

Evaluating Distributed Teams with the Team Multiple Errands Test

**Jamiahus Walton, Stephen Gilbert, Ph.D., Eliot Winer, Ph.D.,
Michael Dorneich, Ph.D., Desmond Bonner**

Iowa State University

Ames, IA

{jwalton, gilbert, ewiner, dorneich, dbonner}@iastate.edu

ABSTRACT

Modern day teams, whether in the military or civilian workplace, have the ability to achieve goals that are otherwise unobtainable by individuals. The timing and characteristics of feedback that teams receive during training are critical. Though there is a solid foundation of research on optimal feedback, there is limited exploration of what constitutes ideal team feedback including addressing the individual team member versus the whole team and whether that feedback is public (visible to the entire team) or private (visible only to one member of the team).

Previous research that studied the effect of feedback on team performance has yielded slightly different conclusions. For example, research focused on the privacy of feedback suggests that public feedback can have a motivational effect that improves performance. The aim of this work is to discover the most effective combination of the target and privacy of feedback.

To accomplish this goal a modified version of the Multiple Errands Test (MET) was developed to evaluate the performance of three-member teams, the Team MET (TMET). The MET, normally used for evaluating cognitive processing, requires that specific rules be followed while completing multiple tasks within a time constraint. Participants performed the TMET while coordinating purchases in a virtual mall. In each of four timed shopping sessions, participants received feedback on their performance as an individual and team. Feedback was given in one of four conditions: individual private, team private, individual public, and team public. Task performance and rule errors were measured as dependent variables. Results did not yield a broadly significant effect of feedback condition on team or individual performance. However, the study did demonstrate the validity of the TMET as a platform for assessing a team's ability to perform under heavy cognitive load.

ABOUT THE AUTHORS

Jamiahus Walton is an Industrial Engineering M.S. student candidate at Iowa State University. He received his physics B.S. degree from Millikin University. His thesis work focuses on developing a team Multiple Errands Test (MET).

Stephen Gilbert, Ph.D., is an Associate Director of the Virtual Reality Applications Center (VRAC) and Assistant Professor of Industrial and Manufacturing Systems Engineering at Iowa State University. His background includes cognitive science and engineering. He leads research on Team Training for the U.S. Army Research Laboratory.

Eliot Winer, Ph.D., is an Associate Director of the Virtual Reality Applications Center (VRAC), Associate Professor of Mechanical Engineering, and a faculty affiliate of the Human Computer Interaction (HCI) graduate program at Iowa State University. He recently led research to develop a next-generation mixed-reality virtual and constructive training environment for the US Army. Dr. Winer has over 15 years of experience working, virtual reality, computer graphics, and simulation technologies.

Michael Dorneich, Ph.D., is Associate Professor of Industrial and Manufacturing Systems Engineering and a faculty affiliate of the human computer interaction (HCI) graduate program at Iowa State University. Dr. Dorneich's research interests focus on creating joint human-machine systems that enable people to be effective in the complex and often stressful environments found in aviation, robotic, learning, and space applications. Dr. Dorneich has over

19 years' experience developing adaptive systems that can provide assistance tailored to the user's current cognitive state, situation, and environment.

Desmond Bonner is an Industrial Engineering MS student candidate at Iowa State University. He received his graphics design B.F.A. degree from Auburn University. His thesis work focuses on increasing female participation in STEM fields through video games.

Evaluating Distributed Teams with the Team Multiple Errands Test

**Jamiahus Walton, Stephen Gilbert, Ph.D., Eliot Winer, Ph.D.,
Michael Dorneich, Ph.D., Desmond Bonner**

Iowa State University

Ames, IA

{jwalton, gilbert, ewiner, dorneich, dbonner}@iastate.edu

INTRODUCTION

Modern teams have the capacity to achieve goals that are difficult to obtain by an individual alone. Providing training to teams will help them maintain their effectiveness. During training, members of a team receive instruction on how to properly complete an action. This instruction can come from peers, teachers, or intelligent tutoring systems (ITSs). ITSs are successful at providing instruction to students individually (Aleven, McLaren, Roll, & Koedinger, 2006; Hategekimana, Gilbert, & Blessing, 2008; Koedinger, Aleven, Hockenberry, McLaren, & Heffernan, 2004). Unfortunately, there are several challenges that arise when designing an ITS to train a team (Sottilare, Holden, Brawner, & Goldberg, 2011).

