# TSTR HITL Results

HITL: April 25-29

Experiment Lead: Savvy Verma

# Outline

- TSTR HITL Overview
- Workload
- Situation Awareness
- Pushback Advisories
- Traffic Realism
- South to North Transition
- Trust
- Usability

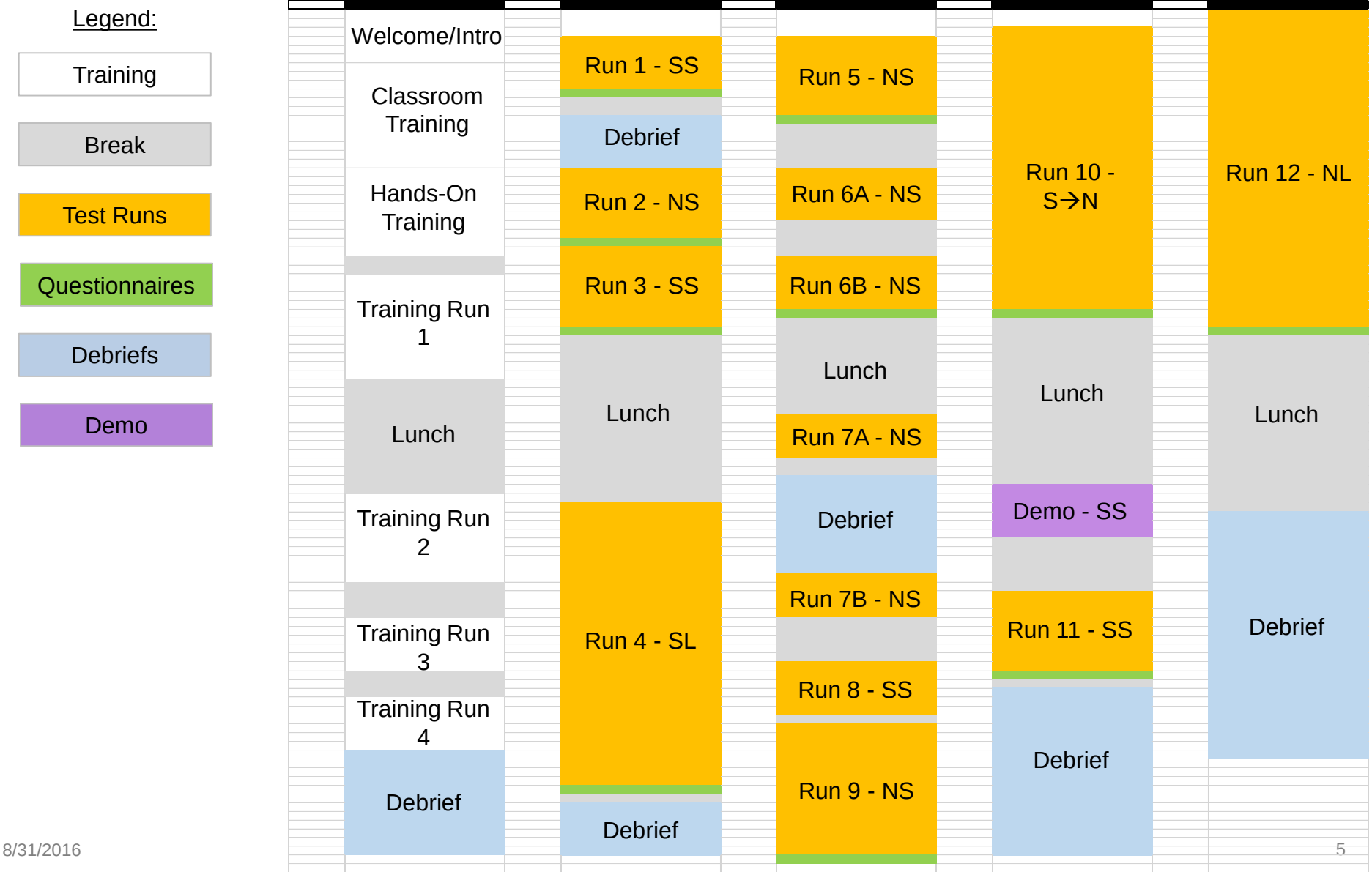
# TSTR HITL

- April 25-29
- 4 Ramp Controller (RC) participants
  - 2 active CLT AAL RCs
  - 1 active DFW AAL RC
  - 1 LAX Tower SME
- 1 Ramp Manager (RM) participant
  - Active CLT RM

# TSTR HITL

- 12 experimental runs consisting of 5 scenarios:
  - South Short (45 min; SS)
  - North Short (45 min; NS)
  - South Long (3 hours; SL)
  - North Long (3 hours; NL)
  - South to North flow change (3 hours; S→N)
- 6 training sessions:
  - 1 classroom training
  - 1 hands-on training
  - 4 training runs
- Demographic, workload, post-run, and post-study questionnaire data collected

# TSTR HITL: Actual Schedule



# Workload

# Workload: WAK

- Collected *during* each run on a tablet
- WAK = Workload Assessment Keypad
- Participants notified by an audible “ding” once every 5 minutes
- Asked push a button to rate their workload on a scale of 1 to 5. Presented as:


(Low Workload) (High Workload)

- Data Collected:
  - Workload Rating
  - Response Times

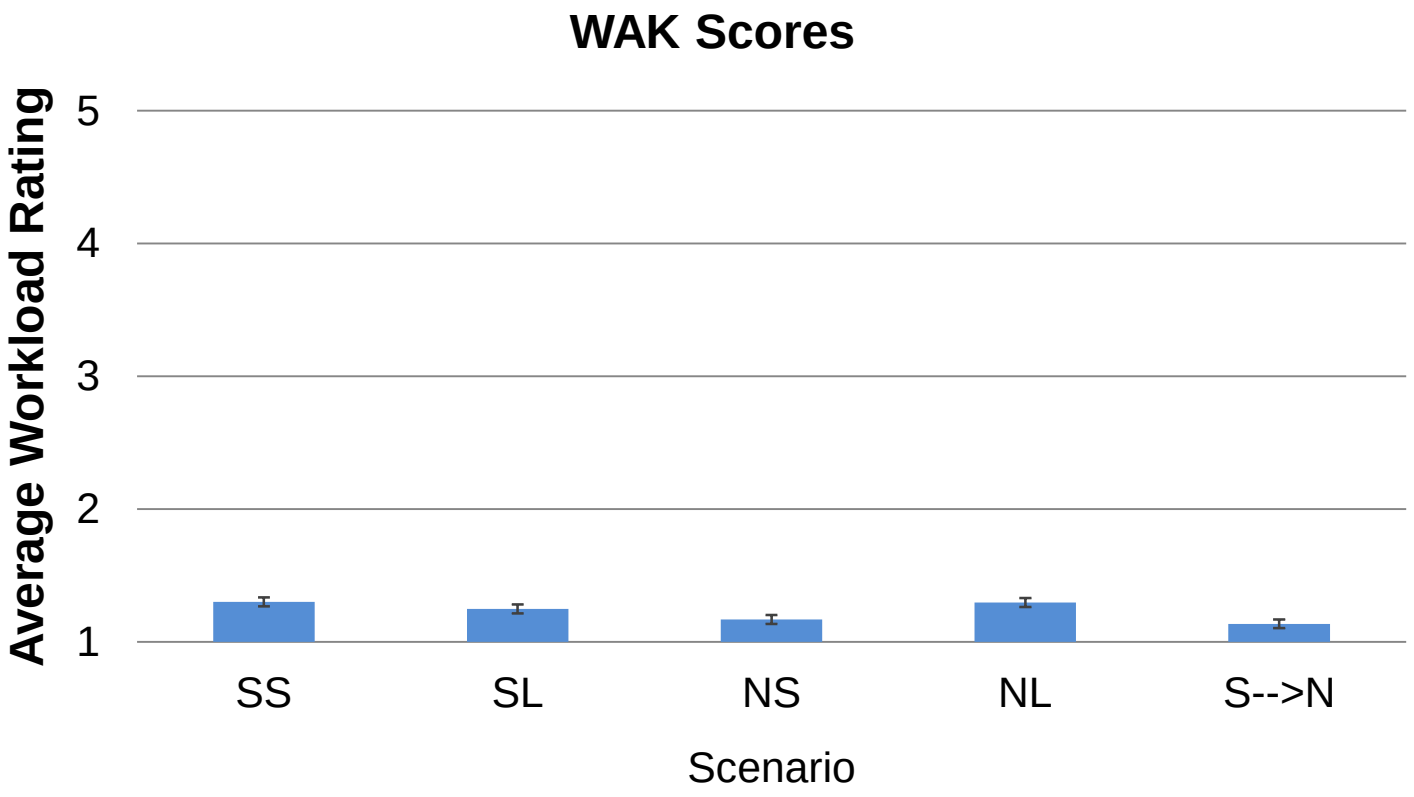
# Workload: NASA-TLX

- Collected *at the end* of each run in the post-run questionnaires
- NASA-TLX assesses workload on 6 dimensions:
  1. Mental Demand
  2. Temporal Demand
  3. Frustration
  4. Performance
  5. Effort
  6. Physical Demand
- Used a rating scale of 1 (low) to 5 (high)
- Performance is inversely coded when calculating a composite TLX score
- Note: Results are low power and not statistically significant

