Human Factors Impact of Different Ramp Controller Scheduling Advisories for ATD-2 Surface Metering in a Human-in-the-Loop Simulation

Bonny K. Parke[1] and Deborah L. Bakowski[2]
San José State University/NASA Ames Research Center, Moffett Field, CA 94035, USA

Yoon C. Jung[3], Hanbong Lee[4], William J. Coupe[5], Lindsay K. S. Stevens[6]
NASA Ames Research Center, Moffett Field, CA 94035, USA

A human-in-the-loop (HITL) simulation took place in 2019 at NASA Ames Research Center's Future Flight Central, an airport simulator with a full-scale 360-degree view that showed real-time airport operations at Dallas/Fort Worth International Airport (DFW). The HITL was part of the NASA/FAA Airspace Technology Demonstration 2 (ATD-2) project which involves an on-going field demonstration at Charlotte Douglas International Airport (CLT). In the field demonstration, metering on the airport surface takes place with one of the goals being to reduce runway queue length by holding aircraft at the gate during busy times, thereby reducing fuel burn and CO_2 emissions. This goal has been successfully accomplished at CLT by showing Ramp Controllers when to release aircraft at the gate so as to reduce aircraft waiting in the queue. Among the questions that the current simulation at DFW was designed to address were, "Which types of advisories to the Ramp Controllers work best, those that show a ±2 minute window for a release time from the gate, or a ±5 minute window for an arrival time at the spot (where Air Traffic Control takes over) or both?" and "How do these different types of advisories affect Ramp Controller workload and situation awareness?" Another question was, "Does a ramp that is less restrictive than the one at CLT affect compliance with these advisories?"

I. Introduction

Surface metering is currently taking place at Charlotte Douglas International Airport (CLT) as part of the Airspace Technology Demonstration 2 (ATD-2) project [1], a combined NASA/FAA effort based on the FAA's Surface Collaborative Decision Making (S-CDM) Concept of Operations [2]. One of the goals of surface metering is to reduce aircraft wait times in the departure runway queue, with its attendant fuel burn and CO_2 emissions. This has been achieved at CLT by Ramp Controllers holding aircraft at the gate until a scheduler-generated time advises them to release the aircraft within a ±2 minute window. During the 28 months from November 29, 2017 to March 31, 2020, it was estimated that 2,881,628 lbs. of fuel were saved at CLT by holding 25,739 departures at the gate for an average

[1]Senior Research Associate, SJSU/NASA Ames Research Center, Mail Stop 262-4, Moffett Field, CA 94035, bonny.parke@nasa.gov

[2]Senior Research Associate, SJSU/NASA Ames Research Center, Mail Stop 262-4, Moffett Field, CA 94035, debi.bakowski@nasa.gov

[3]Aerospace Engineer, NASA Ames Research Center, Mail Stop 210-6, Moffett Field, CA 94035, yoon.c.jung@nasa.gov , AIAA Senior Member.

[4]Aerospace Engineer, NASA Ames Research Center, Mail Stop 210-6, Moffett Field, CA 94035, hanbong.lee@nasa.gov , AIAA Member.

[5]Aerospace Engineer, NASA Ames Research Center, Mail Stop 210-6, Moffett Field, CA 94035, william.j.coupe@nasa.gov , AIAA Member.

[6]Aerospace Engineer, NASA Ames Research Center, Mail Stop 210-6, Moffett Field, CA 94035, lindsay.stevens@nasa.gov

of 5.9 minutes. During the same period, it was estimated that CO_2 emissions were reduced by approximately 8,875,416 lbs., equivalent to planting 65,997 urban trees [3].

II. Research Questions

A requirement for the Terminal Flight Data Manager (TFDM), the FAA program which will provide many of the same functions as ATD-2, is that departure aircraft arrive at the spot, within a ±5 minute window of a time generated by a scheduler. The "spot" is a location at the border between the ramp area and the airport movement area where Air Traffic Control takes control of aircraft.

One question is whether this is best accomplished through providing Ramp Controllers with displays showing:

1) A gate advisory with a ±2 minute window (as takes place at CLT),
2) A spot advisory with a ±5 minute window presented at the gate, or
3) Both gate and spot advisories.