Distribution of feedback is one of several issues that arise when designing an ITS for teams. If an individual makes a mistake while working alone then it is clear that that individual receives individual feedback. If that individual makes a mistake within a team setting, then it is unclear how feedback should be given. Feedback can be given directly to the individual or be given to the whole group. An important question that can be derived from this issue is which level of feedback, individual or group, will be more effective, and under what circumstances.

Feedback privacy is another issue that arises when considering feedback within a team setting. There are two distinct privacy conditions: Public and Private. For example, a squad leader can give feedback in such a way that the entire team can hear it (publicly). On the other hand a squad leader can offer feedback in such a way that only one member of the team can hear it (private). If an individual makes a mistake while in the team setting, should the feedback be given publicly or should the feedback be given privately? In other words, does private or public feedback produce the highest performance overall?

To pursue research questions like these, a robust task domain and research platform is needed that allows a researcher or trainer to observe team performance while varying the difficulty of the task and changing feedback parameters. Part of the current research proposes a new platform for such team assessment called the Team Multiple Errands Task. There are many different methods that can be used to assess teams' performance under heavy cognitive load. This current work attempts to modify a well-established psychological assessment of individuals' cognitive load called the Multiple Errands Test (MET) to create the Team MET, or TMET.

RESEARCH QUESTIONS

In the pursuit of answering these questions for military training, a study was conducted in a simpler, non-military setting to focus on the questions of how to give feedback. It is hoped that these results will generalize to military training contexts. The following work attempts to answer the following research questions:

1. How does the **privacy** (Public vs. Private) component of feedback influence team and individual performance?
2. How does the **audience** (Direct vs. Group) of feedback content influence the team and individual performance?
3. Can the Multiple Errands Task (MET) be adapted to a Team MET as a test for teams' ability to perform under heavy cognitive load?

BACKGROUND

Feedback is content that is given to a receiver (i.e., a person) and given by an agent (Hattie & Timperley, 2007). There are many different agents that can be the sources of feedback. Agents can be teachers, peers, environment, and Intelligent Tutoring Systems (ITSs). There are many different aspects of feedback for teams that need to be considered. There are a number of existing studies that focused on feedback within a team setting (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Dominick, Reilly, & McGourty, 1997; Peñarroja, Orengo, Zornoza, Sánchez, & Ripoll, 2015).

Team feedback is different from individual feedback because of a team's ability to respond to the given information (Nadler, 1979). Many studies that focus on feedback within a team setting discuss two types of feedback: Individual and Team. DeShon et al. (2004) conducted a study that attempted to develop a good conceptual model that describes how goal and performance feedback benefit learning and performance. In this model there are two loops that represent the behavior of individuals: the individual loop and the team loop. According to this model, the behavior choices made by individuals within a team setting are driven by obvious individual and team discrepancies. This work concluded that 1) individual feedback will increase focus on individual performance and 2) team-level feedback will increase focus on team performance. This model dictates that the content of feedback can have a positive influence on performance. The more focus there is on performance, individual or team, the higher the performance.

There are many different studies that focus on feedback in an individual setting and feedback within a team setting. Researchers generally agree that the types of feedback that can be given within a team setting are individual and team feedback (Nadler, 1979). There is still little evidence as to the amount of individual or team-level feedback that promotes the highest increase in performance (Gabelica, Bossche, Segers, & Gijssels, 2012). For the privacy component of feedback, studies agree that feedback needs to have some level of publicness on individual level feedback (Gabelica et al., 2012). However, there is little research on the publicness of team-level feedback. The purpose of this current work is to attempt to address this gap by providing evidence as to how different combinations of the Privacy and Audience component of feedback influence performance.

PREDICTED RESULTS

Consider a team that has three members, names X, Y, and Z. In this team, members X and Y are members that are high performing members (i.e., highly skilled). Member Z is not as skilled as members X and Y. However, member Z's performance increases when conducting a conjunctive task with members X and Y. This effect is known as the Köhler effect (Kerr, Messe, Park, & Sambolec, 2005). Based on current research, it is expected that when studying the influence that the privacy of feedback has on performance, we should find that the performance of team members should increase depending on whether or not the feedback content has information pertaining to team performance. Consequently, the hypothesis of this current work is that the performance of a team that receives feedback that is public and has information regarding the team performance will be higher than both teams that receive public feedback about individual performance and teams that receive private feedback on individual or team performance. This effect is predicted independent of the feedback content, as long as it is task-related.

