Using the Simplified Multiple Errands Test to Assess the Effect of Feedback on the Communication and Performance of a Team

Samantha Mater  
 Oberlin College, Oberlin, OH

VRAC, Iowa State University  
smater@oberlin.edu

Anton Hud

Redlands University, Redlands, CA

VRAC, Iowa State University

anton\_hud@redlands.edu

Kelsey Walker

Central Washington University, Ellensburg, WA

VRAC, Iowa State University

walkerkel@cwu.edu

**Abstract.** Feedback is examined in this paper in respect to its interaction and influence on communication and team and individual task performance in a team tutoring scenario. Public and private, and group and individual feedback are manipulated. When it comes to intelligent tutoring, the field currently only supports teaching one person at a time, because trying to teach more than one person introduces a layer of complexity that can create confusion as to which teaching strategy would be best for a team consisting of individuals who learn differently. We are looking to improve the field of team tutoring by examining the effectivness of our stated feedback types in combination. In our study, participants were randomly assigned to teams of three, each using an avatar, to complete a variety of team and individual tasks adapted from the Simplified Multiple Errands Test in our virtual mall environment. We believe that we will find participants given team feedback to have a lower individual performance rate but higher communication rate than those given inidividual feedback; however, team feedback will led to greater team performance. Additionally, we do not anticipate that the difference between the rest of the feedback types and outcomes to be significant. We conclude with what might impact these results, and implications regarding how this study might inform future work in team tutoring.

**Keywords:** *Team Tutoring, Multiple Errands Test, Virtual Environment, Feedback, Teamwork*

1. Introduction

Effective training is necessary to design, build, and utilize complex products and processes in modern society. Whether designing the next automobile or coordinating troop movements, in-context training scenarios involving teams or groups of individuals is a growing need. While much research has been performed on ways to provide individual feedback during training (Narciss, 2004; Brand-Gruwel, Kester, Kicken, Kirschner, 2014), the questions “Do teams function more efficiently when given individual feedback or group feedback?” and “Do teams function more efficiently when feedback is given publicly or privately?” have not been examined together to find the most effective combination. To identify the optimal pairing, we decided to modify the Simplified Multiple Errands Test (Alderman, Burgess, Knight, Henman, 2003) to use as a tool to see how feedback affected the way the team interacted.

* 1. The Problem

Over the past twenty years, with the increasing ubiquity of the internet and other forms of instantaneous communication, the contemporary world has become increasingly globalized. Currently, trade, investment, migration, culture, social relations, and other social, political, and economic factors are being exchanged at unprecedented levels. This expansion of increased interdisciplinary and intercultural collaboration calls for individuals to team together to solve such complex problems. However, for this to take place, individuals must first be trained to work together. Diversity in viewpoints is necessary for coming up with the best possible solution (Phillips, 2009). Groups and teams make for enriched learning environments, by increasing the diversity of perspectives, enhancing creativity, and drawing on the skills of multiple individuals rather than just one, which is reflected in the increasing popularity of crowdsourcing. Additionally, groups consistently outperform individuals (Hill, 1982; Yager, 1985; Lou, 2001) and it takes significantly less resources to train people in teams or groups compared to individually, so improving the field of team tutoring is important.

* 1. How Studying Feedback Can Help

By increasing knowledge of optimal feedback presentation in the virtual team tutoring environment, we are able to better understand how to teach groups and teams more effectively, by improving learners’ team communication and their individual and team performance. Identifying the best feedback presentation methodologies can improve teamwork by fostering communication and empower teams by enabling group members to proficiently work together toward common goals. This study also increases understanding regarding how people are affected by feedback physiologically, emotionally, in regard to their level of self-esteem, and their personality type – and how these elements factor into their communication and performance. Feedback-related changes in behaviors regarding communication and performance are considered using correlates identified in metrics triangulation statistical analysis.

* 1. Why do this in a Virtual Environment

Like the original MET, our task could be carried out in the real world, but our future lies in technology. Children and toddlers are being given iPads to play on, building their own games to play on them, developing knowledge pathways differently than prior generations – accessing more information, and people, faster. Thus, it only makes sense to incorporate technology into other aspects of their lives, especially education. Already, the military utilizes virtual environments to train soldiers (Smelik, Tutenel, de Kraker, Bidarra, 2010).. Virtual training, in addition to being cheaper, safer, and arguably more efficient for learners, also affords abilities of individuals to interact with one another regardless of geographic distance (Walker, et al. 2010). This paper seeks to contribute to existing team tutoring literature by identifying the best ways feedback can be implemented in team tutoring scenarios in order to optimize team communication and performance.

