

# Assessment of Ramp Times 4 (ART-4) Human-in-the-Loop (HITL) Simulation

## Final Results

September 2019

- Airspace Technology Demonstration 2 (ATD-2) sub-project conducted a Human-in-the-Loop (HITL) simulation to assess Ramp Controllers ability to deliver aircraft to the spot within the compliance window (+/- 5 min) under various metering conditions
- Compliance at the spot was similar between the different metering conditions ranging between 83% - 85% and increased to 92% - 99% when aircraft were initially compliant with gate advisories
- Metering benefits that exist in the field did not appear in the simulation due to simulation artifacts such as gate holding departures in Baseline, which effectively metered the demand
- The combined Target Off-Block Time (TOBT) +Target Movement Area entry Time (TMAT) condition resulted in higher workload on the Workload Assessment Keypad (WAK) than the Baseline and TOBT alone conditions, and lower situation awareness than the Baseline condition
- Metering at Dallas/Fort Worth International Airport (DFW) with TOBT only or TMAT only could be equally effective, and either would be a better option than TOBT + TMAT due to increased workload and reduced situation awareness



- Background
- Simulation conditions and guidance
- Simulation design
- Metrics
- Research questions
- Results
- Conclusion
- Appendix A: Other Performance Metrics (Slide 80)
- Appendix B: Other Subjective Metrics (Slide 105)

# Background



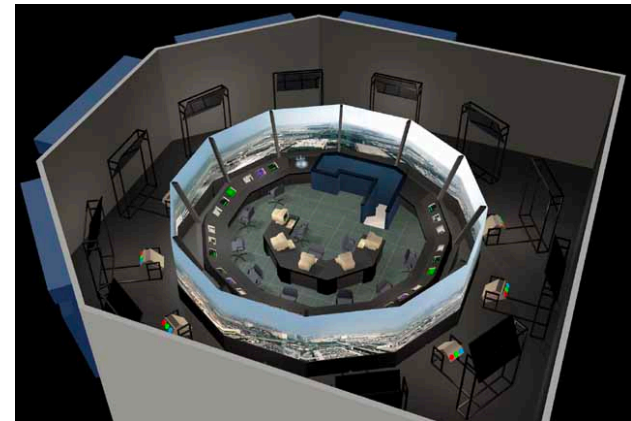
- Surface Collaborative Decision Making (S-CDM) Concept
  - During a Surface Metering Program (SMP), aircraft absorb surface delay in the ramp area, ideally at the gate, instead of at the runway departure queue
    - Saves fuel and lowers carbon dioxide emissions
    - Would improve schedule predictability elsewhere on surface
  - Aircraft must be delivered to the “spot” where Air Traffic Control (ATC) takes control of the aircraft within compliance of a specific time
- Airspace Technology Demonstration 2 (ATD-2) field demo at Charlotte Douglas International Airport (CLT) has shown that these savings occur when Ramp controllers release aircraft at a specific gate pushback time
  - This time is determined by a surface scheduler during an SMP
  - Initial assessment of savings due to surface metering: saved 1.4million pounds of fuel and 4.4million pounds of CO<sub>2</sub> until the end of August, 2019
- The current HITL simulates Dallas/Fort Worth International Airport (DFW)

# HITL Conditions and Guidance

National facility at NASA Ames Research Center which provides a 360-degree full-scale, real-time simulation of an airport



FFC Tower Cab (used as Ramp Tower for ART-4 HITL )





- Test new ATD-2 Ramp tools/features before field deployment in DFW
- Evaluate the impacts of various metering goals on Ramp operations at DFW





- **Baseline** Ramp Controllers instructed to operate as they would in normal, current-day operations
- **TOBT Compliance** During metering, Ramp Controllers instructed to focus on ensuring that flights push from the gate within  $\pm 2$  min. of the Target Off-Block Time (TOBT) presented by the advisory
- **TMAT Compliance** During metering, Ramp Controllers instructed to deliver flights to the spot at their Target Movement Area entry Times (TMAT) within  $\pm 5$  min
- **TOBT + TMAT Compliance** During metering, Ramp Controllers instructed to pushback flights in compliance with TOBT  $\pm 2$  min and to deliver flights to the spot at their TMAT within  $\pm 5$  min



Baseline  
(No advisories)

AAL1332 B738 C  
THHOR TPA  
C10 22 17R 2130

**TOBT (Gate)  
Advisories**  
(+2/-2 minute  
compliance window)

8 min

AAL1332 B738 C  
THHOR TPA  
C10 22 17R 2130

**TMAT (Spot)  
Advisories**  
(+5/-5 minute  
compliance window)

T 16 min

AAL1332 B738 C  
THHOR TPA  
C10 22 17R 2130

**TOBT (+2/-2) plus  
TMAT(+5/-5)  
Advisories**

8 min

T 16 min

AAL1332 B738 C  
THHOR TPA  
C10 22 17R 2130



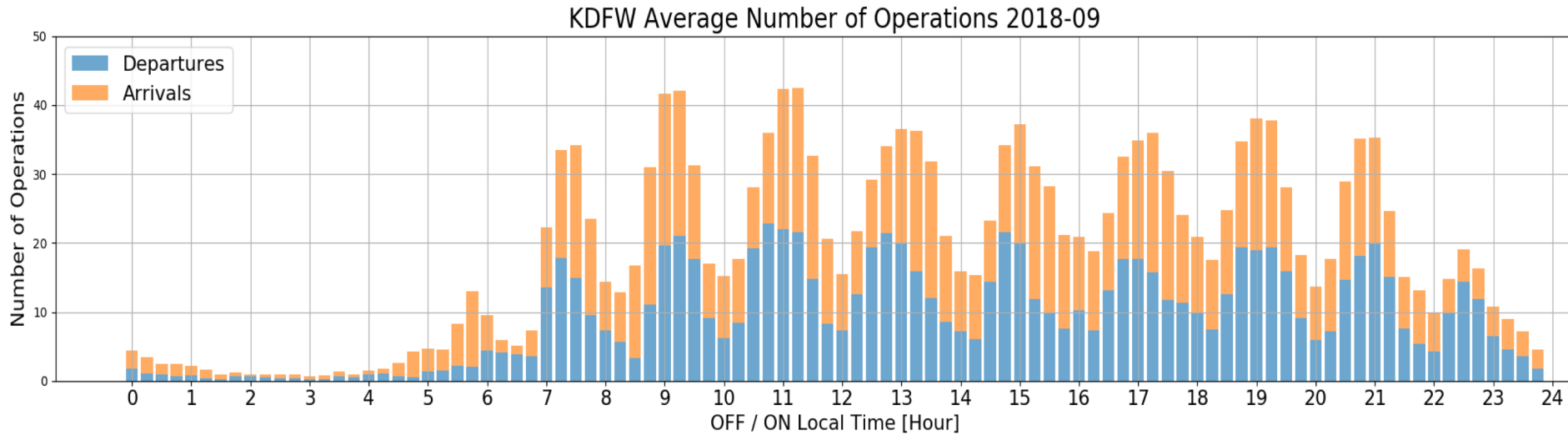
- Ramp Controllers were told to use best judgement and company policy for determining hold procedures at a gate
  - For example, if a departure flight with a TOBT gate hold advisory had a gate conflict with an arrival flight, a Ramp Controller might push the departure off the gate early to free the gate for the arrival
  - Spot assignments could be changed for a flight
- Other HITL participants such as ATC Tower Ground and Local Controllers were instructed to operate as they would in normal, current-day operations.

# Simulation Design



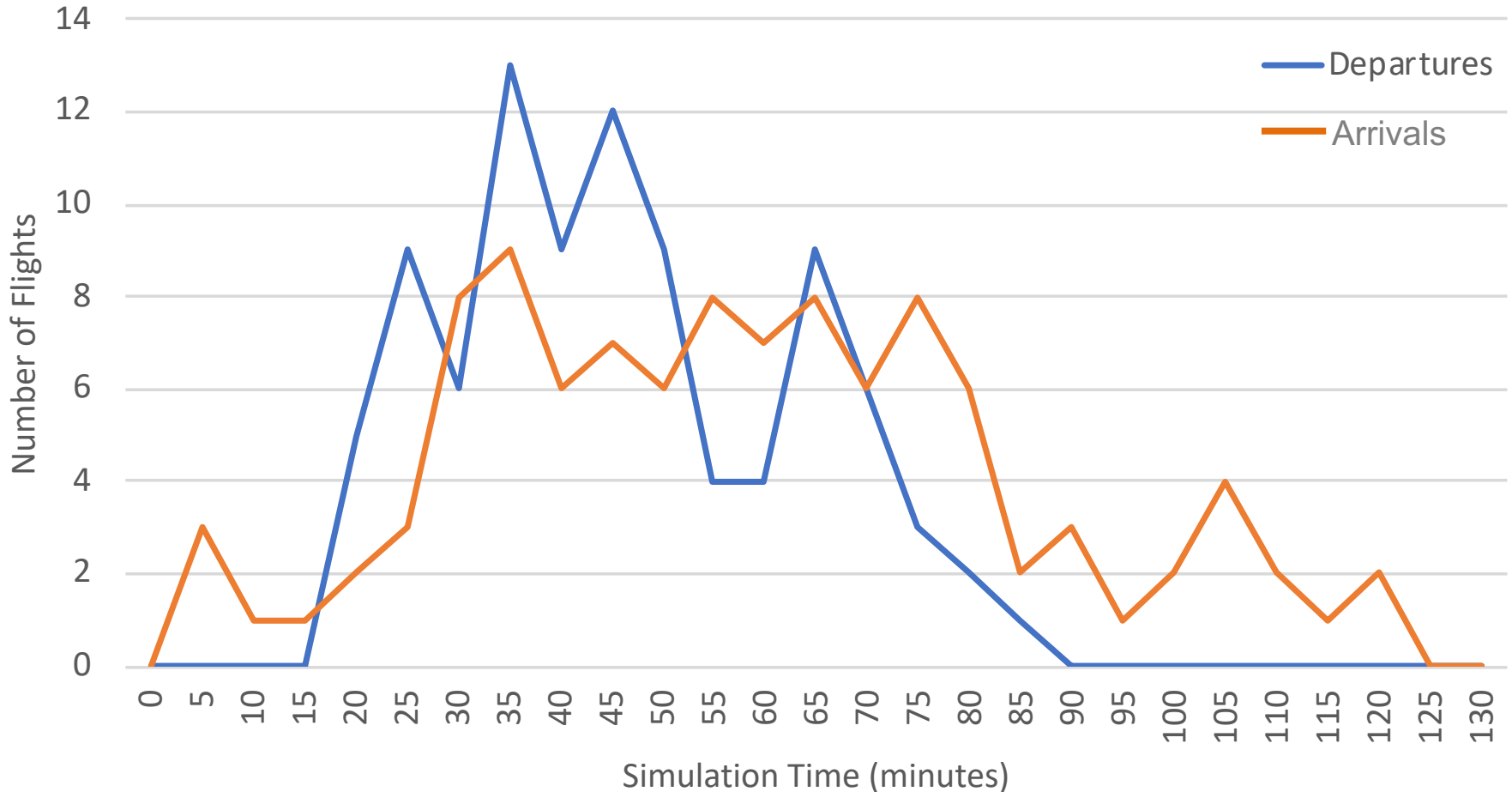
- Used the simulation output data from a total of 28 runs for data analysis
  - Five different scenarios
  - Seven runs for each metering condition
- Grouped the analysis results by metering condition
  - Quantitative metrics
  - Qualitative metrics

Condition	Run #	Remark
Baseline	02, 04, 08; 20, 25, 28, 31	1 run for Scenario #1, #2, and #3; 2 runs for Scenario #4 and #5
TOBT only	06, 11, 14; 19, 22, 26, 29	
TMAT only	03, 07, 16; 17, 21, 23, 27	
TOBT+TMAT	05, 10, 12; 18, 24, 30, 32	



Used real traffic from 80-90 minutes of DFW Bank 2 traffic to build HITL scenarios.

HITL Scenario Demand Profile - 8/12/2018





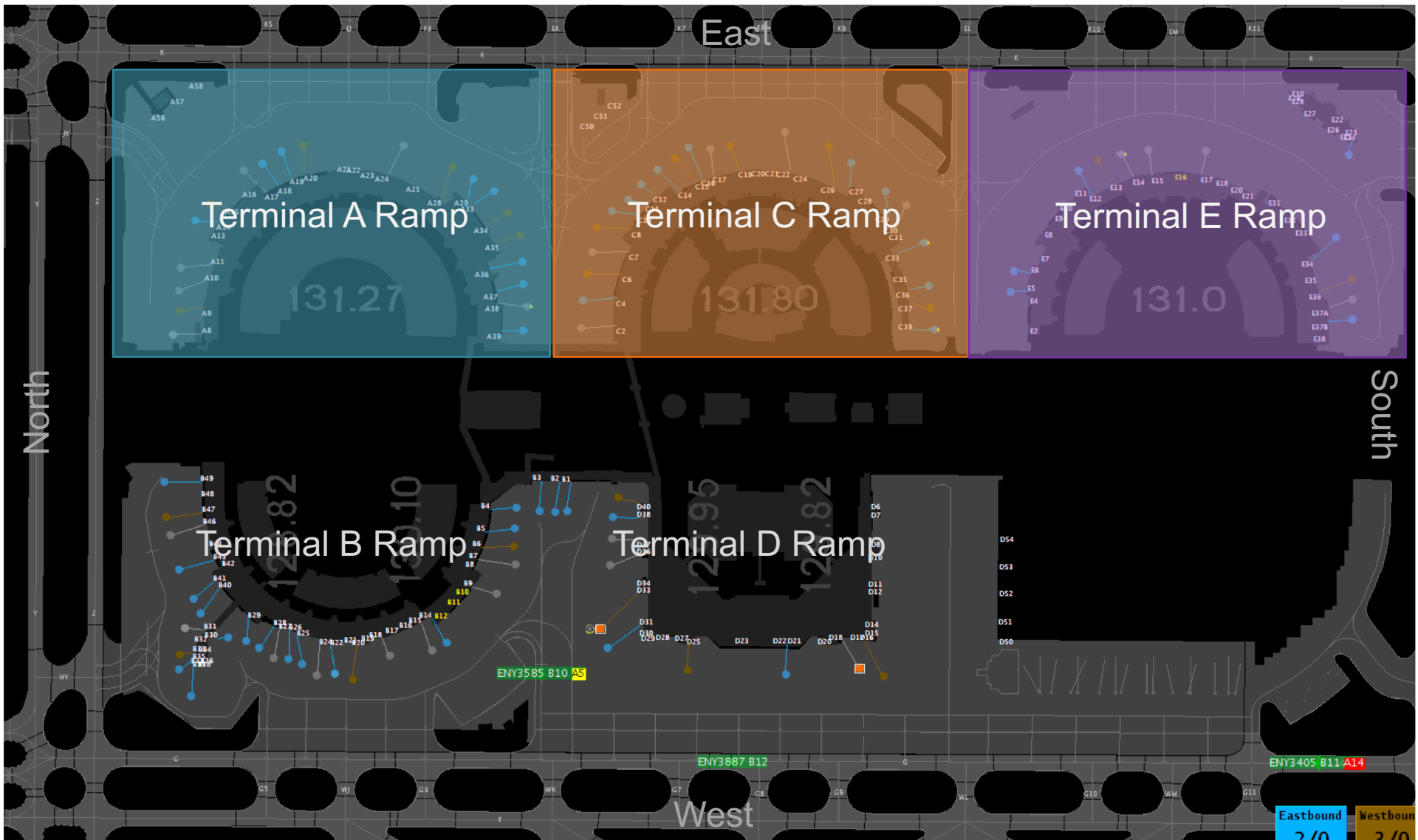


- Scenarios
  - Clear weather, no wind
  - South operations
  - No compression or decompression
  - Duration:
    - Scenario: 130 min
    - Run: 80-90 min
  - GA flights: 4-7
  - Cargo flights: 1-8
  - Example (Scenario #1): Total No. of Flights: 208
    - Departures: 91
    - Arrivals: 117
  - Traffic was increased by 15%, reflecting projected increases in live traffic operations at DFW



- Traffic Management Initiatives (TMIs)
  - EDCTs: 1 flight
- Built-in scenario uncertainties
  - Pseudo Pilots' time to respond to pushback clearances from Ramp Controllers
  - Taxi speeds
  - Time between pushback and engine spooling
- Surface metering values
  - Predetermined at 8 min for target excess queue time





Note: Gates C2 through C12 are Ramp A controller's responsibility

## Participants

Count	Participant	Qualification
1	Ramp Manager	Operational AAL DFW Ramp
2	Ramp Controllers	Operational AAL DFW Ramp
1	Ramp Controllers	Operational DFW Airport Ramp
1	ATC-Tower TMC	Operational DFW ATC-Tower
4	ATC-Tower Controllers	Retired DFW ATC-Tower

## Confederates

Count	Participant	Qualification
8	Ghost East Airport Pseudo-pilots	ATG (simulation software) experience
4	Ghost West Airport pseudo-pilots	ATG experience
1	Ghost controller (ATC & Ramp)	RTC/RMTC experience

# Metrics

- Taxi out/in time in the Airport Movement Area (AMA) and Ramp
- Excess taxi out/in time (AMA and Ramp)
- Gate pushback hold
- TOBT compliance
- TMAT compliance
- Surface counts (AMA, departure queues and Ramp)
- Runway throughput



- During each run
  - Workload Assessment Keypad (WAK) tablets collected workload ratings on a 1-5 scale every 5 minutes
  - Audio and video recording of each station
- Post-run surveys
  - Workload ratings via 5 NASA Task Load Index (TLX) items
  - Situation Awareness (SA) ratings via adapted 3-D Situational Awareness Rating Technique (SART)
  - Acceptability ratings
  - Operational efficiency ratings
- Post-study survey & debriefs



- Five different traffic scenarios
  - Different demand profiles and runway balance determined when and whether metering turned on
- Variable pushback direction given by Ramp Controllers
- Variable Ramp Controller-initiated spot changes
- “FAA Holds” for arrivals expecting a gate conflict
  - Simulation work-around
- Local Tower Controller variation
  - Tactical takeoff sequencing from multiple departure queues
  - Departure runway separation times
  - Arrival runway crossing times at Runway 18L
  - 18L departure runway separation times affected by Arrival Departure Window (ADW) restriction

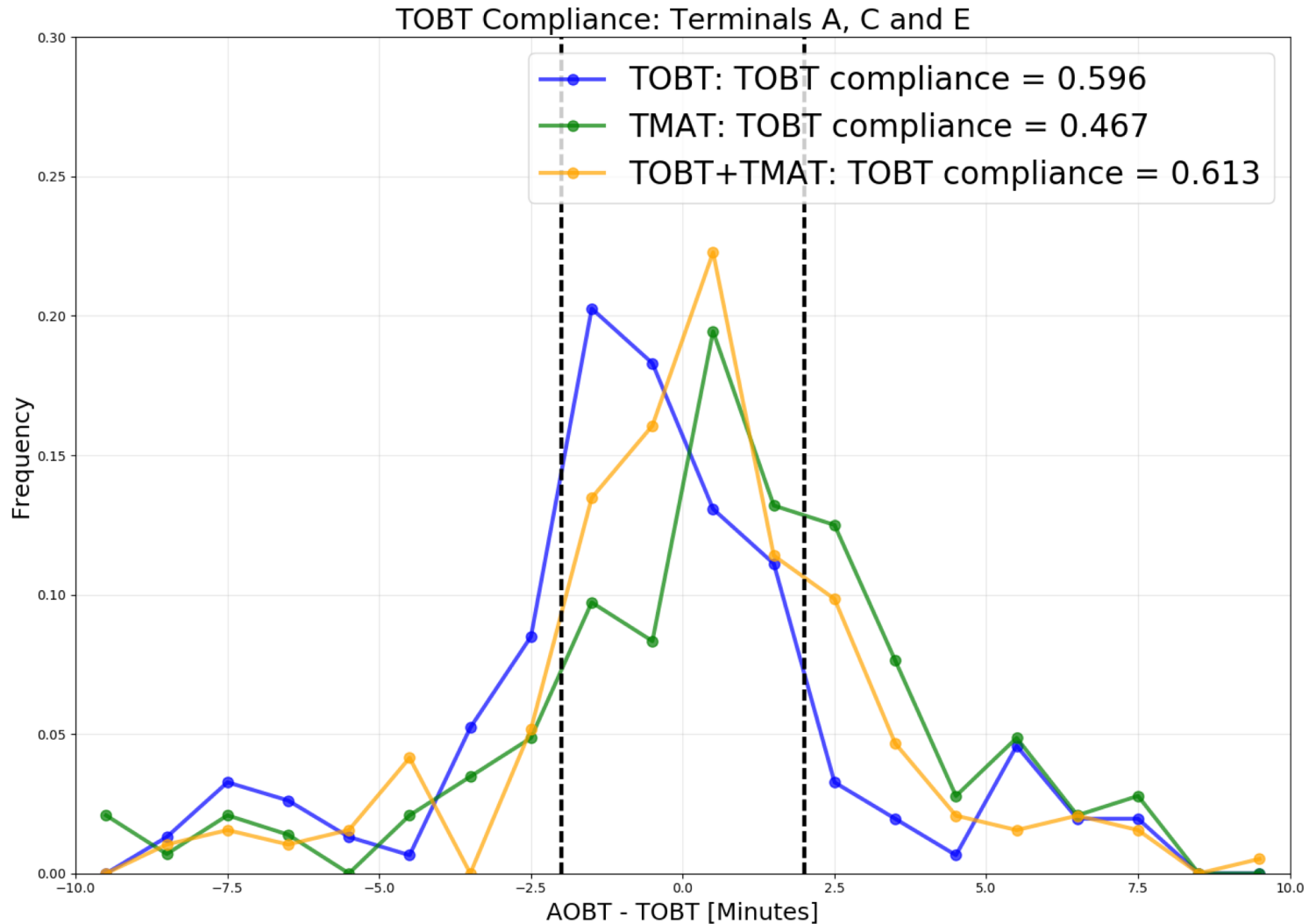
# Research Questions

- **Compliance with pushback advisories (TOBTs) and spot times (TMATs)**
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- **Were the metering performance benefits that were observed in the field also observed in the HITL?**
- **What factors reduced the ability to measure the benefits of metering?**
- **What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?**
- **What were the processes Ramp Controllers used to meet TOBTs and TMATs?**
- **Were there differences between conditions regarding perceived operational efficiency?**
- **What were the main suggestions from Ramp personnel on improving RTC?**
- **What can we conclude?**

# Results

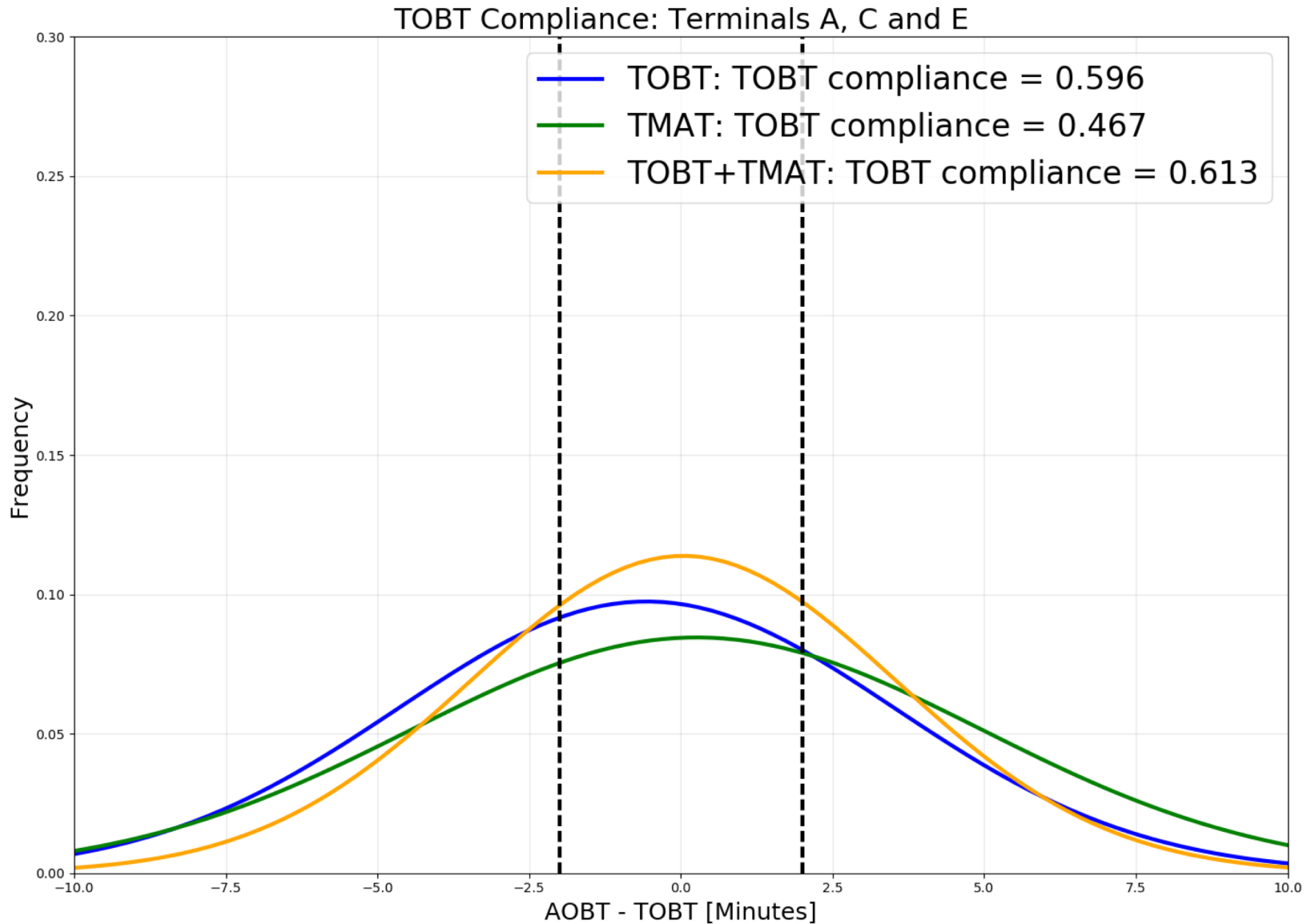
- **Compliance with pushback advisories (TOBTs) and spot times (TMATs)**
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- Were the metering performance benefits that were observed in the field also observed in the HITL?
- What factors reduced the ability to measure the benefits of metering?
- What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?
- What were the processes Ramp Controllers used to meet TOBTs and TMATs?
- Were there differences between conditions regarding perceived operational efficiency?
- What were the main suggestions from Ramp personnel on improving RTC?
- What can we conclude?

- **Compliance with pushback advisories (TOBTs) and spot times (TMATs)**
  - **Were the Ramp Controllers able to comply with the TOBTs?**
    - In the TOBT only and TOBT+TMAT conditions, Ramp Controllers' compliance with the TOBT was similar at 60% and 61%, respectively.
  - **Were the Ramp Controllers able to comply with the TMATs?**
    - In the TOBT only, TMAT only, and TOBT+TMAT conditions, Ramp Controllers' compliance with the TMAT was similar at 83%, 85%, and 85%, respectively. Compliance at the spot increased to 93%, 99%, and 92%, when aircraft were initially compliant at the gate.
  - **How did the conditions affect the timing of compliances at the spot?**
    - Overall compliance at the spot was similar, but the data showed a tendency for Ramp Controllers to release aircraft at the gate earlier in the TOBT only and TOBT+TMAT conditions compared to the TMAT only condition.