MULTIPLE ERRANDS TEST (MET)

This current study used an adopted version of a task used in psychology called the Multiple Errands Test (MET). The MET was originally introduced by Shallice and Burgess (1991). The purpose of MET was to examine the performance of patients with prefrontal brain injuries as they perform tasks with variable cognitive load. The patients were given eight tasks to conduct within a shopping context. Of the eight tasks given to the patients, two were challenging tasks and the other six were simple tasks. The patients had to follow a set of instructions, or rules, while completing the eight tasks. If this task could be adapted to a team context, then it would offer a platform for assessing teams' ability to perform tasks with variable cognitive load.

Throughout the years there have been modified versions of the original MET. First, there was the MET – Hospital Version (MET-HV) (Knight, Alderman, & Burgess, 2002). This version of the MET was designed specifically for a hospital setting that consisted of patients that could not be studied in a public setting. Second, there was a version called the MET – Simplified Version (MET-SV) (Alderman, Burgess, Knight, & Henman, 2003). The MET-SV was

designed to examine a wide range of individuals that one may find in a different location, other than a hospital. Lastly, there was the Virtual MET (VMET) that was created as modified version of the MET-HV that was completed in a Virtual Mall (Rand, Basha-Abu Rukan, Weiss, & Katz, 2009). This shows that it is not unheard of to modify the original version of the MET for different foci. The following section describes a modified version of the original MET that was created specifically for teams.

TEAM MULTIPLE ERRANDS TEST (TMET)

This modified version of the original MET is called Team Multiple Errands Test (TMET). In this current work, the TMET was conducted within a Virtual Mall, similarly to the VMET, but with a team of three participants. In the TMET, each team member had two different shopping lists: Team and Individual List. The Individual List consisted of six items that were unique to the team member with which it was given. In other words, any items that are on Player 1's list will not be found on Player 2's or Player 3's list. The team as a whole was responsible for collecting all 18 of the items on the Team List, and no member was responsible for any specific item on the Team List. It was up to the team to communicate to decide how to collect all of the items on the list. The objective of this task is to buy all of the items on the list as quickly as possible. The participants had seven rules that they needed to follow while completing the TMET. The seven rules are as follows.

1. Do not spend over your allotted amount of money
2. If you enter a store, you must buy something.
3. You must only pick up one item from each store.
4. You can only visit a store once during the duration of a task.
5. You must buy only items that are on your list.
6. Meet up with your teammates at the fountain by when the timer is at 0:30 (30 seconds remaining) or earlier, after all your items are purchased.
7. Signal the researcher when you are finished.

The individual and team scores are based on the number of correct items collected. The teams and each of its members were not aware of their score while completing the task. For each session the teams had eight minutes to complete the shopping task.

METHOD

The purpose of the following experiment is to test the hypothesis that performance will be highest when feedback is given publicly and is directed to the team. The objective of this experiment is to better understand how the privacy (Public vs. Private) and audience (Direct vs. Group) of feedback influences team performance. This experiment is a 2 X 2 within-subject design. The independent variables in the current work are privacy and audience (**Figure 1**). The dependent variables are the number of correct items collected, the errors committed, and the time remaining.

Participants

Participants were recruited from a broad population because the TMET is a general shopping task. In this particular experiment the participants were all from a large Midwestern University. Confederates were used to make sure teams conducted the experiment with three members. A confederate is a person that has knowledge of the experimental objective and knows the purpose of the experiment. Confederates are not considered in data analysis. In total there were 10 teams and 30 participants (26 real participants and 4 confederates).

The 26 non-confederate participants were made up of 16 males (62%), 10 females (38%). The age range of the participants was as follows: 62% (16) were 18-21

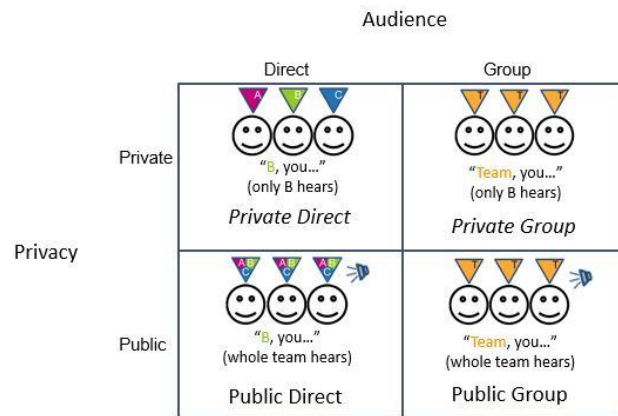


Figure 1. Experimental conditions. Used with permission from Walton et al. (2015). Created by student interns Hud, Mater, and Walker.