1. Related Work

This paper builds on previous research in the areas of team tutoring, feedback, teamwork measurements, and the Multiple Errands Test. We begin by briefly outlining research done into team versus individual feedback in a team tutoring setting.

**Team Tutoring.** Cognitive overload of feedback was observed by DeShon et al. in 2004, when participants, given two types of feedback on both team and individual tasks, were unable to optimally act on their different types of tasks. DeShon et al. found that in a team setting the individuals only given feedback on individual tasks improved on individual performance, while individuals who were only given feedback on group tasks improved on group performance. Furthermore, participants who only received individual feedback were unable to accurately estimate the group’s collective abilities, which led to the conclusion that individuals ultimately allocate more effort into what they receive feedback on. This is the first study to examine the effects of feedback on a team when both individual and group goals were present, and the first to look at team performance as a multi-level concept. Building off of this study, we aim to integrated duality of goals into our adaptation of the MET and consider team and individual performance and communication to be a layered concept influenced by multiple variables.

**Real Time Language Feedback.** As mentioned previously, feedback has been shown to be a powerful factor in allocation of priorities (Northcraft, Schmidt, Ashford, 2011). Because of this, studying feedback is incredibly vital, and indeed it has been widely studied. Experimentation has already been done regarding feedback timing, the effects of feedback on communication, and types of feedback (Smith, Kimball, 2010; Smits, Boon, Sluijsmans, Van Gog, 2008; Tausczik, Pennebaker, 2013; Lou, Abrami, d’Apollonia, 2001). Real time language feedback has been shown to be more effective than after action review (AAR) (Rasker, Post, Schraagen, 2000) which can be given through the use of Intelligent Tutoring Systems. Because of this, we are implementing real time language feedback in our study, which we hope to someday use to improve intelligent tutors.

**Metrics.** Outlining a construct for teamwork has been an ongoing struggle for researchers, as Salas, Sims and Burke describe, “a clear definition of teamwork continues to be elusive” (p.556, 2005). At the present, there still seems to be no particular construct for measuring teamwork, but rather a variety of approaches (Aguado, Rico, Sánchez-Manzanares, and Salas, 2014; Entin, Entin, 2001; Shanahan, Best, Finch, and Sutton, 2007; Tausczik, Pennebaker, 2013). For a comprehensive, but by no means complete, list see Shanahan et al. (2007). Lists of skills shown effectively correlated with good teamwork, such as the ACT WorkKeys, do exist; however, WorkKeys is a theoretical test with a multiple choice answer key, allowing no room for creative responses, and it is only useful if team members have predefined roles and goals both at the team and individual level. This does not work for our study, because participants in our study do not have roles, may not contribute well to the team, or work effectively together. Also WorkKeys only observes four areas, Applied Mathematics, Applied Technology, Locating Information, and Reading for Information. Our study will be examining significantly more, and diverse, variables than these.

Aguado et al. (2014) put forward the Teamwork Competency Test (TWCT) and also examines the Teamwork Knowledge, Skill, Ability Test (TWKSAT). Both of these would be able to help target specific areas to tutor in, as well as provide us with a metric that’s been tried before, however, in the study, the participant demographics were skewed in gender and age, which could have an impact on results, and needs further testing before extensive use.

Tausczik and Pennebaker explored how real time language feedback affected group communication (2013). Their study looked at group processes likely related to performance and found Information Exchange to be a relevant measure. While the study is limited in scope, as it only provides generalized group feedback on a single aspect of communication, we believe that Information Exchange as a metric was sound, and we have modified it for our own use. Tausczik’s study, like DeShon’s (2004) also demonstrated that giving people more than one type of feedback in a single scenario could cause cognitive overload, depleting participants’ ability to optimize on the material they are trying to learn.