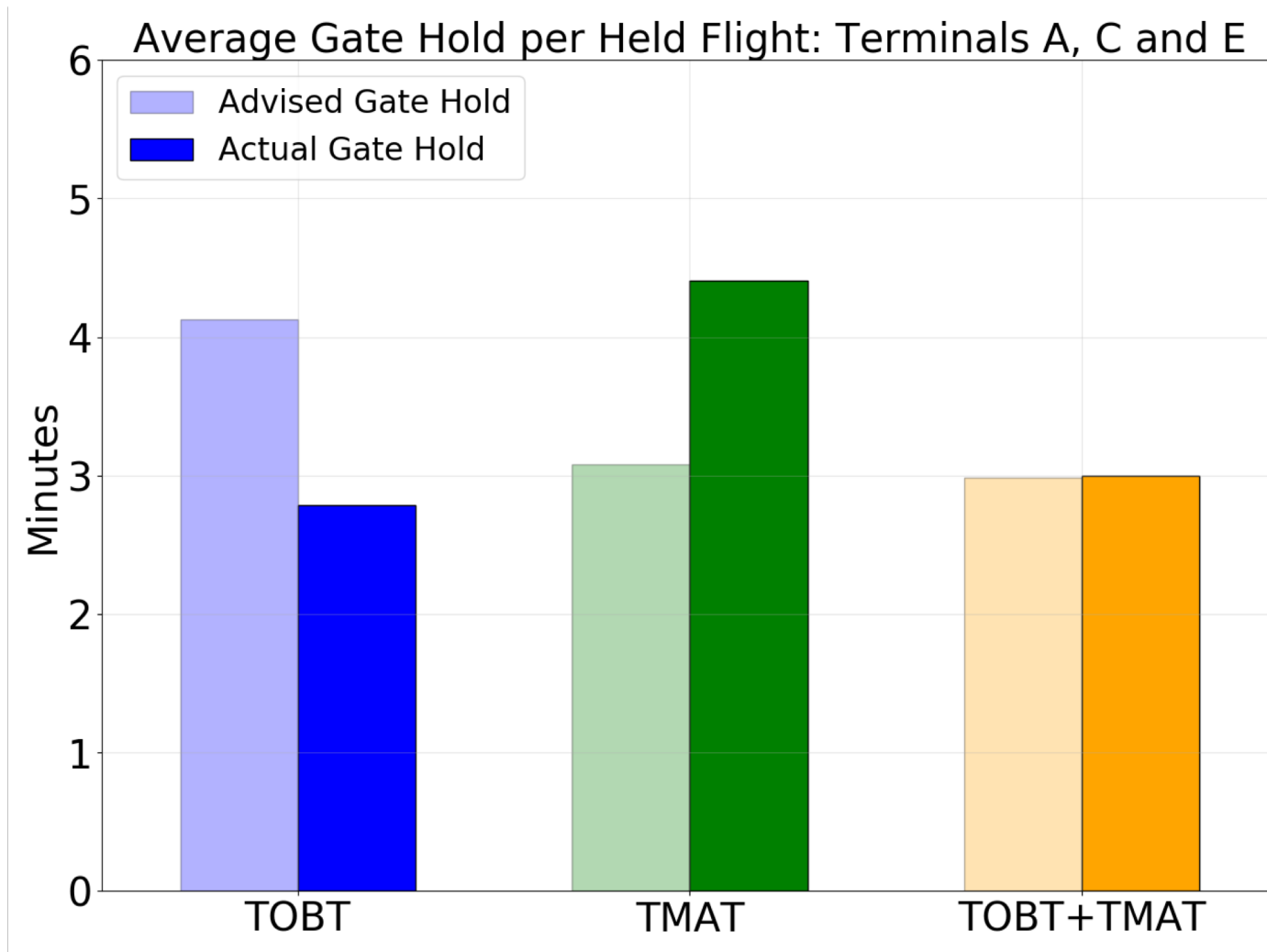
TOBT only and TOBT + TMAT conditions showed similar TOBT compliance (60% & 61%) which was higher than in the TMAT only condition (47%). For TMAT only condition there was no TOBT advisory displayed to ramp controllers but in the remaining analyses we measure against the TOBT advisory that would have been generated by the system.

# TOBT Compliance: Terminal A, C and E Fitted to a Normal Distribution

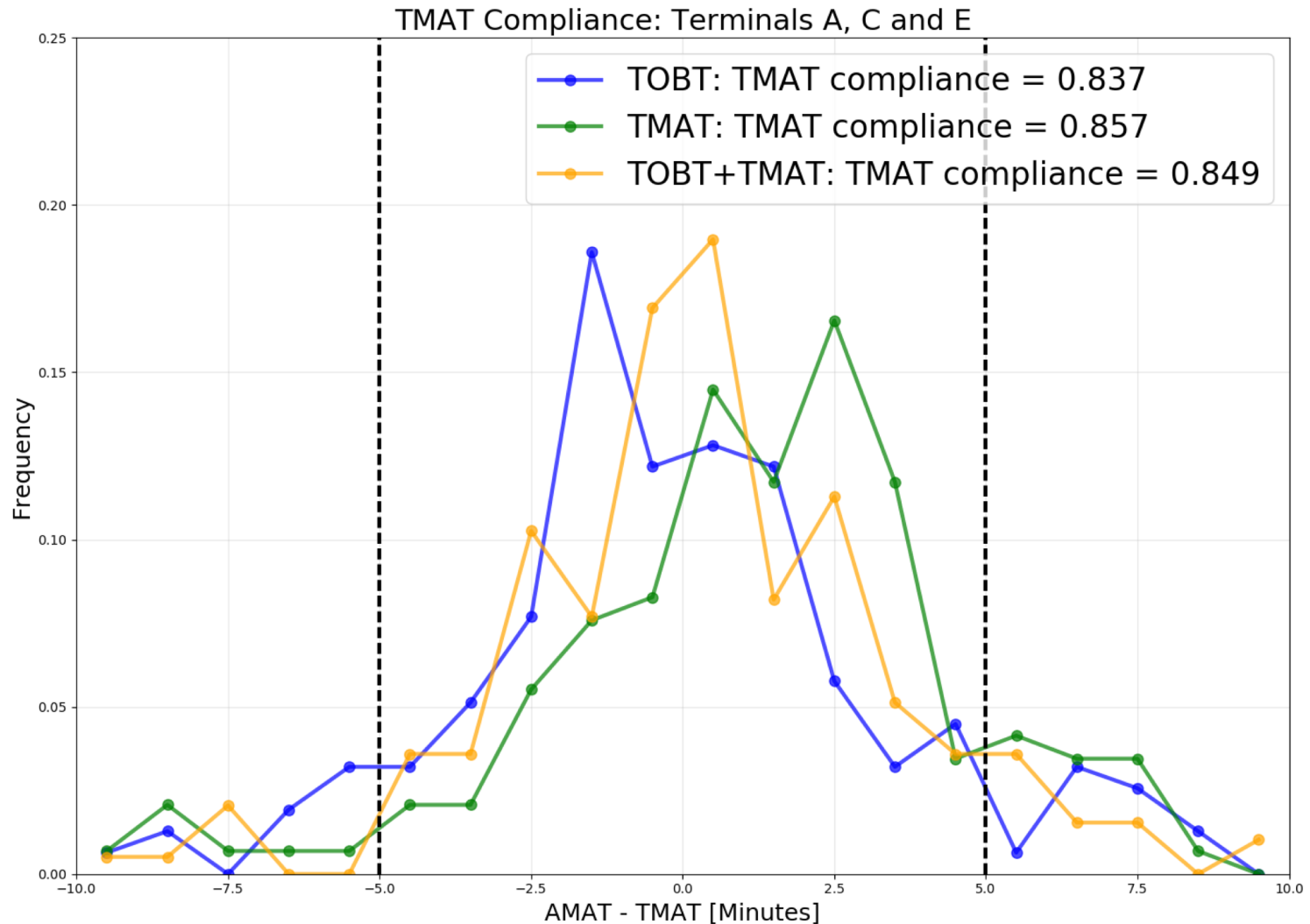




# Average Minutes of Recommended and Actual Gate Holds by Condition: Terminals A, C and E

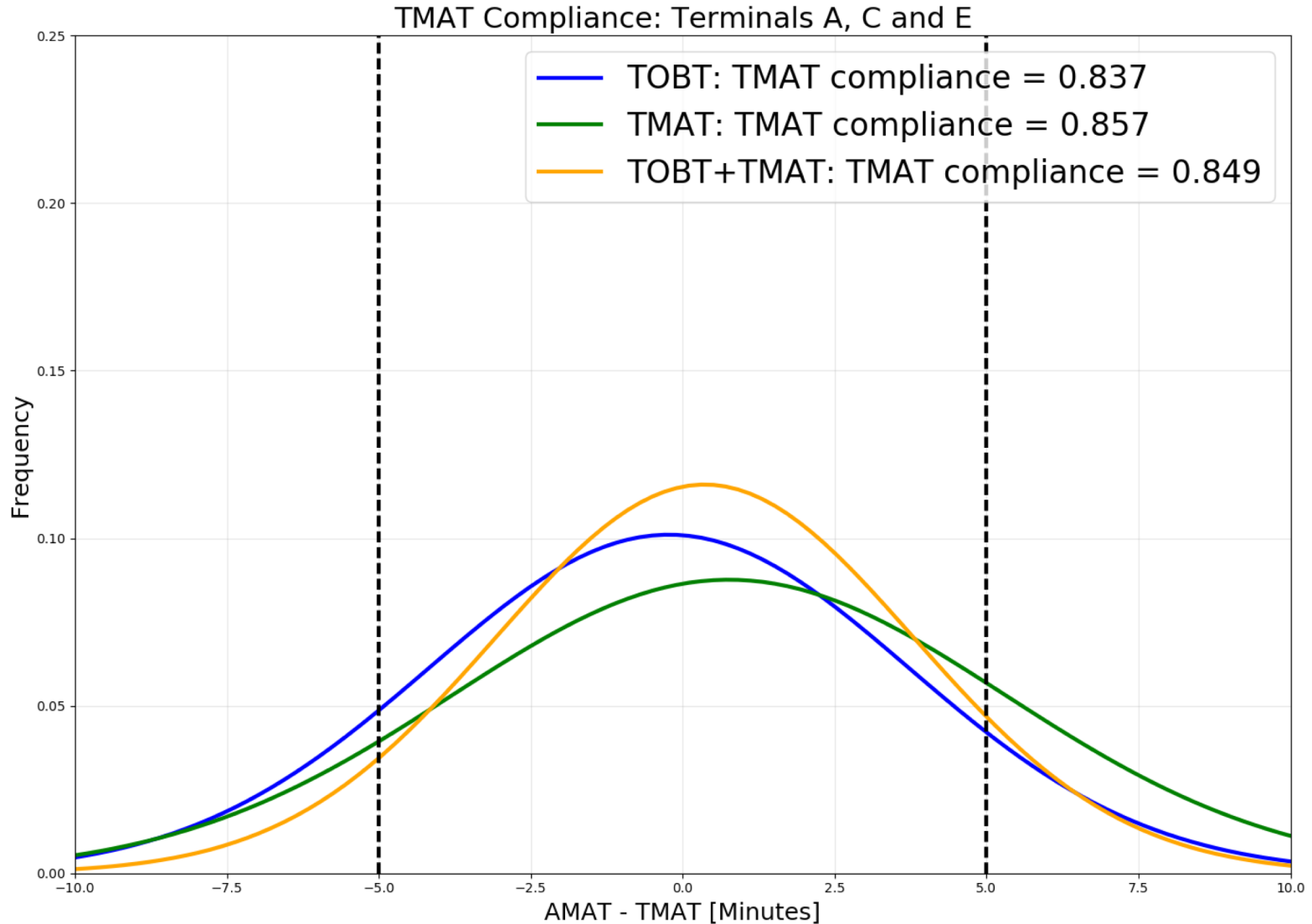


Ramp Controllers tended to release aircraft earlier than the gate advisory in the TOBT only condition and later than would be recommended for them to meet the  $\pm 5$  minute TMAT advisory in the TMAT only condition.

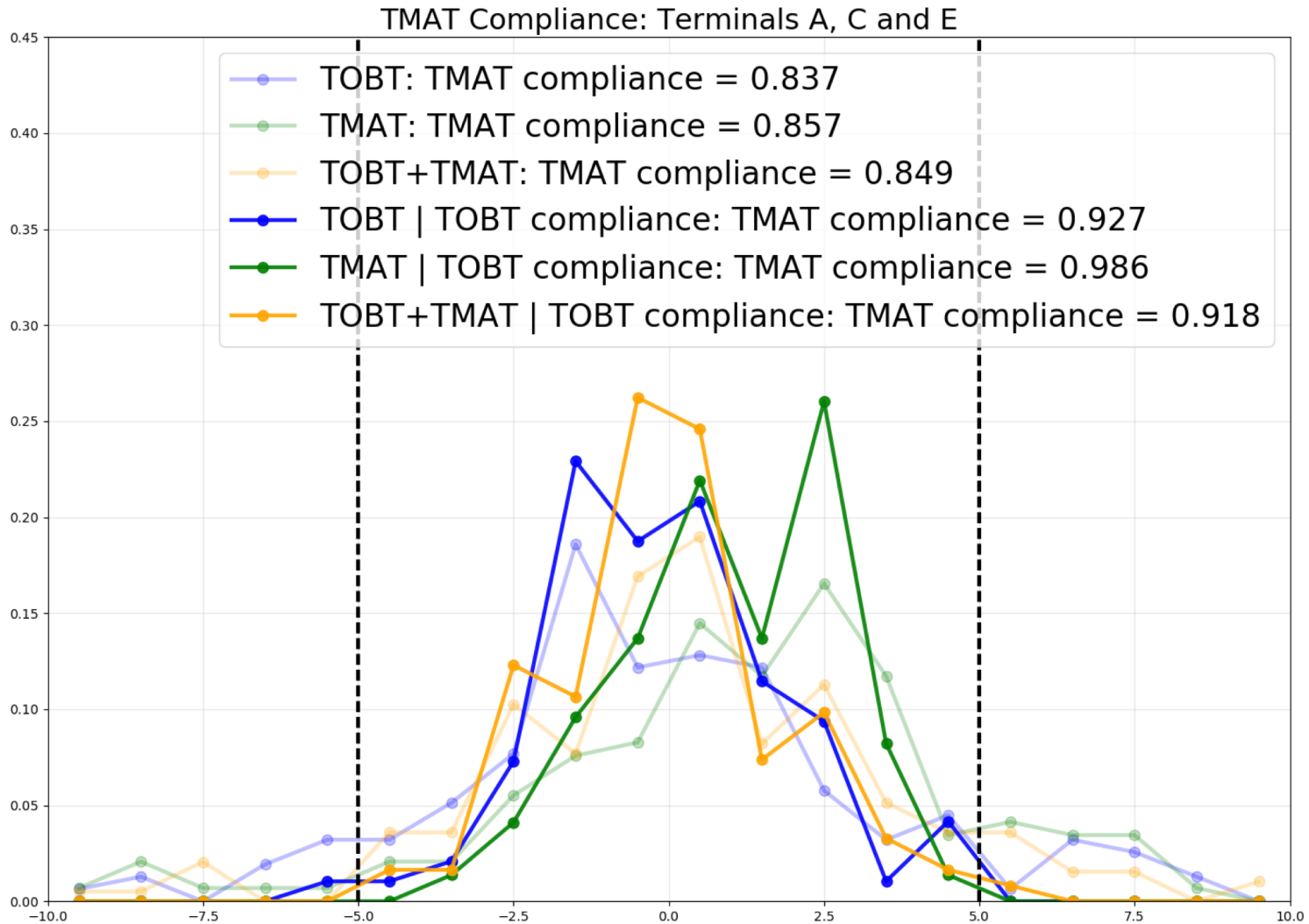


TMAT compliance was similar across metering conditions (83%-85%) and was on the later side of the +/- 5 compliance window in the TMAT only condition (green line).

# TMAT Compliance: Terminal A, C and E Fitted to a Normal Distribution

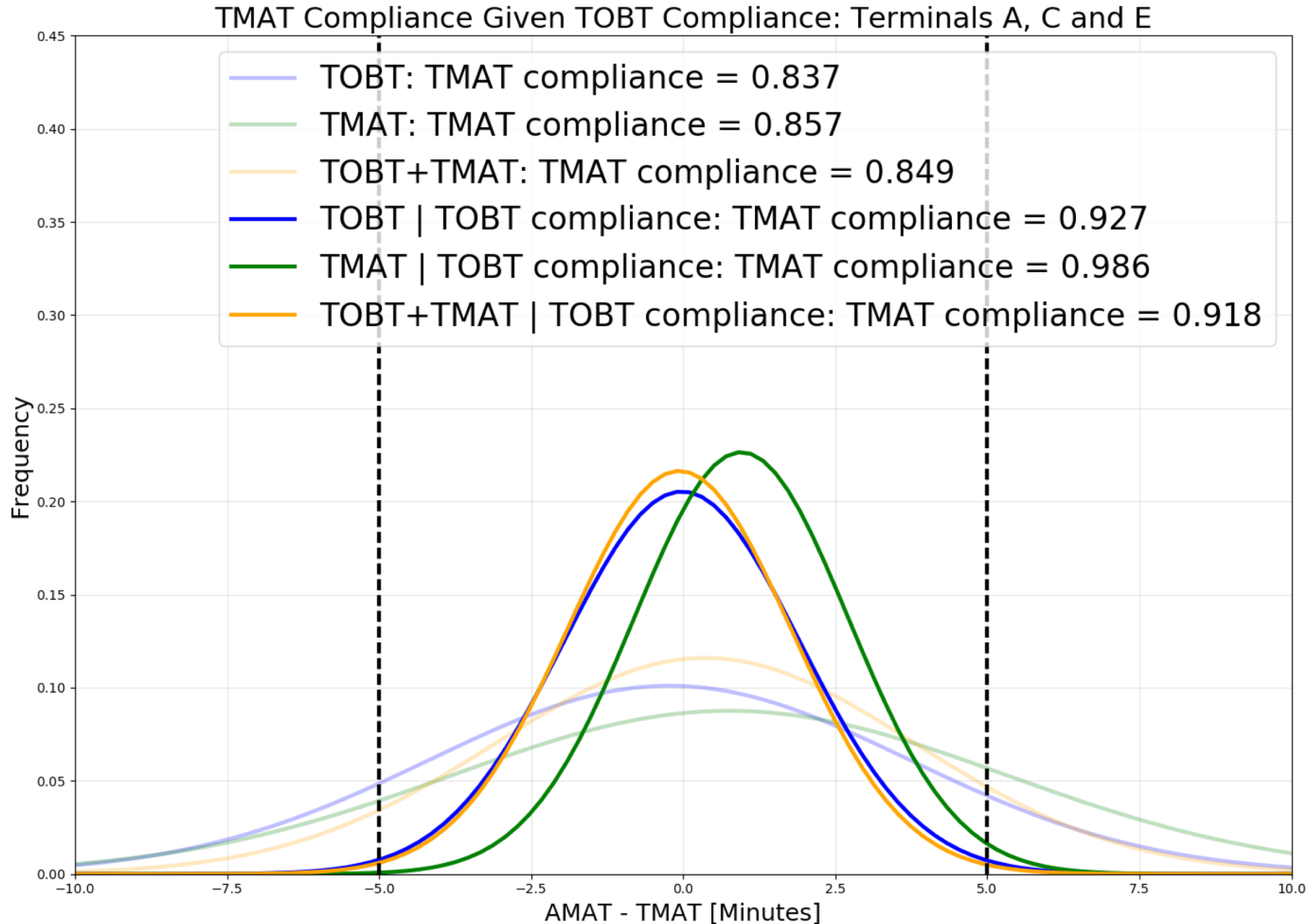


# TMAT Compliance Given TOBT Compliance: Terminal A, C & E



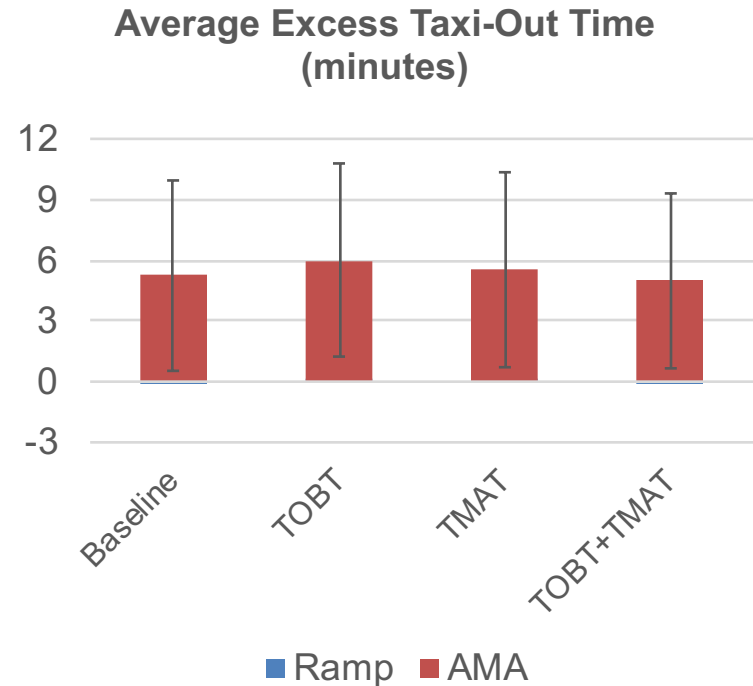
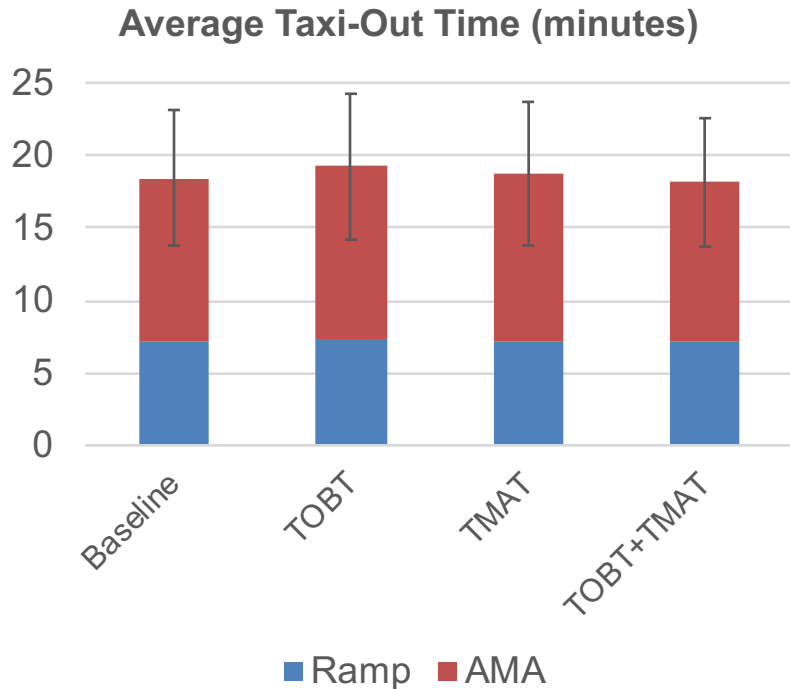
TMAT compliance increased in all metering conditions from 83%-85% to 92%-99% when aircraft were initially compliant with the TOBT.

# TMAT Compliance Given TOBT Compliance: Terminal A, C & E Fitted to a Normal Distribution



- Compliance with pushback advisories (TOBTs) and spot times (TMATs)
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- **Were the metering performance benefits that were observed in the field also observed in the HITL?**
- What factors reduced the ability to measure the benefits of metering?
- What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?
- What were the processes Ramp Controllers used to meet TOBTs and TMATs?
- Were there differences between conditions regarding perceived operational efficiency?
- What were the main suggestions from Ramp personnel on improving RTC?
- What can we conclude?

# Average Taxi-Out Times – Terminal A, C and E

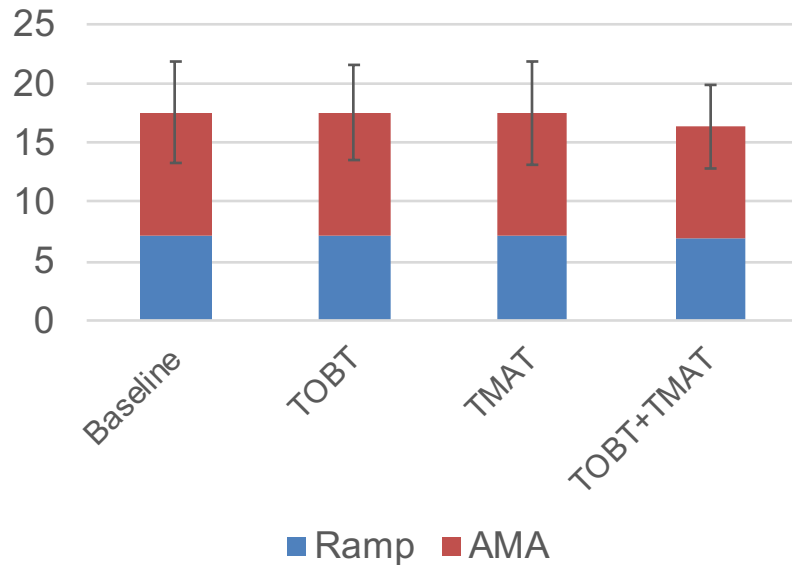


Average taxi-out time and excess taxi-out times were not different in the four conditions, hence the benefits of metering were not captured in this HITL mainly due to simulation artifacts (see slides 46 - 49).