years old and 38% (10) were 22-30 years old. The participant weekly gaming experience was as follows: 27% (7) played zero hours per week, 27% (7) played zero to one hour per week, 4% (1) played one to two hours per week, 23% (6) played two to five hours per week, 12% (3) played five to ten hours per week, 4% (1) played 10 to 15 hours per week, and 4% (1) played 15 to 20 hours per week. The highest educational degrees received by participants were as follows: 57% (17) had a High school degree, 7% (2) had an Associate's degree, 7% (2) had a Bachelor's degree, 13% (4) had a Master's degree, and 3% (1) had a PhD. The team experience for participants was as follows: 31% (8) worked in teams daily, 50% (13) work in teams once or twice a week, 4% (1) worked in teams once or twice every two weeks, 4% (1) worked in teams once a month, 4% (1) worked in teams once or twice every year, and 8% (2) rarely worked in teams.

Procedure

Before arriving to the experiment participants complete an online informed consent form. After the participants were introduced to the experiment, they were then trained on how to navigate the virtual environment, designed in Unity. The training session ranged from 5 to 8 minutes. The goal of the training session was to ensure that each member was familiar with all of the controls within the environment and understood how to interact with the environment. The rules that participants had to follow while completing the shopping task were presented to them. After the training was complete the participants started the main task. There were four different sessions for the four different conditions (see **Figure 1**). Each session lasted no longer than eight minutes, depending on how much time the team needed to complete the task. After each session each participant completed a post-session survey. Also, the participants were given five minutes to reflect with one another on how they performed as a team in the previous session, a form of internal After Action review. This reflection time occurred after each session. Once all four sessions were complete, the participants were each given an overall post-session survey. The entire team was debriefed once each participant completed the overall post-session survey. The entire time it took to complete a session for each team was approximately 1.75 hours.

Confederates

There are a number of studies that, especially in psychology, that have used confederates (Goethals & Reckman, 1973; Kuhlen & Brennan, 2013; Lockridge & Brennan, 2002). Each confederate attempted to be a “neutral” player each time he or she participated in a session. To accomplish this goal, the confederate followed a script designed to make the confederate an average player who did not initiate specific team strategies. If the non-confederate players devised a strategy, then the confederate player went along with that strategy.

Feedback

Throughout each session the participant had the chance to receive feedback structured to the two conditions of privacy (Public or Private) and audience (Direct or Group). For example, in the Group condition, members of a team may receive feedback that says “Team, remember not to spend more money than budgeted.” This feedback was given publicly or privately, depending on the privacy. An example in the Direct condition was “Player 2, remember to always buy an item from every store you enter.” This feedback could also be given publicly or privately. Feedback within a given session was given in a specific mode, e.g., Public Direct, and the four treatment conditions were counterbalanced across the four sessions, such that each team experienced each feedback condition, but in different orders. Table 1 is a complete list of all the possible feedback a participant could receive and its corresponding rule. The first two feedback messages in the Related Feedback section would be given the first time a participant broke the rule. The last two feedback messages were given when the rule was broken three or more times. Feedback messages were printed visually on screen at the upper right. If the participant received multiple feedback messages, then the feedback would stack on top of one another with a maximum of three messages, most recent on top.

Limitation / Assumptions

This current work, like many other studies, made some assumptions and has some limitations. Based on results from pilot studies it was assumed that each session had a similar level of difficulty. If the sessions had different difficulty levels, e.g., a learning curve, then that could have influenced the participant’s performance. A limitation to this

experiment is that there were a small number of participants and an even smaller number of teams. This means that the statistical power of the results presented in the next section are weak due the small sample size.

Table 1. List of rules and the related feedback

Rule	Related Feedback
1. Do not spend over your allotted amount of money.	<ul style="list-style-type: none"> • Player X, remember not to spend more money than budgeted. • Team, remember not to spend more money than budgeted. • Player X, you have spent more money than budgeted. • Team, at least one of your members has spent more money than budgeted.
2. If you enter a store, you must buy something	<ul style="list-style-type: none"> • Player X, remember to always buy an item from every store you enter. • Team, remember to always buy an item from any store you enter. • Player X, you have forgotten Y times to buy something from a store you enter. • Team, your team has forgotten to buy something from a store you enter Y times.
3. You must only pick up one item from each store.	<ul style="list-style-type: none"> • Player X, remember you may buy only 1 item from each store. • Team, remember you may buy only 1 item from each store.
4. You can only visit a store once during the duration of a session.	<ul style="list-style-type: none"> • Player X, remember that you can only visit a store once during the duration of this session. • Team, remember that you can only visit a store once during the duration of this session. • Player X, Y times you have visited a store more than once. • Team, Y times members of your team have visited a store more than once.
5. You must buy only items that are on your list.	<ul style="list-style-type: none"> • Player X, remember to buy only items that are on your list. • Team, remember to buy only items that are on the list. • Player X, Y times now you have bought an item that was not your list. • Team, Y times members of your team have bought an item that was not on the list.
6. Meet up with your teammates at the fountain when the timer is at 0:30 (30 seconds remaining) or earlier, after all your items are purchased.	<ul style="list-style-type: none"> • Player X, remember to meet at the fountain with at least 30 seconds remaining. • Team, remember to meet at the fountain with at least 30 seconds remaining • Player X, you have X minutes remaining to return to the fountain. • Team, you have X minutes remaining to return to the fountain.

