histogram of Residuals (CTI_{team})

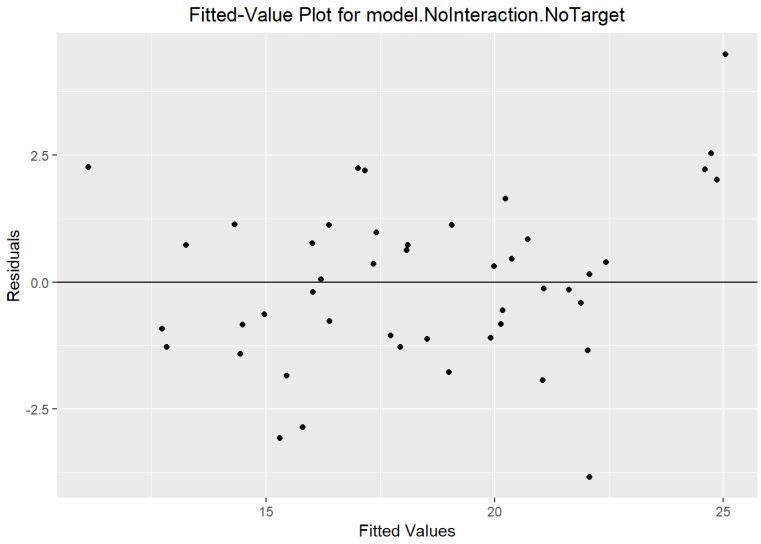


Figure 99 - Residual fitted plot (CTI_{team})

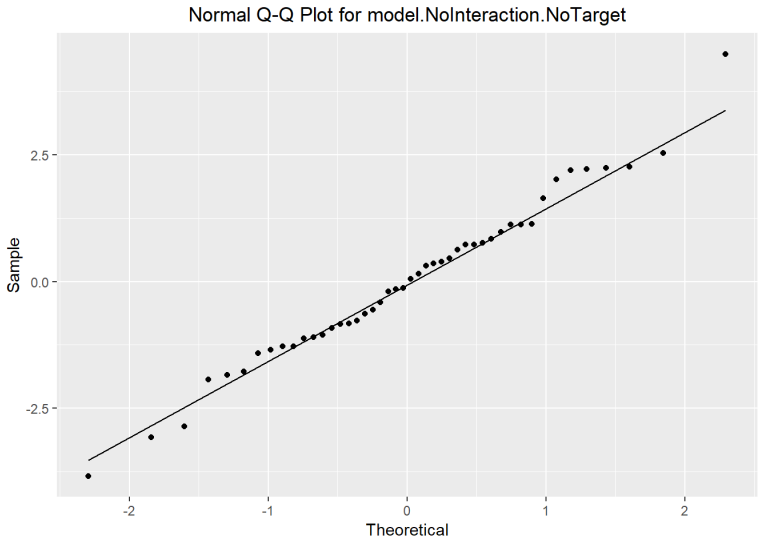


Figure 100 - Normal Q-Q plot for residuals (CTI_{team})

Table 137 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	21.6	21.3
SessionOrder: 3	-2.97	-2.72
SessionOrder: 4	-4.68	-4.50

Restricted Maximum Likelihood Estimation (CTI_{team})

Go Alone

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 138. The results showed that the effect of Session 4 is significant in reference to Session 2, the reference category.

Table 138 - Fixed effect result of REML Estimation (CTI_{team})

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	18.9	0.619	38	30.5	< .05*
Session: 3	-0.726	0.508	45	-1.43	.16
Session: 4	-2.31	0.516	45	-4.48	< .05*

Go Together

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 139. The results showed that the effects of Session 3 and 4 are significant in reference to Session 2, the reference category.

Table 139 - Fixed effect result of REML Estimation (CTI_{team})

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	21.6	1.04	26	20.7	< .05*
Session: 3	-2.97	0.803	26	-3.70	< .05*
Session: 4	-4.68	0.800	26	-5.85	< .05*

Evaluating Effect size (CTI_{team})

Go Alone

The values for R_m^2 and R_c^2 are presented in Table 140. The results suggest that the fixed variables have a small effect size (i.e., explain 8% of the variance) and the fixed and random variables have a large effect size (i.e., explain 74% of the variance).

Table 140 - Effect size for LMM (CTI_{team})

R^2 Type	Value
R_m^2	0.084
R_c^2	0.735

Go Together

The values for R_m^2 and R_c^2 are presented in Table 99. The results suggest that the fixed variables have a medium effect size (i.e., explain 17% of the variance) and the fixed and random variables have a large effect size (i.e., explain 80% of the variance).

Table 141 - Effect size for LMM (CTI_{team})

R^2 Type	Value
R_m^2	0.173
R_c^2	0.798

Time Remaining (Individual) by Strategy

The Time Remaining (Individual), or T_{ind}, was a dependent variable used to measure performance.

Distribution Overview (T_{ind})

The overall distribution of T_{ind} is skewed for teams using Go Alone or Go Together (Figure 101). The distribution of the T_{ind}, when grouped by Feedback, are similar to the overall distribution for teams using either Go Alone or Go Together (Figure 102). Over sessions, the researcher noticed no obvious pattern in the distribution for teams using Go Alone (Figure 103). In session 4 for teams using Go Together, the distribution in the Ind condition seems to skew to the left (Figure 104).