# Workload: Results

- WAK by Scenario Type

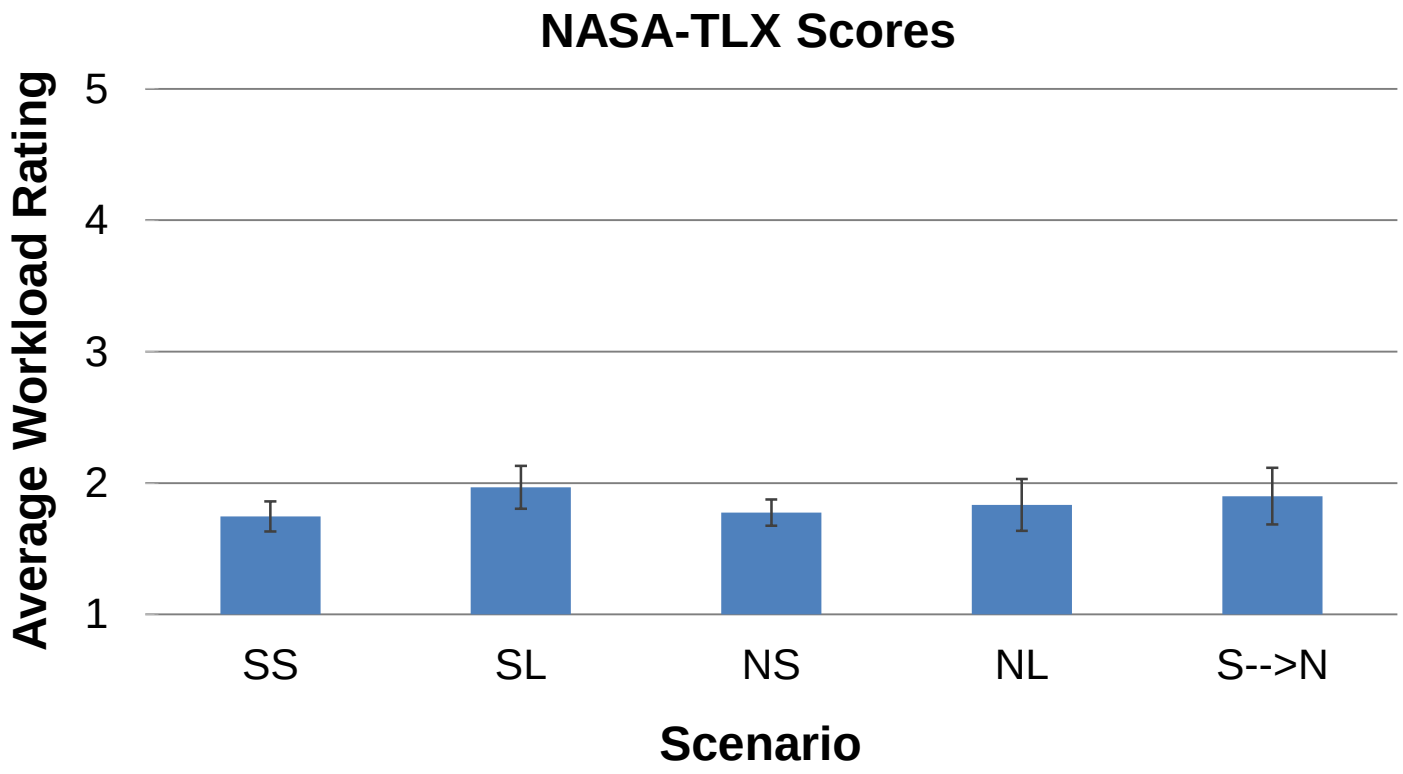
| | Scenario | | | | |
|---|---|---|---|---|---|
| | SS | SL | NS | NL | S-->N |
| Average WAK Score | 1.30 | 1.25 | 1.17 | 1.30 | 1.14 |
| Standard Error | 0.053 | 0.044 | 0.034 | 0.037 | 0.026 |



# Workload: Results

- NASA-TLX by Scenario Type

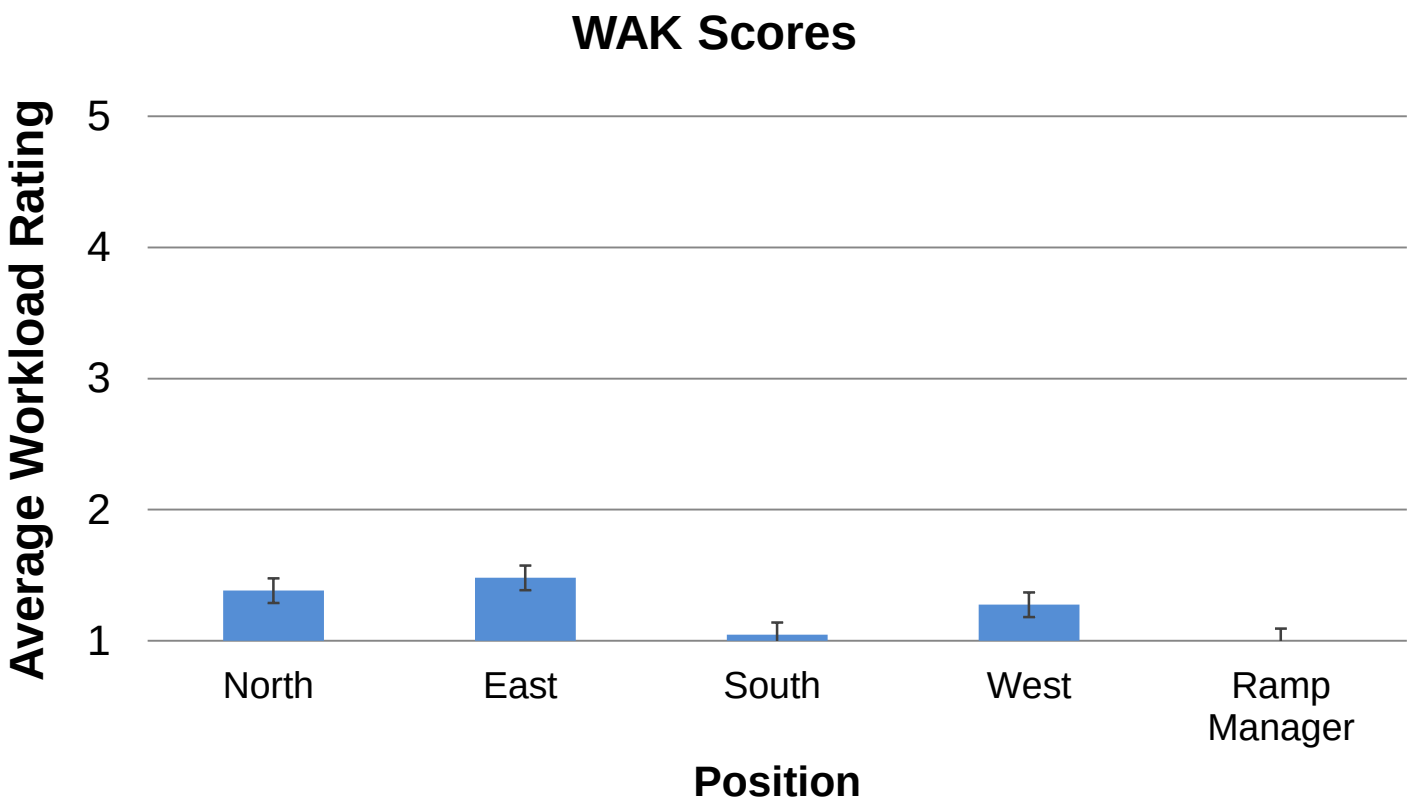
| | Scenario | | | | |
|---|---|---|---|---|---|
| | SS | SL | NS | NL | S-->N |
| Average TLX Score | 1.74 | 1.97 | 1.78 | 1.83 | 1.90 |
| Standard Error | 0.115 | 0.162 | 0.100 | 0.198 | 0.216 |



# Workload: Results

- WAK by Position

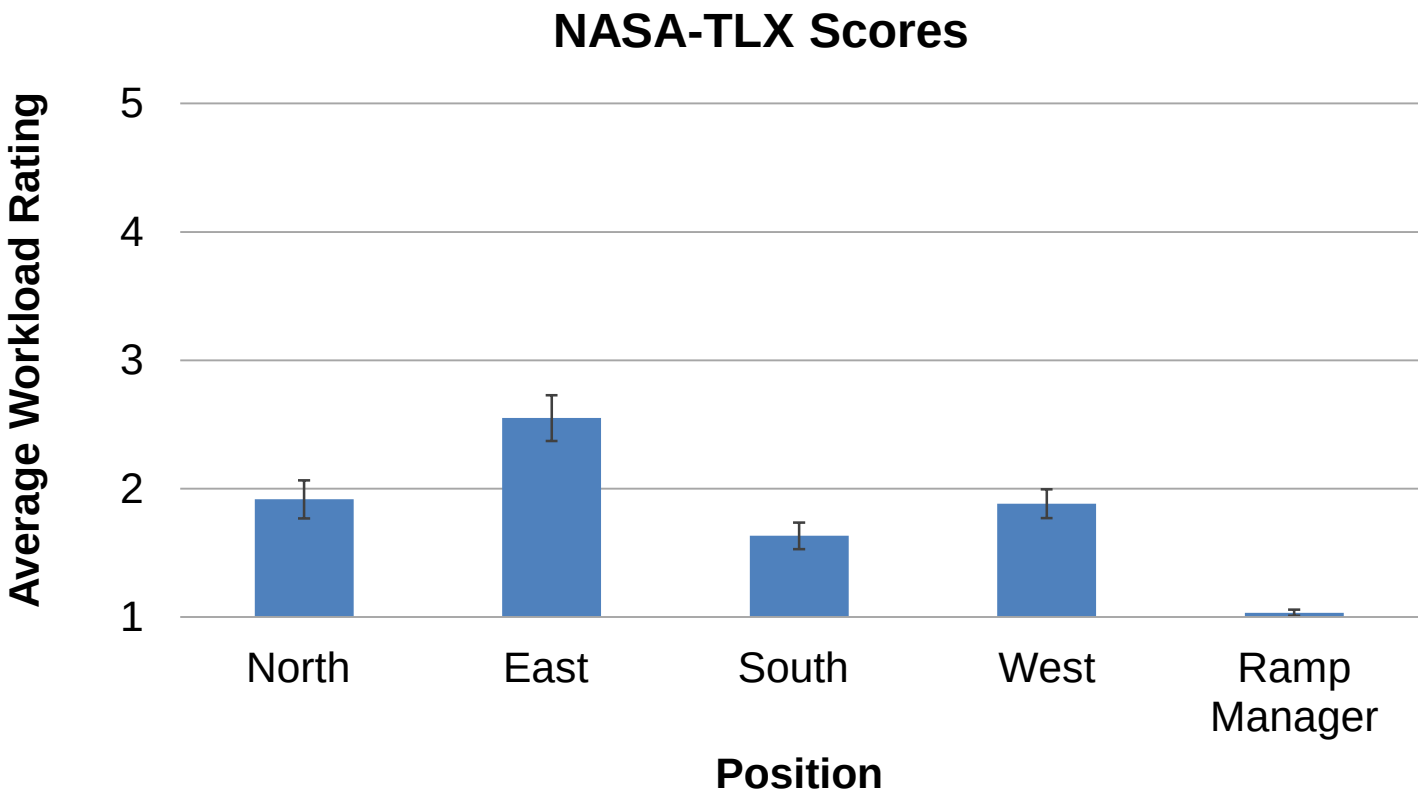
| Position | | | | | |
|---|---|---|---|---|---|
| | North | East | South | West | Ramp Manager |
| Average WAK Score | 1.38 | 1.48 | 1.05 | 1.28 | 1 |
| Standard Error | 0.050 | 0.048 | 0.016 | 0.040 | 0 |