RESULTS

The individual and team performance was based on the correct items collected, errors committed, and time remaining at the end of a session. It was discovered that 70% of the participants strongly agreed that their own individual performance improved over time and 77% of the participants strongly agreed that the team's performance improved over time. Consequently, session order was significant for some variables. When the session order was significant, the data were separated by session and then analyzed.

Time Remaining - Individual

The amount of time remaining for each participant was calculated as a percentage of the total amount of time they were given. Teams were given a total of eight minutes (480 seconds) to complete each session. There was no significant difference on Time Remaining with respects to Privacy, Audience, or the interaction after running a two-way within-subject ANOVA test. A one-way within-subject ANOVA test showed that there was a statistically

significant difference among the session's orders. Specifically, there is a statistically significant difference between Sessions 1 and 3 ($p < .001$), and 1 and 4 ($p < .001$) (pairwise t-test). This result suggests that the participants experienced a learning curve. The data were grouped by session for further analysis.

For Session 1, the privacy component had a significant ($F(1,18) = 53.0$, $p < .001$, $\eta^2 = .060$) influence on the time remaining at the end of each session. The Public Group condition produced the highest percentage of time left, $1.18\% \pm .590\%$. For Session 2, the Audience component of feedback had a significant influence on the time remaining at the end of each session ($F(1,18) = 4.91$, $p = .039$, $\eta^2 = .156$). The Private Direct condition produced the highest percentage of time left, $6.8\% \pm 1.4\%$. For Session 3, the interaction of the privacy and audience ($F(1,18) = 96.9$, $p < .001$, $\eta^2 = .001$), the privacy component ($F(1,18) = 116.1$, $p < .001$, $\eta^2 = .195$), and the audience component ($F(1,18) = 4.039$, $p = .059$, $\eta^2 = .011$) of feedback had a significant influence on the time remaining at the end of each session. The Private Group condition produced the highest percentage of time left, $14.9\% \pm 4.2\%$. For Session 4, there was no significant difference for the interaction of the privacy and audience feedback component, privacy component, or the audience component. **Figure 2** shows the interaction plots for all four sessions.