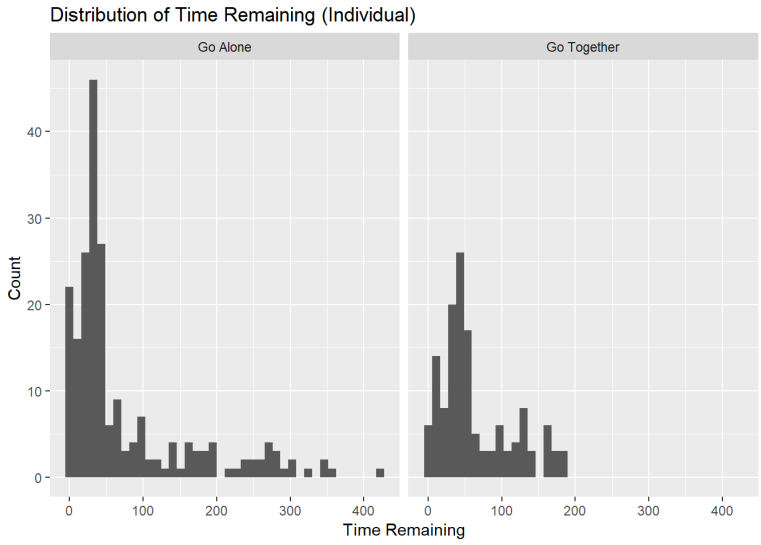


Figure 101 - Distribution overview of T_{ind} divided by strategy

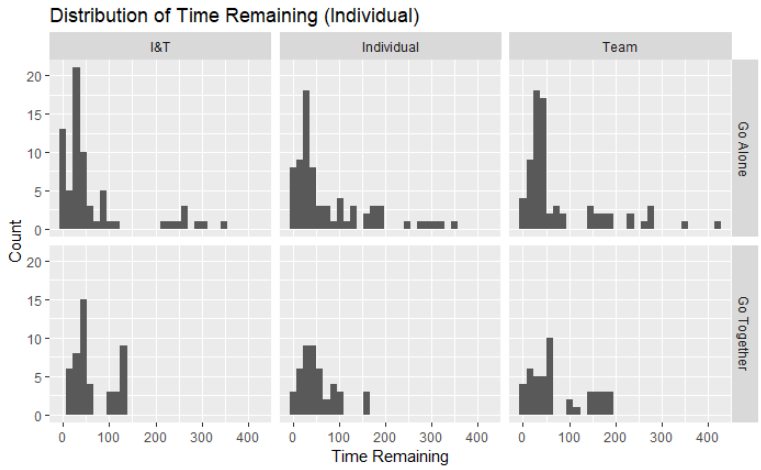


Figure 102 - Distribution of T_{ind} grouped by Feedback and divided by strategy

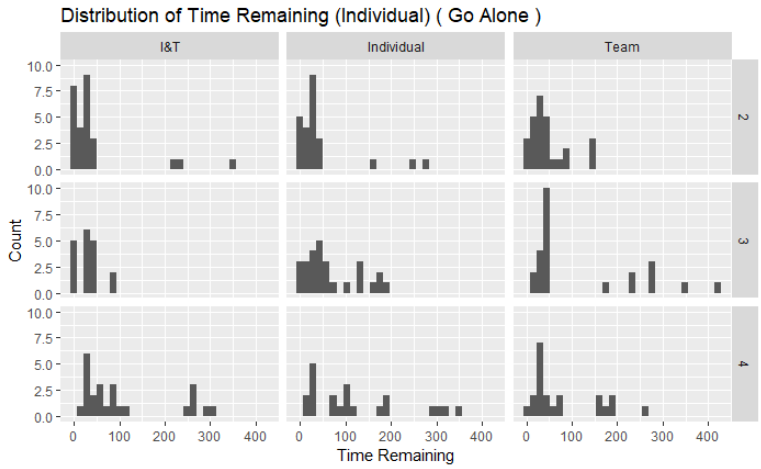


Figure 103 - Distribution of T_{ind} grouped by Feedback and Session Order for teams using Go Alone

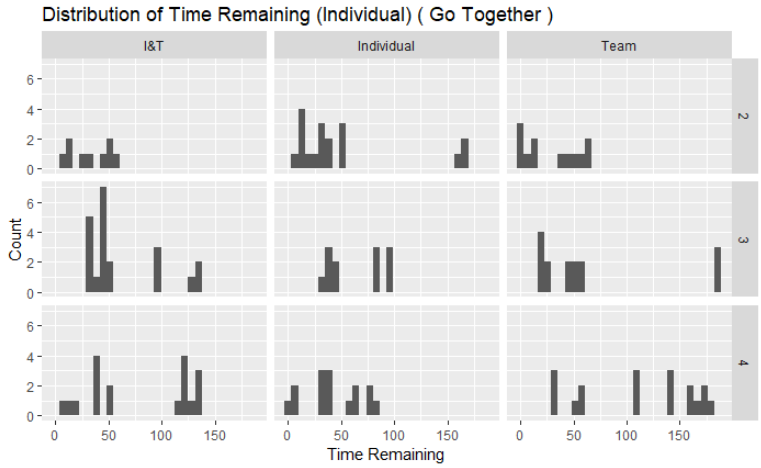


Figure 104 – Distribution of T_{ind} grouped by Feedback and Session Order for teams using Go Together

Model Selection (T_{ind})

Go Alone

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 142. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect was negligible. The No Interaction

model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible, and the simplest model that described the data best was the No Interaction No Feedback model.

Table 142 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	2285.73	2326.07
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	2279.08	2305.97
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**2280.51**	**2300.68**

Go Together

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 143. The All model was significantly different from the No Interaction model, No Interaction No Session, and the No Interaction No Feedback model, indicating that the interaction effect was not negligible.

Table 143 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$\boldsymbol{y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}}$	**1336.55**	**1371.68**
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	1349.93	1373.35
No Interaction No Feedback	$y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}$	1354.77	1372.34
No Interaction No Session Order	$y_{ijt} = \mu + \alpha_i + \gamma_t + \epsilon_{ijt}$	1395.05	1412.61

Testing Assumptions: Residuals (T_{ind})

Go Alone

The histogram of the residuals (Figure 105) appeared to be approximately distributed. The residual fitted-value (Figure 106) showed a violation of constant variance. The points on

the residual normal Q-Q plot (Figure 107) do not lay in a roughly straight line. The classic LMM and the robust LMM produced different coefficient values (Table 144), so the researcher used square root to transform the data.

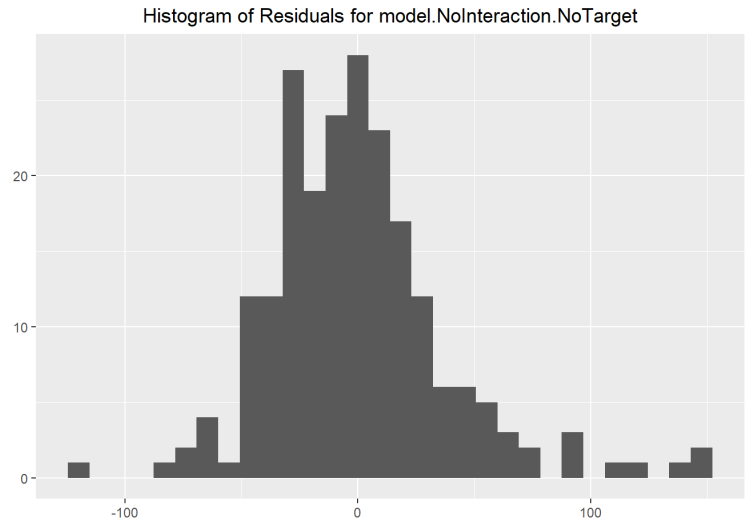


Figure 105 - Histogram of Residuals (T_{ind})

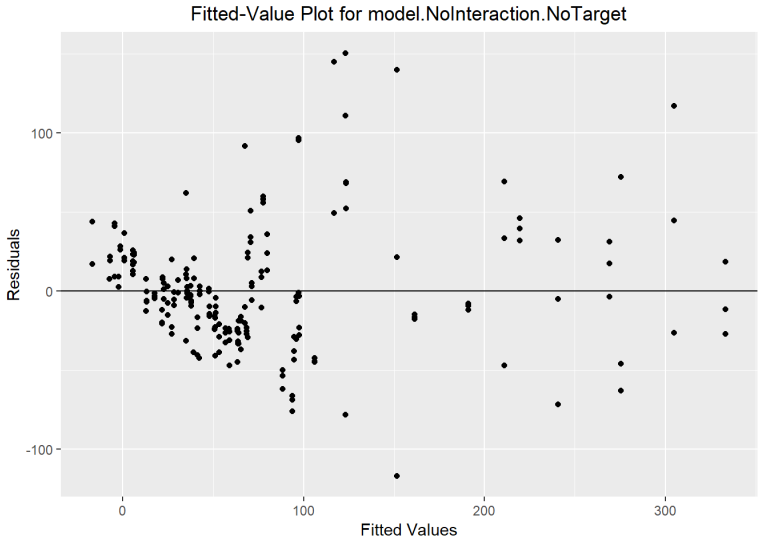


Figure 106 - Residual fitted plot (T_{ind})