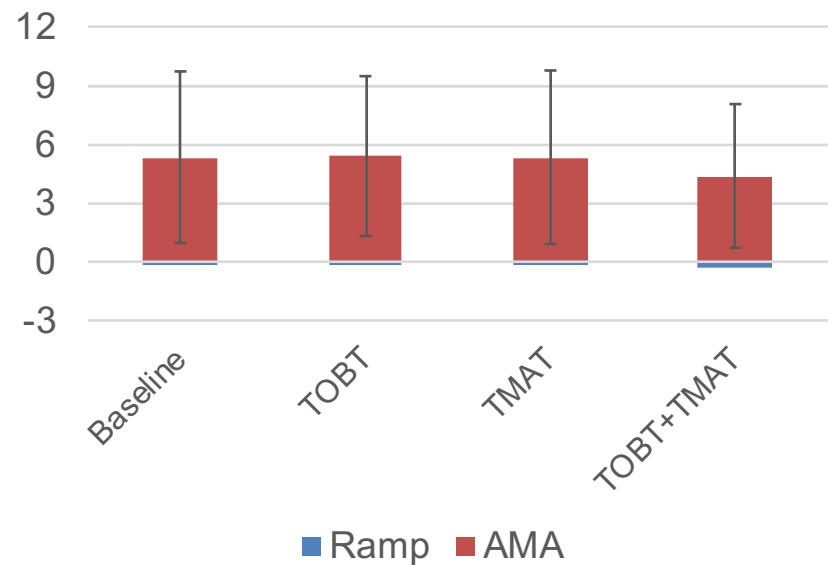
# Average Taxi-Out Times for Runway 17R – Terminal A, C and E



Average Taxi-Out Time for 17R  
(minutes)



Average Excess Taxi-Out Time for  
17R



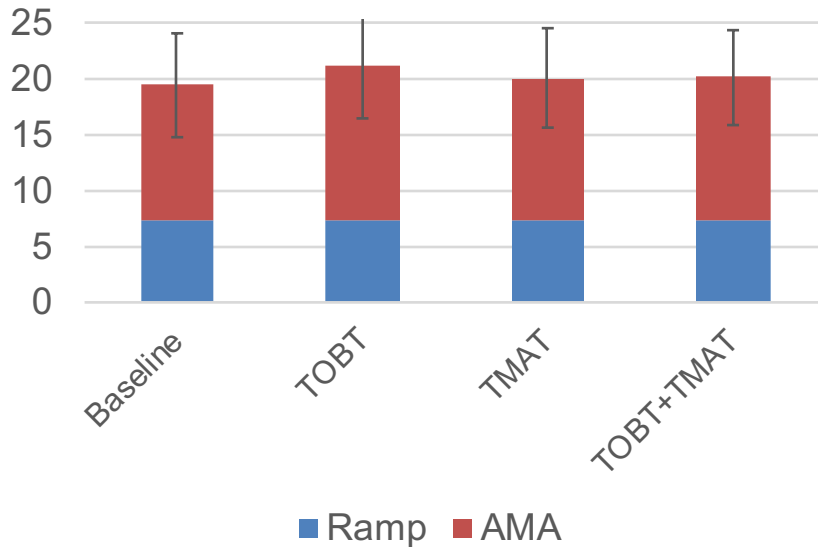
Nor were there differences between conditions on these two measures for departures going to Runway 17R.



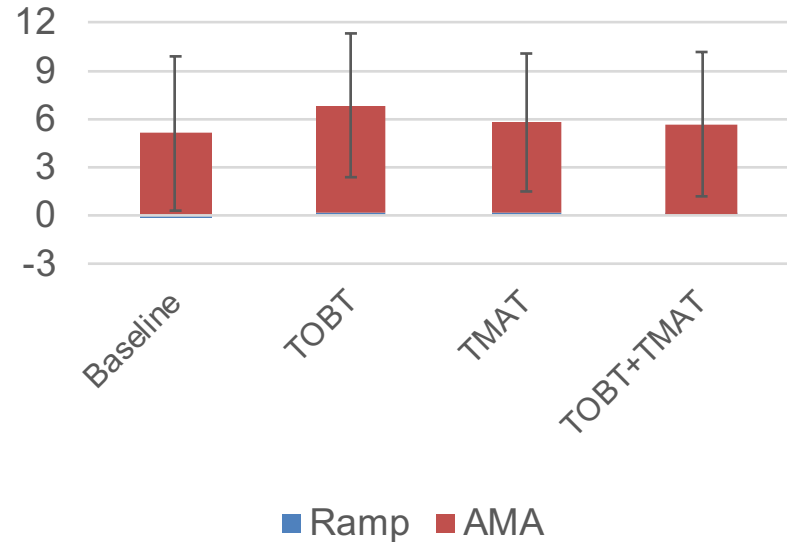
# Average Taxi-Out Times for Runway 18L – Terminal A, C and E



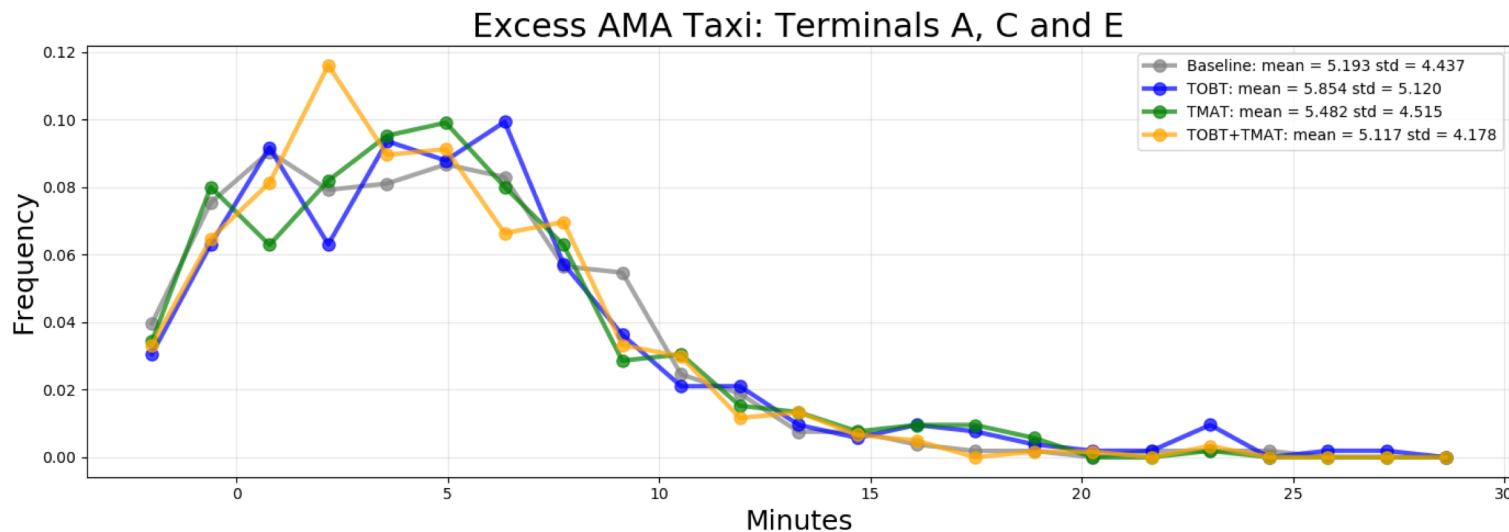
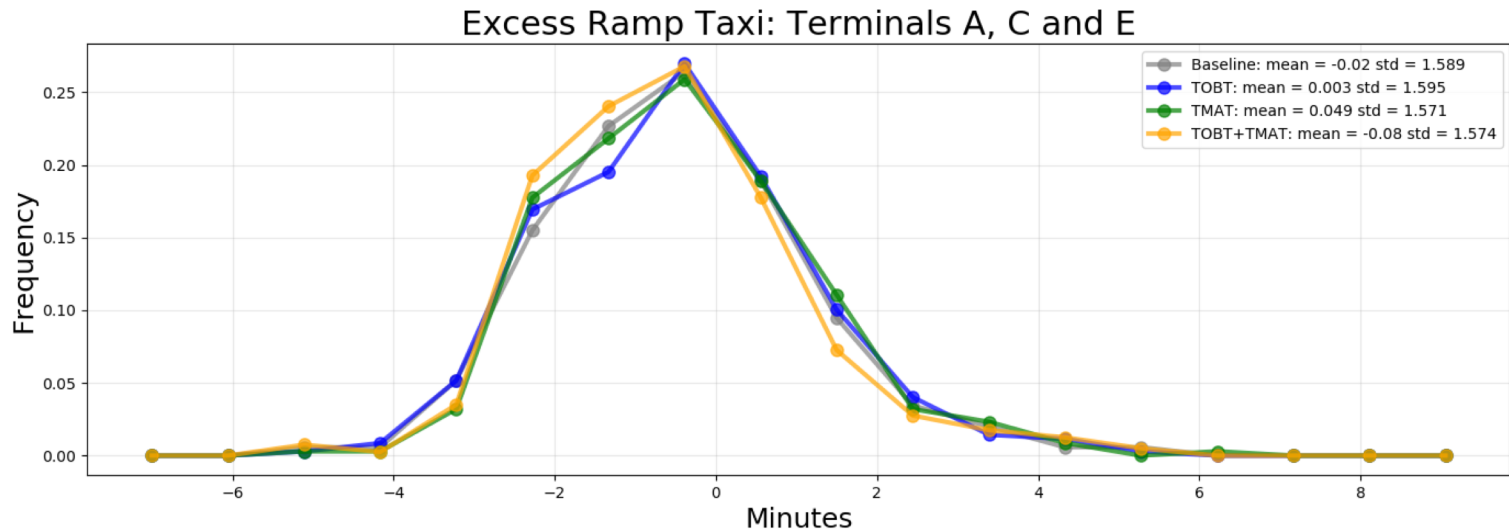
Average Taxi-Out Time for 18L  
(minutes)



Average Excess Taxi-Out Time for  
18L



This was also the case for departures going to Runway 18L.

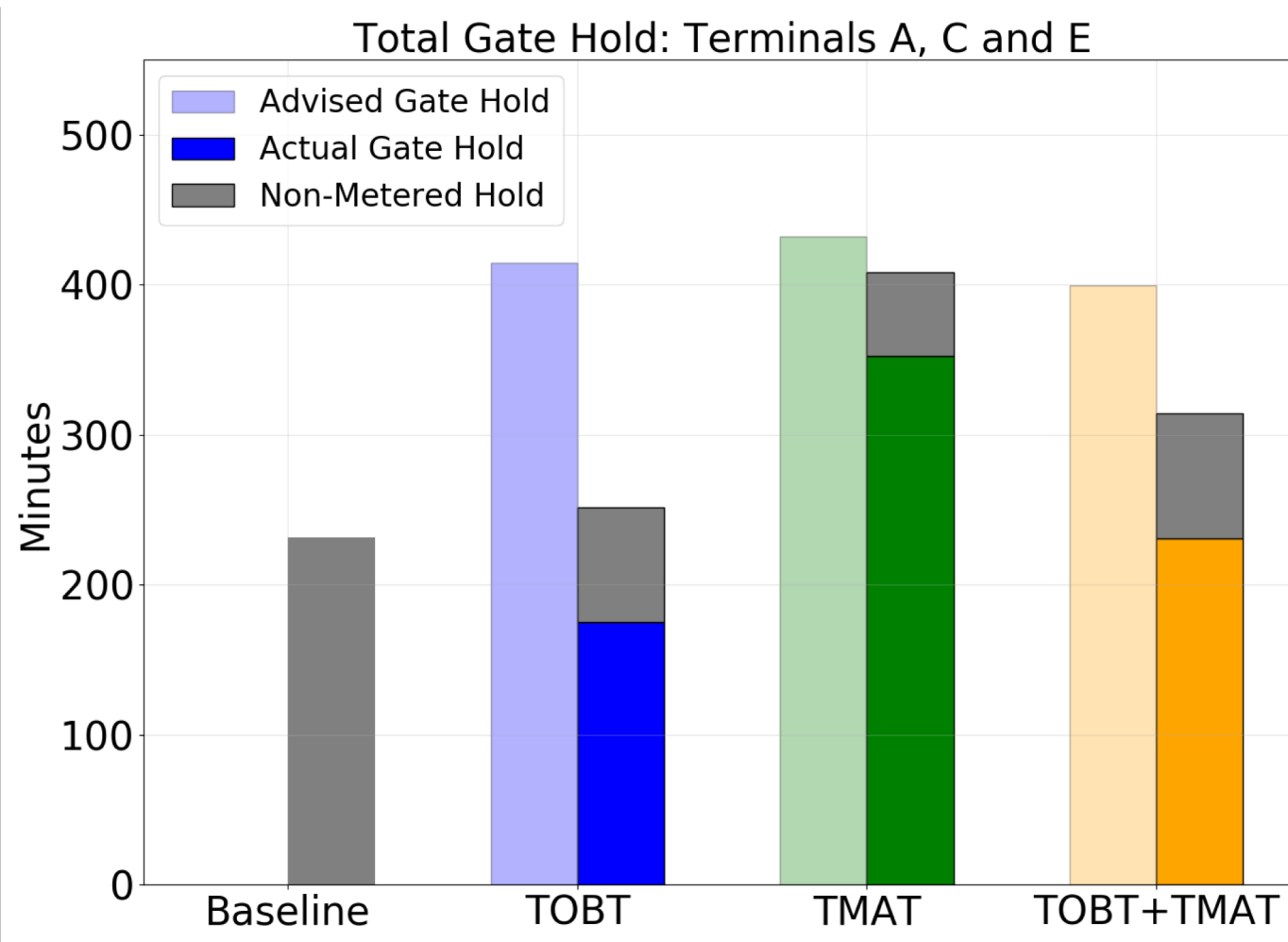


The distribution of excess Ramp and AMA taxi-out time was similar for Baseline and all three metering conditions

- **Were the metering performance benefits that were observed in the field also observed in the HITL?**
  - Excess taxi-out time across Baseline and the three different metering conditions showed similar results, so the benefits of metering were not captured in this HITL.
  - Surface counts also showed comparable results and are available in Appendix A.

- **Compliance with pushback advisories (TOBTs) and spot times (TMATs)**
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- **Were the metering performance benefits that were observed in the field also observed in the HITL?**
- **What factors reduced the ability to measure the benefits of metering?**
- **What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?**
- **What were the processes Ramp Controllers used to meet TOBTs and TMATs?**
- **Were there differences between conditions regarding perceived operational efficiency?**
- **What were the main suggestions from Ramp personnel on improving RTC?**
- **What can we conclude?**

# Total Minutes of Gate Hold by Condition: Terminals A, C and E



Total minutes of non-metered gate hold (gray) in Baseline was comparable to gate holding in most metering conditions.

# Example Surface Metering Display Plot for a Metering Run: TOBT+TMAT (Run 24, Scenario 5)

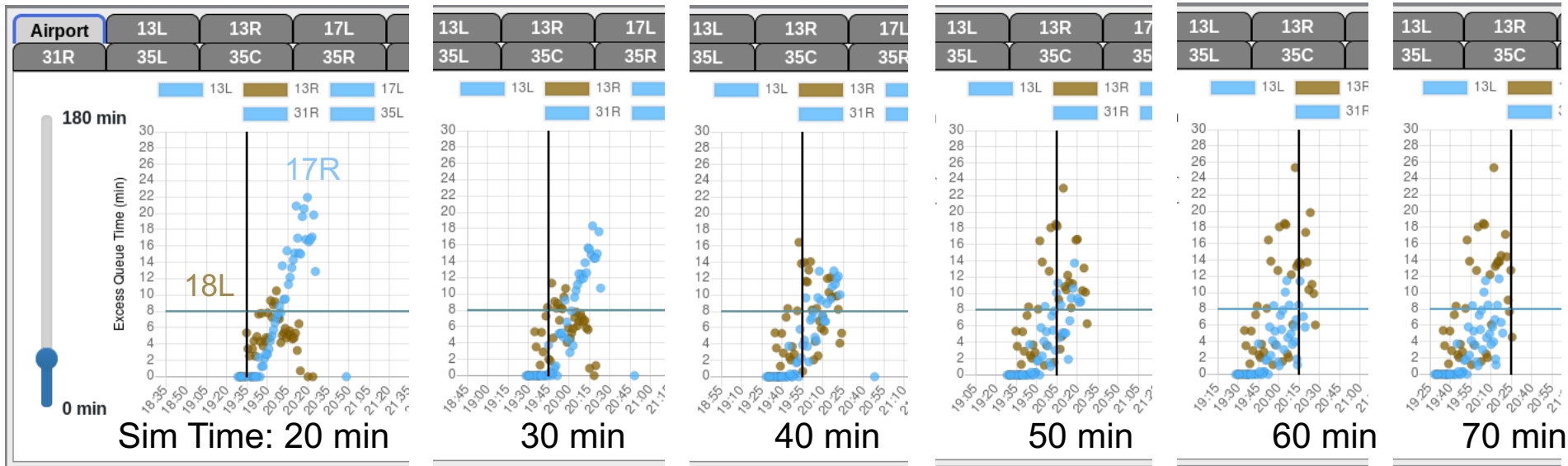


Sim Time: 20 min

Status	Runway	Start	End	Ft Count	Avg Hold	Max Hold	Action
AFFIRMED	17R	01/19:37	01/20:18	22	6.5 min	14.0 min	✓ ✗
ACTIVE	18L	01/19:33	01/20:06	28	< 1 min	5.0 min	✓ ✗

Sim Time: 70 min

Status	Runway	Start	End	Ft Count	Avg Hold	Max Hold	Action
COMPLETED	17R	01/19:35	01/20:13	30	1.3 min	6.9 min	✓ ✗
COMPLETED	18L	01/19:33	01/20:20	25	< 1 min	6.0 min	✓ ✗



As expected, excess taxi-out (queue) time during this metering condition run was reduced due to scheduler-based holds.

# Example Surface Metering Display Plot for a Baseline Run (Run 25, Scenario 5)

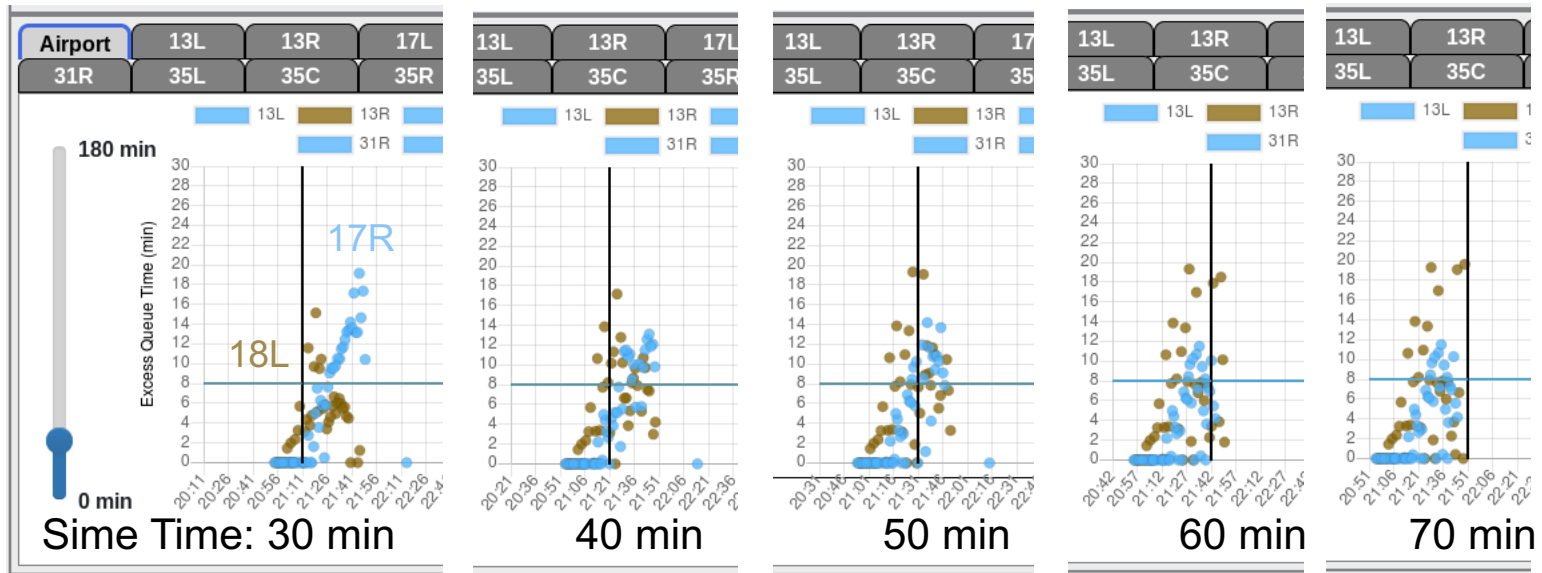


Sim Time: 30 min

Status	Runway	Start	End	Flt Count	Avg Hold	Max Hold	Action
No Current SMPs							

Sim Time: 70 min

Status	Runway	Start	End	Flt Count	Avg Hold	Max Hold	Action
No Current SMPs							



As not expected, excess taxi-out (queue) time was also reduced during this Baseline run likely due to gate holds introduced via simulation artifacts.

- **What factors reduced the ability to measure the benefits of metering?**
  - The Baseline condition showed more gate holds than expected
  - These gate holds appeared to reduce the excess taxi-out time on the airport surface during the Baseline runs in the same manner as in the metering conditions
  - The Baseline condition thus appeared to be effectively metering the departures
- Simulation artifacts that might have contributed to this increased gate-holding in Baseline were
  - Unrealistically long pushback (spool-up) times based on Charlotte Airport data, which resulted in aircraft blocking pushback in the ramp area
  - Limited options in the simulation compared to DFW regarding
    - Pushback direction
    - Taxi routes
    - Ramp taxi speed



- **Compliance with pushback advisories (TOBTs) and spot times (TMATs)**
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- **Were the metering performance benefits that were observed in the field also observed in the HITL?**
- **What factors reduced the ability to measure the benefits of metering?**
- **What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?**
- **What were the processes Ramp Controllers used to meet TOBTs and TMATs?**
- **Were there differences between conditions regarding perceived operational efficiency?**
- **What were the main suggestions from Ramp personnel on improving RTC?**
- **What can we conclude?**

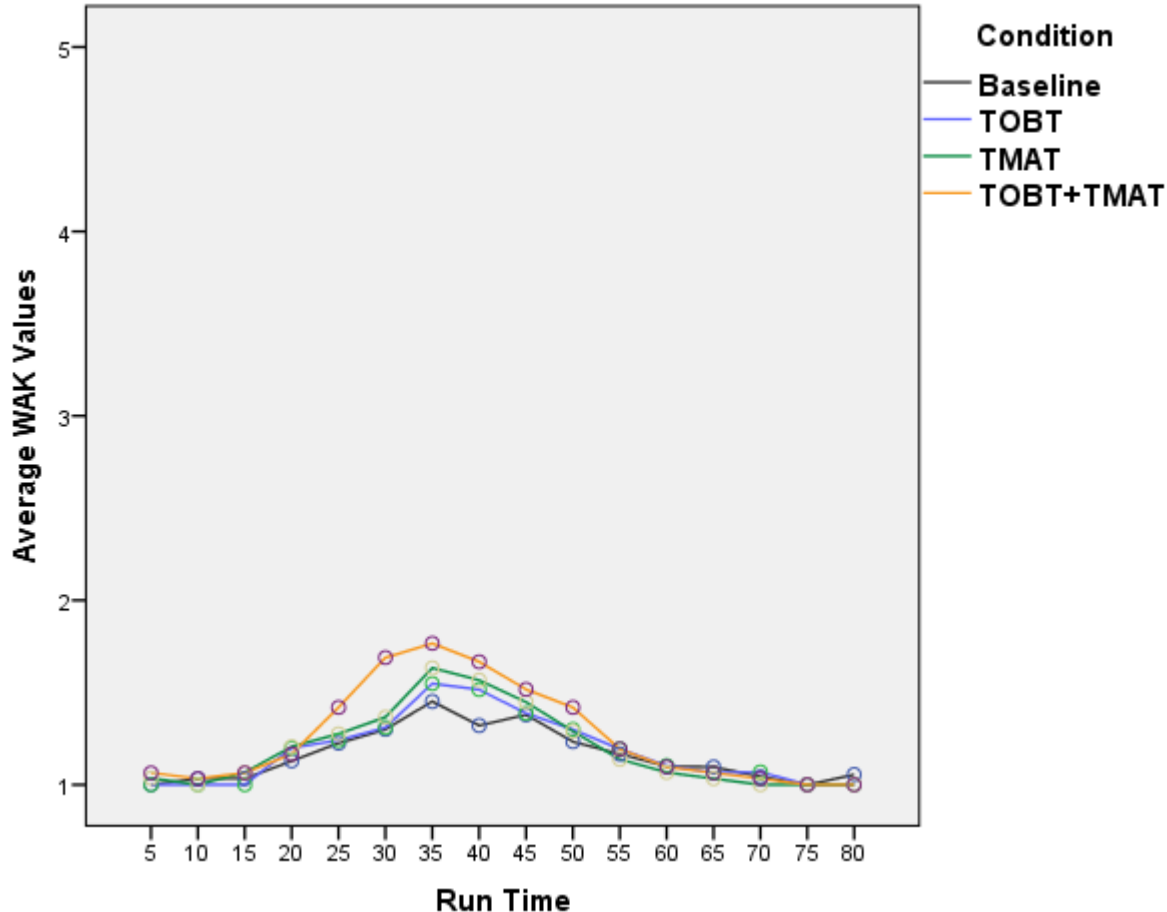
## Workload Assessment Keypad (WAK)

- WAK collected individual data points of workload during each run for the Ramp personnel once every 5 minutes
- Ramp personnel were asked to consider a workload rating of “3,” or moderate, as comparable to nominal operations at DFW

High workload

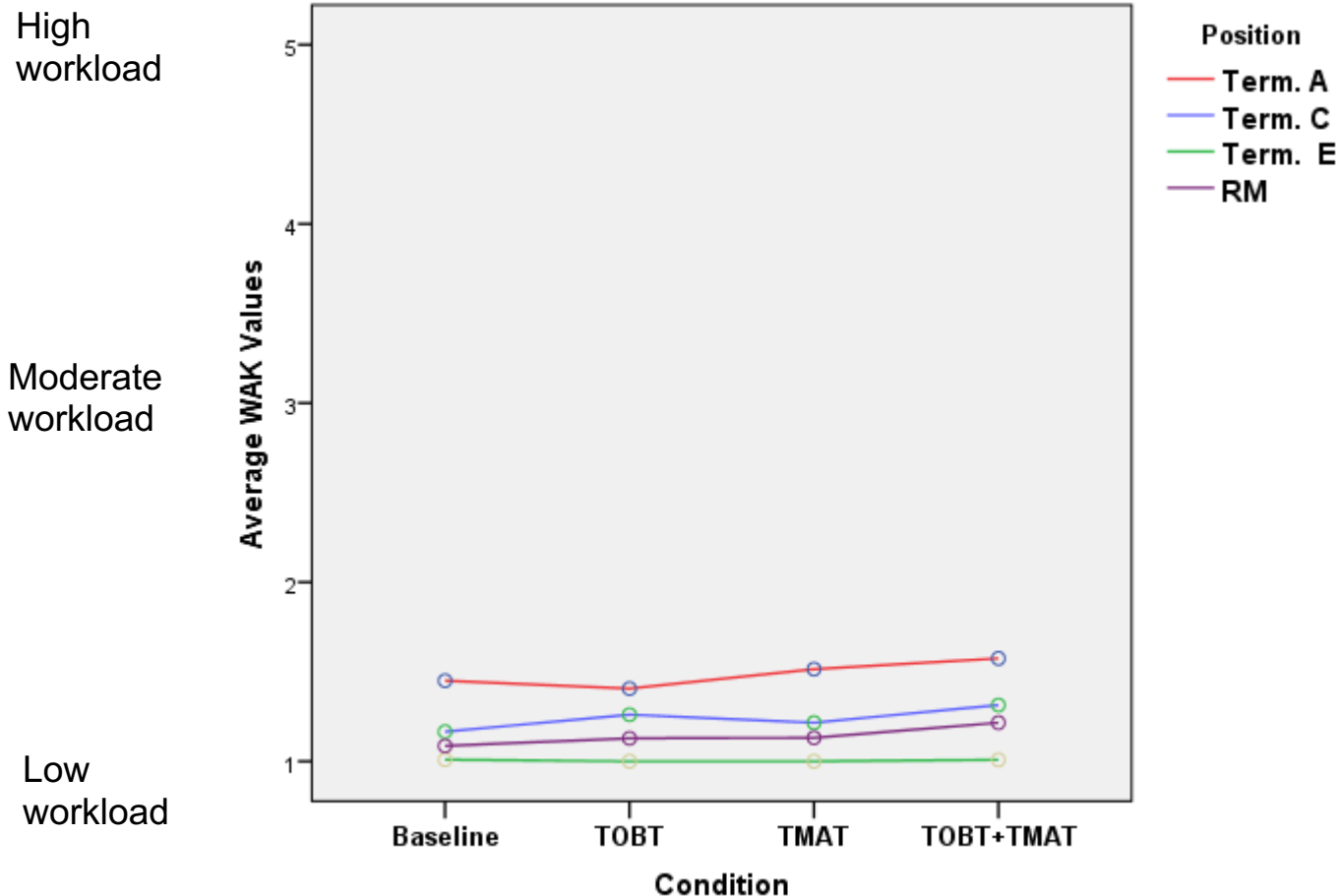
Moderate workload

Low workload



$N = 1818$  (~455 in each condition), ANOVA condition significant at  $p = .04$ ,

Workload was very low overall but statistically significantly higher in the TOBT+TMAT condition than in Baseline and the TOBT condition.