**Multiple Errands Test.** The MET, (Knight, 2002; Alderman, 2003; Raspelli, 2012; Shallice and Burgess, 1991) has been used to identify the differences in executive functions between cognitively healthy individuals and those suffering from brain damage who seem otherwise cognitively functional. In the Simplified MET, individuals are given a list of tasks to complete in an indoor shopping mall along with a list of rules to follow. The evaluators of the simplified MET measured errors made, and used this metric to identify which types of errors were unique to brain damaged participants versus mistakes made by healthy participants (Alderman, 2003). In 2009, Rand et al. tested the ecological validity of performing the MET in a virtual environment (vMET). Pending future research, Rand et al. found a virtual environment to be an ecologically valid option after using it to find the difference between healthy participants and healthy post-stroke participants. The vMET study demonstrated necessary ecological validity of the virtual environment, and since we are not examining healthy versus damaged brains, but rather the executive functions carried out by teams in a virtual environment, it was useful in setting the technological precedent, guiding the design of our virtual environment.

We are changing the simplified MET substantially in three ways. The first, as previously mentioned, we are carrying out this experiment in a virtual environment. In this regard, there are errors present in non-virtual METs, such as posting a personal letter, and stopping to get a coffee*,* which we do not anticipate occurring in our study.

The second substantial change to the simplified MET is having groups of three complete the MET as a team, rather than having one person complete the MET individually. As stated earlier, the literature on team tutoring points to groups outperforming individuals. In order to capitalize on this, however, we need to train groups effectively. To understand how we can best train individuals, we are investigating the executive functions of the group as a whole using the MET, which has never been done before.

The third change implemented is participants are provided with real-time feedback through the use of the “Wizard-of-Oz” technique (Maulsby, Greenberg, Mander, 1993) to help them accomplish the task. The Wizard-of-Oz technique involves the experimenter giving participants proactive and reactive feedback (i.e. when they are running low on money, or when they have not communicated with the team in three minutes). We are coaching participants to learn how to perform the task and communicate with their teams more effectively. We will run participants through the modified MET twice in order to identify changes in participants’ communication and performance. While we expect typical participant learning to occur through multiple trials of the same scenario, we do not expect this to be especially significant, as we will have already trained participants in the environment prior to the first trial of the MET.

1. Method

**Participants.** We will recruit participants 18 years old and older who do not have any seizure disorders or any implanted medical devices. We will primarily be recruiting from a large, public research university.

* 1. Scenario

**Practice Task.** Participants will be familiarized with the controls in the virtual environment and then will be asked to make a purchase in the virtual environment. Once they have demonstrated an understanding of how the interface, controls, feedback, and game works, they will proceed to the modified MET.

**Modified MET.** Upon launching the virtual environment, the participant is connected to a server hosted by the test administrator, who will be able to see everything that’s happening in real time, to give appropriate feedback. As participants move around the environment, they are given prompts to interact with stores. Upon approaching a store, participants can open its inventory and double click an item to purchase it (see Fig. 1). The cost is subtracted from the total amount of “money” participants have, appearing numerically top right hand corner of their screen. The item then appears in the user’s inventory. Each item a participant purchases is time stamped and logged, along with any of their individual rule breaks, then output to a file upon reaching the user’s time limit, which will be fifteen minutes. At the end of this time, the users should have met up at a location specified in the instructions. The local machine time is also displayed in the top right corner of the screen, so participants can keep track of how long they’ve been in the environment and how long they have left before they need to meet up.

The administrator will give one of the four feedback types as she or he observes communication according to the previously agreed upon protocol for observation of specific communication behaviors. Training to determine inter-rater reliability between experimenters’ ability to identify specific behaviors and give appropriate feedback, according to the feedback conduct in the experimental booklet, will be done with an 80% rater agreement prior to the experiment.



**Fig. 1.** Purchasing a Hard Hat at the hardware store.

* 1. Independent Variables.

In giving feedback, we manipulated two variables—whether the feedback was publicly or privately given, one being over a loudspeaker and the other over a headset, and whether the feedback was about the team or the individual.

* 1. Dependent Variables.