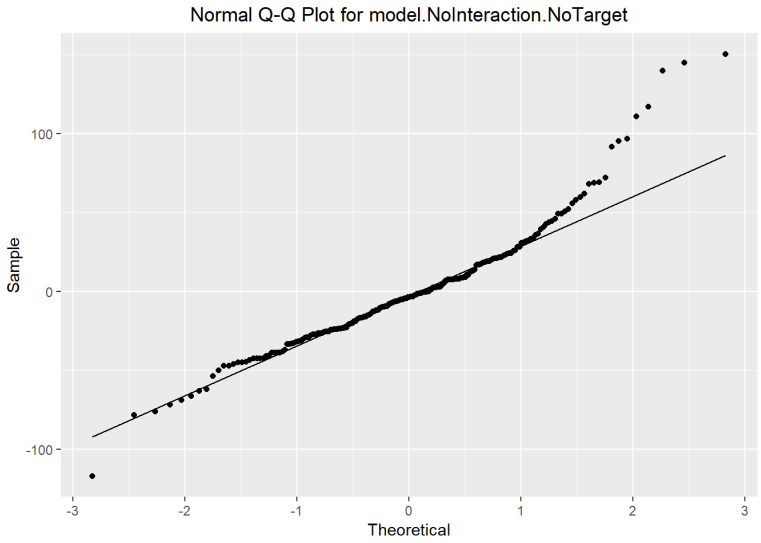


Figure 107 - Normal Q-Q plot for residuals (T_{ind})

Table 144 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	45.7	29.2
SessionOrder: 3	29.6	20.7
SessionOrder: 4	58.1	45.9

Go Together

The histogram of the residuals (Figure 108) appeared to be approximately distributed. The residual fitted-value Figure 109) showed a slight violation of constant variance. The points on the residual normal Q-Q plot (Figure 110) do not lay in a roughly straight line. The classic LMM and the robust LMM produced different coefficient values (Table 145), so the researcher used square root to transform the data.

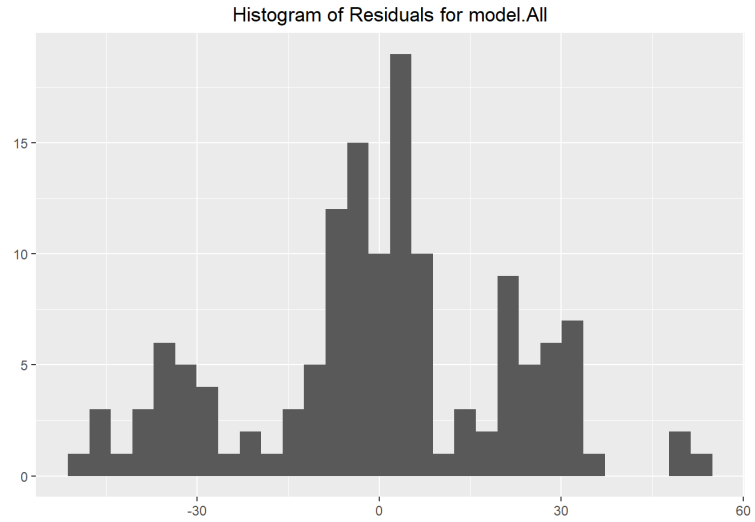


Figure 108 - Histogram of Residuals (T_{ind})

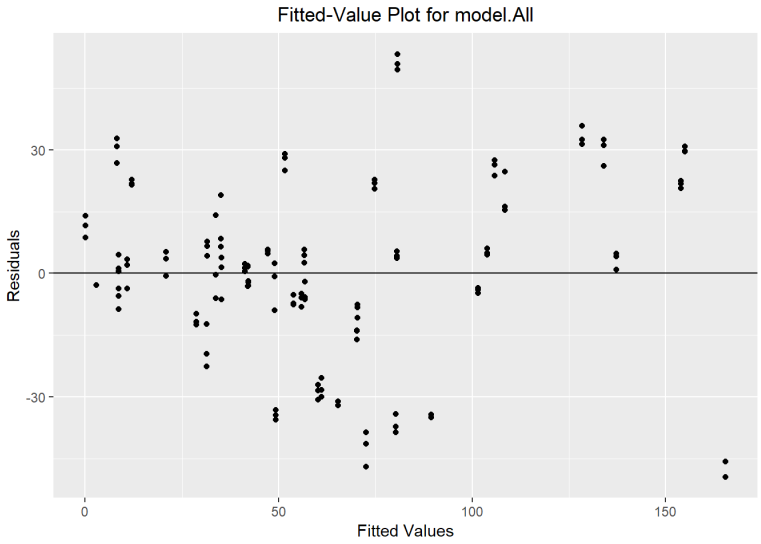


Figure 109 - Residual fitted plot (T_{ind})

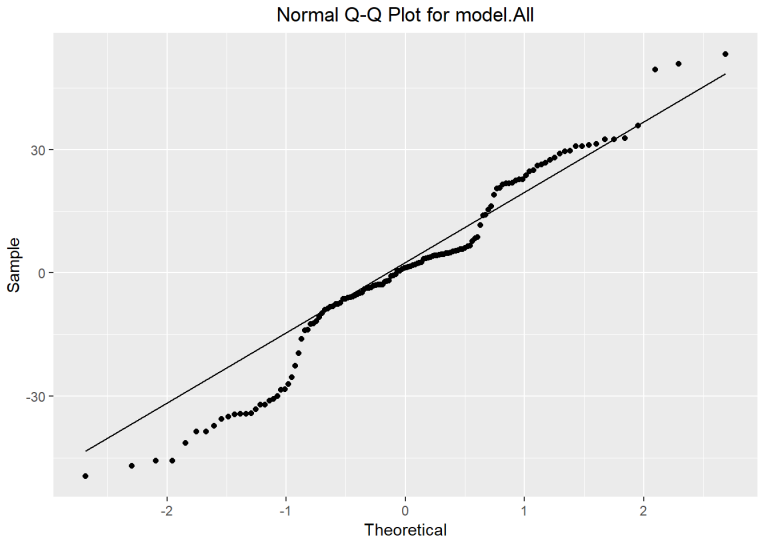


Figure 110 - Normal Q-Q plot for residuals (T_{ind})

Table 145 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	32.2	31.7
I&T	0.740	-1.39
Team	-6.64	-11.0
SessionOrder: 3	24.5	20.4
SessionOrder: 4	23.7	23.2
I&T : 3	7.91	10.3
Team : 3	3.03	9.85
I&T : 4	7.00	4.90
Team : 4	64.2	63.8

Restricted Maximum Likelihood Estimation (T_{ind})

Go Alone

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 146. The results showed that the effect of Session 3 and 4 are significant in reference to Session 2, the reference category.

Table 146 - Fixed effect result of REML Estimation (T_{ind})

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	5.47	0.747	31	7.34	< .05*
Session: 3	2.07	0.400	188	5.17	< .05*
Session: 4	3.75	0.406	188	9.24	< .05*

Go Together

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 147. The results showed that the effect of Session 3 and 4 are significant in reference to Session 2, the reference category. The results also show that the effect of Team in session 4 is significant in reference to Individual condition in session 2, the reference category.

Table 147 - Fixed effect result of REML Estimation (T_{ind})

Fixed effects					
	Estimate	**Standard Error**	**Degree of Freedom**	**T-value**	**P value**
Intercept	5.14	0.767	28	6.70	< .05*
I&T	0.028	0.729	117	0.038	.970
Team	-0.450	0.727	122	-0.618	.538
Session: 3	1.86	0.692	120	2.69	< .05*
Session: 4	1.85	0.671	121	2.76	< .05*
I&T : 3	0.187	1.09	120	0.172	.864
Team : 3	0.137	1.14	122	0.120	.905
I&T : 4	0.381	1.11	121	0.343	.732
Team : 4	3.12	1.08	123	2.90	< .05*

Evaluating Effect size (T_{ind})

Go Alone

The values for R_m^2 and R_c^2 are presented in Table 148. The results suggest that the fixed variables have a small effect size (i.e., explain 8% of the variance) and the fixed and random variables have a large effect size (i.e., explain 77% of the variance).