# Workload: Results

- NASA-TLX by Position

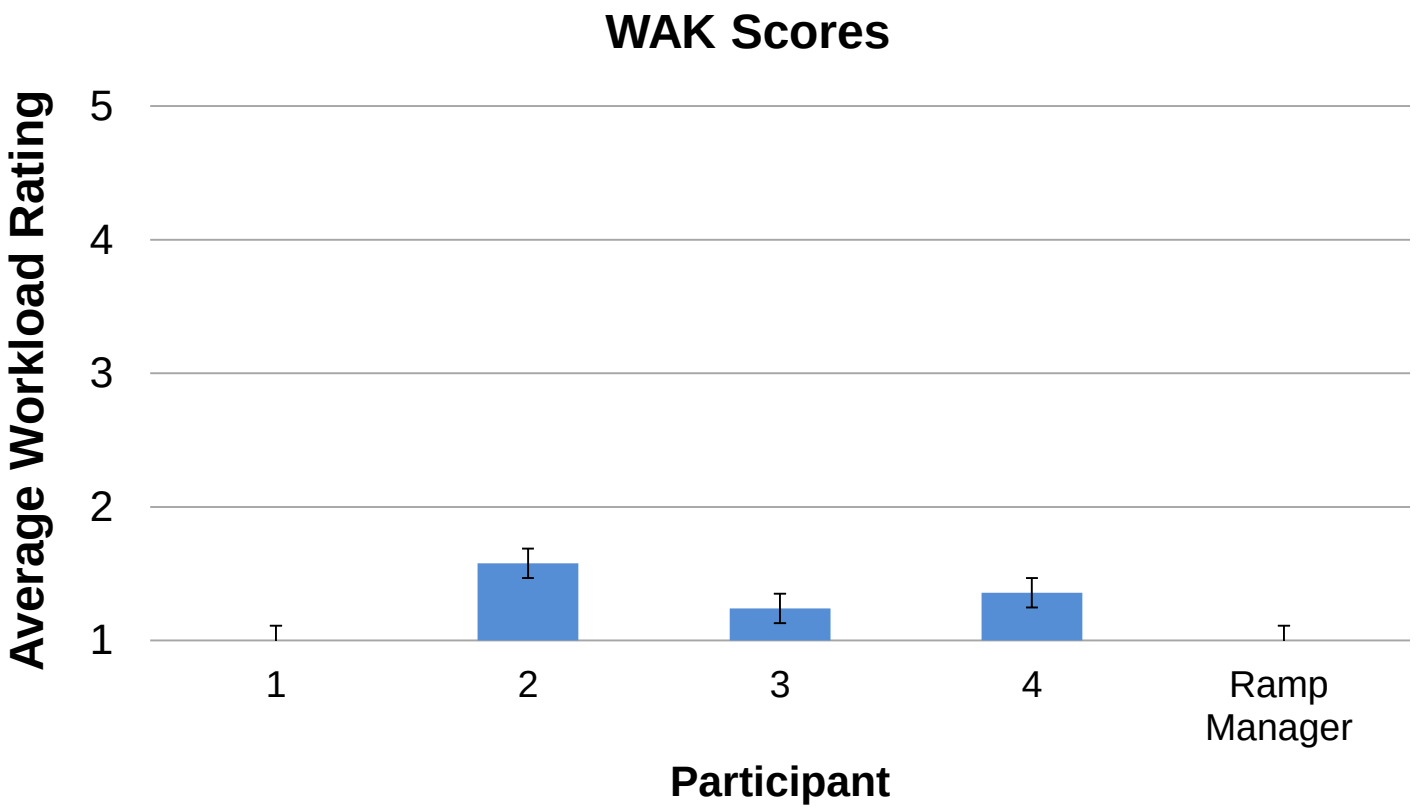
| | Position | | | | |
|---|---|---|---|---|---|
| | North | East | South | West | Ramp Manager |
| Average TLX Score | 1.92 | 2.55 | 1.63 | 1.88 | 1.03 |
| Standard Error | 0.149 | 0.179 | 0.104 | 0.112 | 0.023 |



# Workload: Results

- WAK by Participant

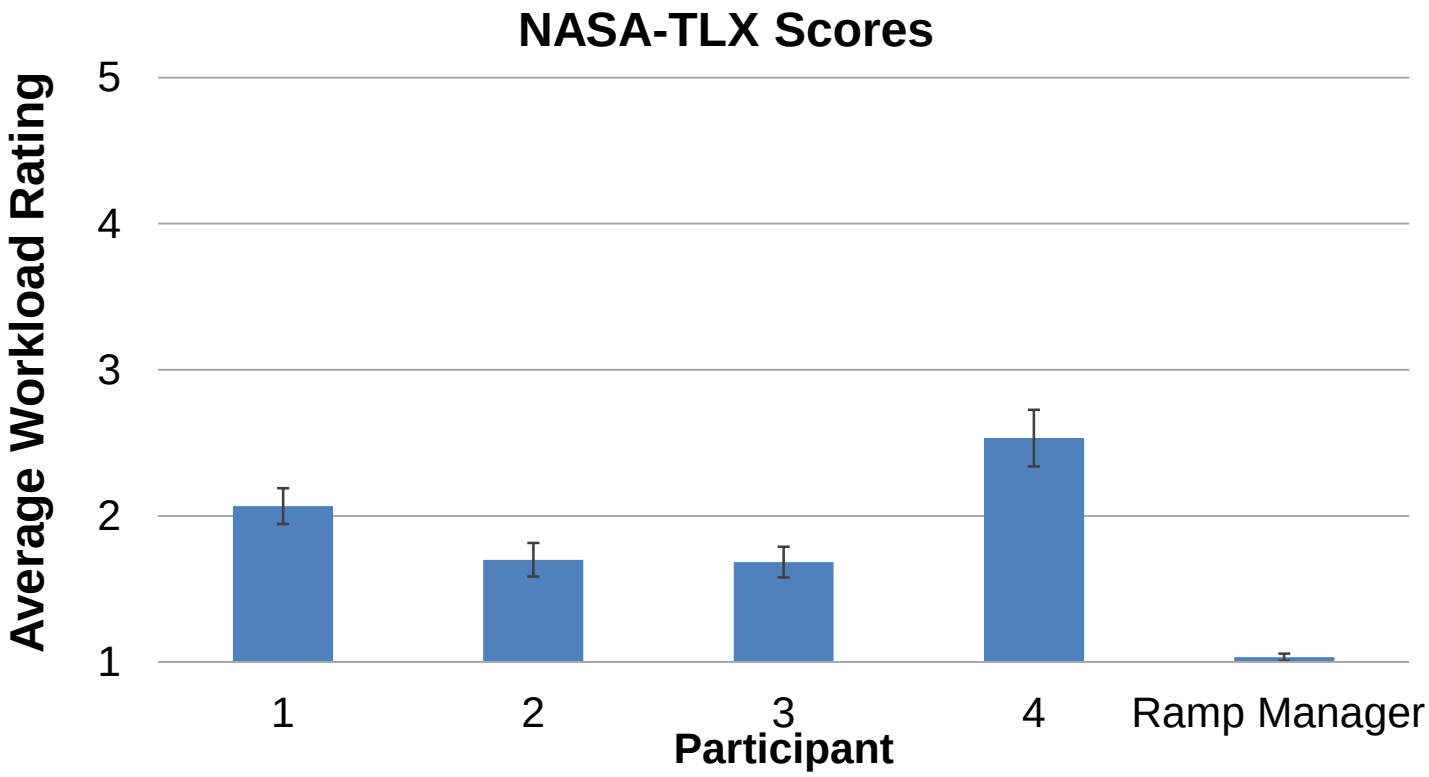
| | Participant | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Ramp Manager |
| Average WAK Score | 1 | 1.58 | 1.24 | 1.36 | 1 |
| Standard Error | 0 | 0.050 | 0.038 | 0.044 | 0 |



# Workload: Results

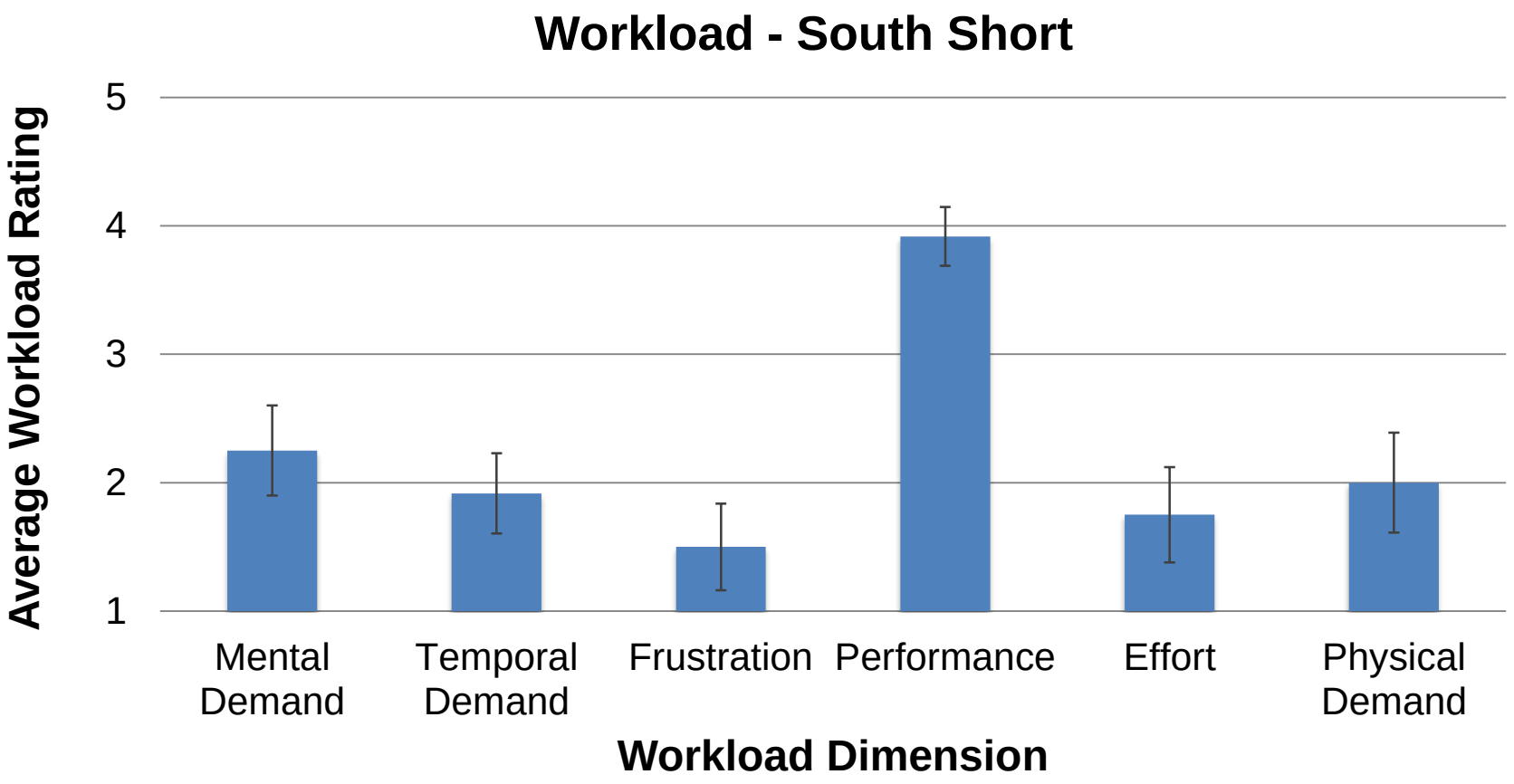
- NASA-TLX by Participant

| | Participant | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Ramp Manager |
| Average TLX Score | 2.07 | 1.70 | 1.68 | 2.53 | 1.03 |
| Standard Error | 0.123 | 0.115 | 0.105 | 0.194 | 0.023 |