We tested options #1 and #3 in a previous Human-in-the-Loop (HITL) simulation of CLT in June 2018. Results showed that option #3, with both advisories, was associated with a lower compliance at the spot than option #1 (with just the gate advisory). Although the small sample size in this HITL simulation made the significance of this finding unclear [4], Ramp Controllers did report a higher workload and increased time pressure with both advisories than in the other condition, suggesting that it was difficult for them to keep track of both times at once [5]. An unanswered question is whether Ramp Controllers could comply with spot-only advisories presented at the gate (#2 above), as this condition did not exist in the previous simulation.

Another unanswered question is whether the layout of the ramp at the simulated airport, CLT, made it difficult to deliver aircraft to the spot in compliance with an assigned spot time. Not only does CLT have limited ramp real estate, it also has ramp areas where traffic flow is reduced to a single lane, so traffic can flow in only one direction at a time. Once an aircraft enters the flow of traffic in these ramp areas, few, if any, options exist to adjust the aircraft's rate of travel to the spot or to change the sequence in which they arrive at the spot.

III. Current Simulation

In the current simulation, which took place in April-May 2019, both the simulated airport and the conditions were changed from the 2018 HITL simulation. The simulated airport was changed from CLT to Dallas/Fort Worth International Airport (DFW), which has a larger ramp area relative to the traffic it handles than does CLT and has more easily accessible spots. The conditions included a spot-only advisory condition (option #3 below), which resulted in the following four conditions:

1) Baseline, or no metering, where the Ramp Controllers were advised to handle aircraft as they would normally,
2) Metering with instructions for Ramp Controllers to comply with gate departure advisories only (Target Off-Block Times or TOBT) within a ±2 minute window,
3) Metering with instructions for Ramp Controllers to comply with spot arrival advisories only (Target Movement Area entry Times or TMAT) within a ±5 minute window, and
4) Metering with instructions to comply with both gate departure advisories and spot advisories (TOBT+TMAT).

The displays presented to the Ramp Controllers in each of these four conditions are shown in Fig. 1.

Fig. 1. TOBT and TMAT Advisories were shown in the form of countdown timers on the new ATD-2 ramp controller decision support tool, the Ramp Traffic Console (RTC). From left to right they are 1) baseline, no advisories, aircraft information only, 2) gate advisory (TOBT) shown in cyan ("8 min"), 3) spot advisory (TMAT) shown in green ("T 16 min"), and 4) both gate and spot (TOBT+TMAT) advisories displayed.

The HITL simulation was conducted at Future Flight Central, a national facility at NASA Ames Research Center that provides a 360-degree full-scale, real-time simulation of an airport, shown in Fig. 2.

Fig. 2. Future Flight Central tower cab was used as Ramp Tower for simulation.

Fig. 3 shows the Terminal A ramp area at DFW with the location of gates and the spots. The spots are at the border between the ramp area and the airport movement area and are where Air Traffic Control takes control of aircraft.

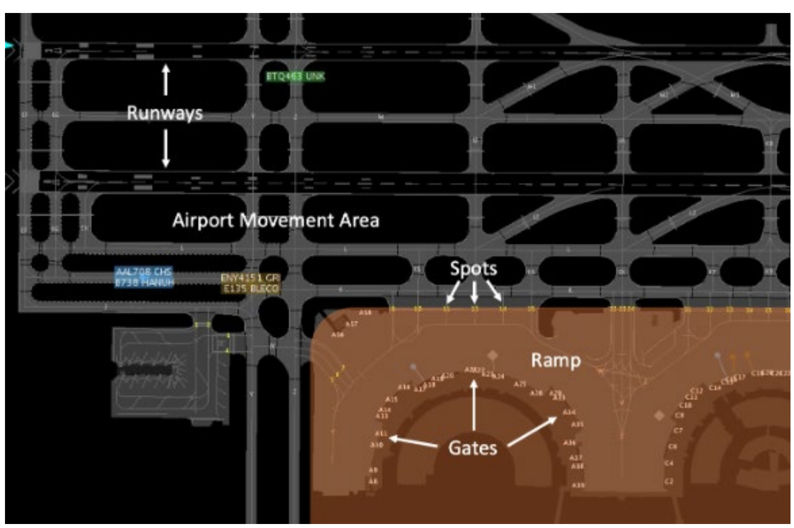

Fig. 3. Depiction of location of gates and spots in Terminal A ramp area at DFW.