Table 148 - Effect size for LMM (T_{ind})

R^2 Type	Value
R_m^2	0.078
R_c^2	0.769

Go Together

The values for R_m^2 and R_c^2 are presented in Table 149. The results suggest that the fixed variables have a medium effect size (i.e., explain 24% of the variance) and the fixed and random variables have a large effect size (i.e., explain 79% of the variance).

Table 149 - Effect size for LMM (T_{ind})

R^2 Type	Value
R_m^2	0.236
R_c^2	0.789

Collection Time per Item (Individual) by Strategy

The Collection Time per Item (Individual), or CTI_{ind}, was a dependent variable used to measure performance.

Distribution Overview (CTI_{ind})

The overall distribution of CTI_{ind} shows a relatively normal distribution for teams using Go Alone or Go Together (Figure 111). The distribution of the CTI_{ind}, when grouped by Feedback, is similar to the overall distribution for using either Go Alone or Go Together

(Figure 112). Over time, the researcher noticed no obvious pattern in the distribution for teams using Go Alone (Figure 113). Over time, the researcher noticed that the distribution for CTI_{ind} skews to the left more noticeably in the Team and I&T condition than compared to the Individual condition (Figure 114).

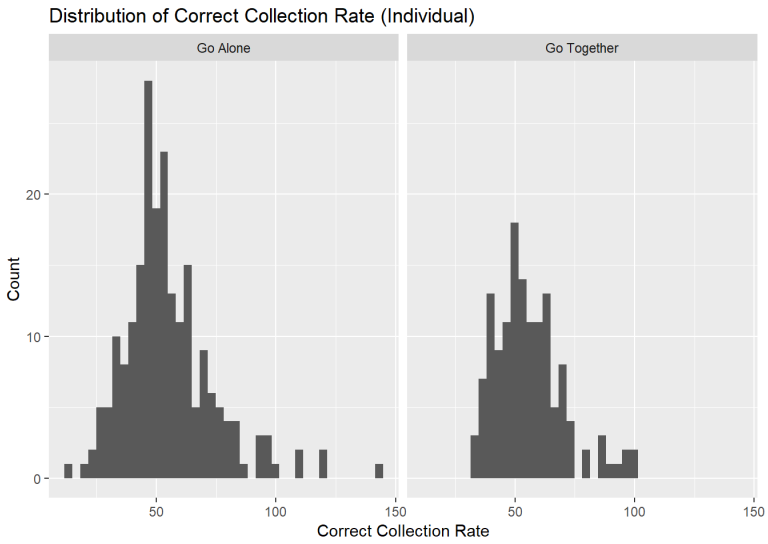


Figure 111 - Distribution overview of CTI_{ind} divided by strategy

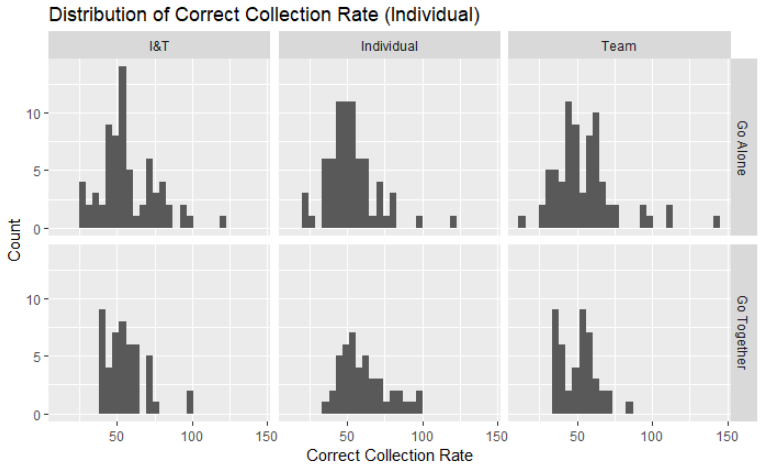


Figure 112 - Distribution of CTI_{ind} grouped by Feedback and divided by strategy

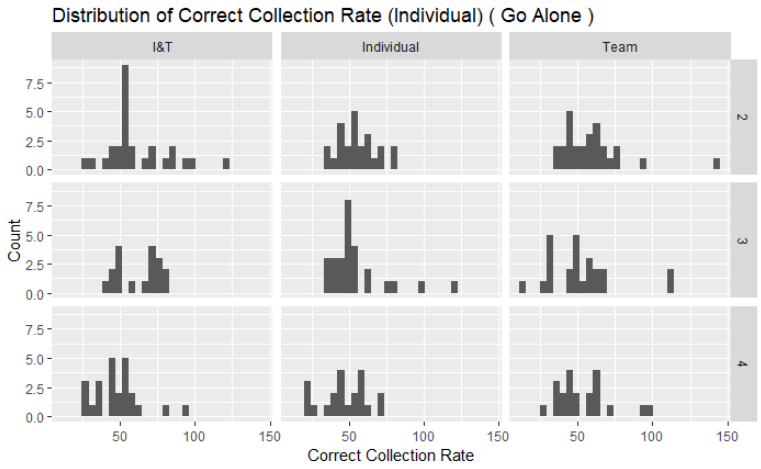


Figure 113 - Distribution of CTI_{ind} grouped by Feedback and Session Order for teams using Go Alone

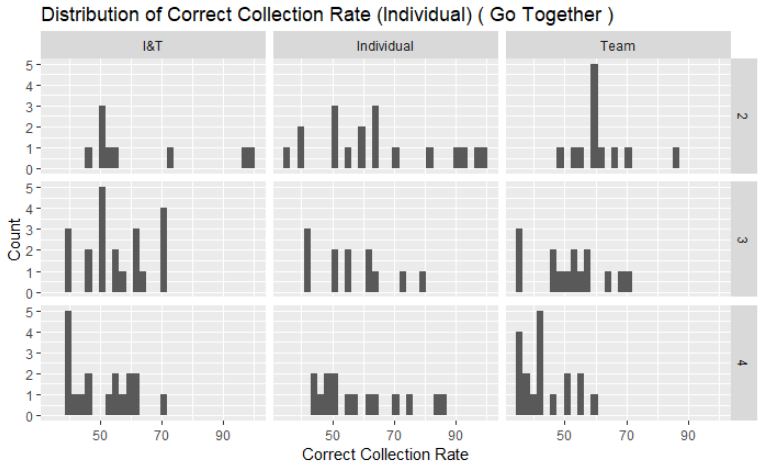


Figure 114 – Distribution of CTI_{ind} grouped by Feedback and Session Order for teams using Go Together

Model Selection (CTI_{ind})

Go Alone

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 150. The All model was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the interaction effect was negligible. The No Interaction

model was not significantly different from the No Interaction No Feedback model, indicating that the Feedback effect is negligible, and the simplest model that described the data best was the No Interaction No Feedback model.

Table 150 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	1792.432	1832.768
No Interaction	$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}$	1787.267	1814.157
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**1787.813**	**1807.981**

Go Together

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 151. The All model was significantly different from the No Interaction No Session model, but it was not significantly different from the No Interaction model or the No Interaction No Feedback model, indicating that the model should not only include the effect of the feedback conditions. The No Interaction model was significantly different from the No Interaction No Feedback model and the No Interaction No Session, indicated that the interaction effect was negligible, and the simplest model that described the data best was the No Interaction model.