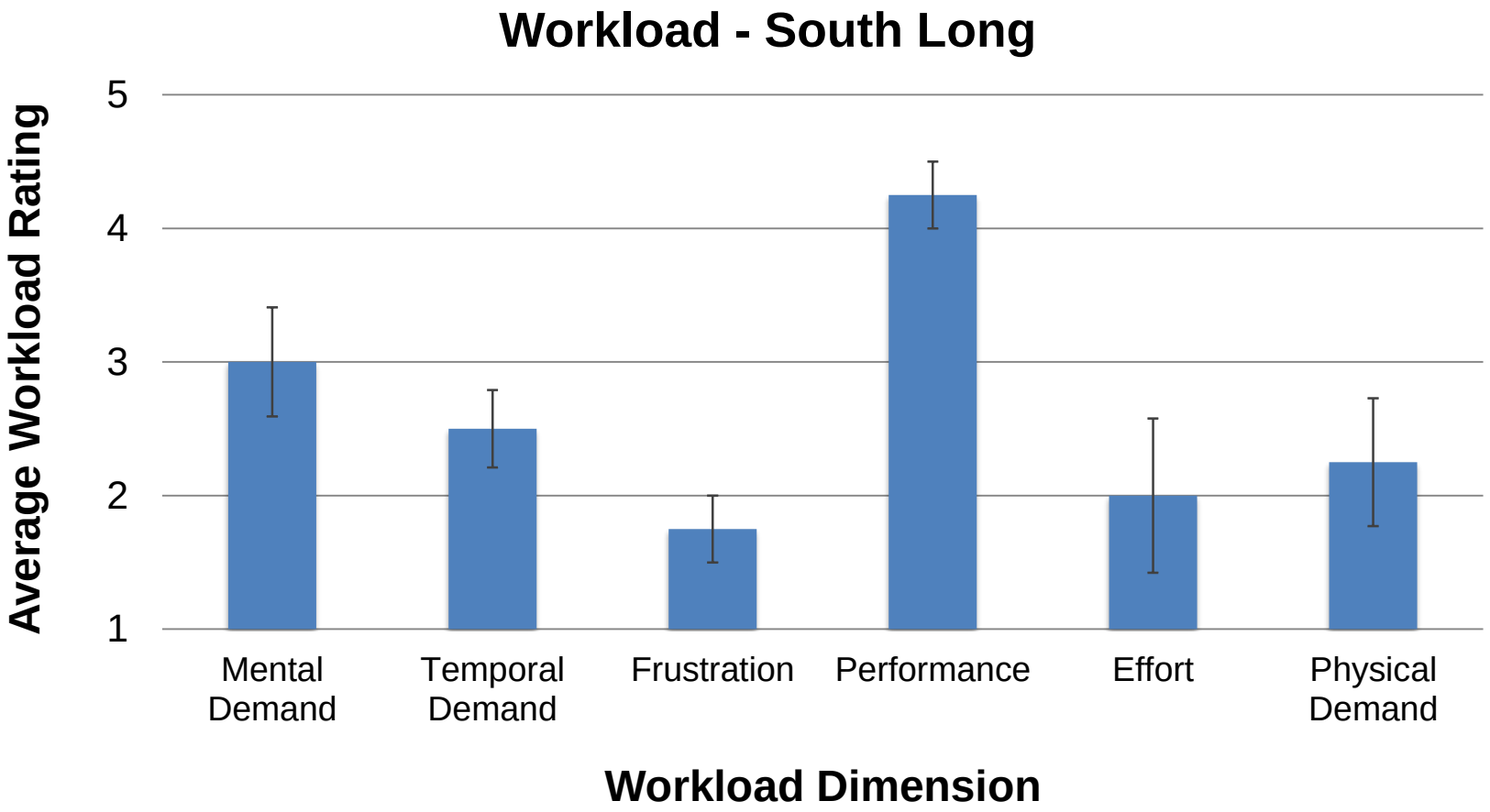
# Workload: Results

- NASA-TLX by individual workload dimensions



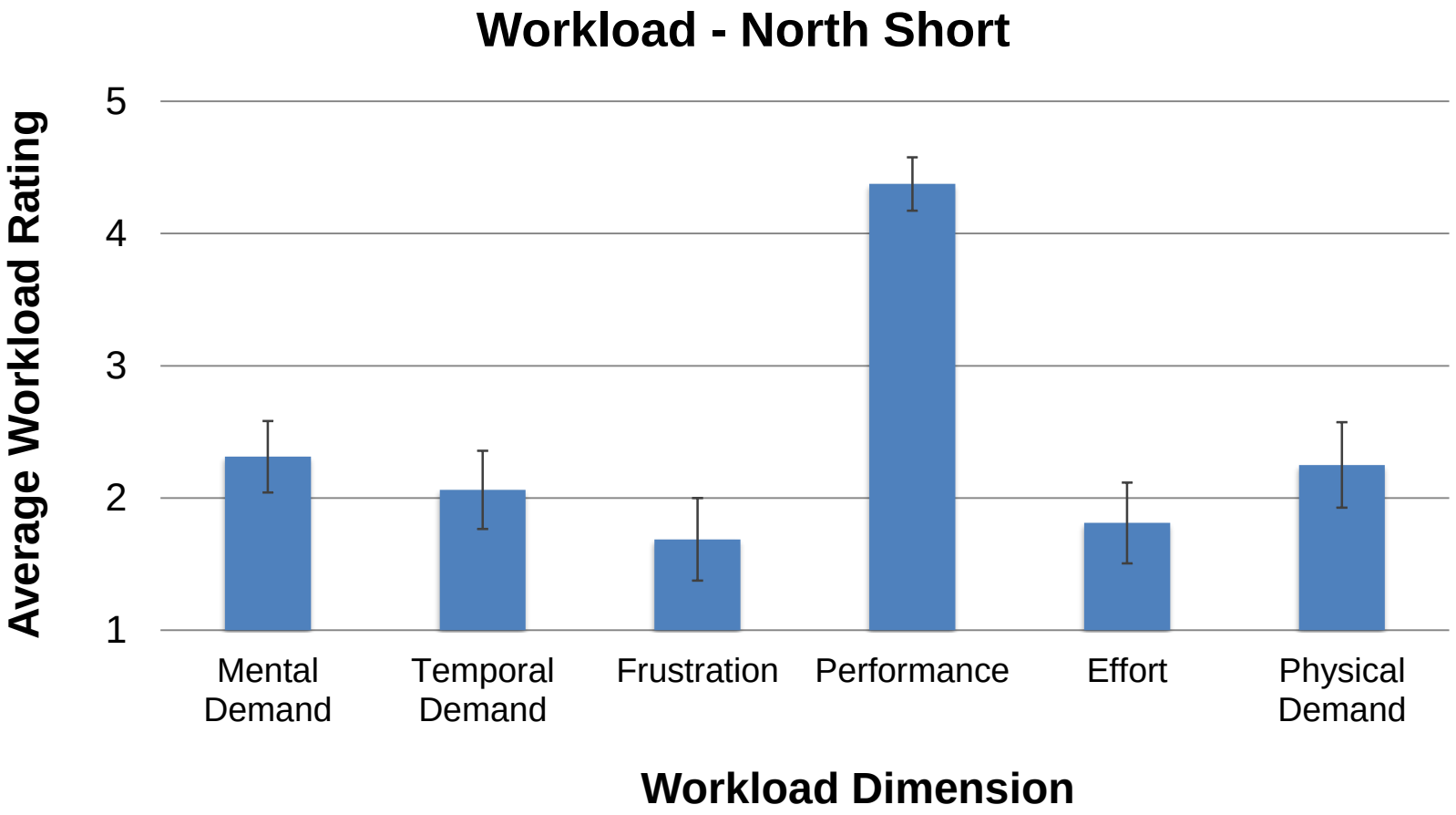
# Workload: Results

- NASA-TLX by individual workload dimensions



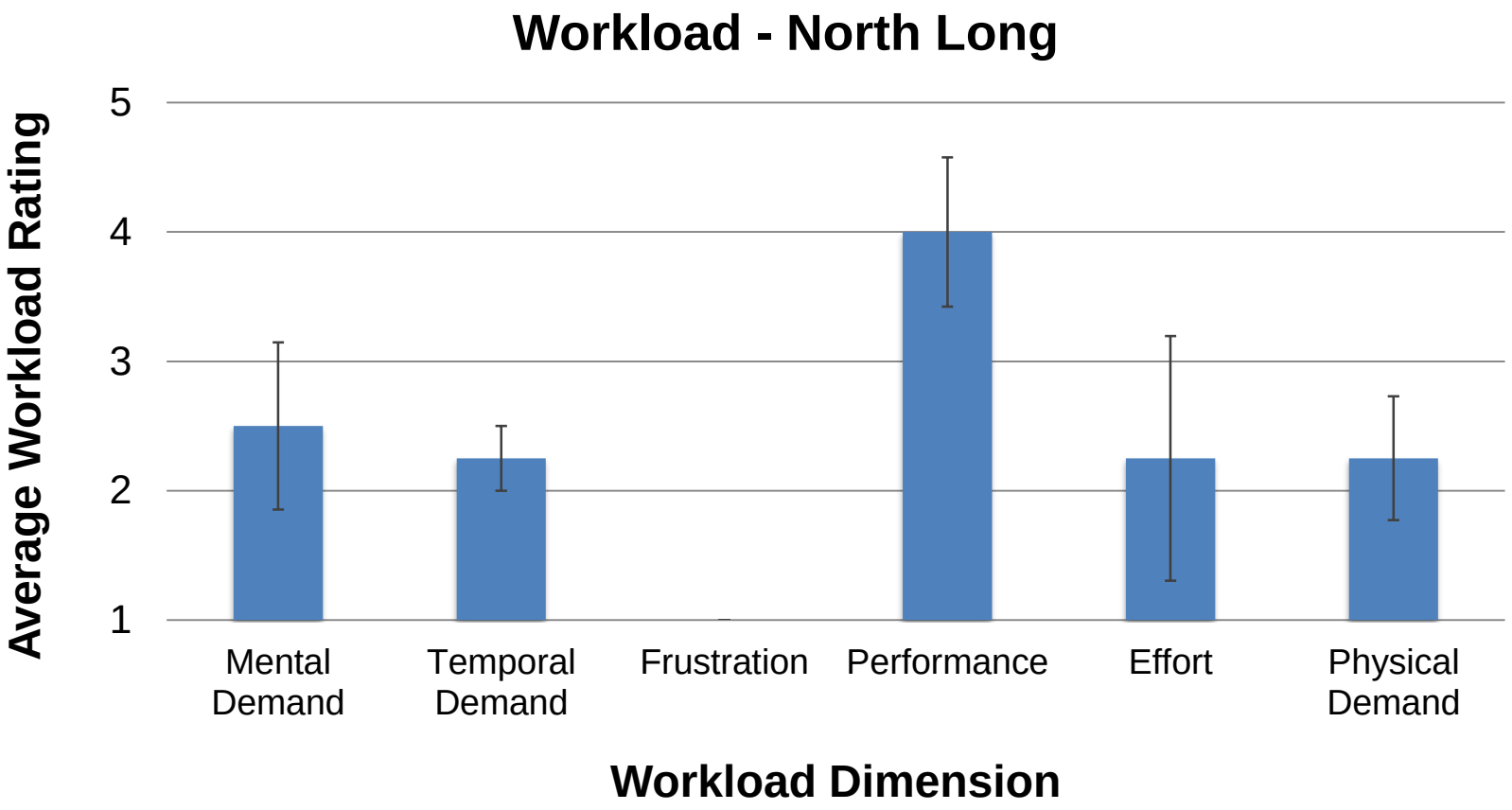