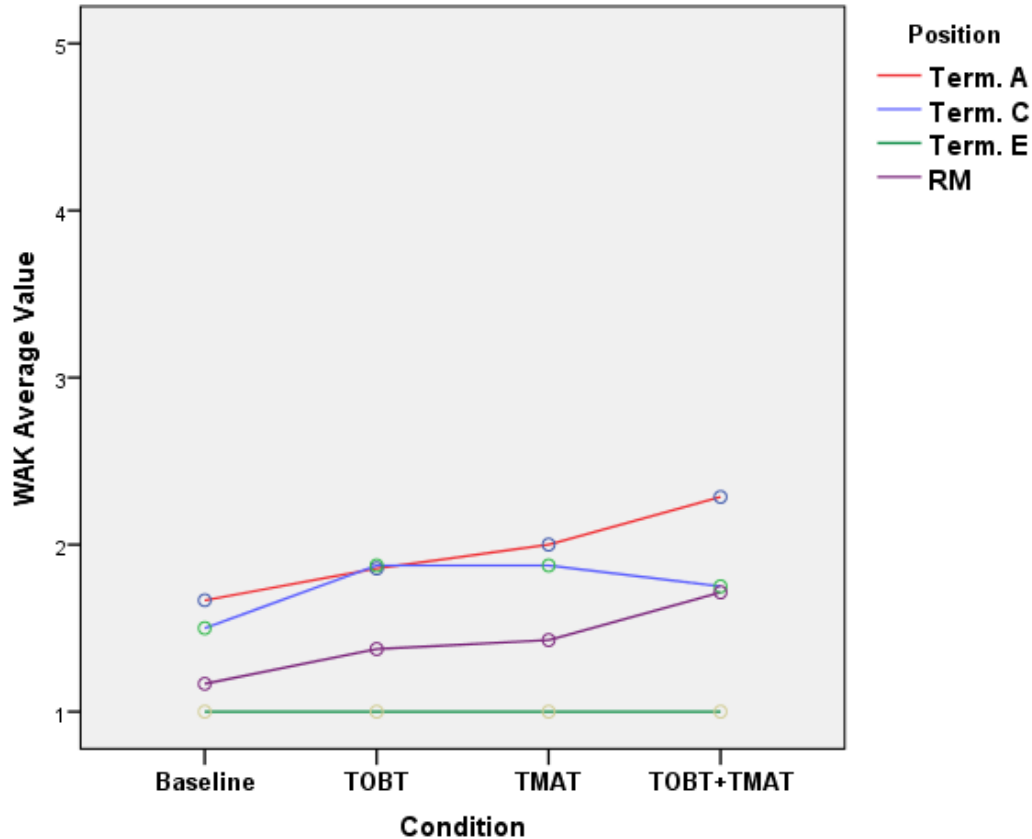
*N* = 1818, ~455 in each position, position significant at  $p < .000$

All had low workload, but Terminal A had the highest, followed by Term. C, the Ramp Manager, and Term. E. All were statistically significantly different from each other.

High workload

Moderate workload

Low workload



$N = 119$  (~30 in each condition),  
Position differences  
 $F(3, 103) = 8.3, p = <.000$ .

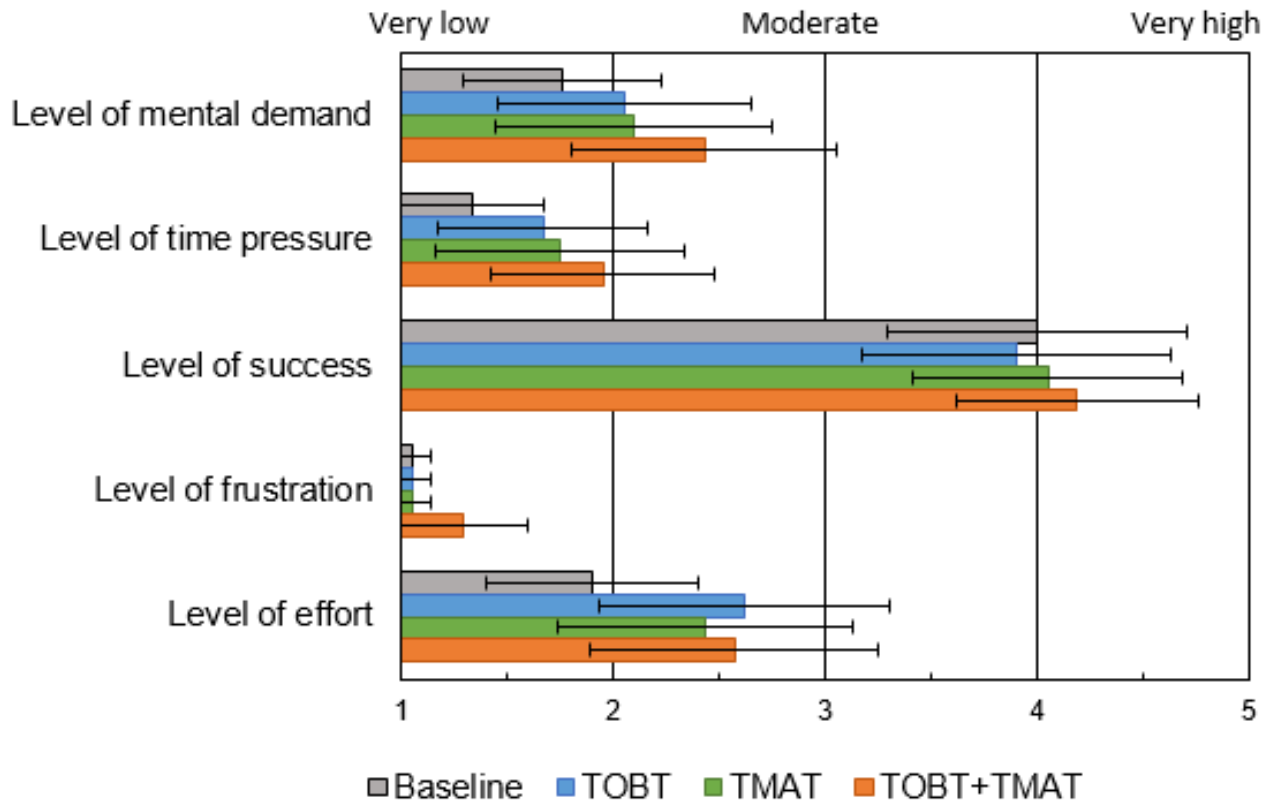
Terminal E's WAK score at 40 minutes into the run was statistically significantly lower than all others, and that of the Ramp Manager (RM) was statistically significantly lower than Terminal A's.

# Ramp Controller Post-Run Workload Ratings on NASA TLX Items



Please rate the following based on when you were busiest during this run:

Error bars are 95% CIs.  $N = 84$  ratings, 21 in each condition for each item.



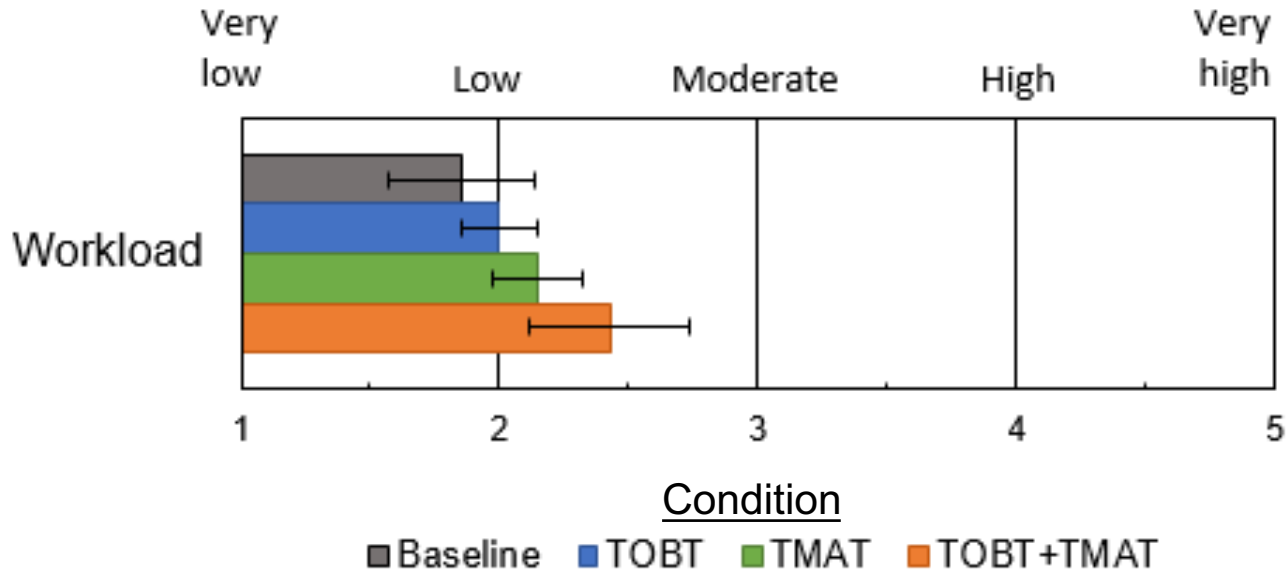
Ramp Controllers' ratings indicated low workload, high success, with no statistically significant differences between conditions, although TOBT+TMAT trended slightly higher on mental demand, time pressure, and frustration.

# Ramp Controller Post-Simulation Ratings of Workload by Condition



Please describe your workload at the busiest times in each of the conditions in this simulation.

Error bars = 95% CIs (Loftus correction applied).  $N = 7$ . Repeated measures ANOVA,  $F(3,18) = 3, p = .058$ .

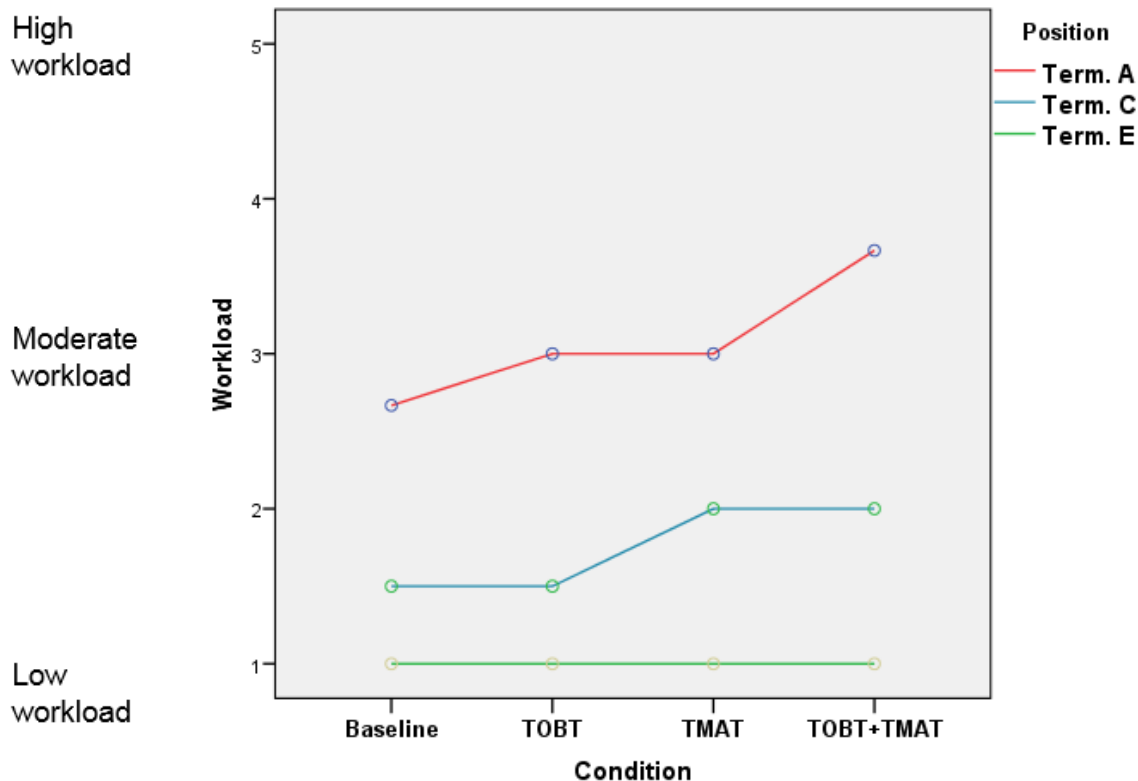


Ramp Controllers' ratings of workload in the conditions were low. TOBT+TMAT condition was nearly statistically significantly higher than both Baseline and TOBT conditions ( $p \leq .10$ ). (By comparison, the two Ramp Manager's post-sim. ratings for Ramp Controllers' workload were: 3 in Baseline, 4 in TOBT, 4.5 in TMAT, and 5 in TOBT+TMAT.)

# Ramp Controller Post-Sim. Ratings of Workload by Position and Condition



Please describe your workload at the busiest times in each of the conditions in this simulation.



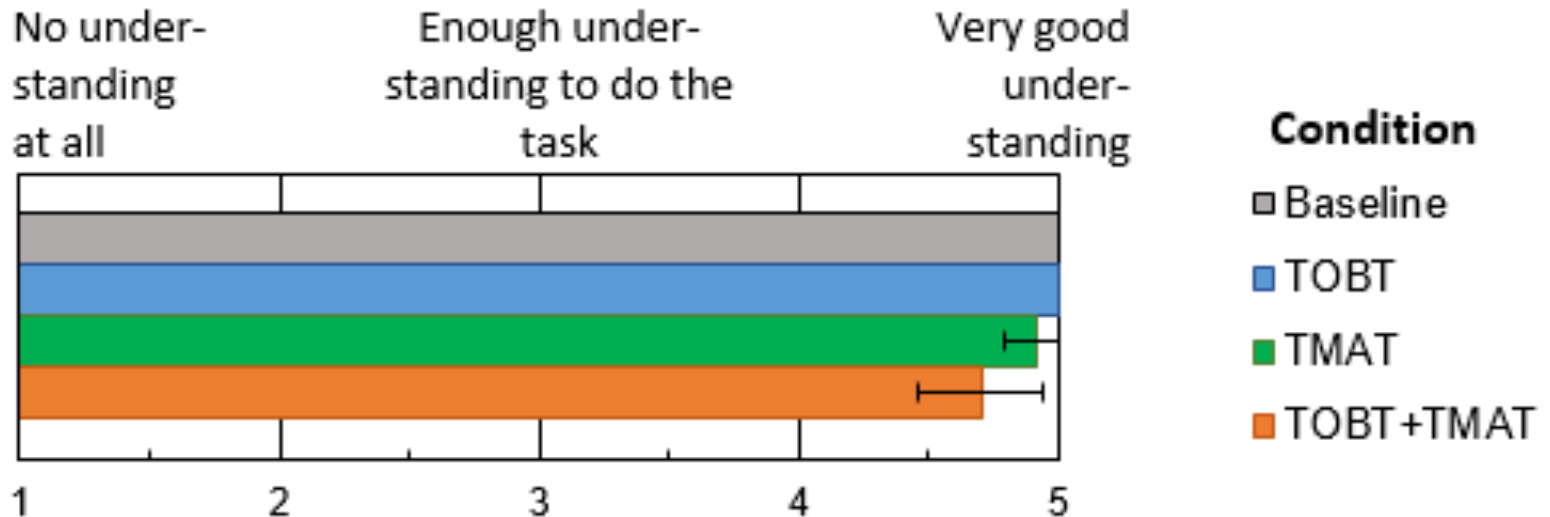
$N = 7$ . Repeated measures ANOVA, position  $F(2,4) = 3.0$ ,  $p = .16$

Terminal A described the highest workload, especially in the TOBT+TMAT condition.



Please rate your understanding of the traffic (all your flights) *when you were busiest* during this run.

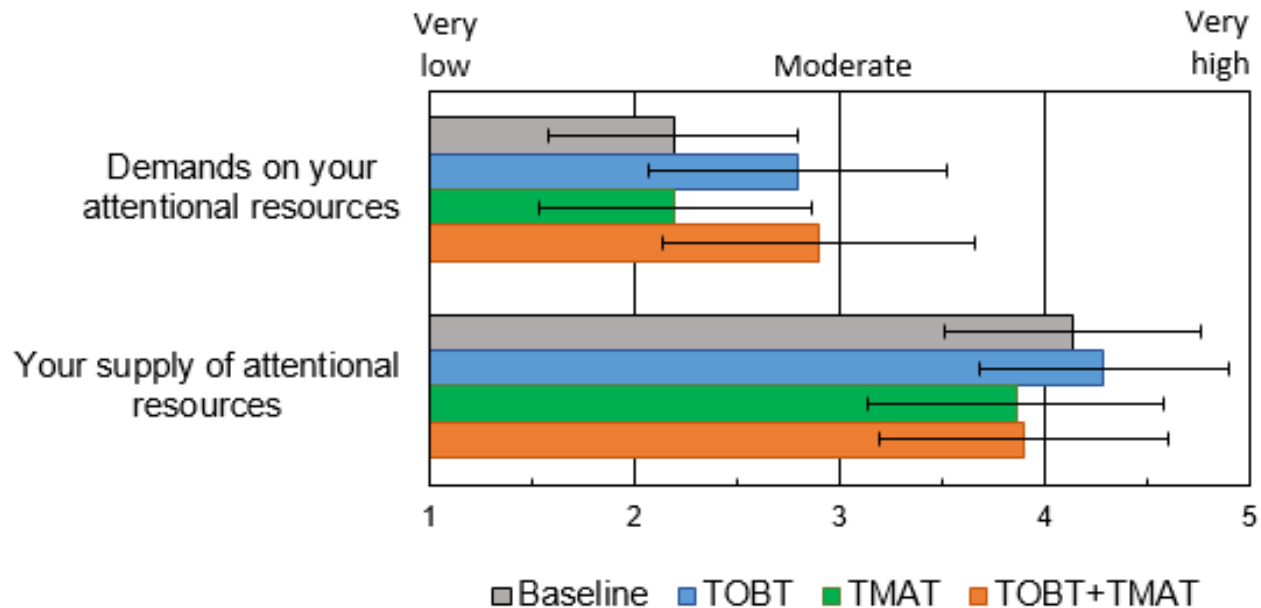
Error bars = 95% CIs;  $N_s = 20-21$  in each condition.  $MS = .38$ ,  $F(3,79) = 3.7$ ,  $p = .015$



The “understanding” component of the SART (Situational Awareness Rating Technique) indicated slight but statistically significantly less understanding in the TOBT+TMAT condition than the other conditions.

Please rate the following when you were busiest during this run:

Error bars = 95% CIs; *Ns* = 20-21 in each condition for each item.



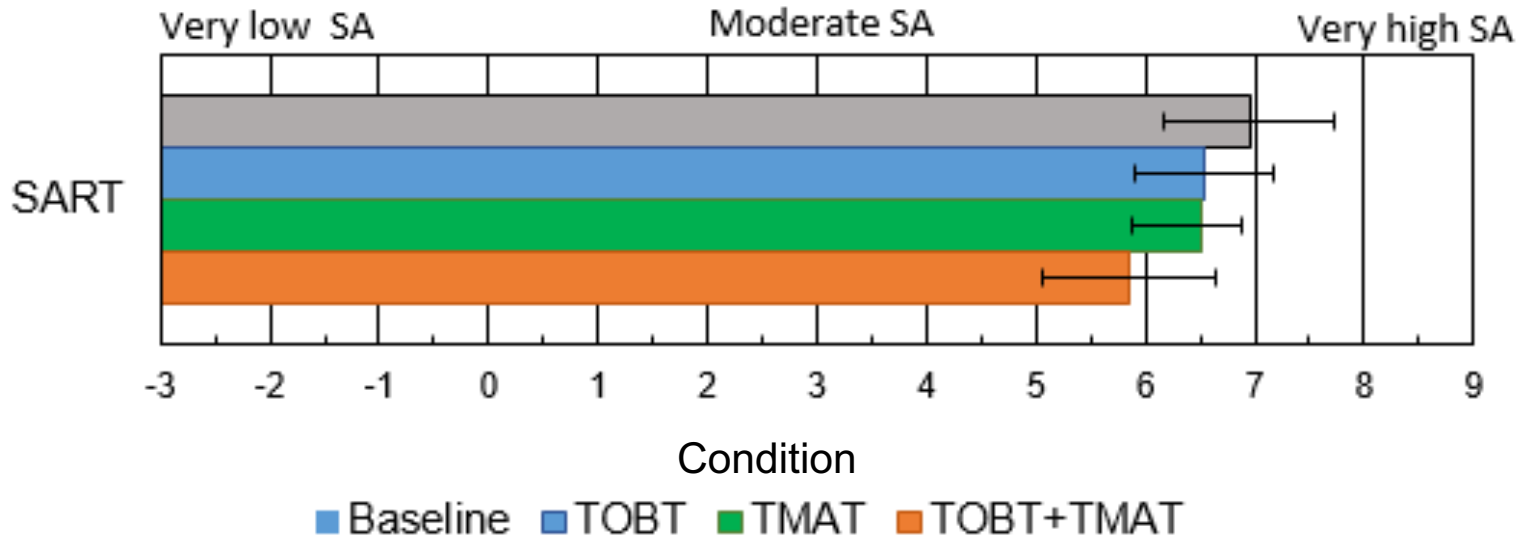
There were no statistically significant differences between conditions in the ratings of the “Demand” and “Supply” of attentional resources parts of the SART.

# SA: Ramp Controller Final Post-Run Scores on SART



SART Rating Formula = Understanding + (Supply – Demand) of Attentional Resources,  
i.e., SA = U + (S-D)

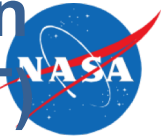
$N = 80$ , ~20 in each condition ANOVA significant at  $p = .28$ , error bars are 95% CIs.



Ramp Controllers' ratings of situation awareness were statistically significantly lower in the TOBT+TMAT condition than in Baseline ( $p = .05$ ).

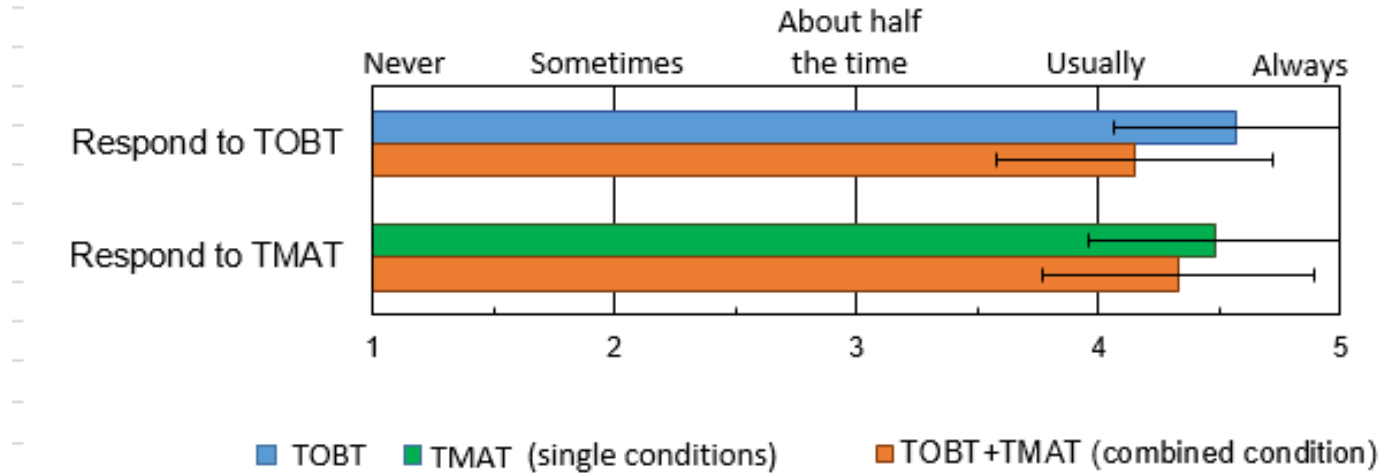
- **What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?**
  - **Workload**
    - WAK ratings were low overall, but statistically significantly higher in the TOBT+TMAT condition than in the Baseline and the TOBT conditions
    - Average WAK ratings were statistically significantly different between all positions, with Terminal E being the lowest, followed by the Ramp Manager, Terminal C, and Terminal A, which had the highest
    - Reasons that the controllers gave for low overall workload were simulation artifacts
      - Gate conflicts handled by Ramp Manager with aid of software
      - Pilots not calling in immediately after landing to check on gate
      - Pilots calling one-by-one on frequency instead of many at once
      - No disruptions and anomalies that occur in the field
  - **Situation Awareness (SA)**
    - SA was statistically significantly higher in the Baseline than in the TOBT+TMAT condition

- Compliance with pushback advisories (TOBTs) and spot times (TMATs)
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- Were the metering performance benefits that were observed in the field also observed in the HITL?
- What factors reduced the ability to measure the benefits of metering?
- What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?
- **What were the processes Ramp Controllers used to meet TOBTs and TMATs?**
- Were there differences between conditions regarding perceived operational efficiency?
- What were the main suggestions from Ramp personnel on improving RTC?
- What can we conclude?



How often were you able to respond to (act on) [gate][spot] hold advisories in this run?

*N*'s = 20-21 ratings in each condition. Error bars = 95% CIs.

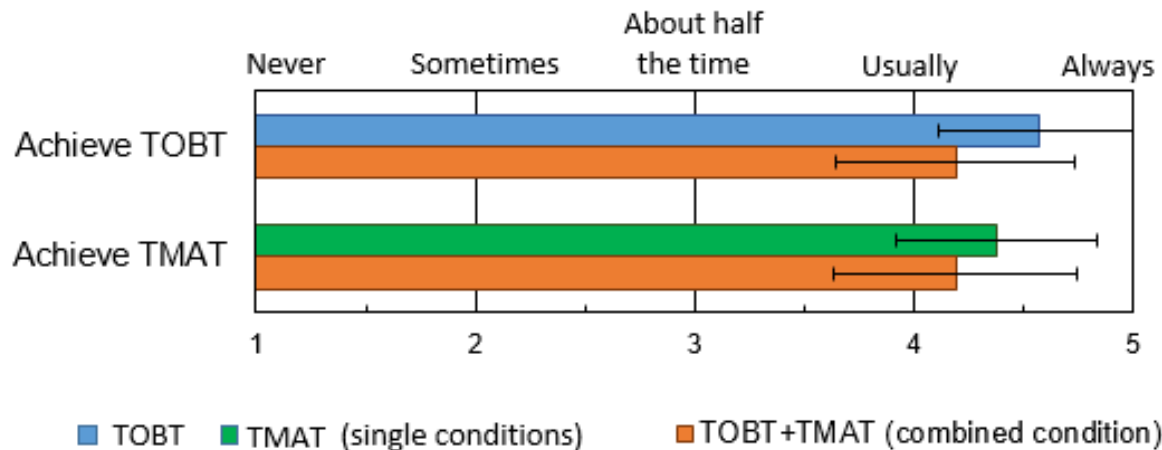


Ramp Controllers “usually” responded to TOBTs and TMATs in both the Single and Combined conditions. They responded that they did so slightly less frequently, but not statistically significantly so, in the Combined condition.