We will compare performance across feedback types through the amalgamation of time to make a selection, selection of the best plan, number of times instructions consulted, path, rule breaks, and task completion to measure the success of both groups and individuals at their respective tasks. We will also compare communication by looking at the measures of information requests and number of communications of each team. Though primarily we are using performance and communication to measure the effectiveness of the different feedback types, you can see from Figure 2 that they are not the only dependent variables we will be looking at. We will also be looking at skin conductance, blink rate, and facial affectation to see if there are correlations between those metrics and feedback.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dependent Variable** | **Metric** | **Measurement / Unit** | **Frequency of Collection** | **Data Type** |
| Performance | Time to make a selection | seconds | once per trial | ratio |
| Performance | Number of times policy statements consulted | count | once per trial | discrete |
| Performance | Avatar Path | % of optimal path followed | once per trial | percentage |
| Performance | Rule Breaks | count | once per trial | discrete |
| Performance | Task Completion | yes / no | once per trial | binary ordinal |
| Workload | TLX measures:  a) Mental Demand  b) Physical Demand  c) Temporal Demand  d) Performance  e) Effort  f) Frustration | 0 – 10 | once per trial | subjective ordinal |
| Workload | Self-Report Survey | 7 point Likert scale | once per trial | discrete |
| Emotional Arousal | Skin Conductance | microSiemens | once per minute | ratio |
| Emotional Arousal | Affect | Facial Emotion | once per minute | subjective |
| Attention | Blink Rate | count | count per  minute | ratio |
| Communication | Information Requests | count | count per  minute | ratio |
| Communication | Total Communications | count | count per  minute | ratio |
| Communication | Coordination Requests | count | count per  minute | ratio |
| Communication | Action Requests | count | count per  minute | ratio |

Fig. 2. Dependent variable chart.

* 1. Procedure

Participants are randomly assigned into groups of three, and are first asked to take the Jung Typology Test; then they are fitted with a BodyMedia armband, which measures skin conductance. For metrics triangulation and accuracy, one of the three members in each group will also be fitted with Affecta Qsensor wristband to wear throughout the study as well as the BodyMedia sensor. The other two participants will only wear the BodyMedia sensor without the Qsensor throughout the experiment. Next, participants are taken into separate rooms and given the practice task to complete to familiarize them with the controls of the virtual environment and the feedback system. Once they demonstrate understanding of the controls, participants are given an instruction sheet, clearly outlining the modified MET tasks. Participants are given fifteen minutes to complete the scenario, and then are given the NASA-TLX to complete as well as the demographics survey.

* 1. Testing Environment

The virtual environment uses Unity for the game engine, and it is programmed primarily with C#. The three desktop computers used run windows, and each monitor has a webcam attached at the top, which utilizes a facial affect reading software, Facereader, that measures blink rate over time. We have four computers in total: one server, hosting the virtual environment, and the three, connecting to the server and used by participants, displaying the third person perspective of each participant’s avatar. Users communicate with one another over headsets using Ventrilo voiceover internet protocol (VOIP). At the start of the experiment, participants are told in which condition they will be receiving feedback, whether privately and individually, privately and as a team, publicly and individually, or publicly and as a team. Private feedback is directed to individuals through their headphones, and on the participants’ screen, and public feedback will be given over a loudspeaker system, as well as on the participants’ screen. The audio and the visual feedback will include the same words, spoken or in print.

* 1. Data Analysis Plan

The data collected using the metrics in Figure 2 will be analyzed to determine significance amongst feedback groups. The specific techniques we will use are still being decided.

1. Results

At the moment, we have not yet run our study. However, when we do, we anticipate that our results will look similar to Figures 3. Based on previous research in the field of team tutoring, we anticipate that team members who are given individual feedback will do better on individual tasks, and team members who are given team feedback will do better on team related tasks. Additionally, those given public feedback will perform better than those given private feedback, because they will also be hearing feedback given to their peers. We also believe that teams given public feedback will have higher communication levels than those given private feedback, and the team feedback group will have better communication than the individual feedback group.

**Fig. 3.** Anticipated results.

1. Discussion and Future Work

However, as DeShon wrote, team feedback is a multi-level concept and what impacts performance, and our metrics will not be the only factors influencing outcomes. What that means is that while this study will show us correlation, it might not necessarily mean causation. That being said, we picked these measures because we believe they will have a large impact on our results, and will show us improvements between feedback types. In the future, we are going to run our test. After we obtain and analyze our results, we could continue by looking at the effects of adding positive and negative feedback to our study, as well as expand our demographics to be more inclusive.