# Workload: Results

- NASA-TLX by individual workload dimensions



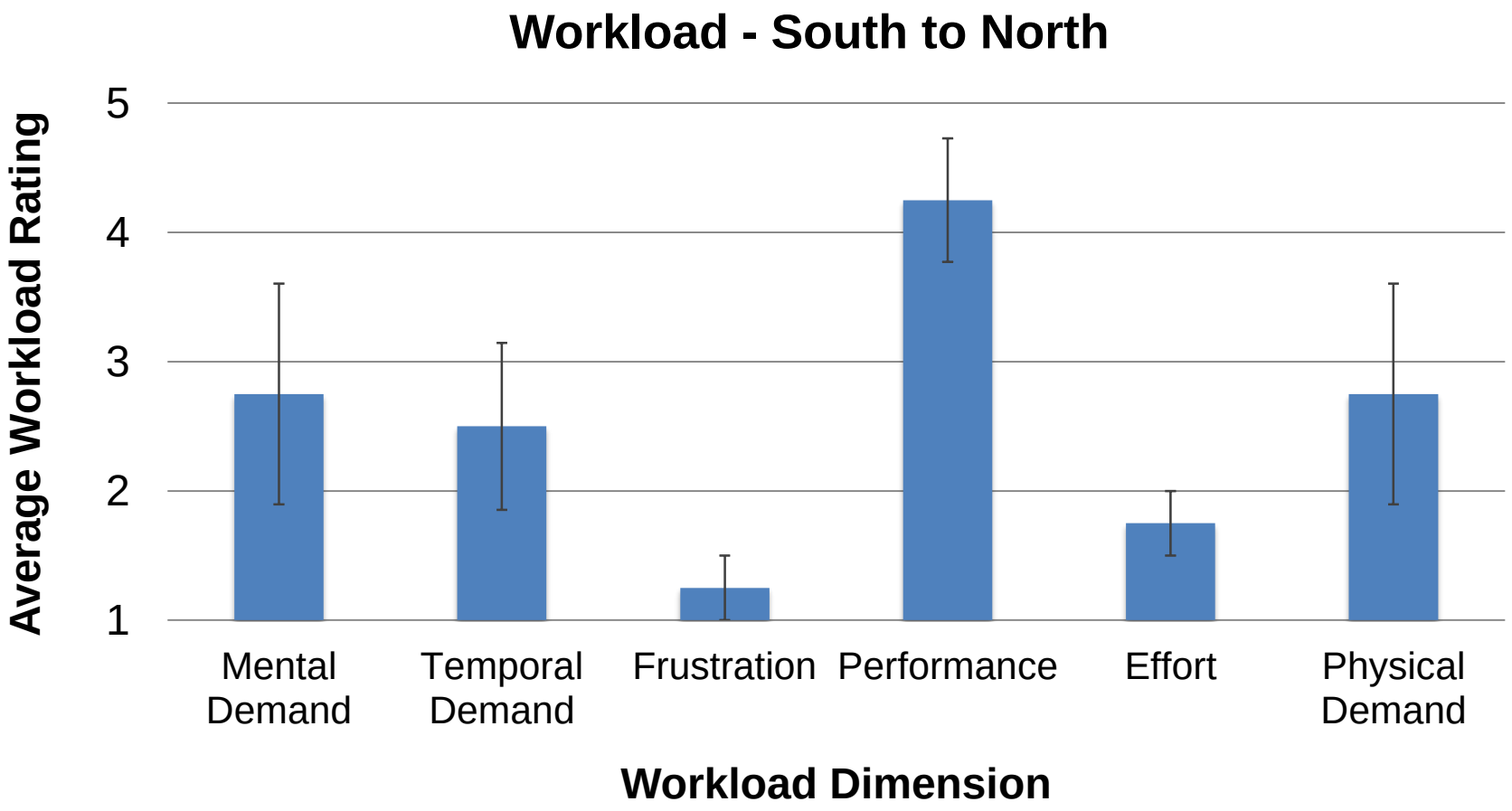
# Workload: Results

- NASA-TLX by individual workload dimensions



# Workload: Results

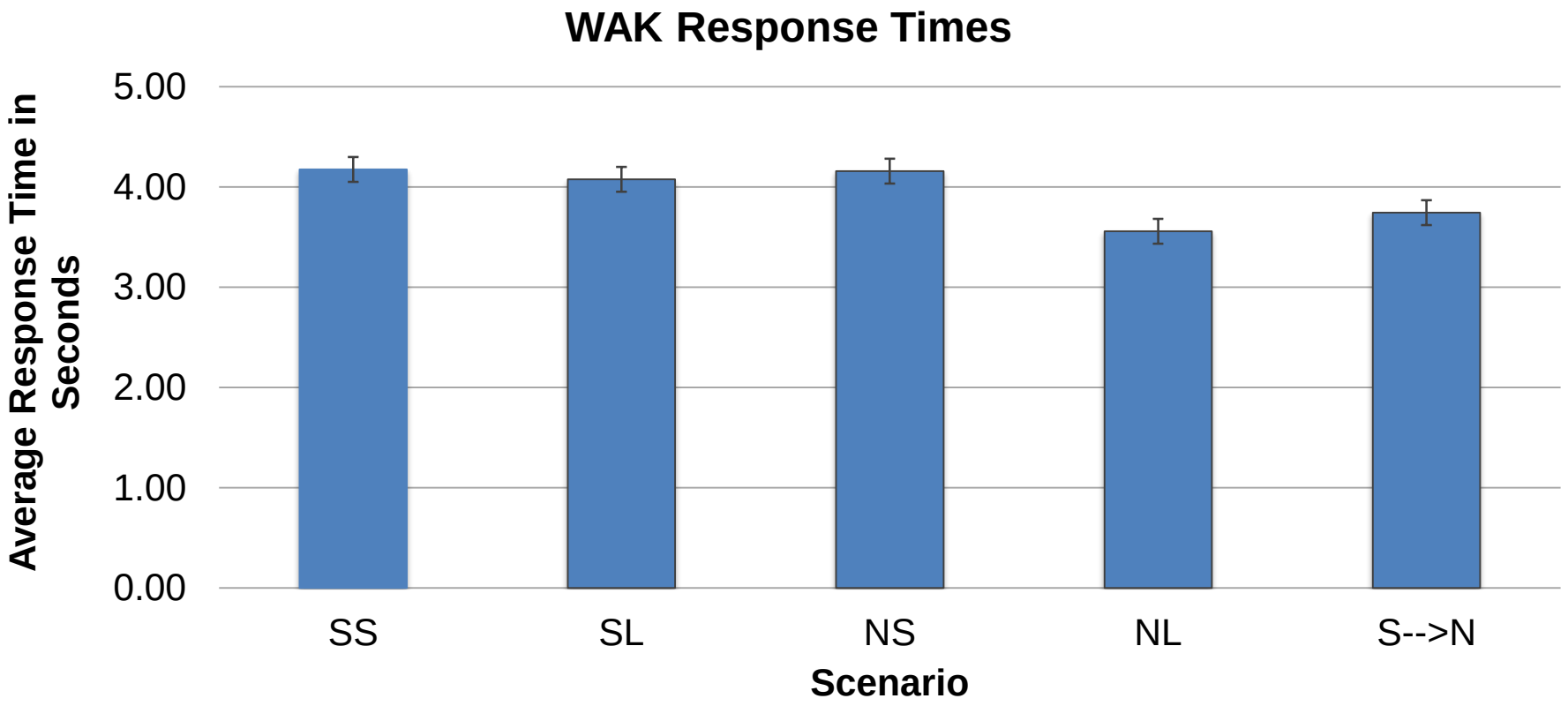
- NASA-TLX by individual workload dimensions