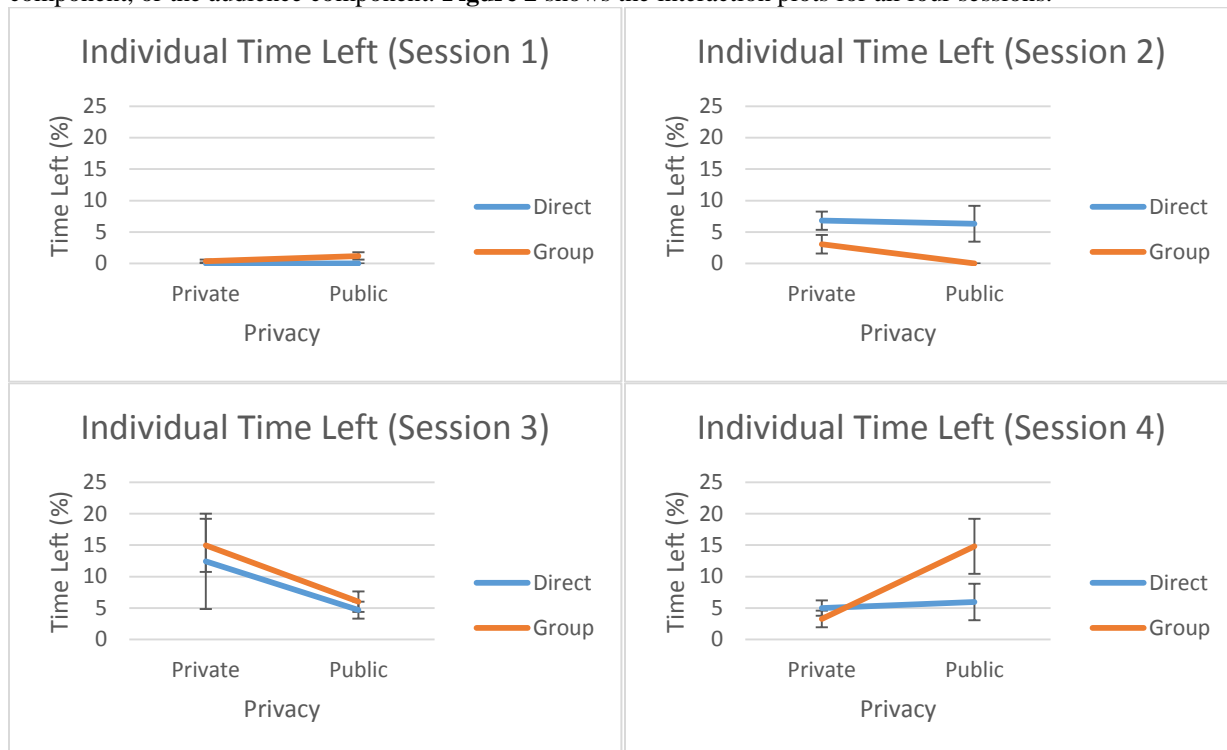


Figure 2. Time Left Interaction plots for individuals. Error bars represent S.E.M. Differences were not significant, in part because even though each team experienced each condition at some point, there were very few teams in a specific session who experienced a specific condition.

Time Remaining - Team

There was no significant difference on Time Remaining of the Team with respect to Privacy, Audience, or the interaction after running a two-way within-subject ANOVA test. As well, the session order did not have a significant influence on the data overall.

Errors – Individual

In this current work, an error occurs when a participant breaks one of the rules. There was no cap to the number of rules that a participant was able to break. There was no significant difference on Errors with respect to Privacy, Audience, or the interaction after running a two-way within-subject ANOVA test. As well, the session order did not have a significant influence on the data overall.

Errors – Team

There was no significant difference on team errors with respect to Privacy, Audience, or the interaction after running a two-way within-subject ANOVA test. As well, the session order did not have a significant influence on the data overall.

Items Collected – Individual

Each player's individual list contained six items. There was no significant difference on items purchased with respect to Privacy, Audience, or the interaction after running a two-way within-subject ANOVA test. A one-way within-subject ANOVA test showed that there was a statistically significant ($F(1,100) = 6.5, p < .05, \eta^2 = .335$) difference among the session's orders. Specifically, there is a statistically significant difference between Session 1 and 3 ($p < .001$), and 1 and 4 ($p < .001$), 2 and 3 ($p < .001$), and 2 and 4 ($p < .001$) (pairwise t-test). This result suggests that the participants experienced a learning curve. The data was grouped by session for further analysis.

For Sessions 1, 3 and 4, there were no significant differences for the interaction of the privacy and audience feedback component, privacy component, or the audience component. For Session 2, the privacy component had a significant ($F(1,18) = 10.9, p = .003, \eta^2 = .001$) influence on the item collection. The Private Group condition produced the highest percentage of correct items, $70.37\% \pm 13.73\%$. **Figure 3** shows the interaction plots for all four sessions.



Figure 3. Average percentage of individual items collected by player. Error bars represent S.E.M. Differences were not significant. Note that individual item percentages rise over time.

Items – Team

For the team list, there were a total of 18 items. There was no significant difference on team list purchases with respect to Privacy, Audience, or the interaction after running a two-way within-subject ANOVA test. Also, the session order did not have a significant influence on the data overall.

Communication

During the sessions, participants' verbal communications were audio-recorded. Utterances within these recordings were counted manually by two researchers. An utterance was defined as a series of words spoken with a single intent, separated from other utterances by a breath or silence. E.g., "Should we all shop together? [breath] I'll buy the napkins" would be two utterances. For several teams, audio recordings were lost due to technical issues.

It was hypothesized that team communication would be predictive of team performance, and that team communication would decrease by Session 4 as the team became more efficient with the task. This decrease is hypothesized based on the theory of implicit coordination (Espinosa, Lerch & Kraut, 2004), i.e., that as team members built common understanding of a task, their need to communicate decreases, allowing cognitive load to be focused on increasing performance of the task itself.