Table 151 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	1045.98	1081.11
No Interaction	$\boldsymbol{y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}}$	**1038.49**	**1061.91**
No Interaction No Feedback	$y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}$	1043.60	1061.17
No Interaction No Session Order	$y_{ijt} = \mu + \alpha_i + \gamma_t + \epsilon_{ijt}$	1079.89	1097.45

Testing Assumptions: Residuals (CTI_{ind})

Go Alone

The histogram of the residuals (Figure 115) appeared to be approximately distributed. The residual fitted-value (Figure 116) showed a violation of constant variance. The points on the residual normal Q-Q plot (Figure 117) do not lay in a roughly straight line. The classic LMM and the robust LMM produced similar coefficient values (Table 152), so the researcher did not transform the data.

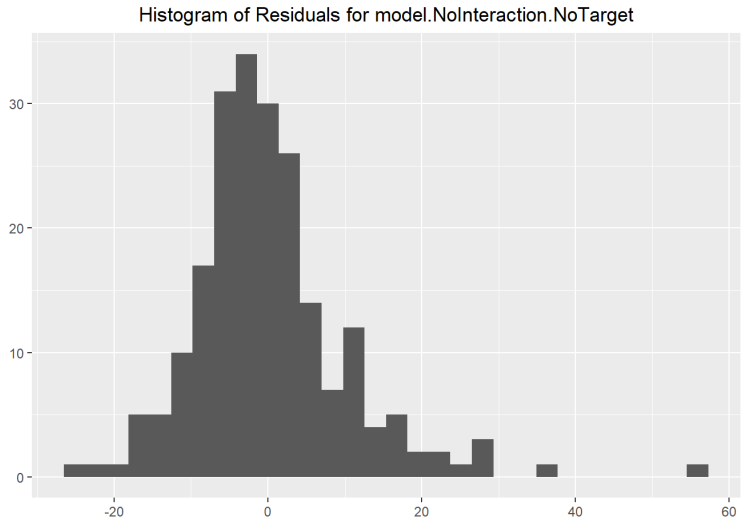


Figure 115 - Histogram of Residuals (CTI_{ind})

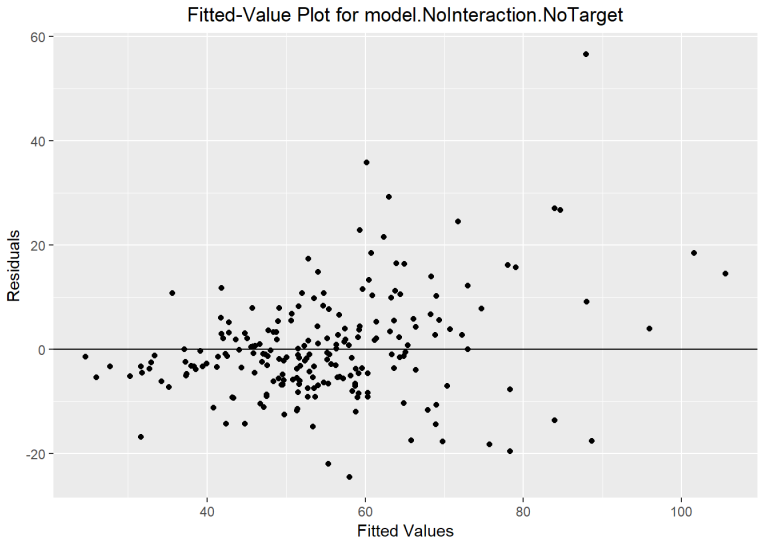


Figure 116 - Residual fitted plot (CTI_{ind})

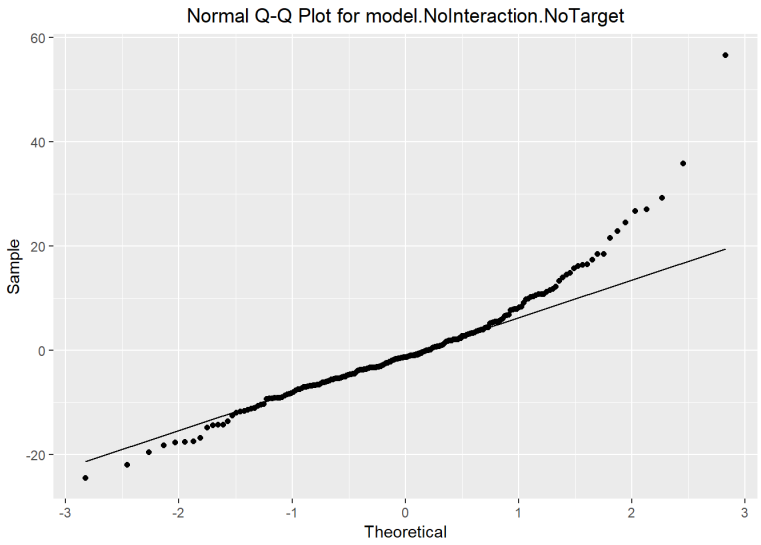


Figure 117 - Normal Q-Q plot for residuals (CTI_{ind})

Table 152 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	59.1	57.3
SessionOrder: 3	-3.98	-3.34
SessionOrder: 4	-9.62	-8.84

Go Together

The histogram of the residuals (Figure 118) appeared to be approximately distributed. The residual fitted-value (Figure 119) showed a constant variance. The points on the residual normal Q-Q plot (Figure 120) do not lay in a roughly straight line. The classic LMM and the robust LMM produced similar coefficient values (Table 153), so the researcher did not transform the data.

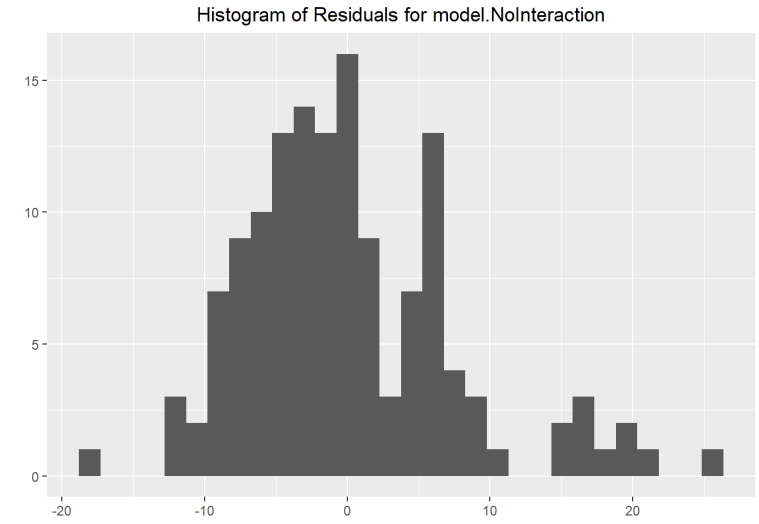


Figure 118 - Histogram of Residuals (CTI_{ind})

Figure 119 - Residual fitted plot (CTI_{ind})

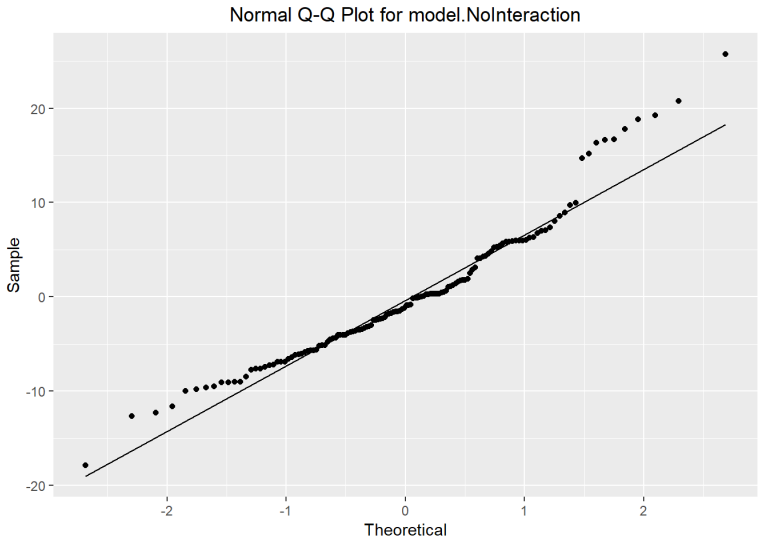


Figure 120 - Normal Q-Q plot for residuals (CTI_{ind})

Table 153 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	67.8	64.5
I&T	-1.49	-1.07
Team	-5.00	-4.10
SessionOrder: 3	-9.04	-7.51
SessionOrder: 4	-14.0	-12.3

Restricted Maximum Likelihood Estimation (CTI_{ind})

Go Alone

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 154. The results showed that the effect of Session 4 is significant in reference to Session 2, the reference category.