Data were analyzed from twenty-eight runs, each 80-90 minutes long, with seven runs for each of the four conditions. The scenarios used actual flight operations data from Bank 2 at DFW, in south flow configuration, with an additional 15% increase in traffic to reflect projected demand at DFW. The HITL lasted two weeks with a different set of three Ramp Controllers (one each for Terminals A, C, and E) and one Ramp Manager for each week. There were therefore a total of six Ramp Controllers and two Ramp Managers, all of whom had operational experience at DFW. Four ATC Tower controllers, who also had operational experience at DFW, were the same throughout the two weeks, as were the 13 Pseudo Pilots, who communicated with controllers and controlled both departure and arrival aircraft in a different room.

In addition to quantitative metrics, the results of which are discussed in another paper at this conference (see Ref. [6]), human factor metrics were collected via the following instruments: workload self-ratings on Workload Assessment Keypads (WAK) administered every five minutes, post-run surveys including five adapted NASA Task Load Index (TLX) ratings [7], post-study surveys, end-of-day debriefs, and two end-of-study debriefs (one for each week of the HITL). On the workload ratings, the Ramp Controllers were asked to rate the workload as a "3" on a 1-5 scale as being equivalent to normal operations at DFW.

IV. Results

A. Compliance

1. At the gate

Compliance with the target gate departure advisory (±2 minutes) was similar in all TOBT-only and TOBT+TMAT conditions in both simulations at about 60%, ranging from 57.1% to 61.7%.

2. At the spot

In the DFW simulation, compliance with a target spot arrival advisory (±5 minutes) was high and similar in all three metering conditions, ranging from 89% to 91%, as shown in the second row of Table 1. (Compliance at the spot increased somewhat when aircraft were initially compliant at the gate, as was the case in the earlier HITL simulation [4]). Comparing compliance at the spot between the previous HITL at CLT and the current HITL at DFW, there are two conditions that are comparable: the TOBT-only[7] and the TOBT+TMAT condition, as shown in Table 1. In the TOBT-only condition, the compliance at the spot is nearly identical. However, in the TOBT+TMAT condition in CLT, only 69% (29/42) of the aircraft were in compliance compared to 91% (169/186) in the current DFW-based HITL ($X^2(df1)$ =14.3, $p < .001$). It is likely that the more restrictive ramp area at CLT made it more difficult to comply with the TMAT advisories, as predicted.

Table 1. Comparison of TMAT Compliance in CLT and DFW HITLs

TMAT Compliance	TOBT-Only Condition	TMAT-Only Condition	TOBT+TMAT Condition	Chi Square Sig. Level
CLT	85.3% (29/34)		69.0% (29/42)	X^2= 2.7(df 1), p = .09
DFW	88.7% (134/151)	86.8% (132/152)	90.9% (169/186)	X^2= 1.4(df 2), p = .50

B. Workload

In general, workload on the measures was rated as low. This finding is most likely linked to differences between the simulation and operations in the field and will be discussed later. Nonetheless, with nearly all measures, there was a higher workload in the combined condition, TOBT+TMAT, as was the case in the previous simulation. Fig. 4 shows an example from the post-simulation surveys administered after each week of the simulation. Ramp Controllers were asked to "Please describe your workload at the busiest times in each of the conditions in this simulation." Although the ratings are not significantly different (N = 7 ratings, repeated measures ANOVA $F(2,4)$ = 3.0, p = .16), they indicate that the workload was highest in the TOBT+TMAT condition, especially for the Ramp Controllers working traffic in Terminal A, the busiest terminal at DFW according to Ramp Controllers. The WAK data also showed low workload overall, but workload was statistically significantly higher in the TOBT+TMAT condition than in the Baseline and TOBT conditions. On the five adapted NASA Task Load Index ratings which were included in the post-run surveys, the TOBT+TMAT condition trended highest in mental demand, time pressure, and frustration, but was not statistically significantly different from the other conditions.

[7]In the TOBT-only condition, TMAT compliance was recorded, but the TMAT advisory was not seen by the Ramp Controllers.