1. Conclusion

With a greater understanding of the most beneficial feedback involved in team tutoring, modern education can continue to develop. Team training scenarios are essential to the contemporary social world, and the aim of this study was to improve these scenarios through increasing the knowledge of the most useful forms of feedback to team communication and performance. We examined team, individual, public, and private feedback together to find the most effective combination. In doing so, we hoped to identify how combinations of feedback affect the way teams interact and perform to improve the field of team tutoring.

1. Acknowledgements

This work was funded by NSF grant CNS-1156841. Thank you to Dr. Michael Dorneich, Desmond Bonner, Jamiahus Walton, Dr. Stephen Gilbert, Dr. Elliot Weiner, Chase Meusel, and Iowa State University’s SPIRE-EIT REU.

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*, *9*(01), 31-44.

Aguado, D., Rico, R., Sánchez-Manzanares, M., & Salas, E. (2014). Teamwork competency test (twct): A step forward on measuring teamwork competencies.

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

Brand-Gruwel, S., Kester, L., Kicken, W., & Kirschner, P. A. (2014). Learning Ability Development in Flexible Learning Environments. In *Handbook of Research on Educational Communications and Technology* (pp. 363-372). Springer New York.

DeShon, R. P., Kozlowski, S. W., 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. *Journal of applied psychology*, *89*(6), 1035.

Entin, E. E., & Entin, E. B. (2001, June). Measures for evaluation of team processes and performance in experiments and exercises. In *Proceedings of the 6th International Command and Control Research and Technology Symposium* (pp. 1-14).

Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one?. *Psychological Bulletin*, *91*(3), 517.

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.

Lou, Y., Abrami, P. C., & d’Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of educational research*, *71*(3), 449-521.

Maulsby, D., Greenberg, S., & Mander, R. (1993). Prototyping an intelligent agent through Wizard of Oz. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 277-284). ACM.

Northcraft, G. B., Schmidt, A. M., & Ashford, S. J. (2011). Feedback and the rationing of time and effort among competing tasks. *Journal of Applied Psychology*, *96*(5), 1076.

Narciss, S. (2004). The impact of informative tutoring feedback and self-efficacy on motivation and achievement in concept learning. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *51*(3), 214.

Rand, D., Rukan, S. B. A., Weiss, P. L., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, *19*(4), 583-602.

Rasker, P. C., Post, W. M., & Schraagen, J. M. C. (2000). Effects of two types of intra-team feedback on developing a shared mental model in Command & Control teams. *Ergonomics*, *43*(8), 1167-1189.

Raspelli, S., Pallavicini, F., Carelli, L., Morganti, F., Pedroli, E., Cipresso, P., ... & Riva, G. (2012). Validating the Neuro VR-based virtual version of the Multiple Errands Test: preliminary results. *Presence: Teleoperators and Virtual Environments*, *21*(1), 31-42.

Shallice, T. I. M., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727-741.

Shanahan, C., Best, C., Finch, M., & Sutton, C. (2007). *Measurement of the behavioural, cognitive, and motivational factors underlying team performance*(No. DSTO-RR-0328). DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION EDINBURGH (AUSTRALIA) AIR OPERATIONS DIV.

Shute, V. J., & Psotka, J. (1994). *Intelligent Tutoring Systems: Past, Present, and Future* (No. AL/HR-TP-1994-0005). ARMSTRONG LAB BROOKS AFB TX HUMAN RESOURCES DIRECTORATE.

Smelik, R. M., Tutenel, T., de Kraker, K. J., & Bidarra, R. (2010). Declarative terrain modeling for military training games. *International journal of computer games technology*, *2010*, 2.

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay–retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 80.

Smits, M. H., Boon, J., Sluijsmans, D. M., & Van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: effects on learning as a function of prior knowledge. *Interactive Learning Environments*, *16*(2), 183-193.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT). *Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED)*.

Tausczik, Y. R., & Pennebaker, J. W. (2013, April). Improving teamwork using real-time language feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 459-468). ACM.

Walker, M. I., Walker, R. B., Morgan, J. S., Bernhagen, M., Markin, N., & Boedeker, B. H. (2010). The use of virtual training to support insertion of advanced technology at remote military locations. *Studies in health technology and informatics*, *163*, 688-690.

Yager, S., Johnson, D. W., & Johnson, R. T. (1985). Oral discussion, group-to-individual transfer, and achievement in cooperative learning groups. *Journal of Educational Psychology*, *77*(1), 60.