How often were you able to achieve [gate][spot] hold advisories in this run?

*N*'s = 21 ratings in each condition. Error bars = 95% CIs.



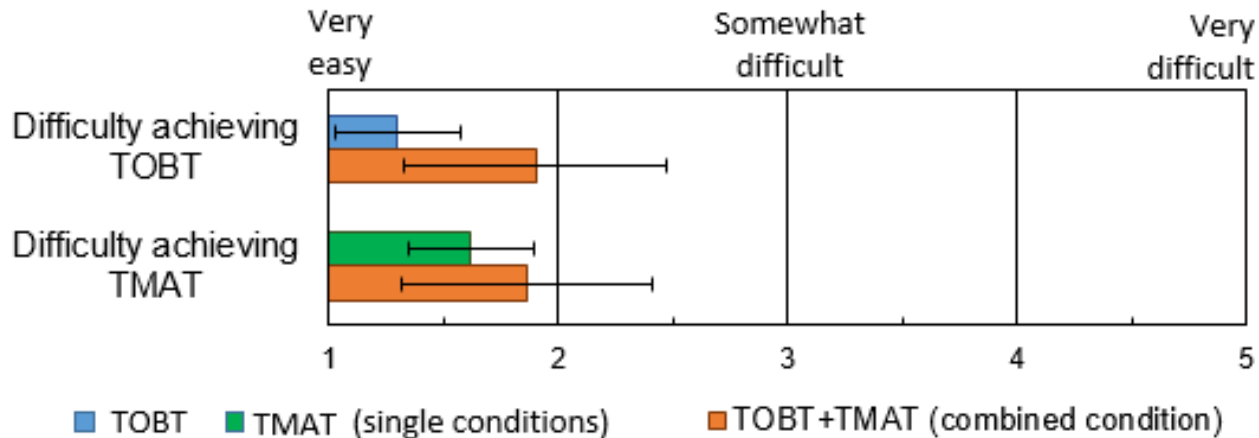
Ramp Controllers “usually” achieved TOBTs and TMATs in both the Single and Combined conditions. They responded that they did so slightly less frequently, but not statistically significantly so, in the Combined condition.

# Post-Run Ratings: Difficulty of Achieving Advisories in Single & Combined Conditions



In this run, how difficult was it to achieve [gate][spot] hold advisories?

*N*'s = 29-21 ratings in each condition. Error bars = 95% CIs.



Ramp Controllers described achieving the advisories as easy in all conditions. They described slightly more difficulty, but not statistically significantly more, in the Combined condition.



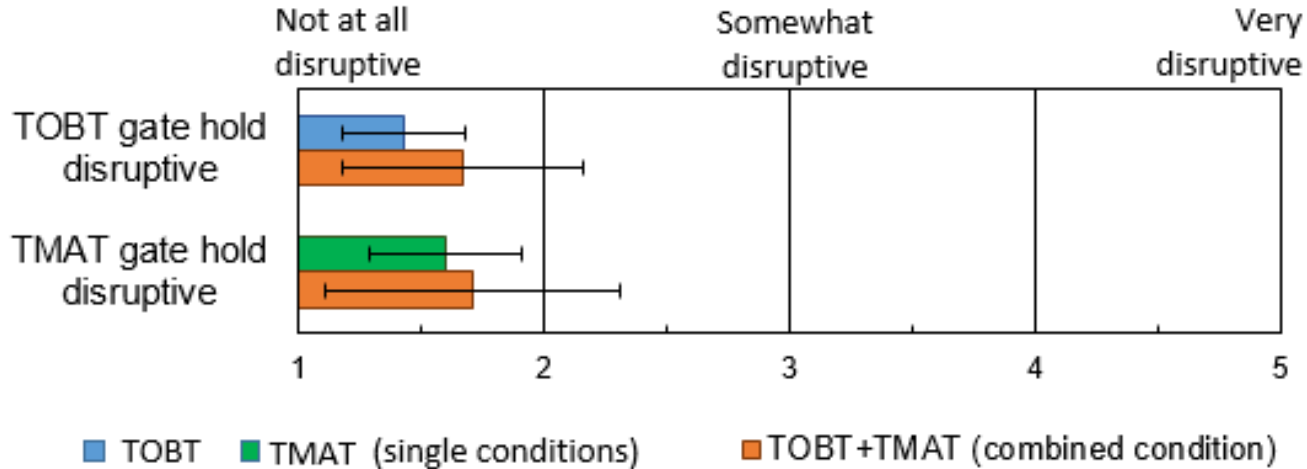


- “It could be confusing, when watching both advisories. The level of success would be greater, when being concerned about 1 at time.” (RM)
- “Too much traffic in DFW to be able to accommodate for [both] gate and spot advisories. With the spot advisories having a +/-5 min window, it helps with hitting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



In this run, how disruptive to the traffic flow in your sector was holding aircraft at the gate to meet [gate][spot] hold advisories?

*N*'s = 17-21 ratings in each condition. Error bars = 95% CIs.



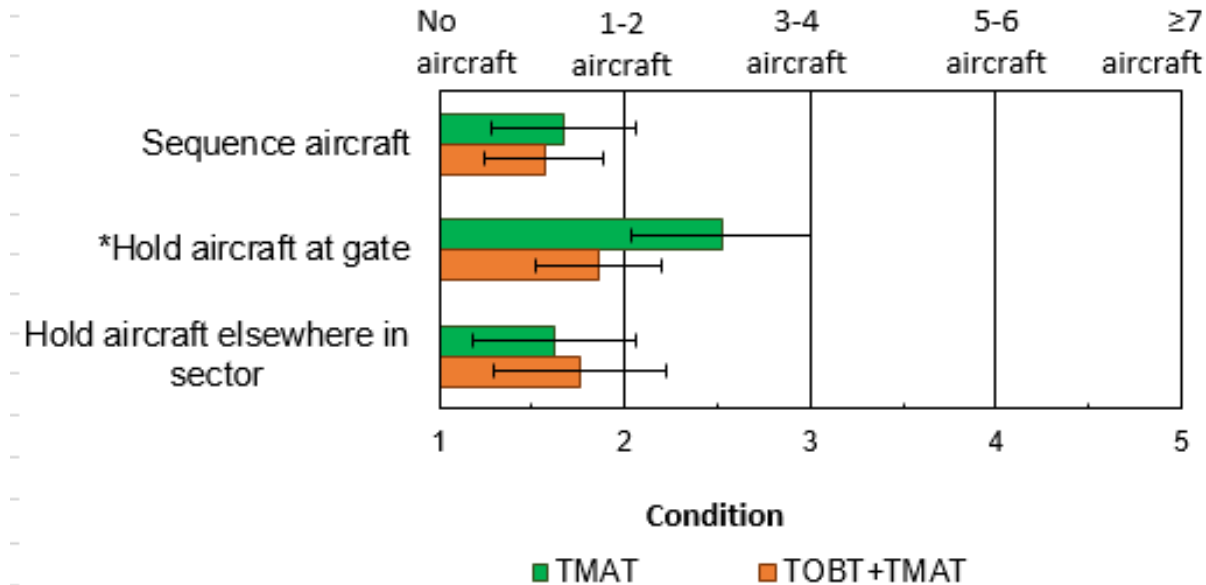
Ramp Controllers found holding aircraft at the gate not disruptive to traffic flow in their sectors in all conditions. They found it slightly more disruptive, but not statistically significantly so, in the Combined condition.

# Post-Run Ratings: Procedures with TMATs in Single and Combined Conditions



How often did you do the following to achieve spot advisories for aircraft in this run?

*N* = 21 ratings in each condition. Error bars = 95% CIs.



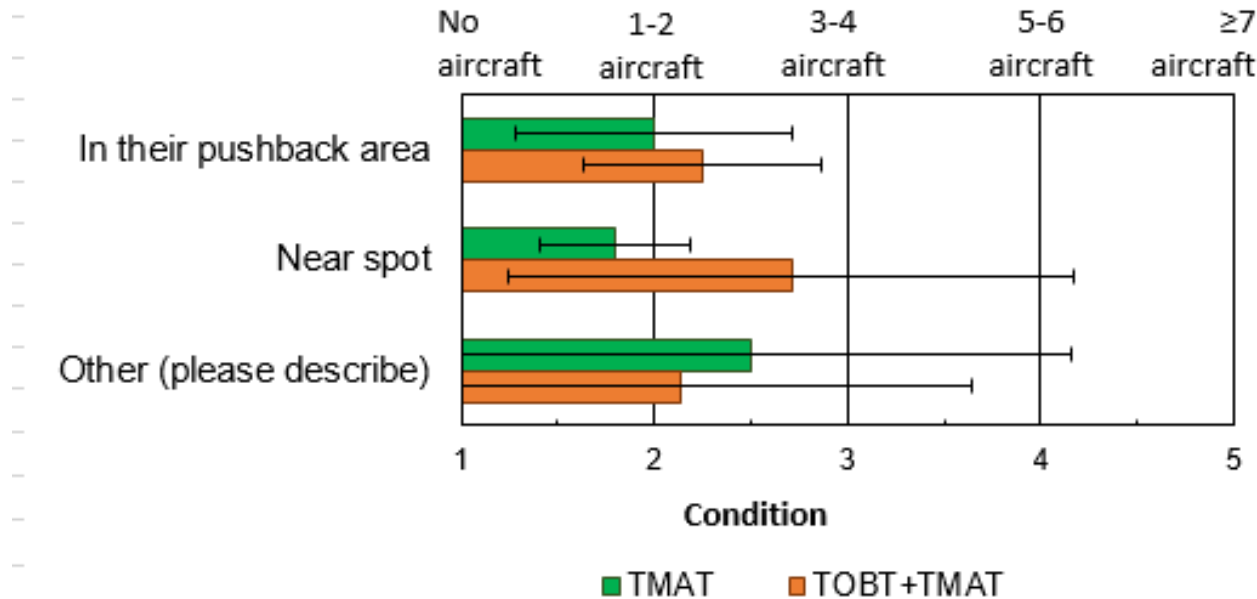
Ramp Controllers said they held aircraft with TMAT advisories at the gate statistically significantly more frequently in the TMAT-only condition than in the TOBT+TMAT condition ( $p = .032$ ). This makes sense, since in the combined condition, they were told to comply with the TOBT advisory, i.e., push back from the gate within  $\pm 2$  minutes.

# Post-Run Ratings: Where, Besides at Gate, were TMATs Held ?



If you held aircraft elsewhere in your sector (other than the gate) to achieve spot hold advisories, where did you hold them?

*N* = 6,8; 5,7; & 6,7 for each condition in each question. Error bars = 95% CIs.

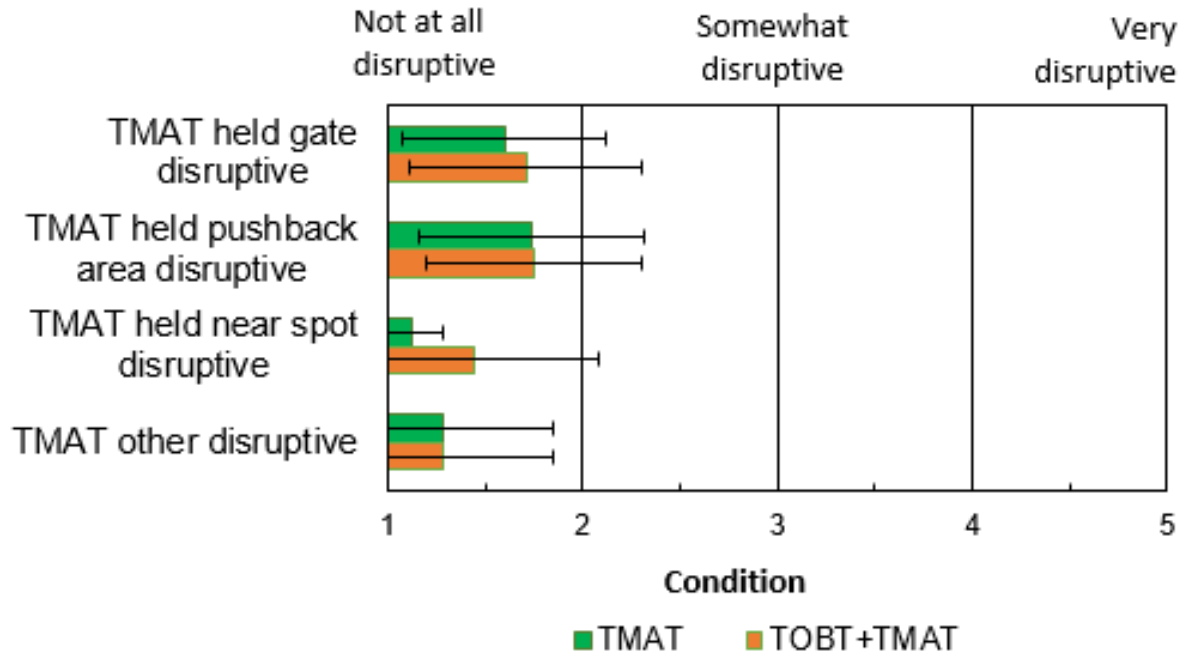


If held elsewhere than at the gate, the TMAT aircraft were held in their pushback area or near their spot. “Other” responses included “in the spot.”



In this run, how disruptive to the traffic flow in your sector was holding aircraft in the following areas to achieve spot advisories?

*N*s = 16-20 (7 for “other”) for each condition in each question. Error bars = 95% CIs.



Aircraft held for TMAT advisories did not appear to be disruptive to sector traffic flow no matter where they were held and in which condition.

# Other Comments on Advisories from RC Post-Sim Surveys



- “Some spot advisories were much too long...normally would not be able to hold an a/c on the gate for 12 mins to be in compliance with spot advisories. Then once off the gate, spooling times were MUCH MUCH too long. If a/c sat on the ramp for as long as they did in the sim, there would be ALOT of congestion.” (Term. C)
- “We really don't have the gate or ramp space to hold planes. We do have space out with ground to help with space.” (Term. A)

- **What were the processes Ramp Controllers used to meet TOBTs and TMA Ts?**
  - In all advisory conditions, both singly and combined (TOBT+TMAT), Ramp Controllers
    - Usually responded to hold advisories,
    - Usually achieved hold advisories,
    - Did not feel it was difficult to achieve hold advisories, and
    - Did not feel that holding aircraft at the gate to achieve advisories was disruptive to traffic flow in their sector.
  - Described holding aircraft at the gate statistically significantly more in the TMAT alone condition than in the TOBT+TMAT condition.
    - This may have been due to having been told to comply with the TOBT advisory, i.e., push back from the gate within  $\pm 2$  minutes.
  - In the TMAT conditions, if not holding aircraft at the gate, Ramp Controllers described holding aircraft in their pushback area, near the spots, or in the actual spot
    - Holding aircraft in these areas was not thought to be disruptive to traffic flow in their sector.
  - Finally, in most of these areas, both the TOBT alone and TMAT alone conditions appeared to be slightly more efficient than in the combined TOBT+TMAT condition, although not statistically significantly so.

- **Compliance with pushback advisories (TOBTs) and spot times (TMATs)**
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- **Were the metering performance benefits that were observed in the field also observed in the HITL?**
- **What factors reduced the ability to measure the benefits of metering?**
- **What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?**
- **What were the processes Ramp Controllers used to meet TOBTs and TMATs?**
- **Were there differences between conditions regarding perceived operational efficiency?**
- **What were the main suggestions from Ramp personnel on improving RTC?**
- **What can we conclude?**

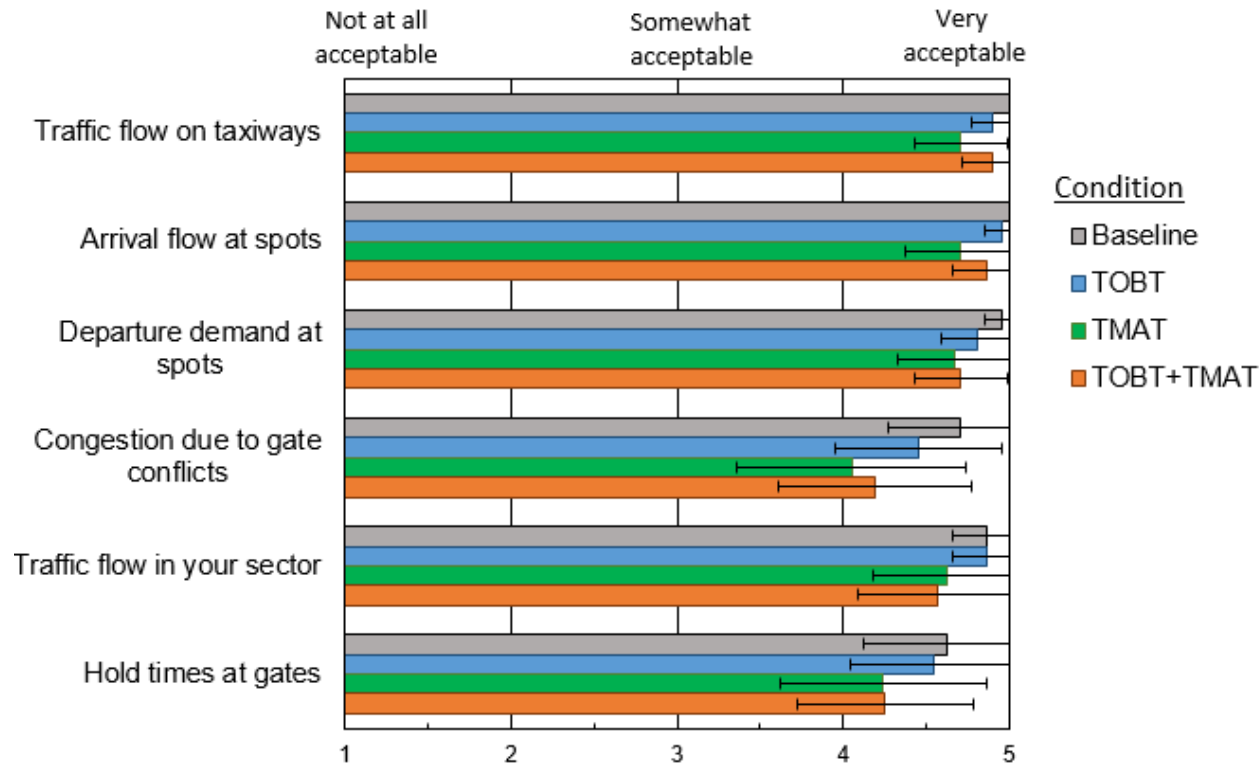


# Ramp Controllers' Post-Run Ratings of Operational Efficiency



During the busiest time in this run, how acceptable were the following in terms of operational efficiency? Comments?

Error bars = 95% CIs; Ns = 19-21 ratings for each item in each condition.



There were no statistically significant differences between the conditions. However, in general, the trend was that Baseline operations and TOBT conditions were seen as most efficient and the two TMAT conditions were seen as least.



- **Were there differences between conditions regarding perceived operational efficiency?**
  - There were no statistically significant differences between the conditions. In general, however, the trend was that Baseline operations and TOBT conditions were seen as most efficient, and the two TMAP conditions were seen as least.

- **Compliance with pushback advisories (TOBTs) and spot times (TMATs)**
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- **Were the metering performance benefits that were observed in the field also observed in the HITL?**
- **What factors reduced the ability to measure the benefits of metering?**
- **What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?**
- **What were the processes Ramp Controllers used to meet TOBTs and TMATs?**
- **Were there differences between conditions regarding perceived operational efficiency?**
- **What were the main suggestions from Ramp personnel on improving RTC?**
- **What can we conclude?**

# Results Summary

## (See Next Slide for Example Suggestions)



- **What were the main suggestions from Ramp personnel on improving RTC?**
  - P-time displayed without blooming on aircraft targets at gates
  - Way to show angle of aircraft once pushed off
  - More informative arrival and departure lists
  - Way to show “reservations” at spots for arrivals that have already called in
  - Quick way to amend pushback direction



- “NEEEEEED to have the posted ETD time displayed (P-time). Without the posted time displayed (without having to bloom the target) it wouldn't allow for the Ramp Controller to pre-plan their pushes accordingly. Also after talking with several other Ramp Controllers in DFW, we need to have something that displays which aircraft I assigned to a certain spot...something called "spot assist" or something that you can turn on and off depending on if you want it or not. Several Ramp Controllers said that aircraft would get forgotten about if they didn't show a visual reminder above the spot that was assigned after the aircraft called "on the ground.”” (Term. C)
- “The icons for aircraft should be changed from boxes to an aircraft icon with the information near the tail of the aircraft. The following information should be the only information needed for ramp control in DFW: Flight number (1492), tail number (3PG), departure time or p-time (18:00), destination city (SJC) with gate information as an option (A21X)” (Term. A)
- “[Need to have] actual departure time and arrival time [list] to move traffic on/off ramp safely without ramp congestion.” (Term. E)
- “Better arrival and departure lists.” (Ramp Manager)
- “For the arrivals list, aircraft stay on list long after the aircraft has parked at the gate.” (Term. E)
- “The ability for the controllers to have a quicker way to amend a push back change.”

## Fewer comments on

- “Need to have the ability on all gates to push a certain tail direction and either ‘tow forward’ OR ‘extending the push further’ ..i.e., ‘American 1234 C29 push tail south and then tow forward and drop your nose on C27.’ This would allow for another aircraft to be parked on C30 without the tail of the C29 aircraft blocking the gate.” (Term. C)
- “I noticed when an aircraft was asked to hold on the gate, there was no counter to alert you as to how long the flight was held on the gate.” (Ramp Manager)
- “Having more views on the console would be helpful, to be able to see different parts of the ramp area, and be able to see those views quickly.” (Ramp Manager)

- Compliance with pushback advisories (TOBTs) and spot times (TMATs)
  - Were the Ramp Controllers able to comply with the TOBTs?
  - Were the Ramp Controllers able to comply with the TMATs?
  - How did the conditions affect the timing of compliances at the spot?
- Were the metering performance benefits seen in the field also seen in the HITL?
- What factors reduced the ability to measure the benefits of metering?
- What were the impacts of the conditions on Ramp Controllers' workload and situation awareness?
- What were the processes Ramp Controllers used to meet TOBTs and TMATs?
- Were there differences between conditions regarding perceived operational efficiency?
- What were the main suggestions from Ramp personnel on improving RTC?
- **What can we conclude?**



- TMAT compliance metrics in all three metering conditions were comparable, ranging from 83% to 85% compliant. When aircraft were initially compliant with gate advisories, compliance at the spot increased, ranging from 92% to 99%.
- There was no noticeable difference between the Baseline and metering conditions in performance metrics mainly due to simulation artifacts.
  - It was challenging to simulate real DFW operations. The simulation had unrealistically long pushback times, and limited options in pushback direction, taxi routes, and taxi speed in the ramp. As a result, the Ramp Controllers were not able to push back aircraft freely even in the Baseline condition, and ended up holding aircraft at the gate, effectively metering departures.
  - The fidelity of simulations will continue to improve due to the increased ability to access operational data from the field.
- The combined TOBT+TMAT condition resulted in higher workload on the WAK than the Baseline and TOBT alone conditions, and lower situation awareness than the Baseline condition. The TOBT condition was seen as closest to Baseline, which was perceived as the most efficient condition on many subjective operational efficiency measures.
- Metering at Dallas/Fort Worth International Airport (DFW) with TOBT only or TMAT only could be equally effective, and either would be a better option than TOBT + TMAT due to increased workload and reduced situation awareness.