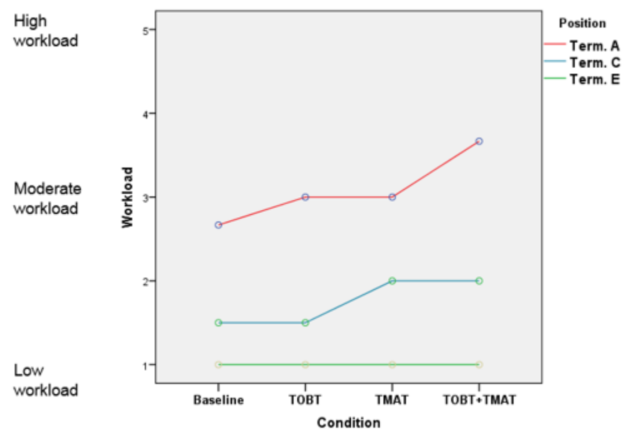

Fig. 4. Ramp Controllers' average self-ratings of their workload by condition at the end of each week of simulation. Terminal A had the highest workload, especially in the TOBT+TMAT condition.

Both Ramp Managers also rated the Ramp Controllers' workload as highest in the TOBT+TMAT condition on the post-simulation surveys, as shown in Fig. 5. However, they differed from the Ramp Controllers' own assessment by rating the Ramp Controllers' workload in this condition as "Very high." (The Ramp Managers rated their own workload as 2.5 in both Baseline and TOBT-only, 3 in TMAT-only, and 3.5 in TOBT+TMAT.)

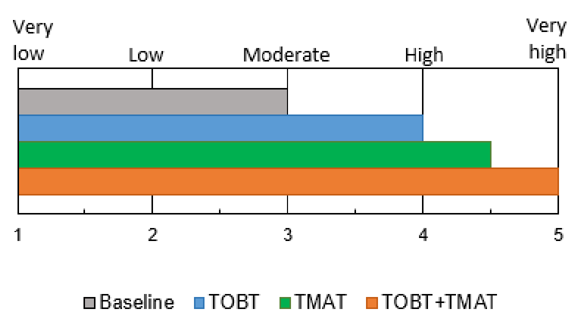

Fig. 5. Ramp Managers' average ratings of Ramp Controllers' workload in each condition at the end of each week of simulation (N = 2).

Example comments from a Ramp Manager and a Ramp Controller on the difficulties inherent in the dual advisory condition (TOBT+TMAT) were:

- "It could be confusing, when watching both advisories. The level of success would be greater, when being concerned about 1 at time." (Ramp Manager)
- "Too much traffic in DFW to be able to accommodate for [both] gate and spot advisories. With the spot advisories having a +/-5 min window, it helps with hitting the target. Holding on the gate would be more ideal in order to achieve a spot advisory than holding on the ramp after pushback." (Ramp Controller, Terminal C)

C. Situation Awareness (SA)

Ramp Controllers' average situation awareness, as measured by an adapted version of the 3D Situation Awareness Technique [8] based on self-ratings on post-run surveys, was statistically significantly lower in the TOBT+TMAT condition than it was in the Baseline condition (paired comparisons, M = 6.95 and 5.85, p = .05, Ns = 21 and 20, 1 missing data point), as shown in Fig. 6. The other two metering conditions were not significantly different from the Baseline or from the TOBT+TMAT condition.

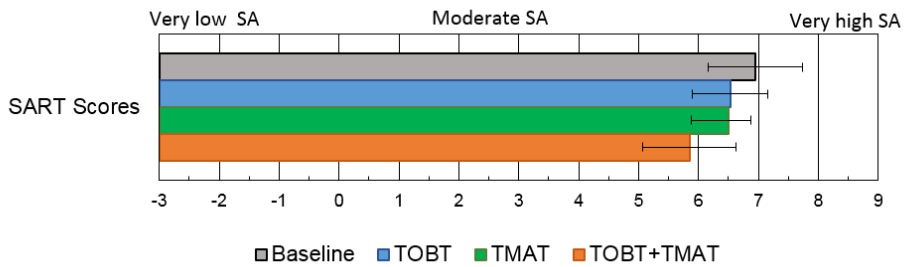

Fig. 6. Ramp Controller average post-run survey scores on the Situational Awareness Rating Technique (SART). *N*s = 20-21 in each condition. Error bars are 95% CIs.