# WAK Response Times: Results

- WAK Response Times by Scenario

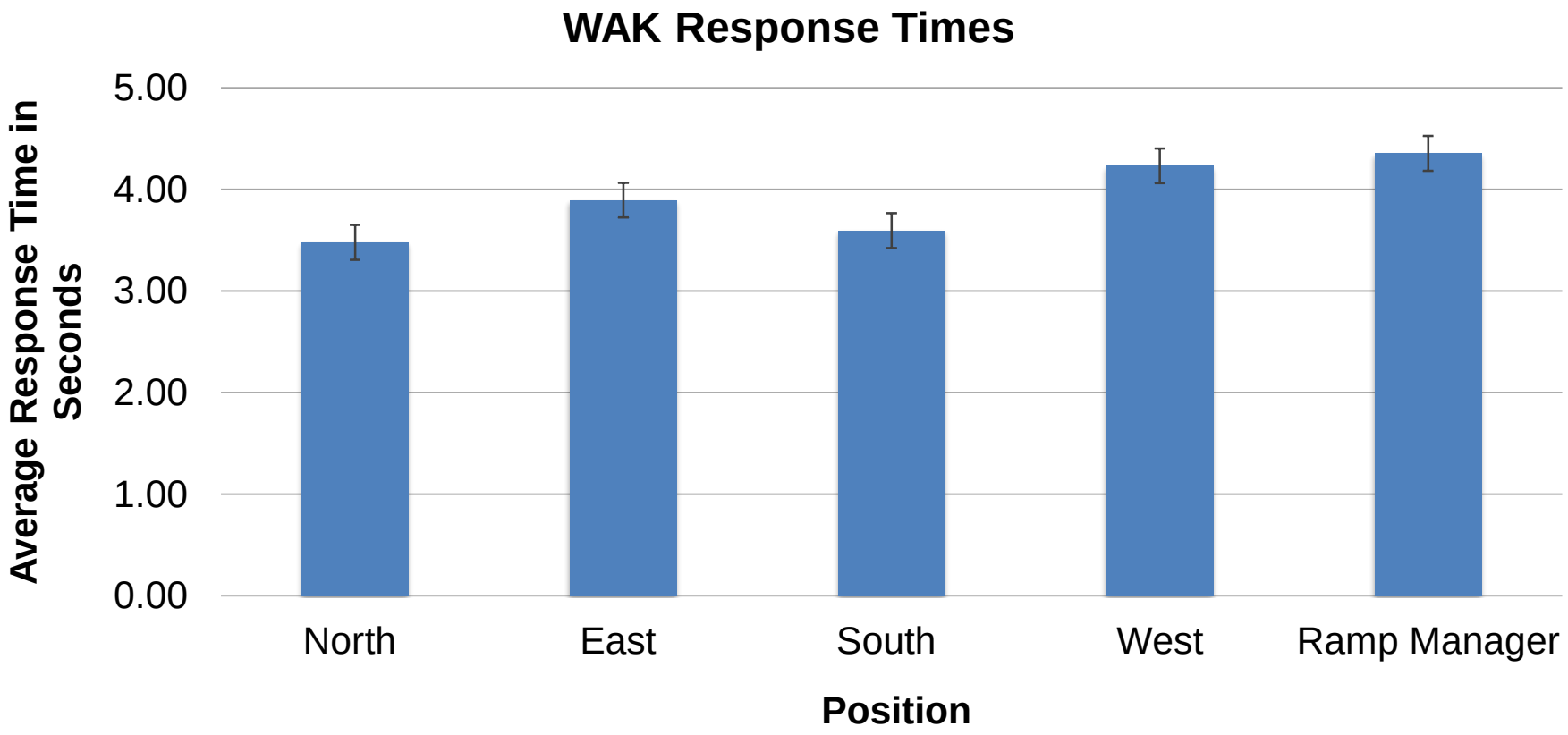
| | Scenario | | | | |
|---|---|---|---|---|---|
| | SS | SL | NS | NL | S-->N |
| Average Response Times | 4.17 | 4.07 | 4.16 | 3.56 | 3.74 |
| Standard Error | 0.23 | 0.18 | 0.15 | 0.08 | 0.12 |



# WAK Response Times: Results

- WAK Response Times by Position

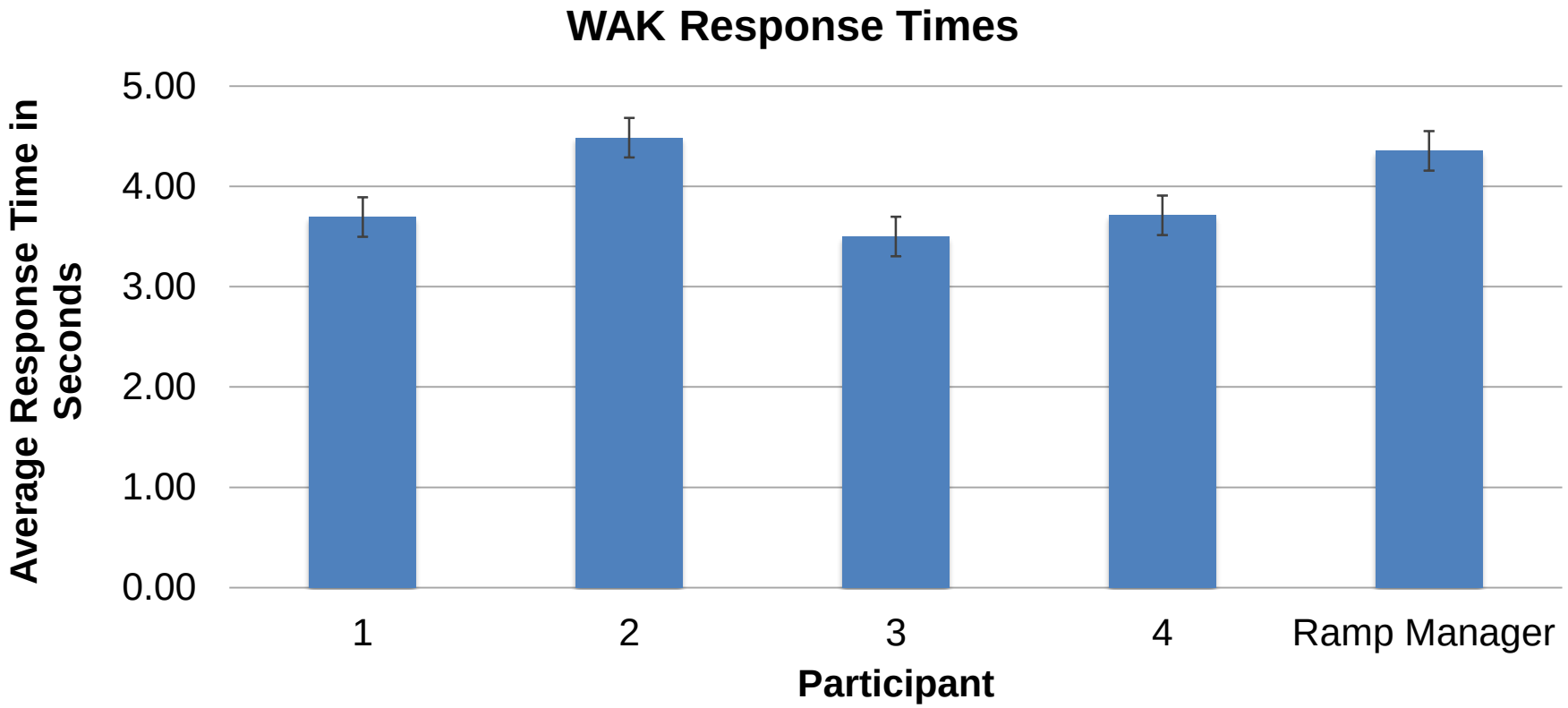
| | Position | | | | |
|---|---|---|---|---|---|
| | North | East | South | West | Ramp Manager |
| Average Response Times | 3.48 | 3.89 | 3.59 | 4.23 | 4.35 |
| Standard Error | 0.11 | 0.15 | 0.13 | 0.16 | 0.17 |



# WAK Response Times: Results

- WAK Response Times by Participant

| | Participant | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Ramp Manager |
| Average Response Times | 3.69 | 4.49 | 3.50 | 3.71 | 4.35 |
| Standard Error | 0.13 | 0.19 | 0.13 | 0.09 | 0.17 |



# Workload: Summary

- Overall, workload scores were very low
  - Performance scores tended to be high – participants rated themselves as performing well
- Overall, response times were small – indicates that workload was low
- Participants commented that the traffic scenarios were very light
- RM WAK response times were likely higher due to the RM's tendency converse often
- West RCs' WAK response times were likely higher due the lack of activity in that sector, which gave them extra time to converse

# Situation Awareness

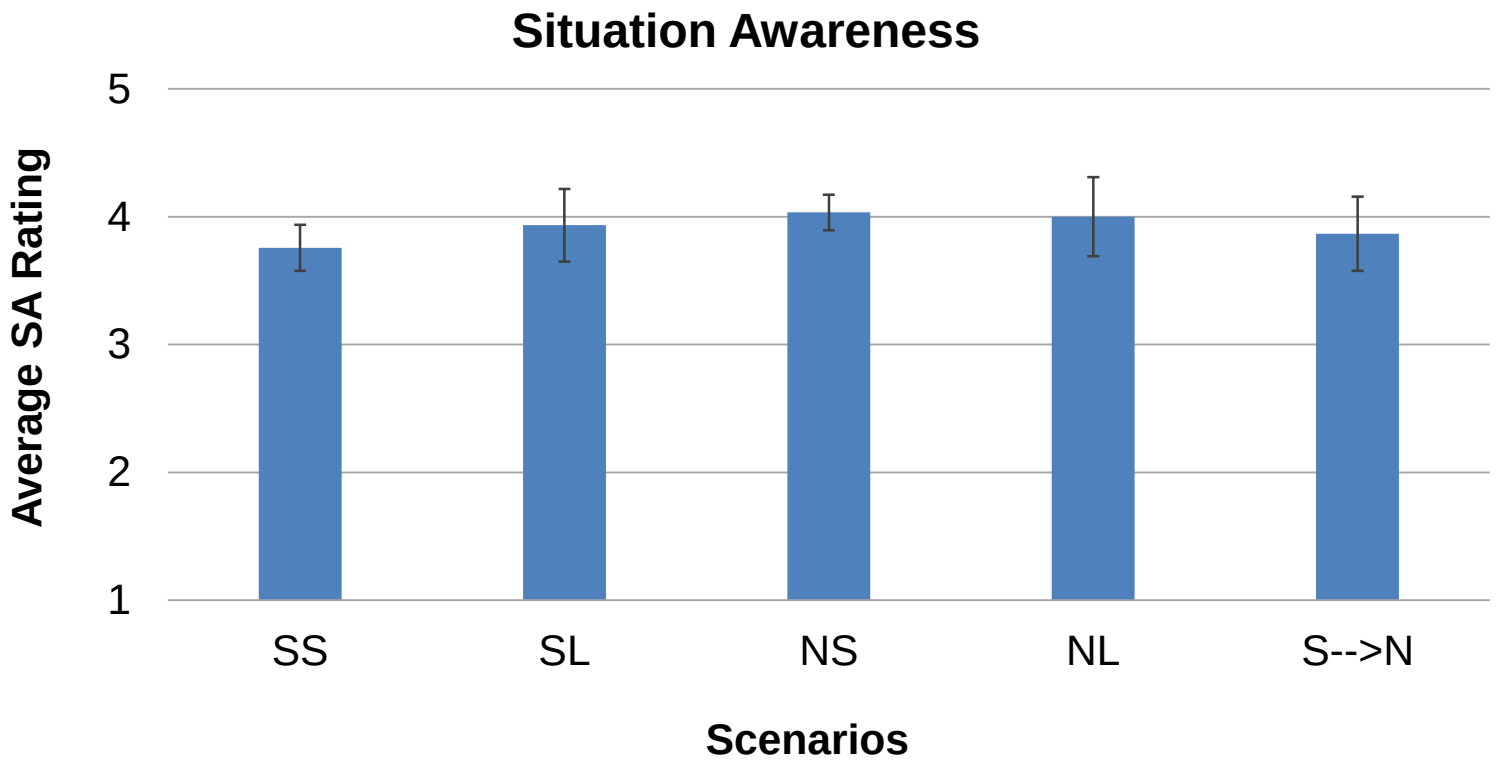
# Situation Awareness (SA)

- Collected at the end of each run in the post-run questionnaires
- 3 Questions from the Situation Awareness Rating Technique (SART)
  - 2.1 Demand on attention
  - 2.2 Level of understanding of the situation
  - 2.3 Available attentional capacity to apply to operations
- Used a rating scale of 1 (low) to 5 (high)
- Question 2.1 is inversely coded for calculating a composite SA score
- Note: Results are low power and not statistically significant