Figure 4 shows a chart of the teams for which utterance data were present, ordered according to the team performance in Session 4. There were too few data points to achieve statistical significance of utterance rate on team performance. However, the lowest performing (Team 5) did also have the lowest communication rates. Teams 8, 2, and 7 do show a pattern of implicit coordination: an initial burst of communication that decreases by Session 4, but that pattern is not consistent across all teams. In a future study with more participants, these results might be stronger.

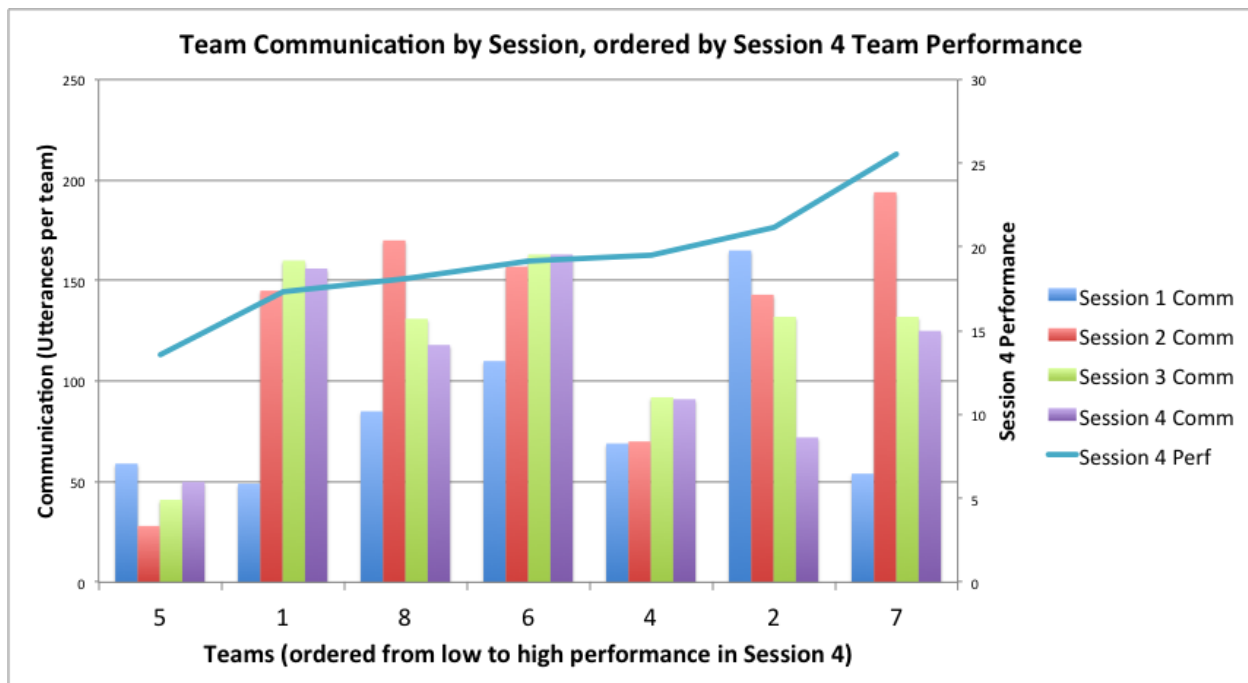


Figure 4: Team Communication (number of utterances) by session, with teams ordered by performance in Session 4 (low to high). Some teams missing due to incomplete audio data. Note evidence of implicit coordination (decreasing communication as teams mature).

DISCUSSION AND CONCLUSION

The purpose of this current work is to study the influence that the privacy and audience component of feedback has on a team's performance. The objective of this experiment is to better understand how the privacy (Public vs. Private) and audience (Direct vs. Group) of feedback influences team performance. Performance was defined in terms of the number of correct items collected, amount of time remaining at the end of each session, and the number of errors committed. The hypothesis of this current work was that the Public Group condition would produce the highest overall performance. The hypothesis of this current work was not supported by the data presented above.

There was no strong conclusion about feedback modality that could be drawn from the data presented overall. There was no condition that consistently produced the highest or lowest overall performance. This may be due to the fact that during the experiment the participants did not seem to pay close attention to the feedback given to them during the task. Based on words spoken by the participants during their internal after action reviews, it was apparent that some ignored the feedback or were so focused on the task at hand that they did not notice the feedback. These factors likely weakened the feedback intervention.

Over time (i.e., across the session order) the influence of the privacy and audience of feedback seemed to change. There was a learning curve for some variables. Because the controls of the game were simple, and players appeared to master the controls during the training session, the authors suggest that the learning curve is not due to growing familiar with the interface itself. Instead, the authors suspect that the improvement over sessions is due to team strategies devised among team members during their internal after action reviews, the reflection times between sessions. Also, because the TMET is a team task that requires significant communication, this improvement over sessions may indicate improving communication or other team skill.