D. Processes Ramp Controllers Used to Handle Advisories

Post-run surveys questions were designed so that Ramp Controllers could rate how frequently, on a five-point scale, they reacted to the particular hold advisories in the previous run. In all three metering conditions average ratings (21 per condition) indicated that the Ramp Controllers:

- Usually responded to (acted on) all hold advisories (averages between 4 "Usually" and 5 "Always"),
- Usually achieved all hold advisories (averages between 4 "Usually" and 5 "Always"),
- Did not feel it was difficult to achieve all hold advisories (averages between 1 "Very easy" and 2 "Slightly difficult"), and
- Did not feel it was disruptive to traffic flow in their ramp sector to hold aircraft at the gate to achieve all hold advisories (averages between 1 "Not at all disruptive" and 2 "Slightly disruptive").

It appears from the quantitative metrics that Ramp Controllers held more aircraft at the gate in the TMAT-only condition than in the other conditions, and for a longer period of time (see Ref. [6]). Also, as shown in Fig. 7, the Ramp Controllers reported that they held more aircraft at the gate in the TMAT-only condition, when gate advisories were not shown, than in the TOBT+TMAT condition ($F(1,40) = 4.7$, $p = .032$). These ratings were in response to the post-run survey question in the TMAT conditions, "How often did you do the following to achieve spot advisories for aircraft in this run?"

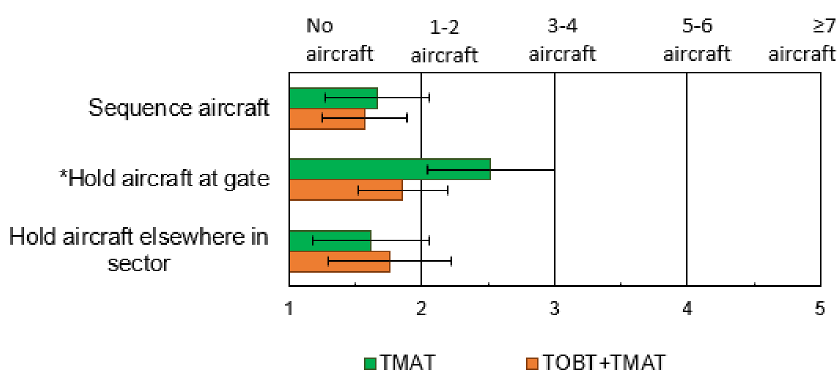

Fig. 7. Ramp Controllers' descriptions of the actions they took to achieve spot advisories in the TMAT conditions. *N* = 21 ratings in each condition. Error bars = 95% CIs.

The higher number of aircraft held at the gate in the TMAT-only condition may have been due to the Ramp Controllers' desire to comply promptly with the TOBT advisory whenever it appeared in the other two conditions, i.e., to push back from the gate within ±2 minutes. Despite holding more aircraft at the gate and for a longer period of time in the TMAT-only condition, the compliance at the spot was similar to the compliance in other conditions, as shown in Table 1.

E. Operational Efficiency

On the post-run surveys, Ramp Controllers were asked to rate the acceptability of several aspects of operational efficiency during the busiest time in the previous run, as shown in Fig. 8. Although there were no significant differences between conditions, and all the conditions showed ratings that averaged 4 or higher, the trend was that the Baseline and TOBT-only conditions were seen as the most efficient and the two TMAT conditions were seen as least efficient.

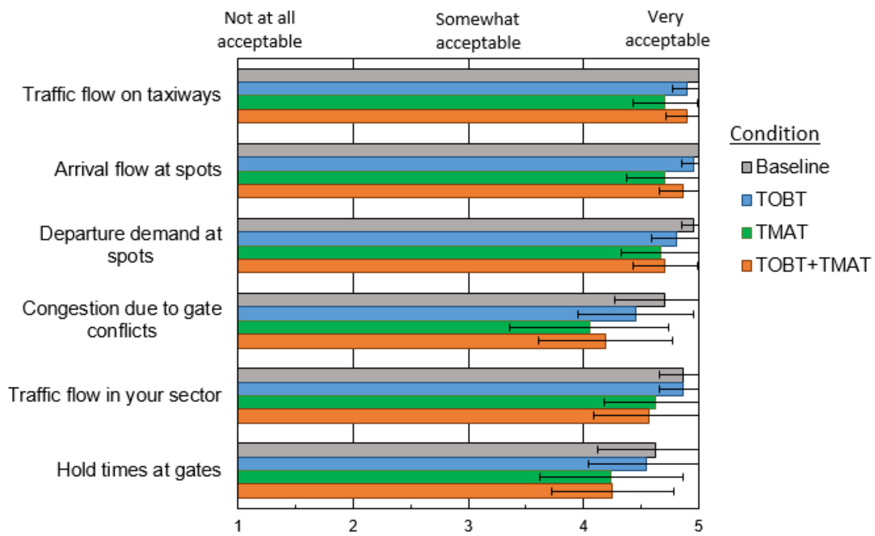

Fig. 8. Ramp Controllers' ratings of the acceptability of operational efficiency at the busiest times post-run. *N*s = 19-21 ratings for each item in each condition. Error bars = 95% CIs.