# Situation Awareness: Results

- Situation Awareness (SA) by Scenario

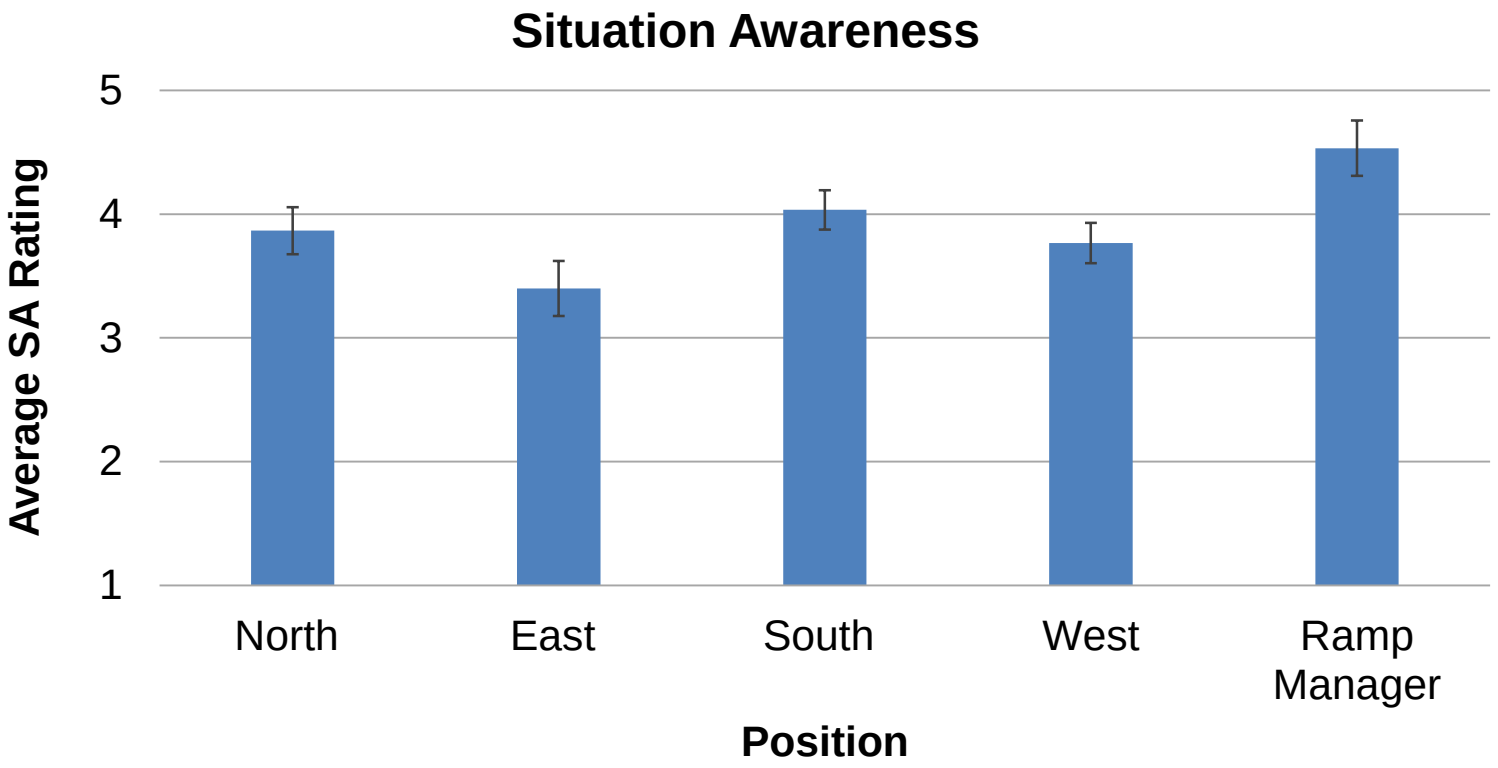
| | Scenario | | | | |
|---|---|---|---|---|---|
| | SS | SL | NS | NL | S-->N |
| Average SA Score | 3.76 | 3.93 | 4.03 | 4.00 | 3.87 |
| Standard Error | 0.180 | 0.284 | 0.139 | 0.309 | 0.291 |



# Situation Awareness: Results

- Situation Awareness (SA) by Position

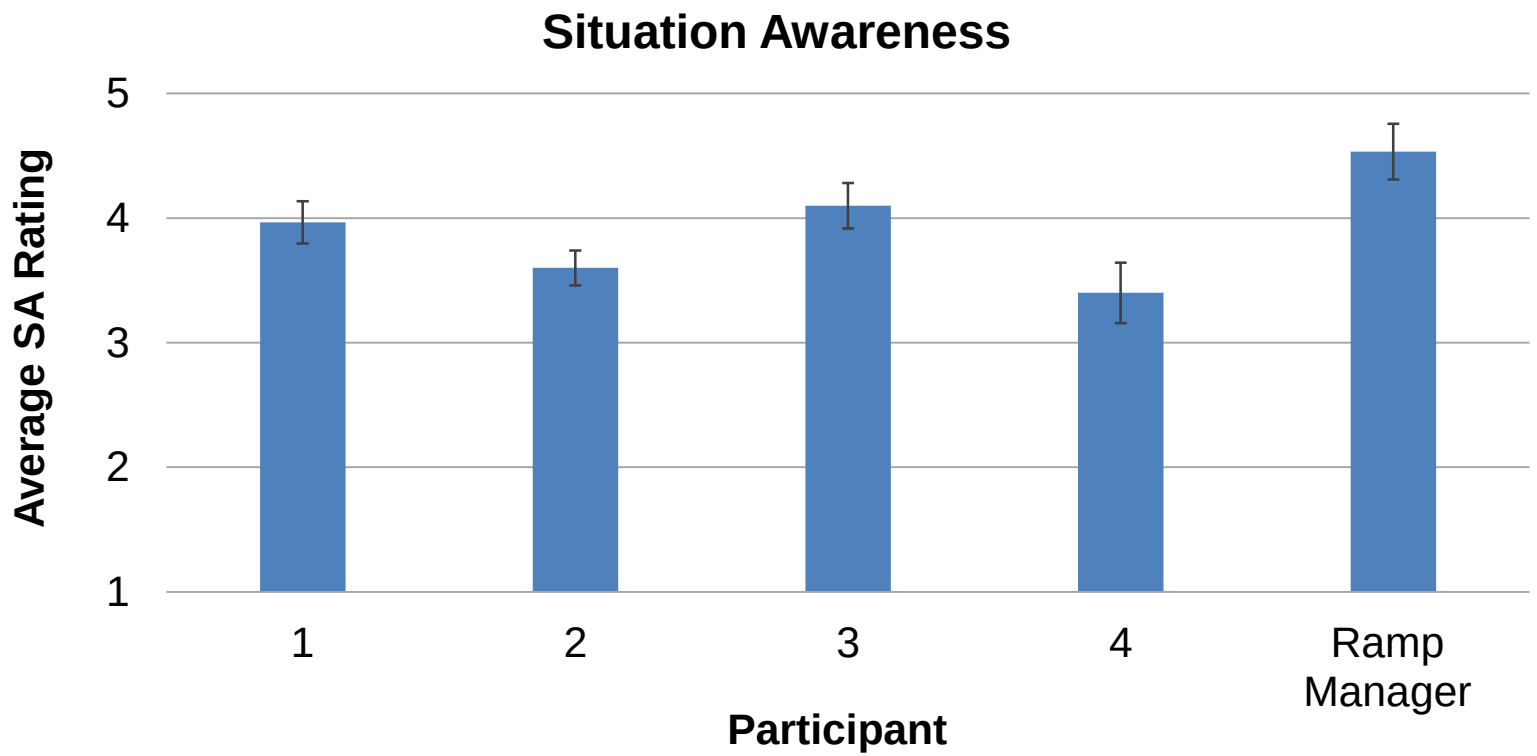
| | Position | | | | |
|---|---|---|---|---|---|
| | North | East | South | West | Ramp Manager |
| Average SA Score | 3.87 | 3.40 | 4.03 | 3.77 | 4.53 |
| Standard Error | 0.190 | 0.223 | 0.160 | 0.164 | 0.224 |



# Situation Awareness: Results

- Situation Awareness (SA) by Participant

| | Participant | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Ramp Manager |
| Average SA Score | 3.97 | 3.60 | 4.10 | 3.40 | 4.53 |
| Standard Error | 0.169 | 0.141 | 0.182 | 0.243 | 0.224 |



# Situation Awareness: Summary

- Overall, SA scores were high

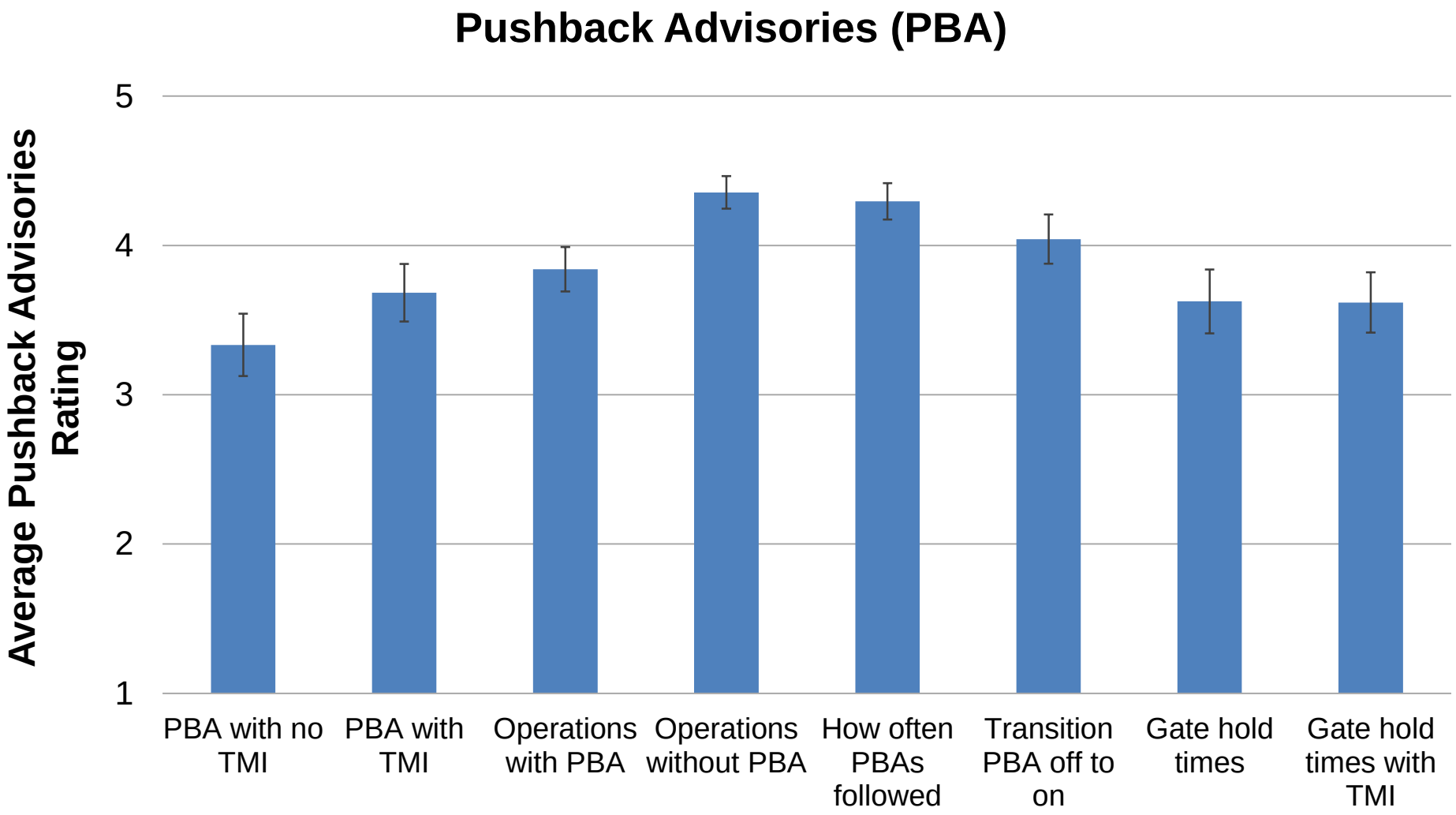
- The RTC display tended to provide adequate information to participants in a way that was easy to understand, which allowed them to manage their sectors without increasing demand on attention or detracting from their attentional capacity.