Table 154 - Fixed effect result of REML Estimation (CTI_{ind})

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	59.1	2.35	42	25.1	< .05*
Session: 3	-3.98	2.04	138	-1.95	.054
Session: 4	-9.62	2.08	139	-4.63	< .05*

Go Together

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 155. The results showed that the effect of the session order is significant in reference to Session 2, the reference category, and the effect of the Team condition is significant in reference to Individual condition, the reference category.

Table 155 - Fixed effect result of REML Estimation (CTI_{ind})

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	67.8	3.07	26	22.1	< .05*
I&T	-1.49	1.79	82	-0.829	.409
Team	-5.41	1.88	84	-2.88	< .05*
Session: 3	-9.04	1.90	83	-4.76	< .05*
Session: 4	-14.3	1.89	84	-7.56	< .05*

Evaluating Effect size (CTI_{ind})

Go Alone

The values for R_m^2 and R_c^2 are presented in Table 156. The results suggest that the fixed variables have a small effect size (i.e., explain 4% of the variance) and the fixed and random variables have a large effect size (i.e., explain 59% of the variance).

Table 156 - Effect size for LMM (CTI_{ind})

R^2 Type	Value
R_m^2	0.044
R_c^2	0.594

Go Together

The values for R_m^2 and R_c^2 are presented in Table 157. The results suggest that the fixed variables have a medium effect size (i.e., explain 17% of the variance) and the fixed and random variables have a large effect size (i.e., explain 71% of the variance).

Table 157 - Effect size for LMM (CTI_{ind})

R^2 Type	Value
R_m^2	0.173
R_c^2	0.712

Frustration by Strategy

The Frustration was a dependent variable used to report the level of frustration experienced by each participant.

Distribution Overview (Frustration)

The overall distribution of Frustration for Go Alone was difficult to visualize because of the high variance and the distribution for Go Together appeared to have a normal distribution (Figure 121). The distribution of the Frustration, when grouped by Feedback, are similar to the overall distribution for using either Go Alone or Go Together (Figure 122). Over time, the researcher noticed no obvious pattern in the distribution for teams using Go Alone (Figure 123). Over time, the researcher noticed that the distribution for Frustration skews to the left more noticeably in the Team and I&T condition than the Individual condition (Figure 124).

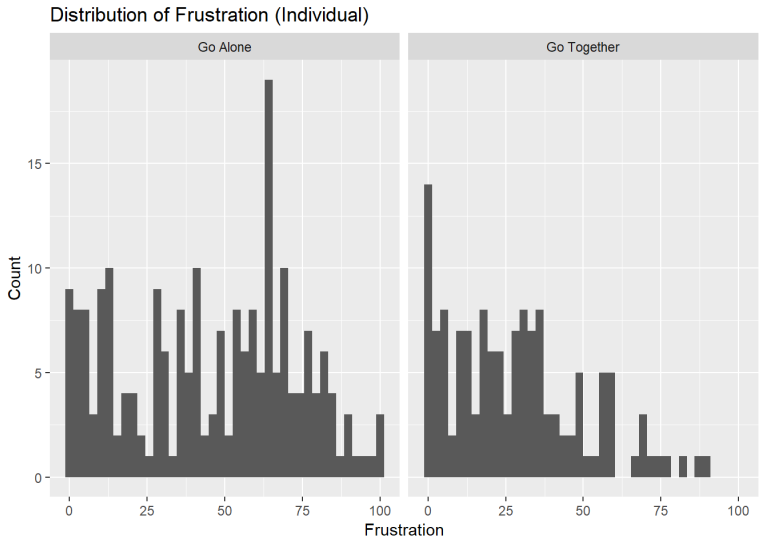


Figure 121 - Distribution overview of Frustration divided by strategy

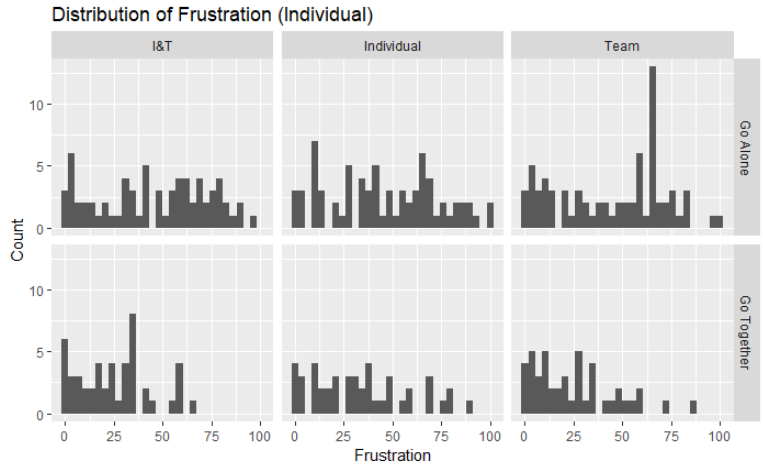


Figure 122 - Distribution of Frustration grouped by Feedback and divided by strategy

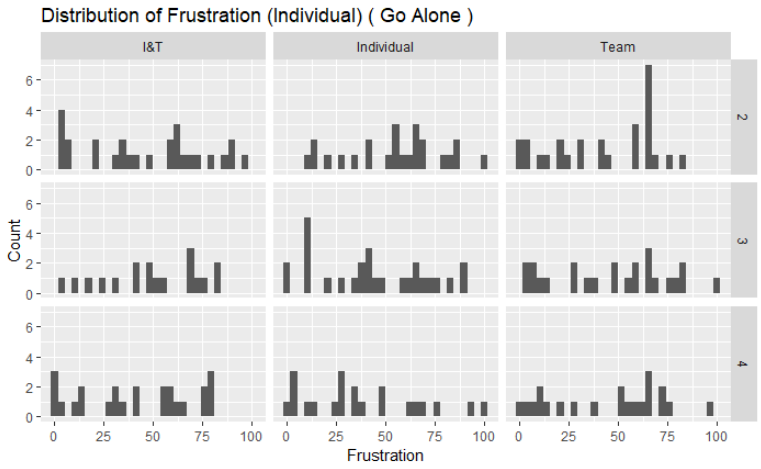


Figure 123 - Distribution of Frustration grouped by Feedback and Session Order for teams using Go Alone

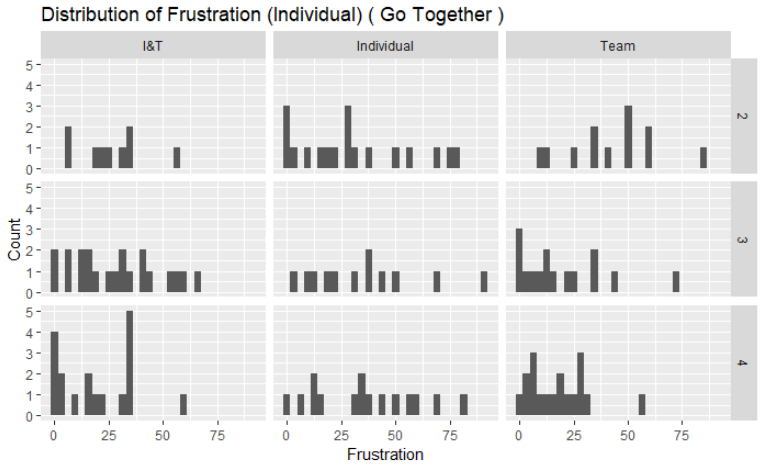


Figure 124 – Distribution of Frustration grouped by Feedback and Session Order for teams using Go Together

Model Selection (Frustration)

Go Alone

The author used the model selection process outlined in Table 16. No model significantly different from the null model but the models that were marginally significant are displayed in Table 158. The simplest model that described the data best was the No Interaction No Feedback model.