V. Discussion

Although there were differences in workload between conditions, with higher workload in the TOBT+TMAT condition, the Ramp Controllers rated their workload overall as low on the WAK and other measures. This was the case even though the Ramp Controllers in Terminals A and C appeared to have a workload that was moderate to high in all conditions to outside observers and to the Ramp Managers (see Fig. 5).

As shown in Fig. 9, the Ramp Controllers rated the workload in the simulation as statistically significantly less realistic than other aspects of the simulation, such as the airport, the ramp, and tools available (Repeated Measures ANOVA, $F(6,36) = 3.1$, $N = 7$, $p = .015$).

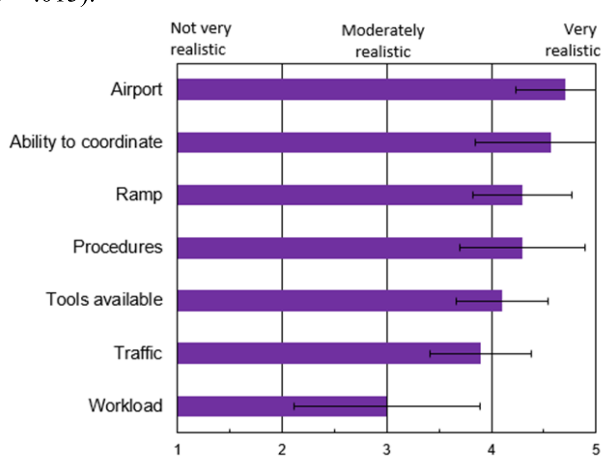

Fig. 9. Ramp Controllers' ratings of simulation realism at the end of each week of the simulation. Error bars = 95% CIs, Loftus correction applied [9].

When the Ramp Controllers were asked in the debriefs why they felt the workload was less realistic, and why they rated the workload as low in the simulation when it appeared to be high, the Ramp Controllers responded that the workload they were used to in the real world at DFW was much higher.

Many of the factors which they felt lowered the workload in the simulation were due to simulation needs of a) having each pseudo-pilot control many aircraft instead of one, and b) having only one radio frequency available for each pseudo-pilot. This resulted in some changes to typical procedures, e.g., the pseudo-pilot could not monitor two frequencies—ramp and ground upon landing, as is typically done in a two-person cockpit. Examples of areas in which these different procedures may have lowered the Ramp Controller workload were:

1. In the simulation, Ramp Controllers did not get a call from pilots immediately after landing to check on gate availability, as happens in the field. After landing, pilots were talking only to Ground Control in the Tower. Pilots instead assumed that the gate assigned to them was open and first contacted ramp on frequency when near the ramp area.
2. In case of a gate conflict, instead of the Ramp Controller having to communicate the conflict to the pilot, who would then communicate with Ground to request that the aircraft be held in the Airport Movement Area (AMA), a Ramp Manager position was created and s/he used the software to do this, after consulting with the Ramp Controller. The Ramp Manager also used software to communicate with Ground to release the aircraft from the hold in the AMA when the gate became available. Ramp Controllers rated the managing of gate conflicts as "Very easy" in all conditions since, as one Ramp Controller wrote, "we did not have to deal with them as we would in real life" (Terminal A). Another wrote,

   > "In normal ops, the amount of workload is low enough to where I could handle gate conflicts and communicate to other controllers on my own, but as responsibility for gate and spot advisories increases, the busier I become and the amount of workload increases to an uncomfortable point if I didn't have a Ramp Manager" (Terminal A).
3. Pilots called their Ramp Controller on the radio frequency one-by-one instead of swamping the frequency with many calls at a time, as is typical at DFW.