# Appendix A

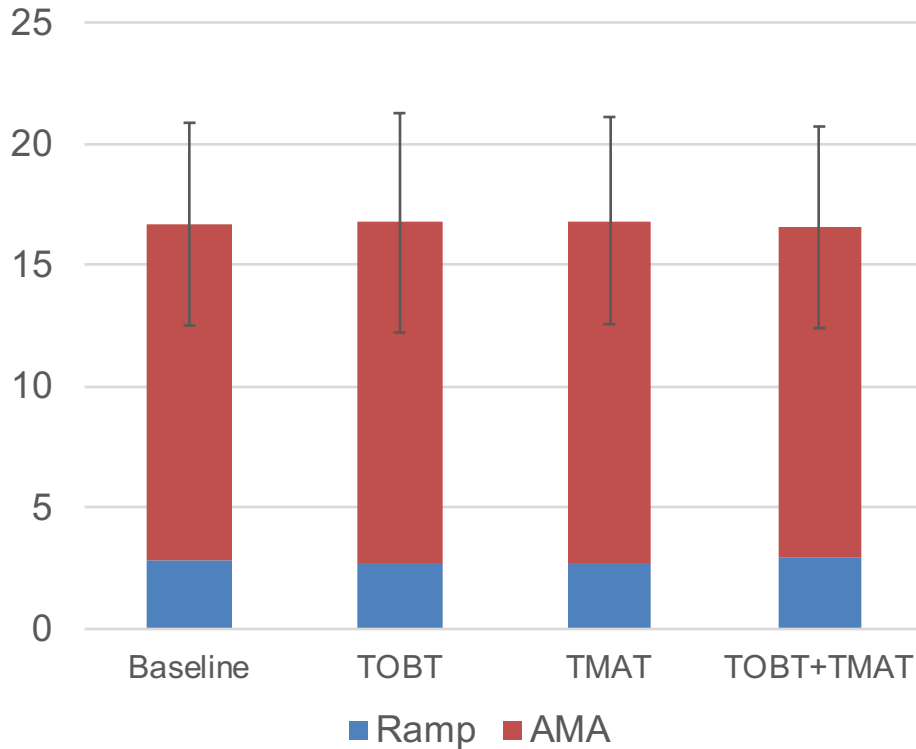
## Other Performance Metrics



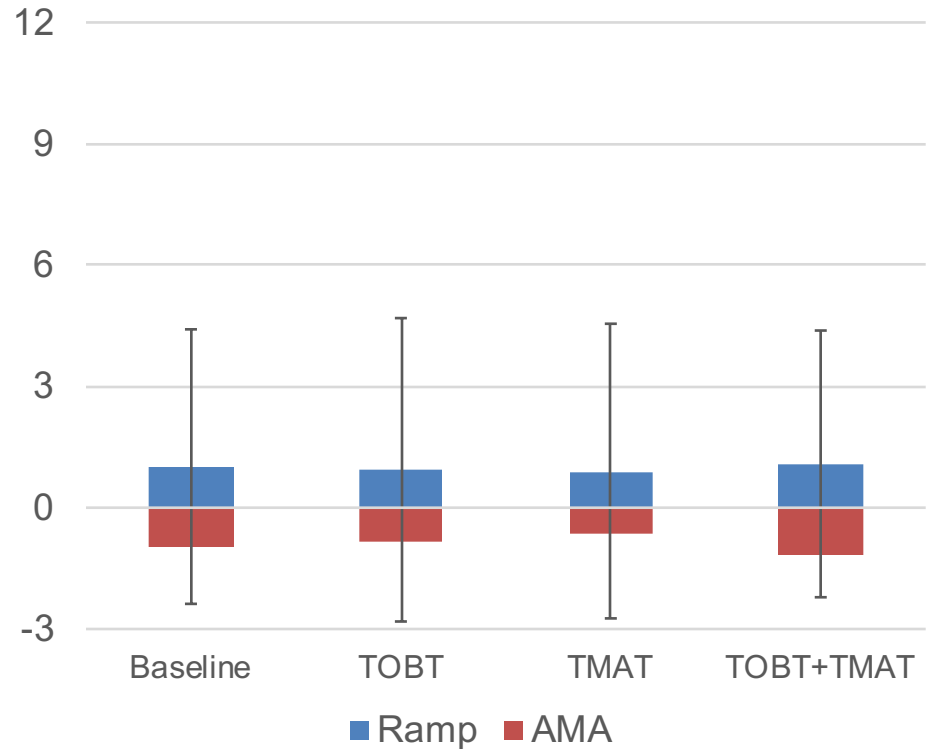
# Average Taxi-In Times – Terminal A, C and E Only



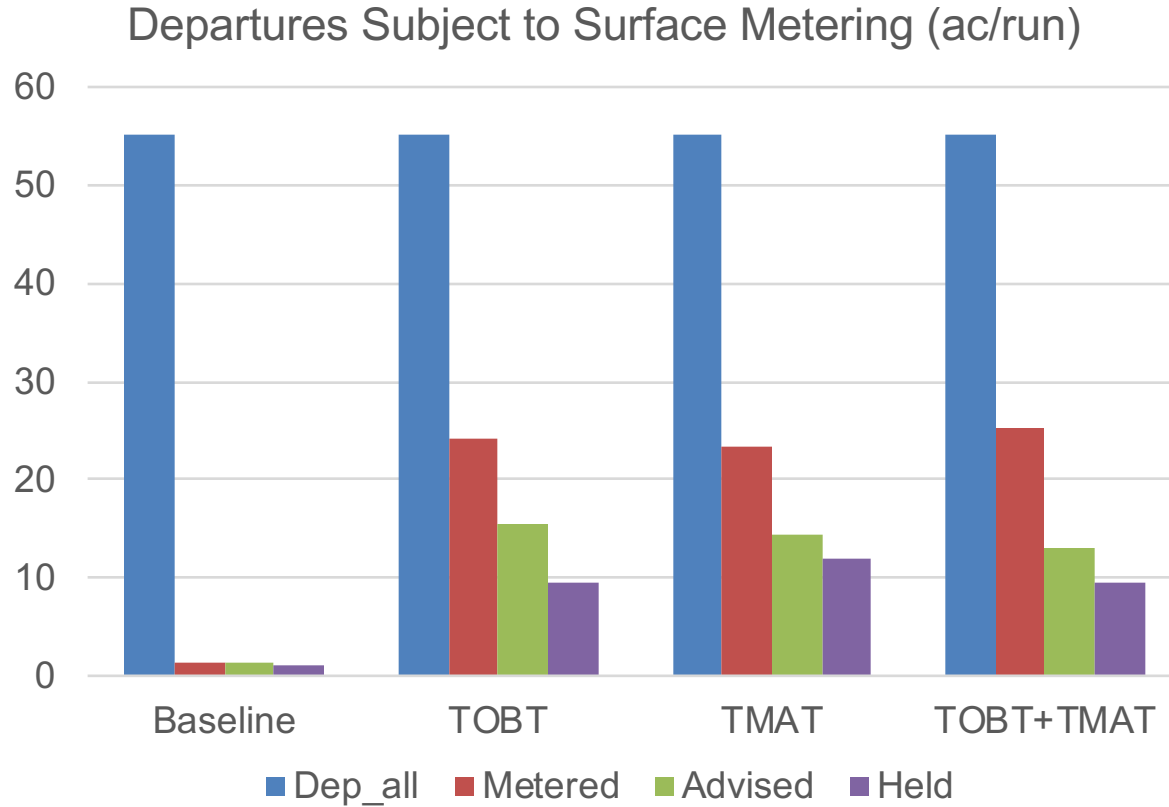
Average Taxi-In Time (minutes)



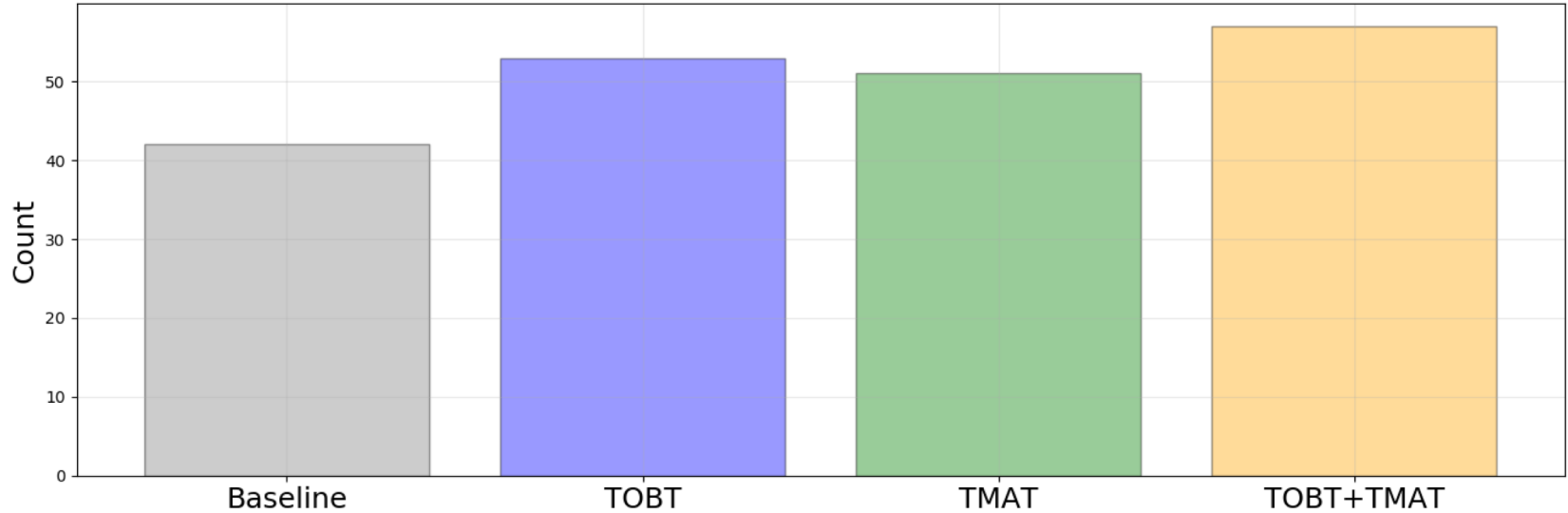
Average Excess Taxi-In Time (minutes)



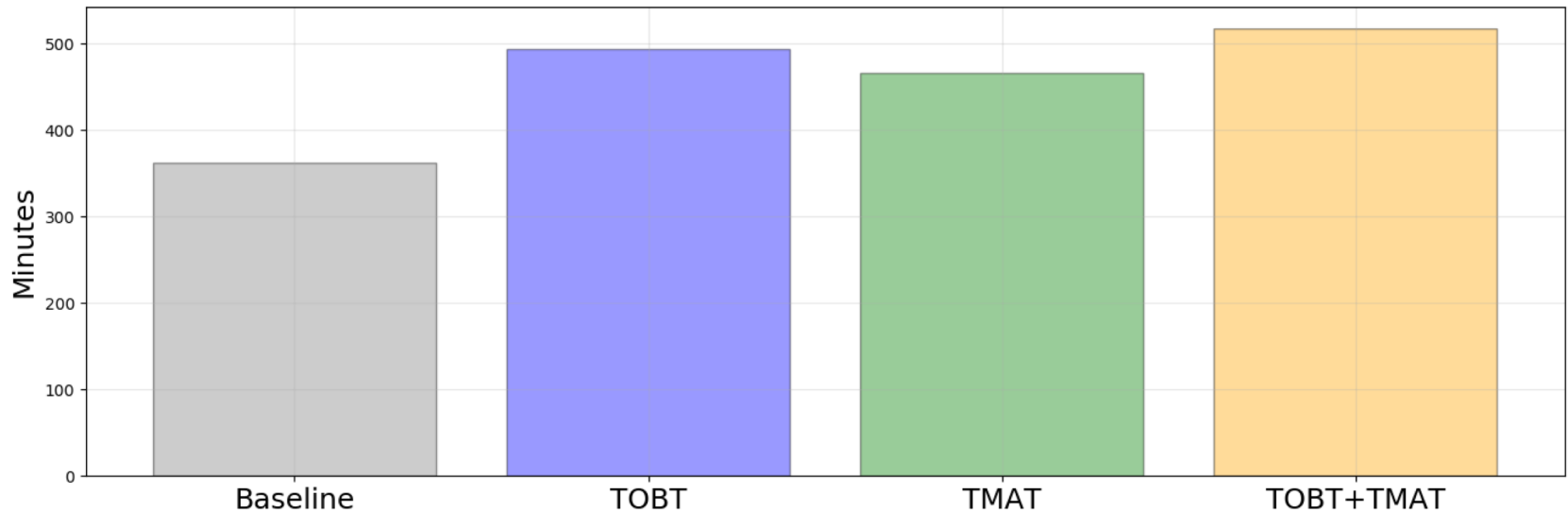
# Departures Subject to Surface Metering – Terminal A, C and E Only



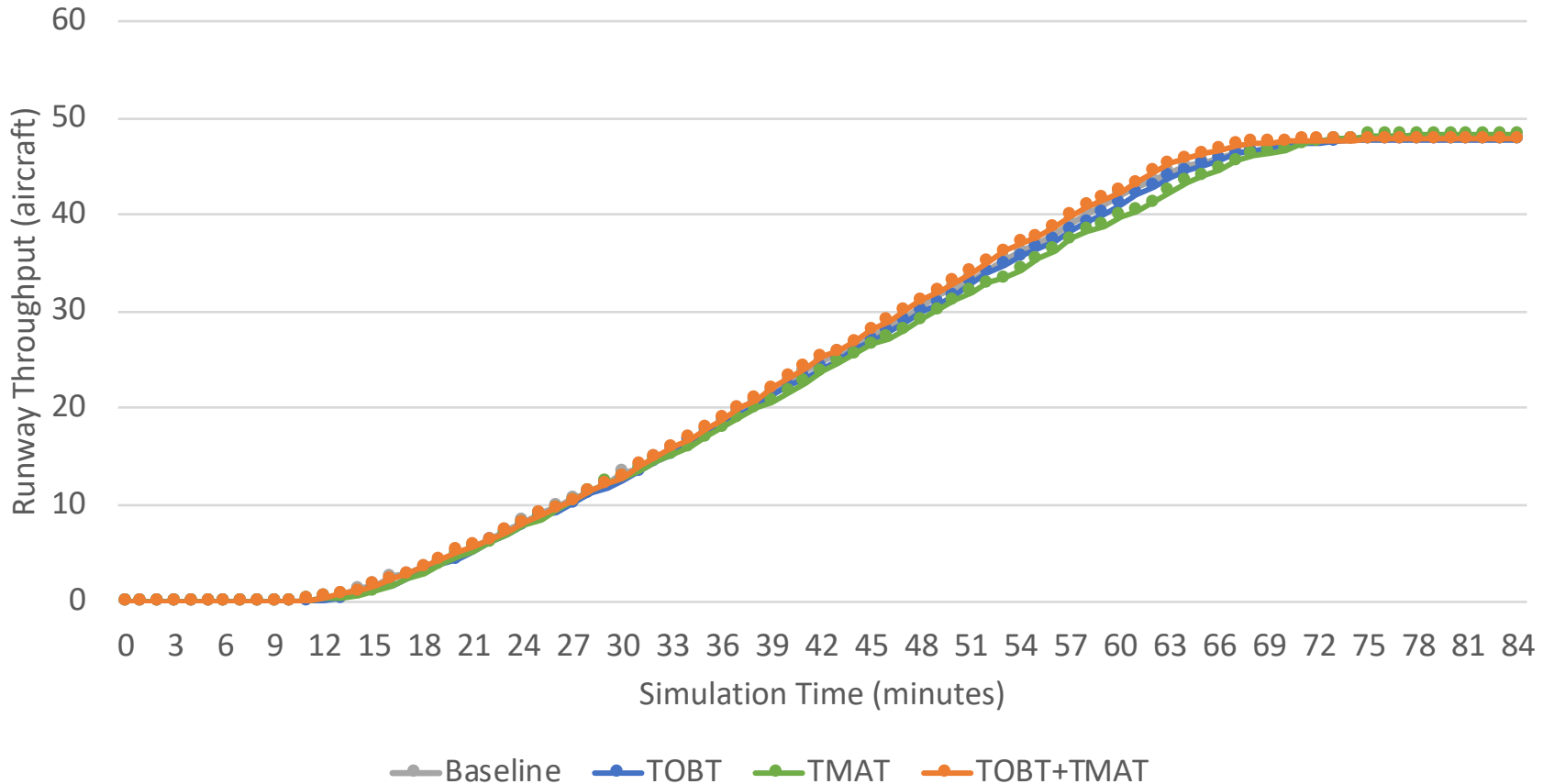
Count of Gate Conflicts



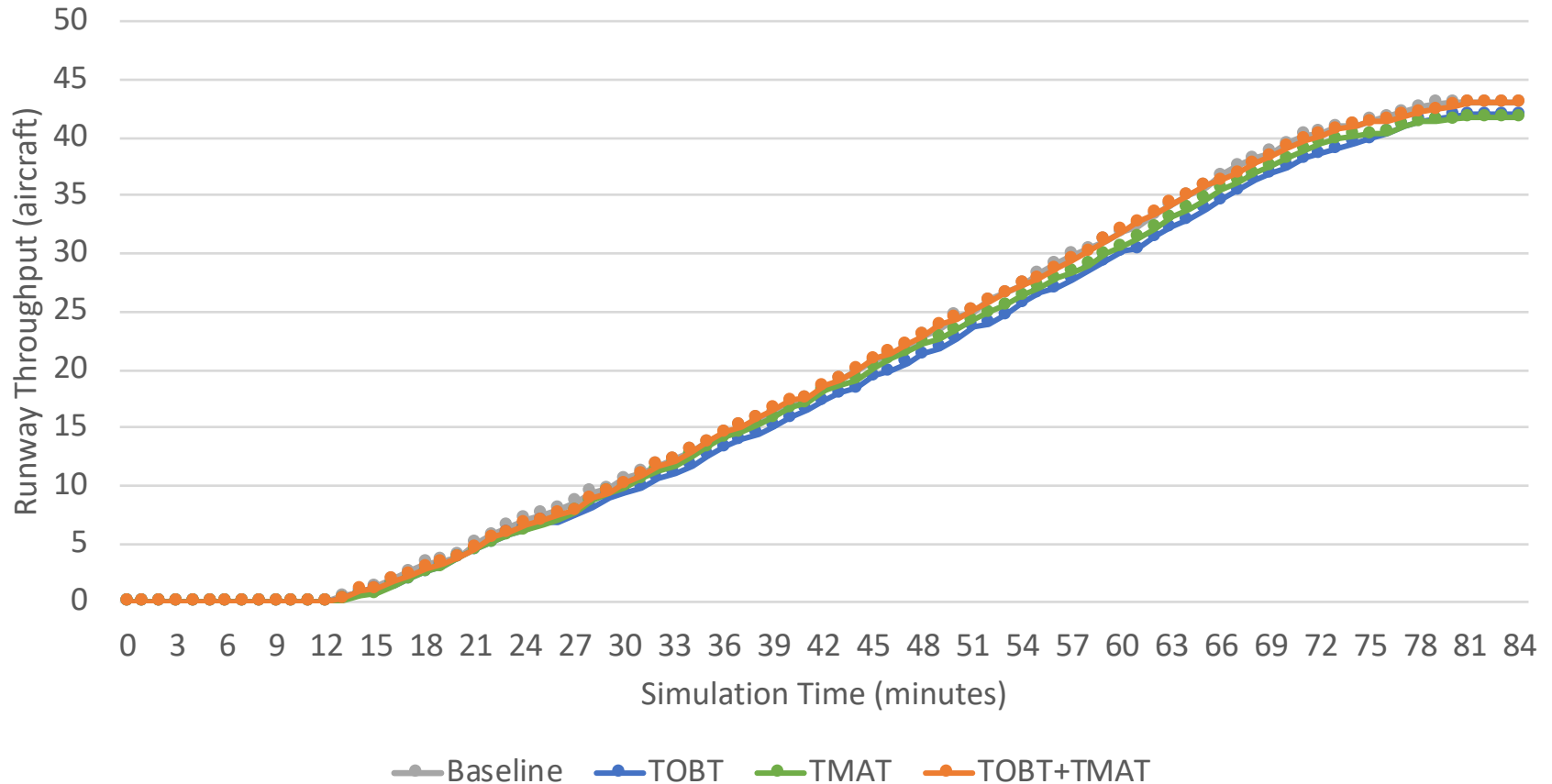
Total Duration of Gate Conflicts



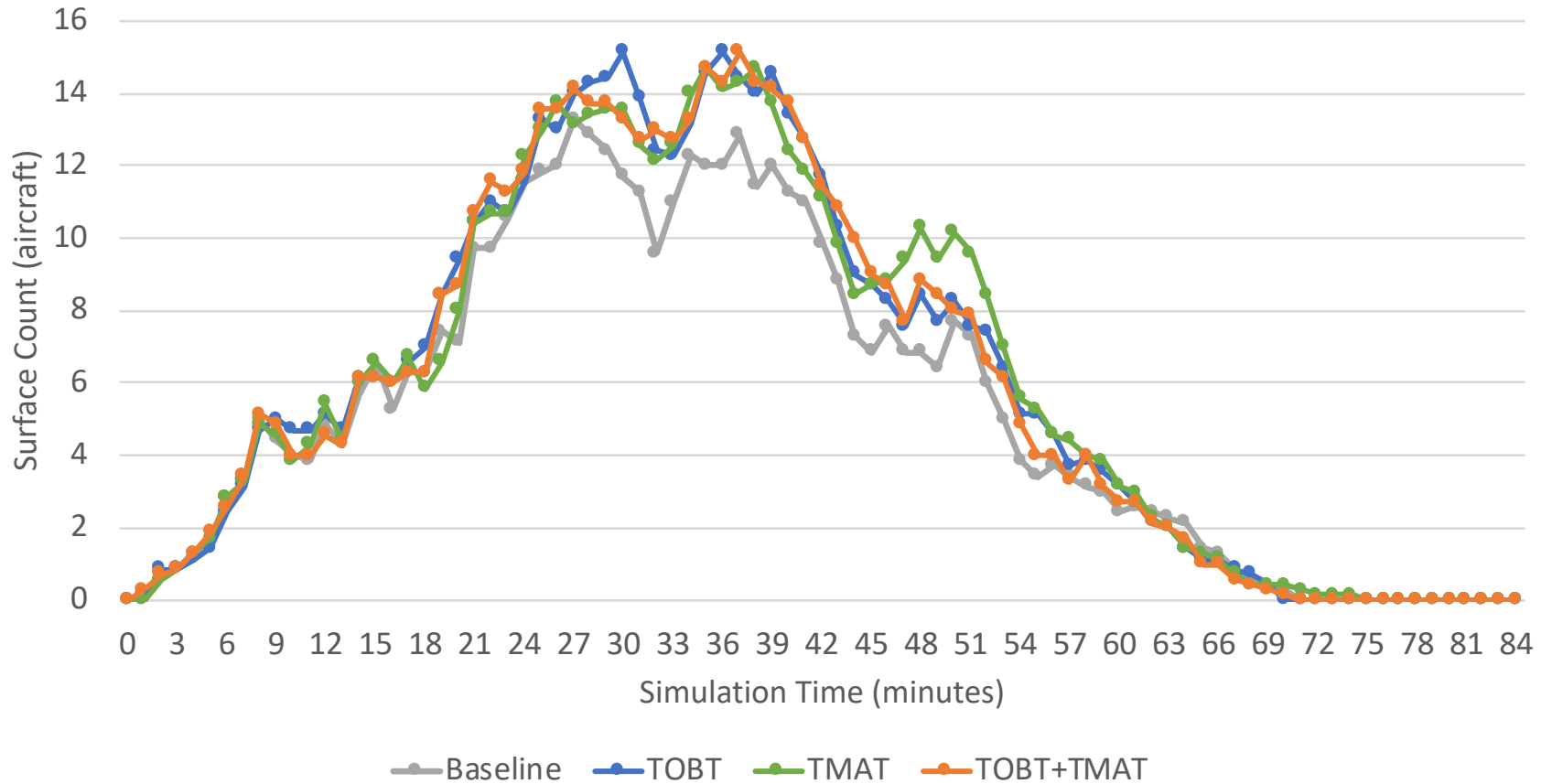
Number of Departures that Completed Takeoffs on Rwy 17R



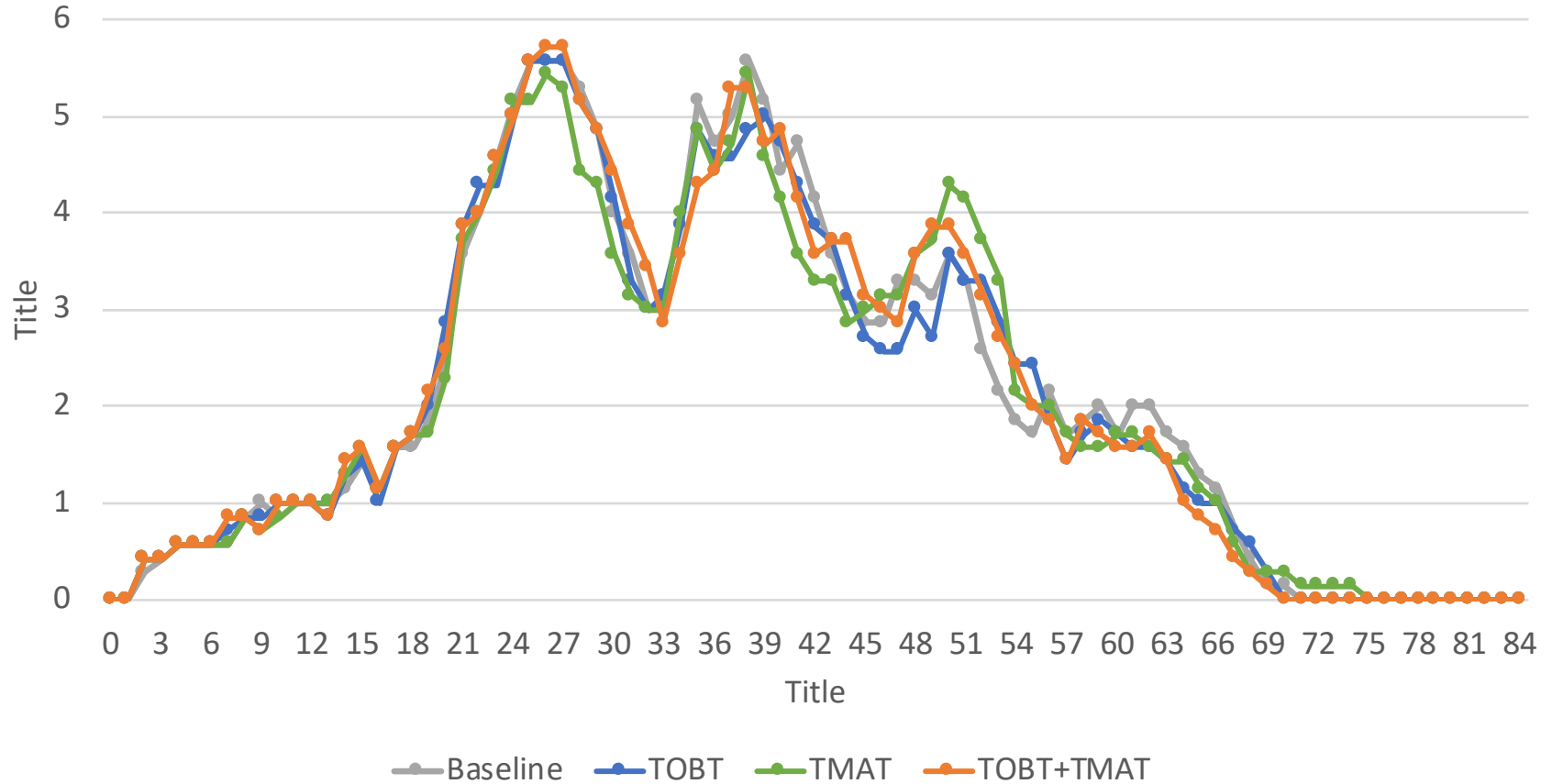
Number of Departures that Completed Takeoffs on Rwy 18L