Table 158 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	1974.756	2015.092
No Interaction No Feedback	$\boldsymbol{y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}}$	**1971.72**	**1991.888**

Go Together

The author used the model selection process outlined in Table 16. The models that were significantly different from the null model are displayed in Table 159. The All model was significantly different from the No Interaction No Feedback model, but it was not significantly different from No Interaction model or the No Interaction No Session model, indicating that the session order effect alone is negligible. The No Interaction model was significantly different from the No Interaction No Feedback model and the No Interaction No Session model, indicating that the interaction effect is negligible, and the simplest model that described the data best was the No Interaction.

Table 159 – Models that are significantly different from the null model. The best fit model is bolded.

Model Name	Equation	AIC	BIC
All	$y_{ijt} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \gamma_t + \epsilon_{ijt}$	1216.254	1251.381
No Interaction	$\boldsymbol{y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \epsilon_{ijt}}$	**1214.304**	**1237.722**
No Interaction No Feedback	$y_{ijt} = \mu + \beta_j + \gamma_t + \epsilon_{ijt}$	1222.749	1240.313
No Interaction No Session Order	$y_{ijt} = \mu + \alpha_i + \gamma_t + \epsilon_{ijt}$	1216.821	1234.384

Testing Assumptions: Residuals (Frustration)

Go Alone

The histogram of the residuals (Figure 125) appeared to be approximately distributed. The residual fitted-value (Figure 126) showed a violation of constant variance. The points on the residual normal Q-Q plot (Figure 127) lay in a roughly straight line. The classic LMM and the robust LMM produced similar coefficient values (Table 160), so the researcher did not transform the data.

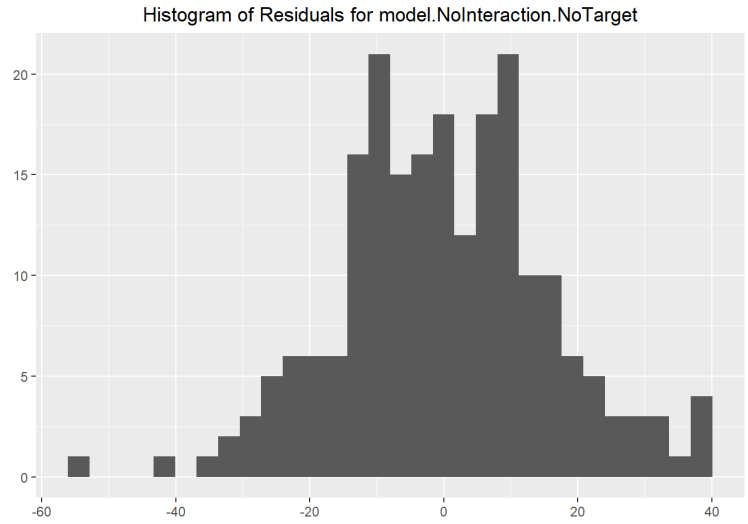


Figure 125 - Histogram of Residuals (Frustration)

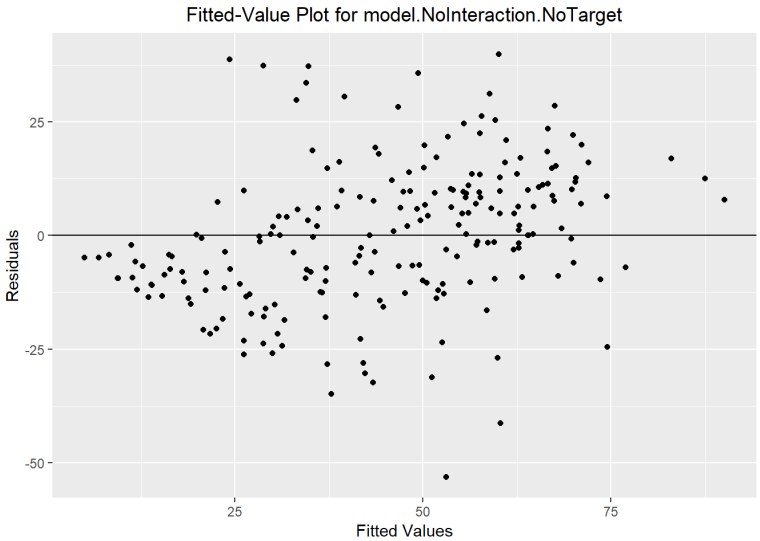


Figure 126 - Residual fitted plot (Frustration)

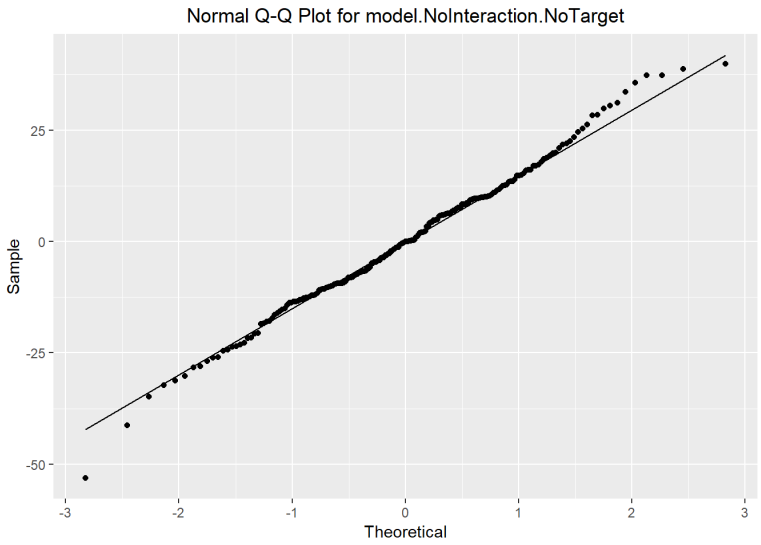


Figure 127 - Normal Q-Q plot for residuals (Frustration)

Table 160 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	47.42	47.02
SessionOrder: 3	-2.57	-2.72
SessionOrder: 4	-7.05	-6.49

Go Together

The histogram of the residuals (Figure 128) appeared to be approximately distributed. The residual fitted-value (Figure 129) showed a violation of constant variance. The points on the residual normal Q-Q plot (Figure 130) do not lay in a roughly straight line. The classic LMM and the robust LMM produced similar coefficient values (Table 161), so the researcher did not transform the data.