Finally, as mentioned elsewhere [6], anomalies that frequently occur in real life are eliminated in simulations so as to be able to compare conditions fairly. These anomalies include baggage dropped in the ramp area, problems with tugs, late passengers, aircraft maintenance, and pilot scheduling issues. The relatively "normal" operations in the simulation also undoubtedly contributed to the lower workload Ramp Controllers experienced in the simulation.

VI. Conclusion

A goal of this investigation was to compare directly the impact of several different advisories on Ramp Controller performance. The results indicate that due to unavoidable limitations in the ability of the simulation to duplicate conditions at the airport, and to the need to develop some work-arounds in the simulation environment to mimic these conditions, the workload was lower than expected in all conditions. Nonetheless, from a human factors perspective, the results suggest that metering at Dallas/Fort Worth International Airport (DFW) could be effective with either gate-only or spot-only advisories, and either would be a better option than both advisories due to the Ramp Controllers' increased workload and reduced situation awareness observed in that condition. Compliance at the spot was similar in all metering conditions. Compared to the other metering conditions, spot advisories alone had more aircraft wait for a longer period of time *at the gate* instead of elsewhere—an advantage if the goal is to reduce wait time elsewhere with its attendant fuel burn and CO_2 emissions. The Ramp Controllers rated operational efficiency at various locations as acceptable in all conditions.

The results also suggested that ramp configuration might be a factor in determining which advisories work best at which airport, with spot advisories more likely to be complied with at airports with less restricted ramp areas. This needs to be followed up with research comparing spot-only advisories at restricted vs. non-restricted ramps like DFW, a comparison not made in this study.

Finally, the creation of the Ramp Manager position who handled gate conflicts via software with Ground Control suggests an area for future development, as it was appreciated by Ramp Controllers, especially in the metering conditions.

References

[1] Jung, Y., Engelland, S., Capps, A., Coppenbarger, R., Hooey, B., Sharma, S., Stevens, L., and Verma, S., "Airspace Technology Demonstration 2 (ATD-2) phase 1 Concept of Use (ConUse)," 2018.

[2] FAA Air Traffic Organization Surface Operations Directorate, U.S. Airport Surface Collaborative Decision Making Concept of Operations (ConOps) in the Near-Term: Application of the Surface Concept at United States Airports, June 2012.

[3] Lee, H., et al., "ATD-2 Benefits Mechanism," April 2020. See also Coupe, J. "Benefits of Surface Departure Metering While 'Doing no Harm' to Other Operational Metrics," *Airspace Technology Demonstration 2 (ATD-2) Industry Workshop, Sept. 4, 2019*. URL: https://www.aviationsystemsdivision.arc.nasa.gov/atd2-industry-workshop/presentations/4C_ATD2_Benefits_Industry_Day_FINAL.pdf

[4] Stevens, L. K. S., Jung, Y. C., Coupe, W. J., Lee, H., Parke, B. K., "Assessment of Ramp Times (ART 2) Human-in-the-Loop (HITL) Simulation – Final Results," March 2019. https://aviationsystemsdivision.arc.nasa.gov/publications/atd2/tech-transfers/4_Evaluation_Results/4.09%20ATD2_ART2_HITL_Results_2019-03-06.pdf

[5] Parke, B., Stevens, L. K. S., Coupe, W. J., Lee, H., Jung, Y. C., Bakowski, D. L., and Jobe, K., "Alternatives for Scheduling Departures for Efficient Surface Metering in ATD-2: Exploration in a Human-in-the-Loop Simulation," Paper presented at the *10th International Conference on Applied Human Factors and Ergonomics,* Washington, D. C., July 2019.

[6] Lee, H., Coupe, W. J., Jung, Y. C., Stevens, L. K. S., Parke, B. K., and Bakowski, D. L., "Objective Measurement Assessment of Departure Advisories for Ramp Controllers from Human-in-the-Loop Simulation." Paper presented at the *AIAA 2020 Aviation Forum*, Reno, Nevada, June 2020.

[7] Hart, S. G., and Staveland, L. E., "Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Human Mental Workload*, edited by P. S. Hancock and N. Meshkati, Elsevier Science Publishers, Amsterdam, 1988.

[8] Taylor, R. M., "Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design," *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations*, *CP478*. Seuilly-sur Seine: NATO AGARD, 1989.

[9] Loftus, R. L., and Masson, M. E. J. "Using Confidence Intervals in Within-Subject Designs." *Psychonomic Bulletin & Review*, Vol. 1 (4), 1994, pp. 476-490.