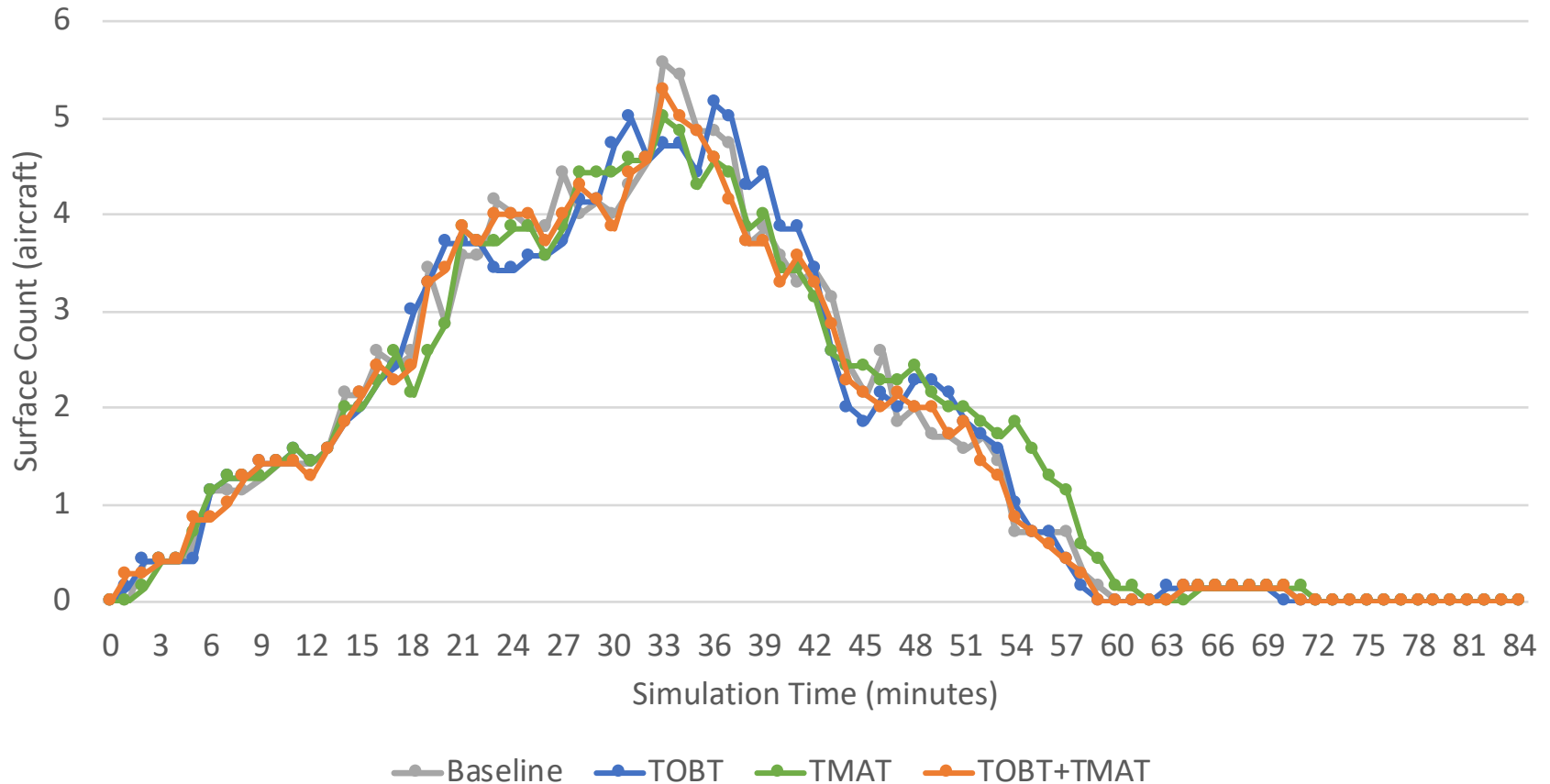
Number of Departures in Ramp Area



Number of Departures in Ramp Area

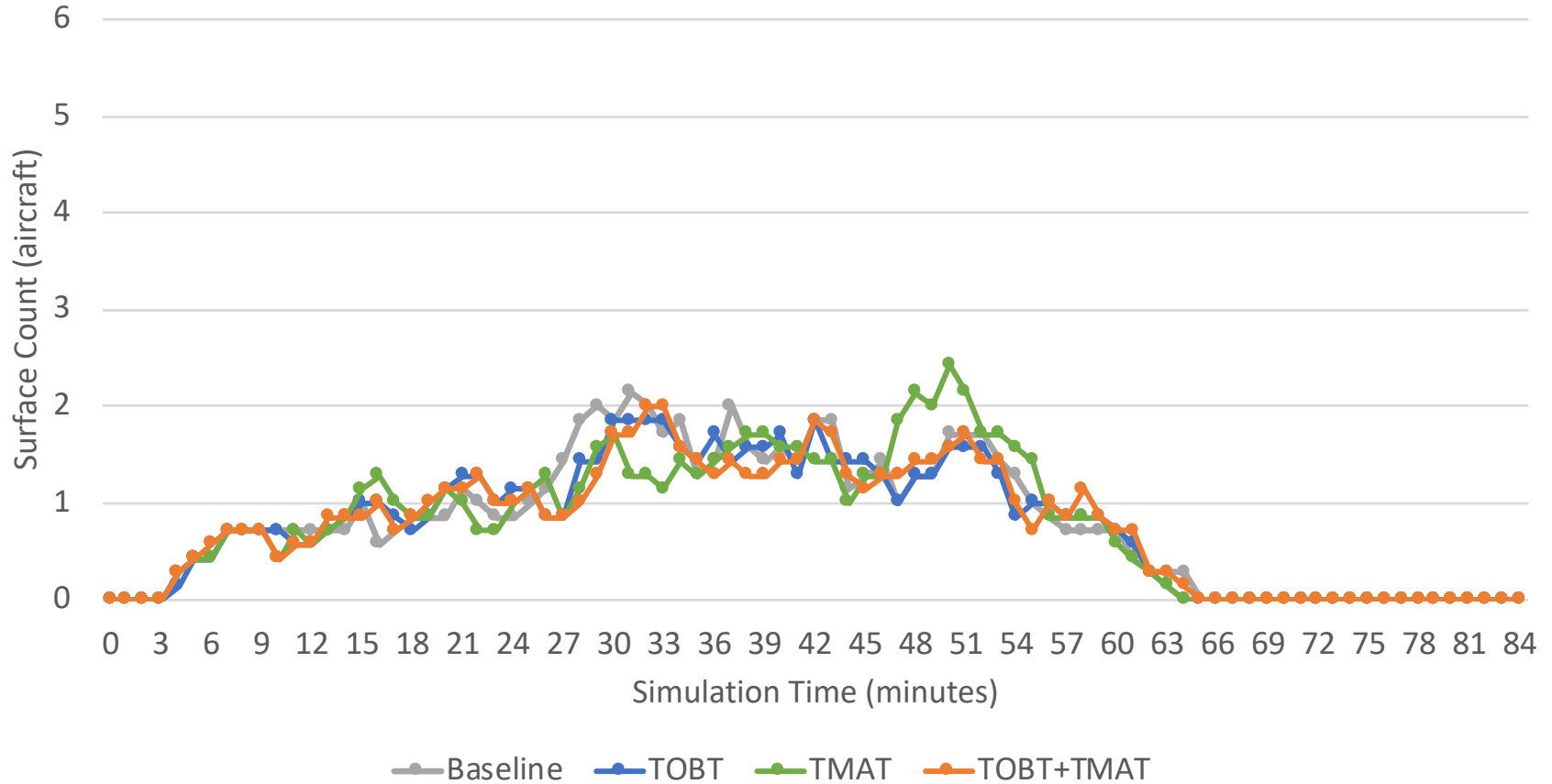


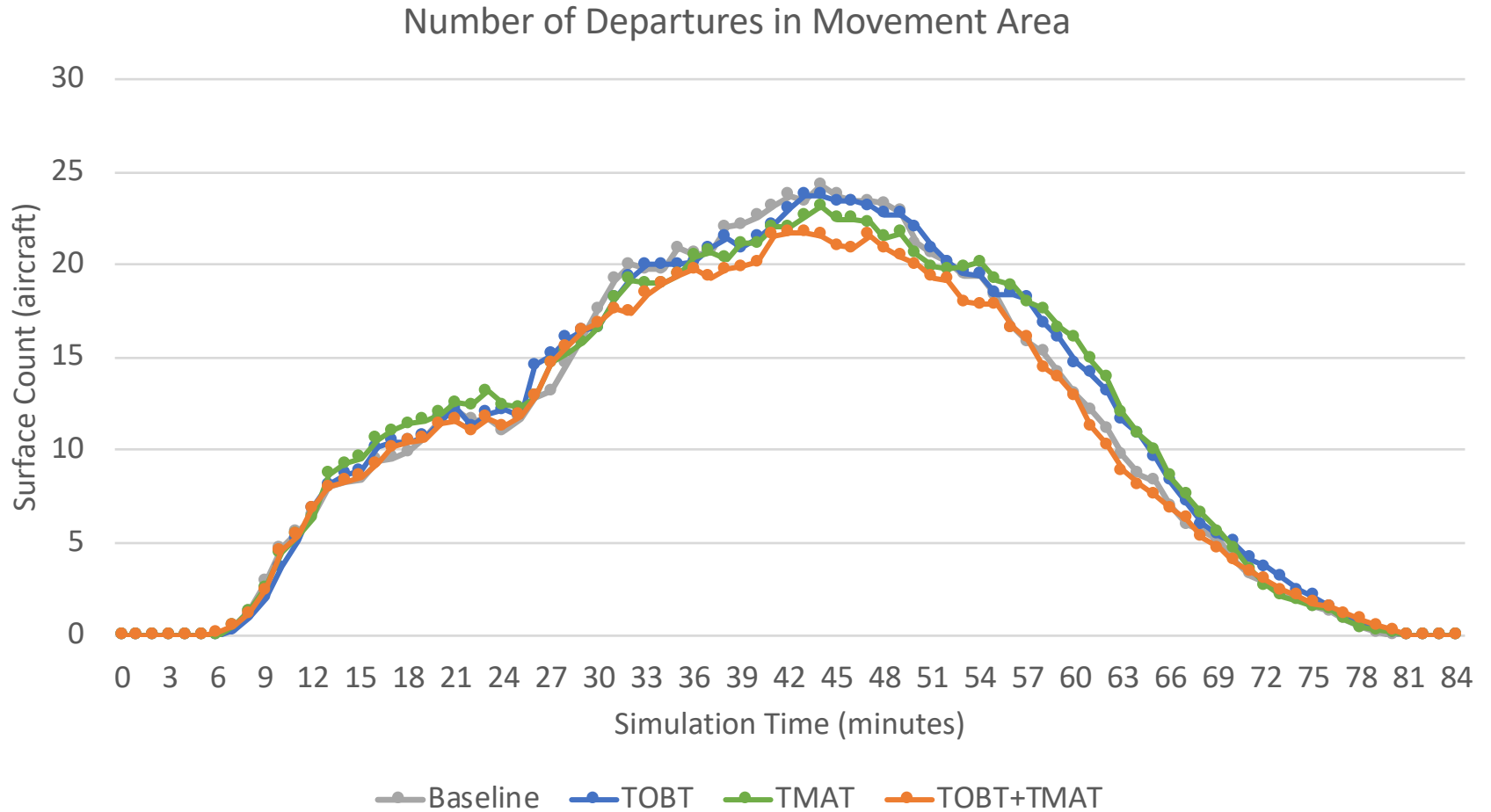
Number of Departures in Ramp Area



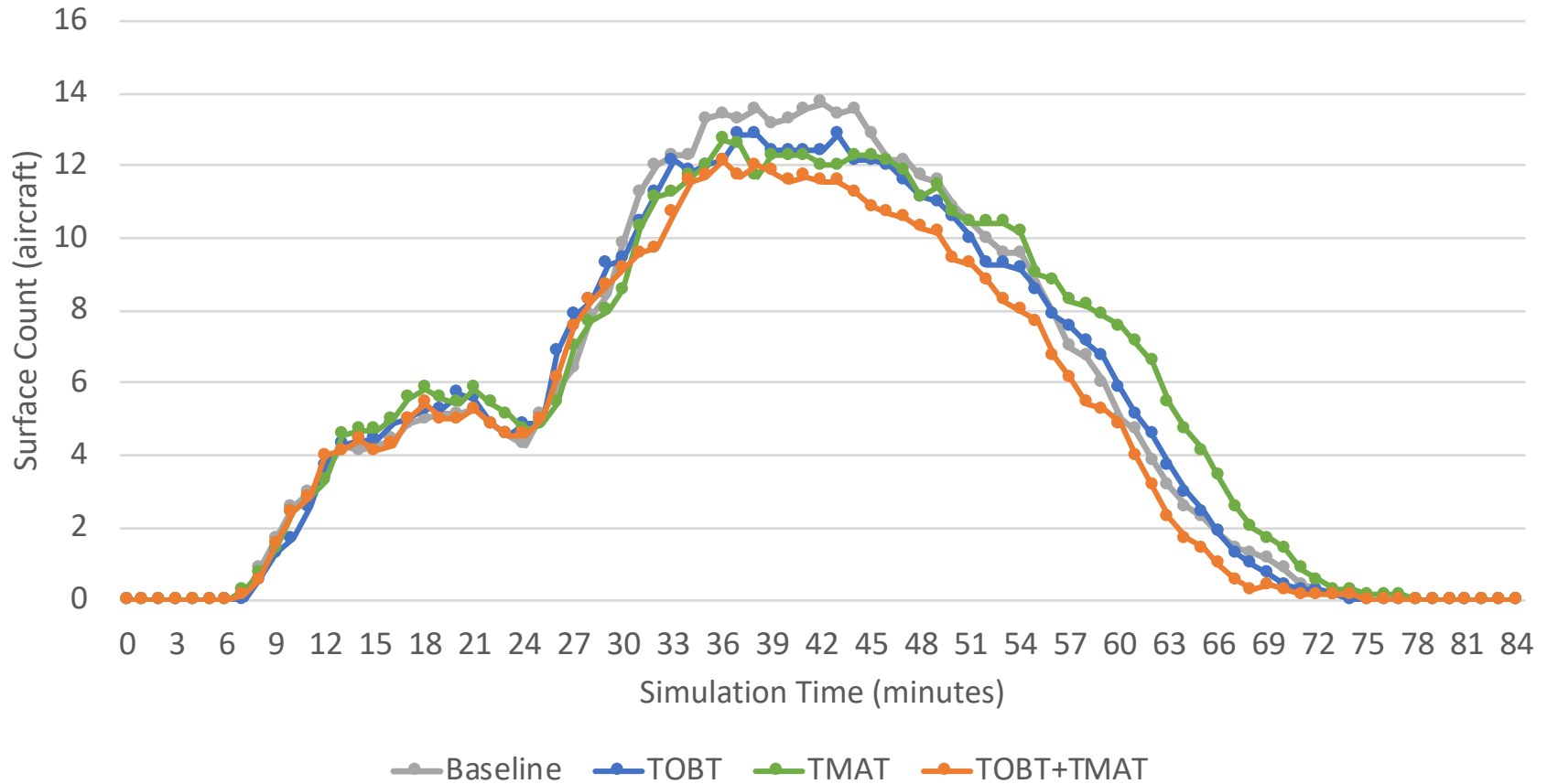


Number of Departures in Ramp Area

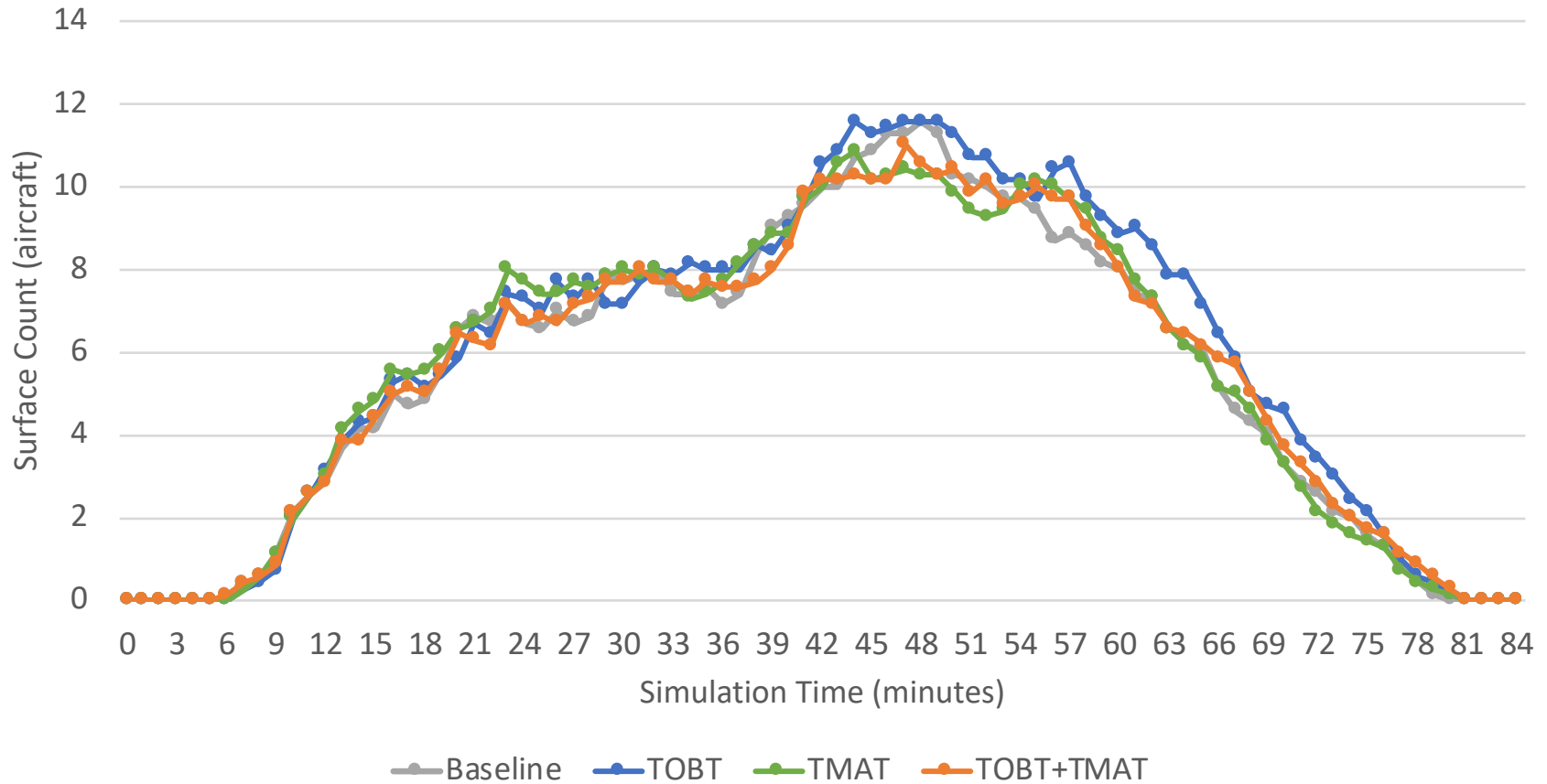




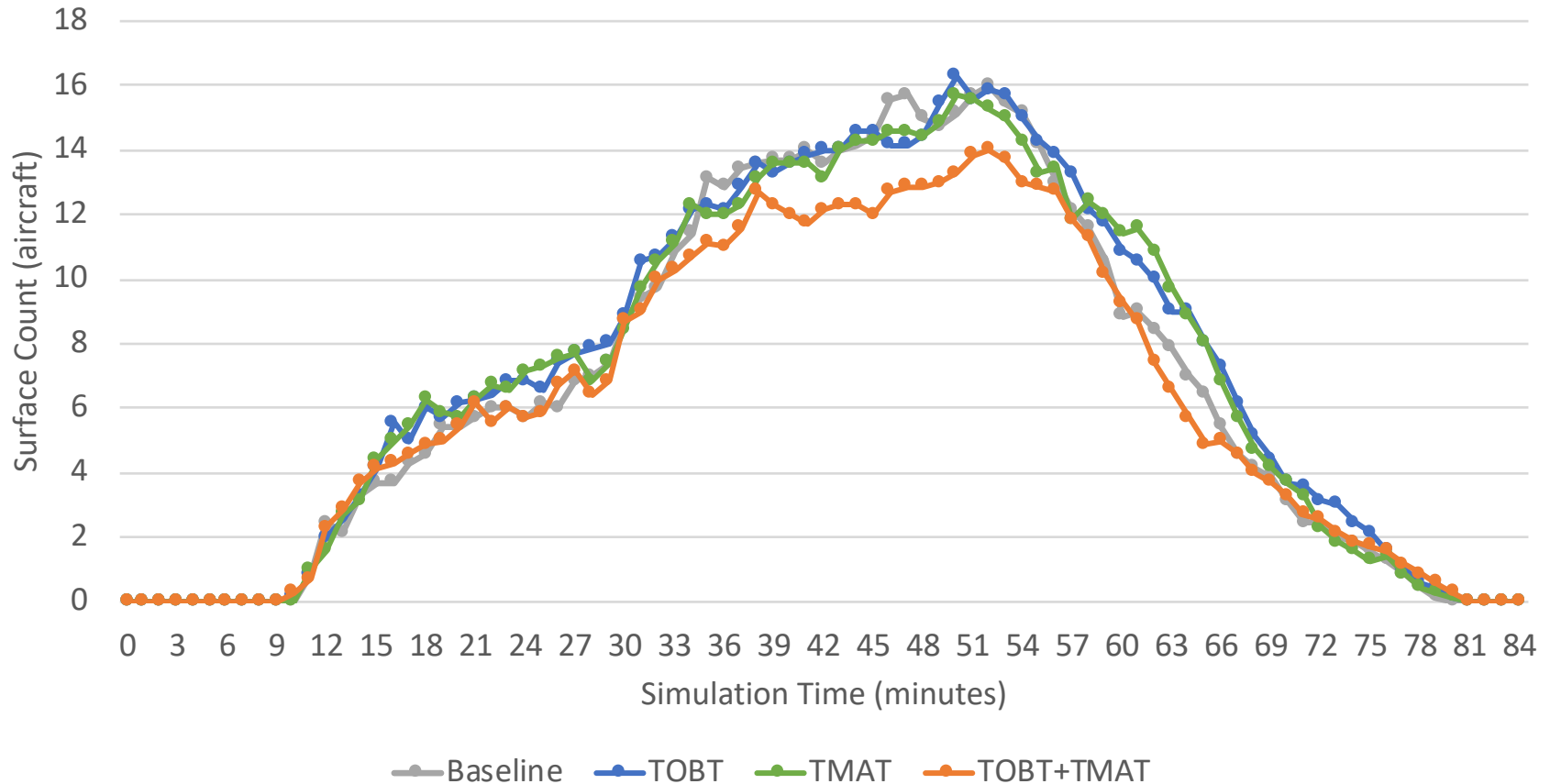
Number of Departures in Movement Area for Rwy 17R



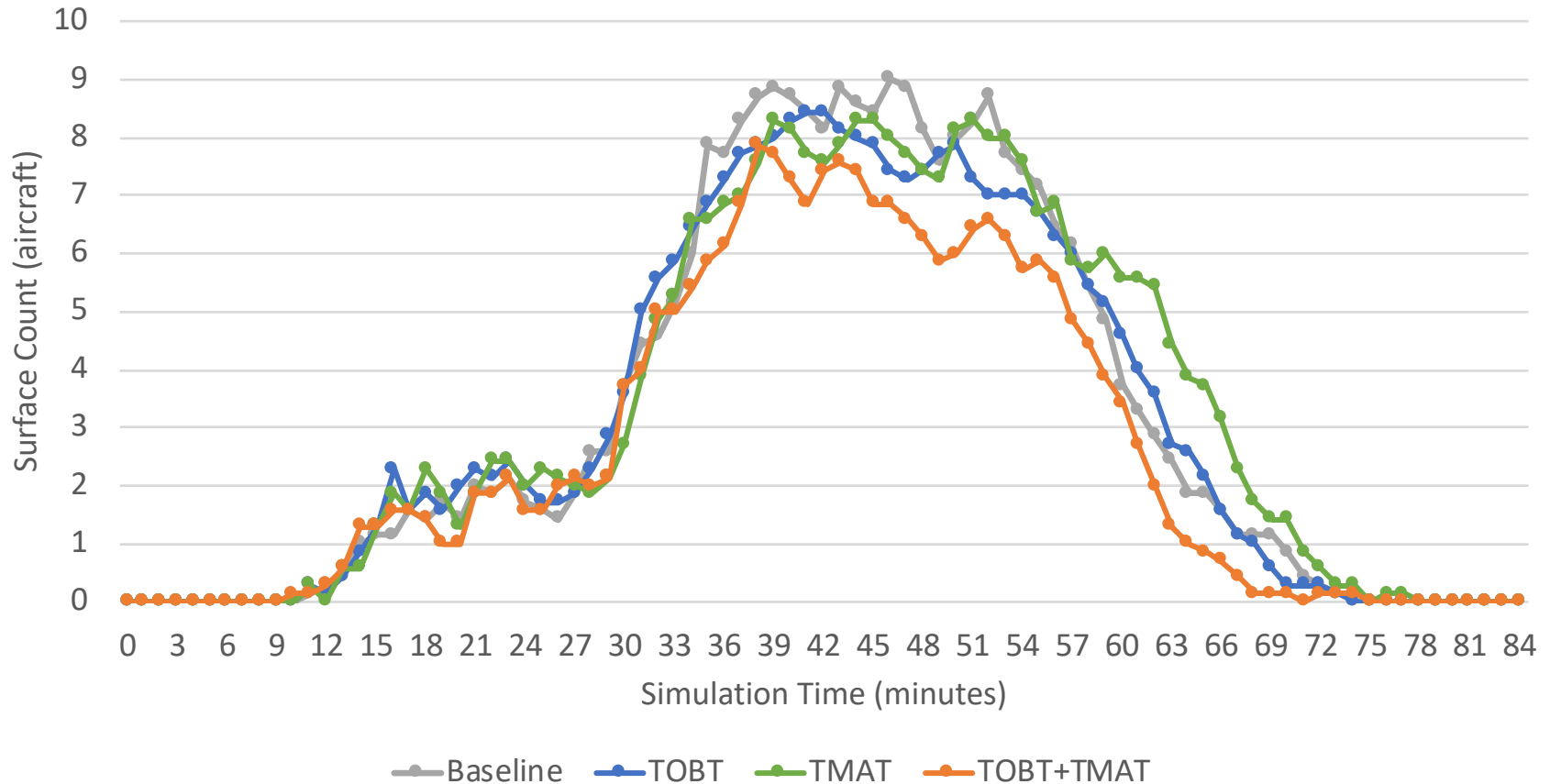
Number of Departures in Movement Area for Rwy 18L