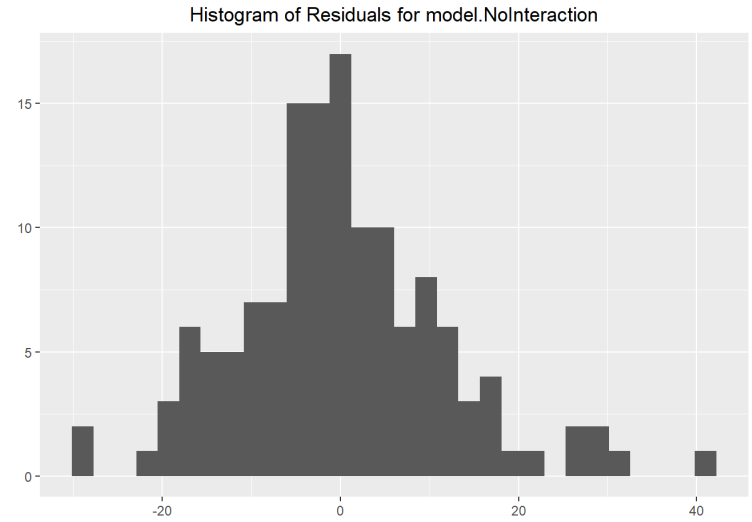


Figure 128 - Histogram of Residuals (Frustration)

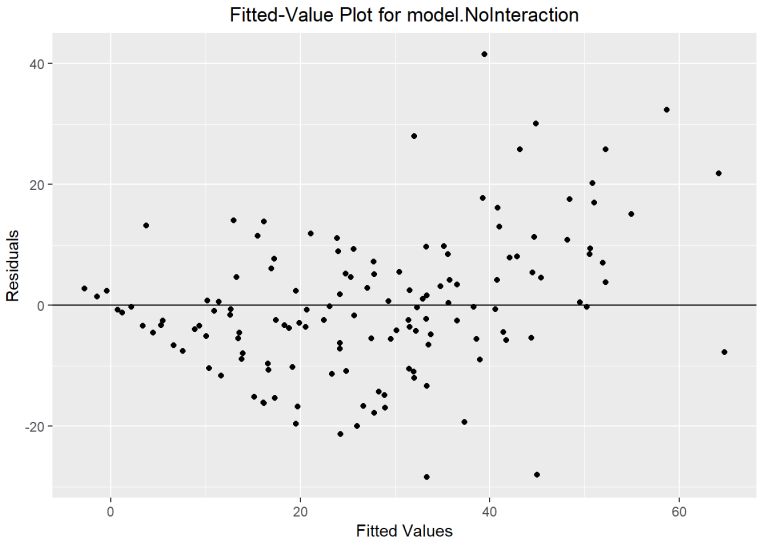


Figure 129 - Residual fitted plot (Frustration)

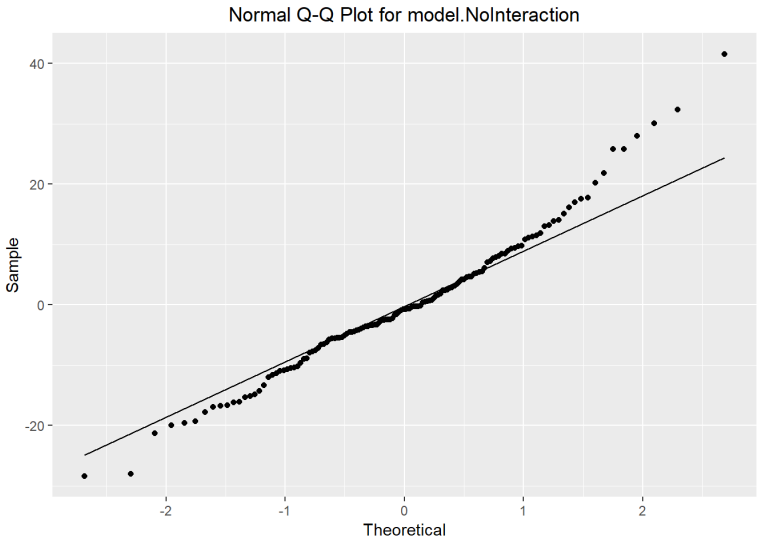


Figure 130 - Normal Q-Q plot for residuals (Frustration)

Table 161 - Comparing classic model to a robust model

Coefficient	No Interaction No Feedback	Robust No Interaction No Feedback
Intercept	39.9	36.3
I&T	-8.72	-5.17
Team	-11.0	-7.34
SessionOrder: 3	-2.91	-2.42
SessionOrder: 4	-7.86	-7.87

Restricted Maximum Likelihood Estimation (Frustration)

Go Alone

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in

Table 162. The results showed that the effect of Session 4 is significant in reference to Session 2, the reference category.

Table 162 - Fixed effect result of REML Estimation (Frustration)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	47.4	3.33	45	14.3	< .05*
Session: 3	-2.57	3.16	138	-0.812	.418
Session: 4	-7.05	3.21	138	-2.19	< .05*

Go Together

The criterion used to fit a model (i.e., generate accurate estimates of the predictors) and evaluate significance in an LMM is the Restricted Maximum Likelihood (REML). The REML method is used instead of the maximum likelihood ratio estimation (MLE) because it is less biased (Gałecki & Burzykowski, 2013). The results for the fixed effect variables are displayed in Table 163. The results showed that the effect of Session 4 is significant in reference to Session 2, the reference category. The results showed that the effect of Team and I&T feedback condition in reference to the Individual condition, the reference category.

Table 163 - Fixed effect result of REML Estimation (Frustration)

Fixed effects					
	Estimate	Standard Error	Degree of Freedom	T-value	P value
Intercept	39.9	3.93	37	10.1	< .05*
I&T	-8.72	3.09	80	-2.83	< .05*
Team	-11.0	3.21	84	-3.44	< .05*
Session: 3	-2.91	3.26	82	-0.89	.375
Session: 4	-7.86	3.24	84	-2.43	< .05*

Evaluating Effect size (Frustration)

Go Alone

The values for R_m^2 and R_c^2 are presented in Table 164. The results suggest that the fixed variables have a small effect size (i.e., explain 1% of the variance) and the fixed and random variables have a large effect size (i.e., explain 56% of the variance).

Table 164 - Effect size for LMM (CTI_{ind})

R^2 Type	Value
R_m^2	0.011
R_c^2	0.555

Go Together

The values for R_m^2 and R_c^2 are presented in Table 165. The results suggest that the fixed variables have a small effect size (i.e., explain 7% of the variance) and the fixed and random variables have a large effect size (i.e., explain 61% of the variance).

Table 165 - Effect size for LMM (CTI_{ind})

R^2 Type	Value
R_m^2	0.069
R_c^2	0.614

APPENDIX I INSTITUTIONAL REVIEW BOARD APPROVAL MEMO

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
2420 Lincoln Way, Suite 202
Ames, Iowa 50014
515 294-4566

Date: 3/8/2018

To: Dr. Stephen B Gilbert
1620 Howe Hall

CC: Jamiahus Walton
407 S. 5th St Apt 118, Ames IA 50010

From: Office for Responsible Research

Title: Teamwork Task in a Virtual Environment

IRB ID: 18-109

Approval Date:	3/8/2018	**Date for Continuing Review:**	3/5/2019
Submission Type:	New	**Review Type:**	Full Committee

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**
- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.
- **Obtain IRB approval prior to implementing** any **changes** to the study by submitting a Modification Form for Non-Exempt Research or Amendment for Personnel Changes form, as necessary.
- **Immediately inform the IRB of (1) all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. Approval from other entities may also be needed. For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 202 Kingland, to officially close the project.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.