Although the results of the different forms of feedback did not yield the expected results, the design of the Team MET (TMET) as a platform for testing team performance was successful. The TMET allows the trainer or researcher to vary the cognitive load of the participants by varying the number of items on the individual and team shopping lists, and by decreasing the time to perform. Success was also measured by two other criteria: 1) team performance and individual performance ranged widely, with no ceiling or floor effects, and 2) performance seemed consistent or improving over sessions, rather than randomly changing. These two criteria suggest that the TMET could be a useful instrument for assessing a team's ability to perform under cognitive load. Lastly, the TMET could be used as a platform for training better team communication, in that the tasks can be done repeatedly until a team maximizes performance. This approach would be useful for integrating a new team member into an experienced team.

ACKNOWLEDGEMENTS

We thank professor Rob West, Ph.D., of Psychology, for pointing us to the Multiple Errands Task originally. We thank NSF REU interns Kelsey Walker, Samantha Mater, Anton Hud for creation of the initial virtual mall and early planning on the experimental design. This material supported in part by the National Science Foundation Grant CNS-1156841.

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. doi:10.1.1.121.9138
- DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A Multiple-Goal, Multilevel Model of Feedback Effects on the Regulation of Individual and Team Performance. *The Journal of Applied Psychology*, 89(6), 1035–56. doi:10.1037/0021-9010.89.6.1035
- Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The Effects of Peer Feedback on Team Member Behavior. *Group & Organization Management*, 22(4), 508–520. doi:10.1177/1059601197224006
- Espinosa, J. A., Lerch, F. J., & Kraut, R. E. (2004). Explicit Versus Implicit Coordination Mechanisms and Task Dependencies: One Size Does Not Fit All. In E. Salas & S. M. Fiore (Eds.), *Team Cognition: Understanding the Factors That Drive Process and Performance*. Washington, D.C.: American Psychological Association.
- Gabelica, C., Bossche, P. Van Den, Segers, M., & Gijssels, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review*, 7(2), 123–144. doi:10.1016/j.edurev.2011.11.003
- Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, 9(6), 491–501. doi:10.1016/0022-1031(73)90030-9

- Hategekimana, C., Gilbert, S., & Blessing, S. (2008). Effectiveness of using an intelligent tutoring system to train users on off-the-shelf software. In *Society for Information Technology & Teacher Education International Conference* (pp. 414–419). Retrieved from <http://www.editlib.org/p/27199/>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Kerr, N. L., Messe, L. A., Park, E. S., & Sambolec, E. J. (2005). Identifiability, Performance Feedback and the Kohler Effect. *Group Processes & Intergroup Relations*, 8(4), 375–390. doi:10.1177/1368430205056466
- Knight, C., Alderman, N., & Burgess, P. W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation*, 12(3), 231–255. doi:10.1080/09602010244000039
- Koedinger, K. R., Alevan, V., Hockenberry, M., McLaren, B. M., & Heffernan, N. (2004). Opening the door to non-programmers: Authoring intelligent tutor behavior by demonstration. In *Proceedings of the Seventh International Conference of Intelligent Tutoring Systems*. Maceio, Brazil. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-30139-4_16
- Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, 20(1), 54–72. doi:10.3758/s13423-012-0341-8
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, 9(3), 550–557. doi:10.3758/BF03196312
- Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance*, 23(3), 309–338. doi:10.1016/0030-5073(79)90001-1
- Peñarroja, V., Orengo, V., Zornoza, A., Sánchez, J., & Ripoll, P. (2015). How team feedback and team trust influence information processing and learning in virtual teams : A moderated mediation model. *Computers in Human Behavior*, 48, 9–16.
- 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 Rehabilitation*, 19(4), 583–602. doi:10.1080/09602010802469074
- Shallice, T., & Burgess, P. W. (1991). Deficits in Strategy Application Following Frontal Lobe Damage in Man. *Brain*, 114, 727–741.
- Sottolare, R., Holden, H., Brawner, K., & Goldberg, B. (2011). Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training. *U.S. Army Research Laboratory, Human Research and Engineering Directorate, Orlando, Florida*.
- Walton, J., Bonner, D., Walker, K., Mater, S., Dorneich, M., Gilbert, S., & West, R. (2015). The Team Multiple Errands Test: A Platform to Evaluate Distributed Teams. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (pp. 247–250).