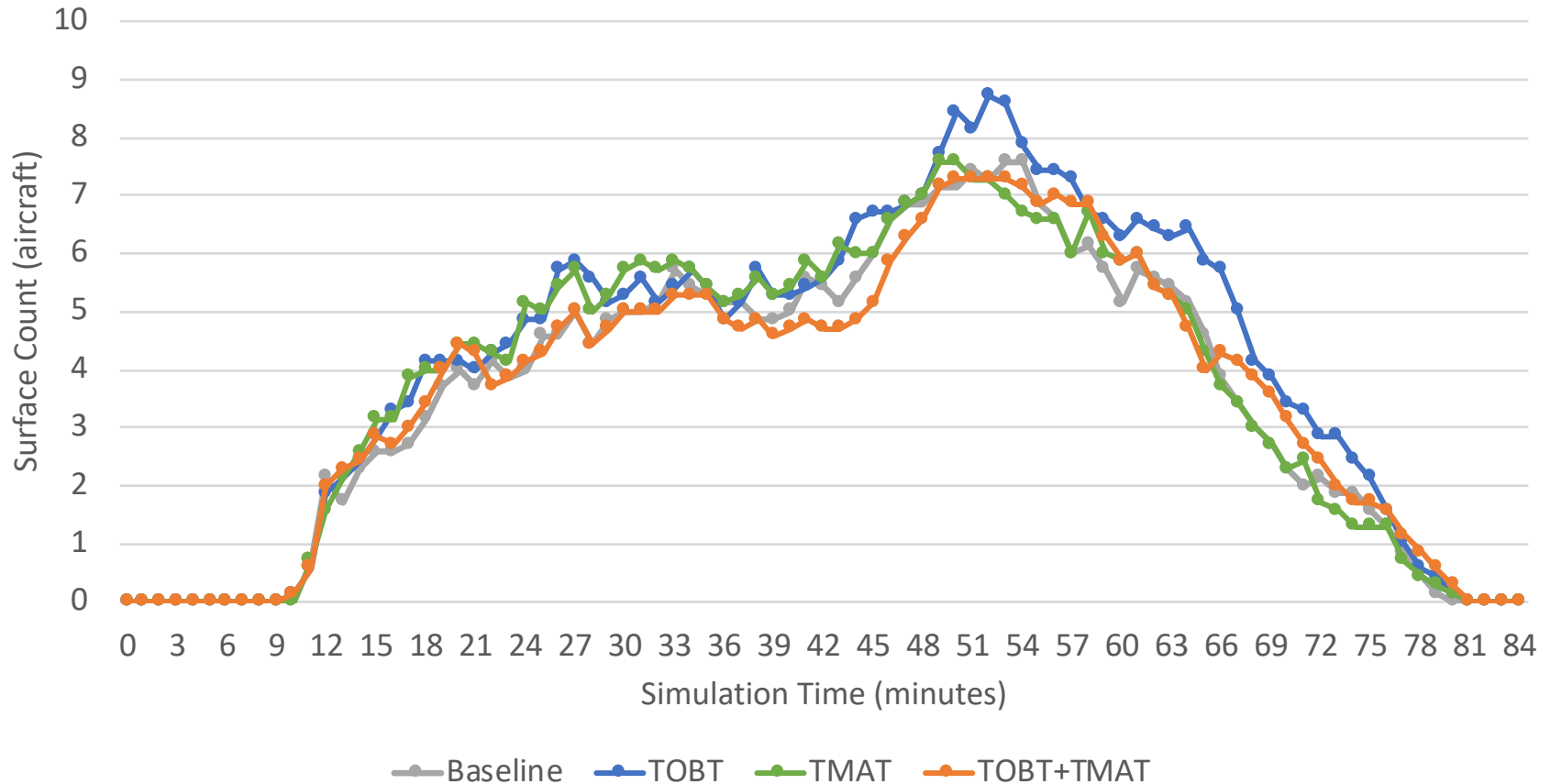
Number of Departures in Queues



Number of Departures in Queues for Rwy 17R



Number of Departures in Queues for Rwy 18L



# Departure Queue Size: Baseline vs. TOBT+TMAT



**Baseline (run25)**

**TOBT+TMAT (run26)**

Sim Time: 50 min

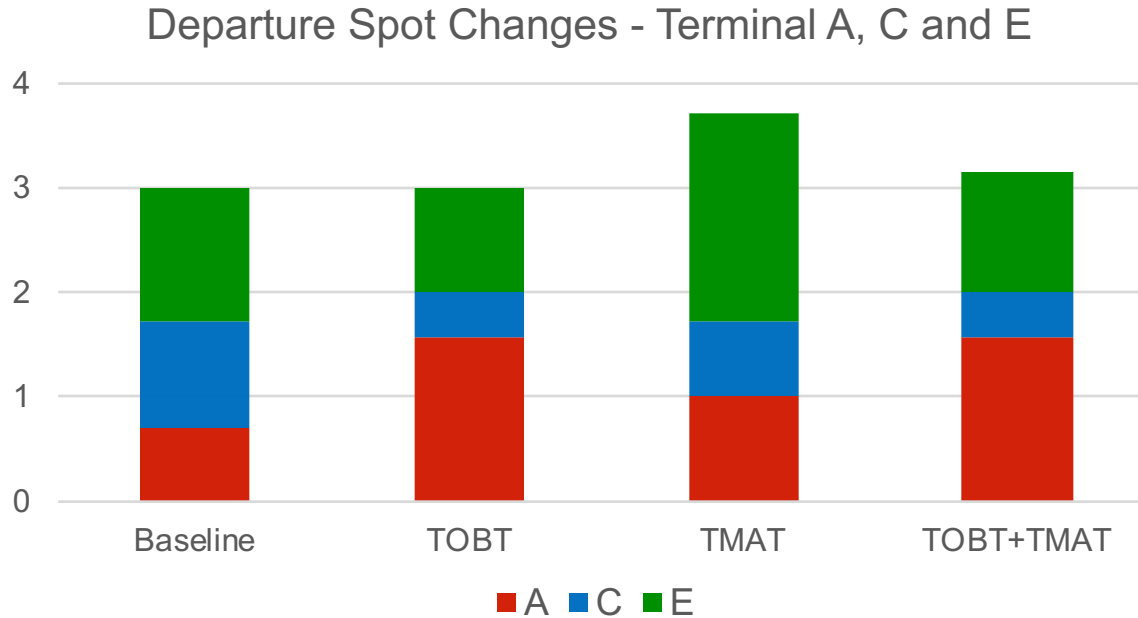


Sim Time: 55 min





- Departure spots tactically changed by Ramp Controllers



Terminal	All	A	B	C	D	E
Baseline	4.7	0.7	1.7	1.0	0.0	1.3
TOBT Only	6.0	1.6	2.9	0.4	0.1	1.0
TMAT Only	6.1	1.0	2.0	0.7	0.3	2.0
TOBT+TMAT	4.9	1.6	1.6	0.4	0.1	1.1

# Appendix B

## Other Subjective Metrics

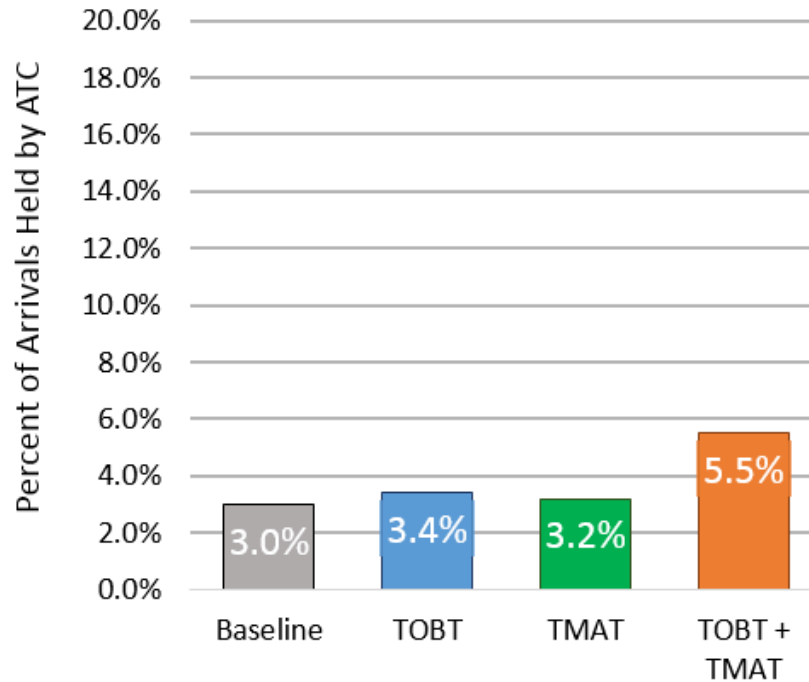


- “FAA Holds” were a work-around in the simulation designed to duplicate pilots’ early call-in’s (on arrival) to the Ramp Controllers to see what gate the pilots were assigned and whether their gates were occupied
- Having the Ramp Manager automatically signal the Ground Controllers to hold gate-conflicted aircraft on the AMA if the Ramp Controllers wished that to happen, was a factor in reducing the workload of the Ramp Controllers from what it is in the field
- However, FAA Holds were an indication of the amount of ramp congestion and Ramp Controller workload, and as such, support the finding of higher controller workload and less acceptable operational efficiency in the TOBT+TMAT condition (see next slide)

# FAA (ATC) Holds on AMA by Condition: Video Analysis



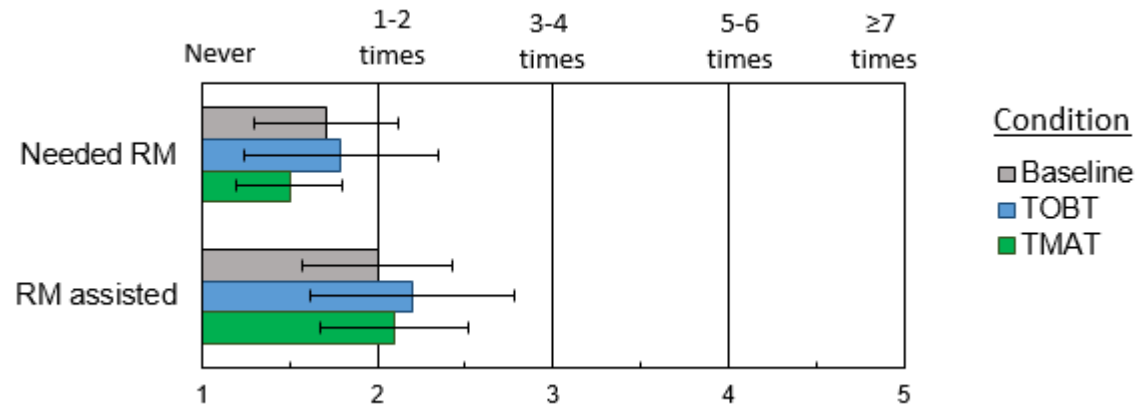
N's = 566 arrivals in Baseline, 668 in TOBT, 663 in TMAT, and 562 in TOBT+TMAT (unequal N's due to missing videos in 2 runs, 1 in Baseline--Run 20, and 1 in TOBT+TMAT--Run 18); Chi square (*df*3) = 6.54, *p* = .09.



There was a slightly higher percentage of aircraft held by ATC on the AMA in the TOBT+TMAT condition than in the other conditions, although this was not statistically significant. This supports more congestion and higher workload in this condition. The average hold time was 6:09 minutes (range 0:08 to 19:36) and did not differ by condition.

A Ramp Manager can help operations by communicating with other Ramp Controllers or Ground, advising on problems, or monitoring metering. In this run, about how often did you need the Ramp Manager to assist in these ways? In this run, about how often did the Ramp Manager assist you in these ways? (Not asked in TOBT+TMAT condition due to length of post-run survey.)

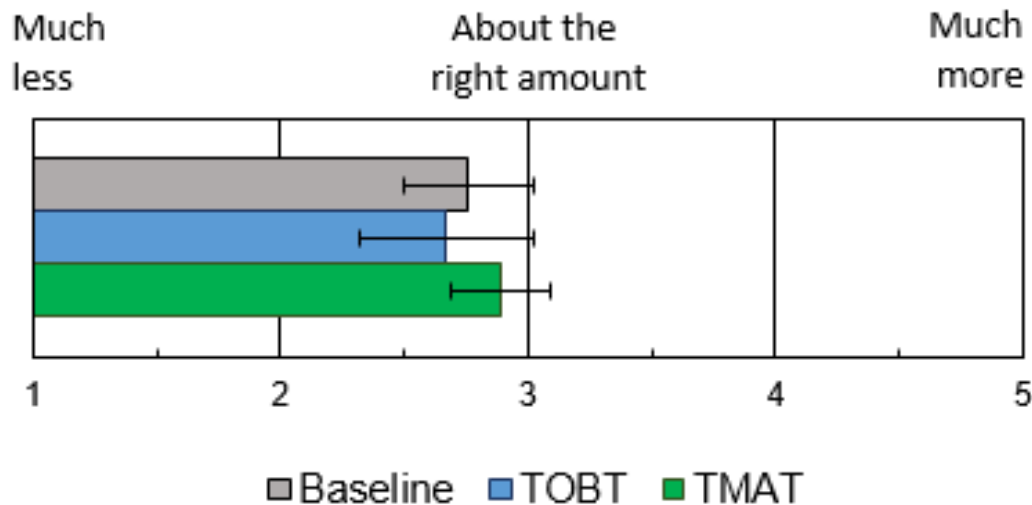
Error bars = 95% CIs; *N*s = 19-21 ratings for each item in each condition.



The RM was rarely needed in these conditions and assisted only a few times. In a paired-t test, the mean of “assist” was 2.1 and of “need” was 1.7,  $p = .001$ ,  $t(59)=4.0$ ,  $N = 60$ . Hence Ramp Controllers got more assistance than they needed in individual cases. There was no difference by condition.

In this run, do you think the Ramp Manager should have assisted you more or less?

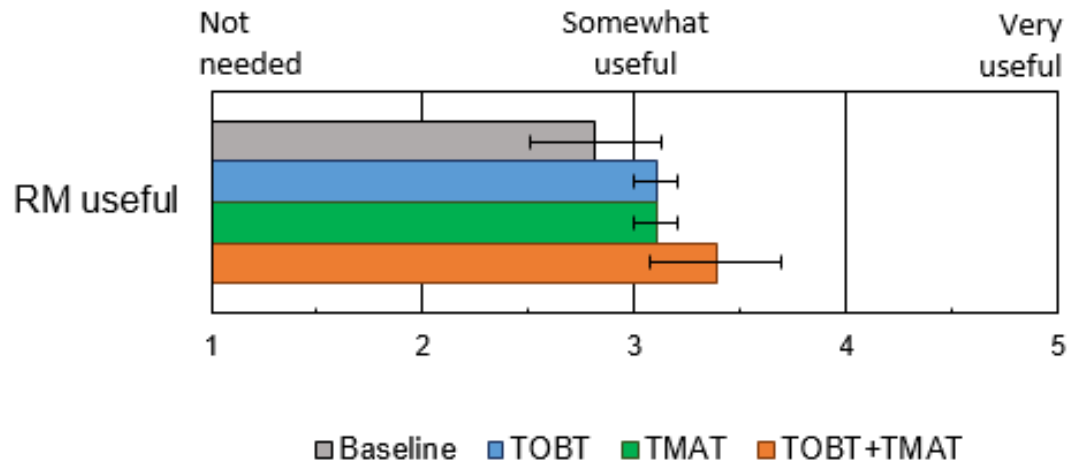
Error bars = 95% CIs; *N*s = 18-21 ratings for each item in each condition.



The Ramp Controllers thought the RM assisted them about the right amount In these conditions.

How useful was it to have a Ramp Manager in the following conditions?

Error bars = 95% CIs (Loftus correction applied).  $N = 7$ . Repeated measures ANOVA,  $F(6,18) = 2.9$ ,  $p = .11$ , Greenhouse-Geisser, sphericity not assumed.



Ramp Controllers rated the Ramp Manager as somewhat useful in all conditions in the simulation. There were no statistically significant differences between conditions. However, in the post-sim surveys, there was a slight tendency for the RM to be seen as least useful in Baseline and most useful in the TOBT+TMAT condition.

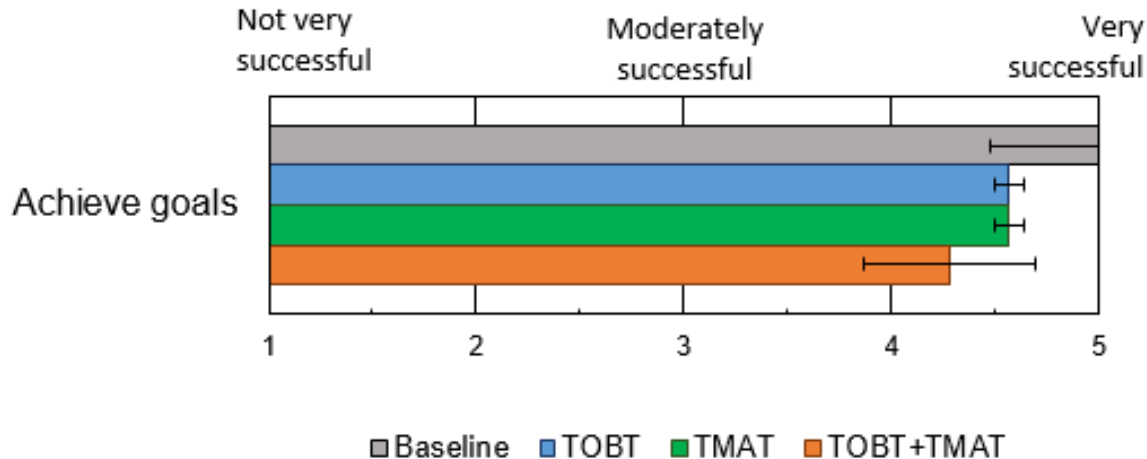


- “In normal ops, the amount of workload is low enough to where I could handle gate conflicts and communicate to other controllers on my own, but as responsibility for gate and spot advisories increases, the busier I become and the amount of workload increases to an uncomfortable point if I didn't have a Ramp Manager.” (Term. A)



Please indicate how successful you were in achieving the goals of each of the four conditions in this simulation.

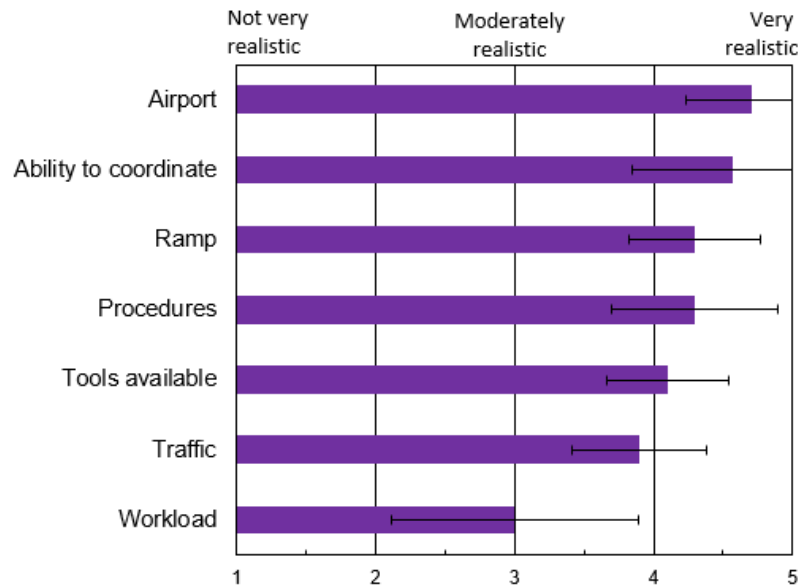
Error bars = 95% CIs (Loftus correction applied).  $N = 7$ . Repeated measures ANOVA,  $F(3,18) = 2.2, p = .12$ .



Ramp Controllers' ratings of success in achieving goals were high in all conditions and not statistically significantly different from each other. Again, it was thought easier to achieve goals in the Baseline condition compared to the other conditions, especially the TOBT+TMat condition, but these differences were not statistically significantly different.

Based on your experience, how realistic was the simulation in terms of the following factors?

Error bars = 95% CIs (Loftus correction applied).  $N = 7$ . Repeated measures ANOVA,  $F(6,36) = 3.1, p = .015$ .



Ramp Controllers' rated all areas on average as being moderately realistic or higher. Realism of workload was statistically significantly lower than realism of ramp, airport, and tools available. Traffic realism was statistically significantly lower than airport. Workload was thought to be not high enough in the simulation, according to RC comments.

Did you have any gate conflicts in this run? (Yes/No)

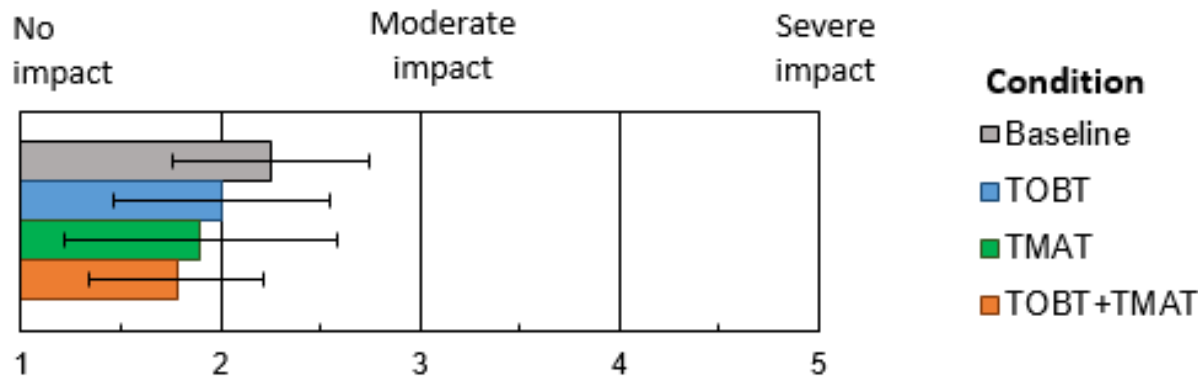
*N* = 82/84 ratings.

	<i>Frequency</i>	<i>Percent</i>
Yes	44	52.4
No	38	45.2
<i>Total</i>	<i>82</i>	<i>97.6</i>
Missing	2	2.4
<i>Total</i>	<i>84</i>	<i>100</i>

In the 28 runs, there were 44 times Ramp Controllers said they had one or more gate conflicts. The number per condition averaged between 9-13 and was not statistically different between conditions.

At the busiest time in this run, what was the impact of gate conflicts on operational efficiency in your sector?

Error bars = 95% CIs; *N*s = 9-13 ratings for each condition, out of the 44 who said who said they had a gate conflict.



The impact of gate conflicts on operational efficiency was judged as small, and was not statistically significantly different between conditions.

# Ramp Controllers' Post-Run Ratings of Difficulty Managing Gate Conflicts



Please rate the difficulty of managing the gate conflicts at the busiest time in this run.

Error bars = 95% CIs; *N*s = 9-13 ratings for each condition, out of the 44 who said they had a gate conflict.



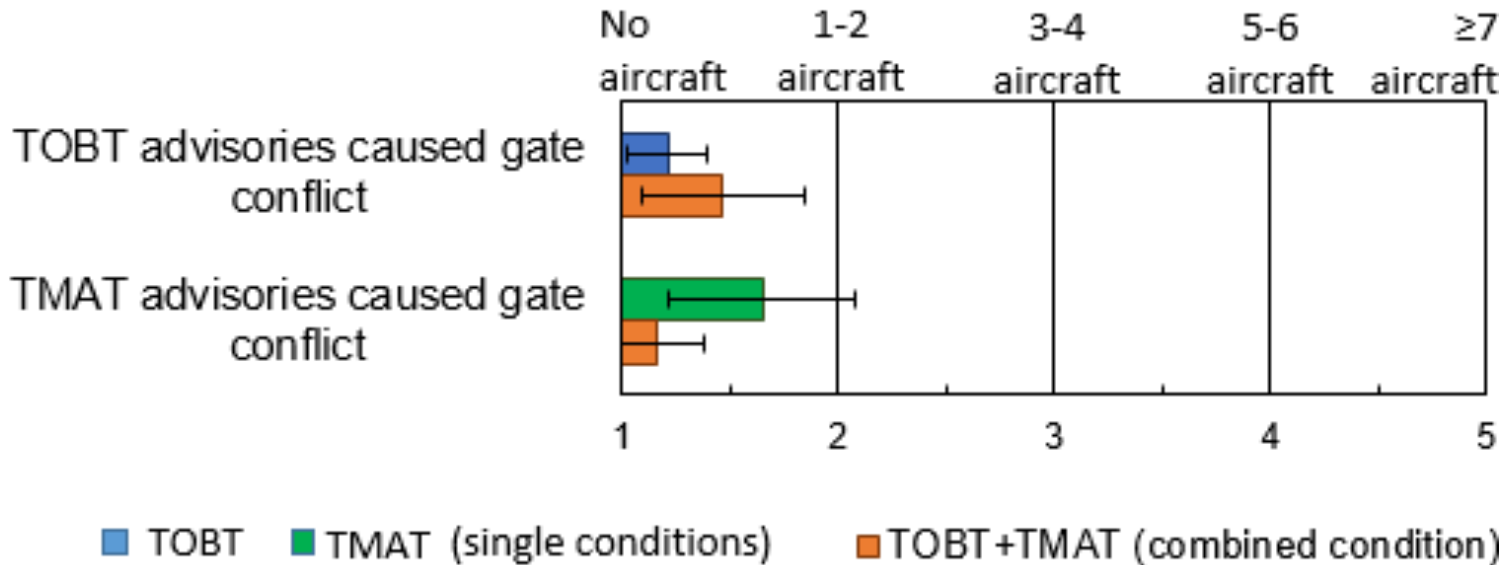
Managing gate conflicts, on average, was not judged as difficult in any condition. Many controllers responded that this was because frequently the Ramp Manager had Ground hold the arrivals involved in a gate conflict “so we did not have to deal with them as we would in real life.”

# Post-Run Ratings: Gate Conflicts Caused by Holding Aircraft to Meet Advisories



How often did holding aircraft to meet [gate][spot] hold advisories cause a gate conflict?

*N*'s = 19-20 ratings in each condition. Error bars = 95% CIs.



Holding aircraft to meet gate advisories were described as not causing many gate conflicts in any condition.

If it were necessary to implement one of the three metering systems listed below at DFW, which would you recommend? Please rank order your choices.

- RC rankings were different by sectors
  - Two Terminal E controllers (low workloads) preferred TOBT+TMATs
  - One Terminal C controller preferred TMATs alone; the other TBOT+TMAT
  - Each of the three Terminal A controllers preferred a different condition (TOBT, TMAT, and TOBT+TMAT)
- Two RM rankings were 1st for TOBT, 2nd for TMAT, and 3rd for TOBT+TMAT

# Ramp Managers' Post-Run Survey (N = 7 Ratings in Each Condition)



- There were no statistically significant differences between conditions on
  - Their workload, which was low on the TLX items
  - Situation Awareness, which was moderately high (SART score average 5.2)
  - Operational efficiency measures, which were high (almost all between 4.7 & 5 on a 5-point scale)
  - Runway 17R queue lengths which were rated as very acceptable (all 5's)





- On a 5-point scale, TOBTs thought to be
  - Achieved only about half the time in the single condition (3.3) and about the same in the TOBT+TMAT condition (3.6)
  - Not difficult—1.4 in the single condition and 1.9 in the combined condition
  - Not disruptive to the traffic flow (1.4 vs. 1.6)
  - TOBT times “about right” (3 vs. 3)
  
- On a 5-point scale, TMA Ts thought to be
  - Achieved “usually” (4) in the single condition and slightly less in the TOBT+TMAT condition (3.4)
  - Not difficult—2.1 in the single condition and 1.9 in the combined condition
  - Not disruptive to the traffic flow (1.6 vs. 1.3)
  - TMA T times “about right” (3 vs. 3)

# Summary of Ramp Managers' Post-Simulation Survey (N = 2)



- Own workload was described on a 5 point scale as 2.5 in Baseline and TOBT, 3.0 in TMAT, and 3.5 in TOBT+TMAT
- SA was described as between 3.5 & 4 in all conditions
- RC workload was described as 3 in Baseline, 4 in TOBT, 4.5 in TMAT, and 5 in TOBT+TMAT
- RC SA was rated as 5 in all conditions
- The RMs rated themselves as useful to the following degree: in Baseline 1, in TOBT as 2, in TMAT as 2, in TOBT+TMAT as 4
- Coordination was thought to be about right between RM and RCs (3 = “about right”), but as 2 (“not enough”) between RCs
- RM’s coordination and workload at the busiest times in the simulation was thought to be acceptable (4) except there was a comment “Too much [workload] for Sector A”
- RM tools were thought to be helpful at 4.5 in terms of supporting RM’s decisions regarding traffic flow in the ramp, departure demand at the spots, length of the runway queue, and airport efficiency
- The helpfulness of RC tools were also rated as 4.5
- The RMs had extensive experience at DFW (5) and rated all aspects of the simulation as very realistic (5s on all items)

# Summary of Tower Controllers' Post-Run Survey



- There were no statistically significant differences between conditions (which the Tower Controllers were unaware of) for
  - Workload and situation awareness
  - Measures of operational efficiency
    - E.g., departure demand at spots



- On a 5-point scale
  - Workload and coordination in the simulation were acceptable at 4.3 each
  - There was about the right amount of coordination (“3” on the scale) between the TMC and the Ramp at 2.9 and 2.8
  - In the field the usefulness of coordination by the tower with the TMC was rated as 4.1, with other Tower Controllers was 4.6, and the Ramp was 3.5
  - The Tower Controllers had extensive experience at DFW (average rating 5 out of 5), and rated all components of the simulation as realistic (between 4 and 5)