# Pushback Advisories

# Pushback Advisories

- Collected at the end of each run in the post-run questionnaires
- 8 Questions
  - 3.1 Pushback advisory ratings with no TMI
  - 3.2 Pushback advisory ratings with TMI
  - 3.3 Ramp control operations when using Pushback Advisories
  - 3.4 Ramp control operations when Pushback Advisories were off
  - 3.5 How often Pushback Advisories were followed
  - 3.7 Transitioning from “advisory-off” to “advisory-on”
  - 3.9 Reasonableness of gate hold times with no TMI
  - 3.10 Reasonableness of gate hold times with TMI
- Used a rating scale of 1 (poor) to 5 (good)
- Note: Results are low power and not statistically significant

# Pushback Advisories: Results



# Pushback Advisory: Summary

- Overall, ratings of the pushback advisories were relatively high

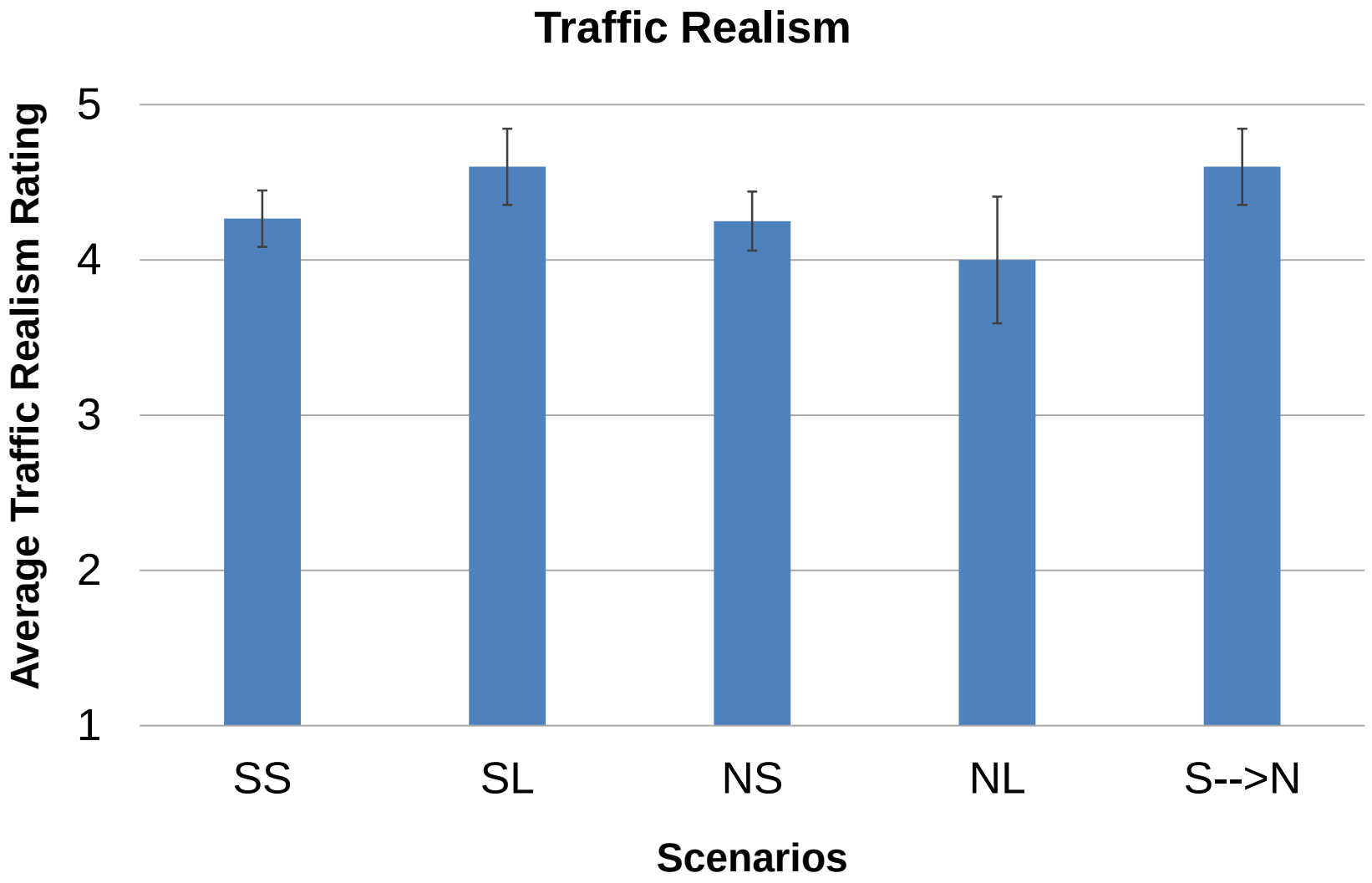
- Participants understood that pushback advisories were being generated by a different scheduler than the one intended for the field.  The new scheduler will provide better advisory times.

# Traffic Realism

# Traffic Realism

- Collected at the end of each run in the post-run questionnaires
- 1 Question
  - 4.1 How realistic was the traffic
- Used a rating scale of 1 (not at all realistic) to 5 (very realistic)
- Note: Results are low power and not statistically significant

# Traffic Realism: Results



# Traffic Realism: Summary

- Overall, traffic realism scores were high

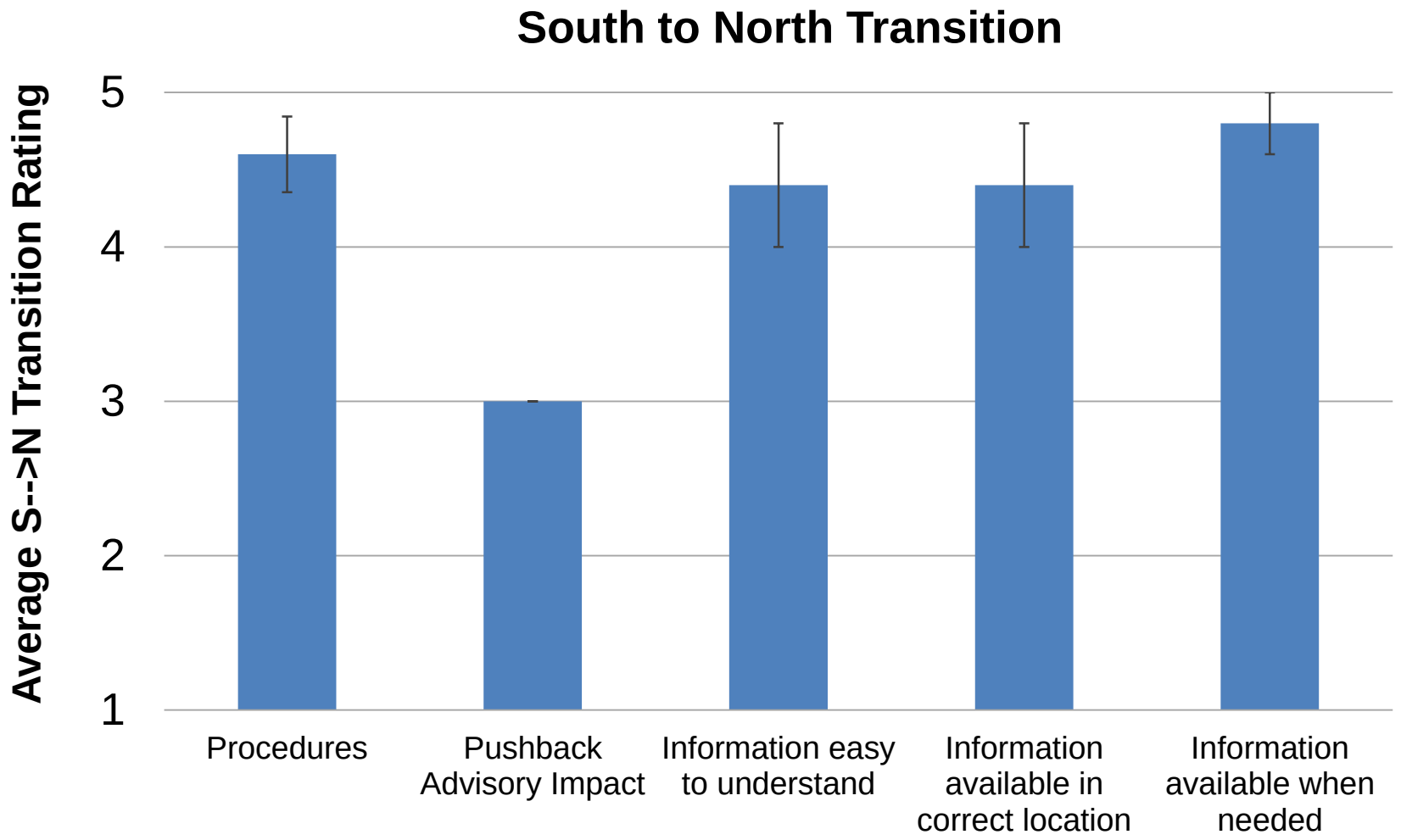
- Participants commented that the traffic was realistic during the beginning of a push. Participants did note that the traffic was very light compared to their typical operations. They suggested improvements to the scenarios by increasing traffic volume and expressed a need to update the outbound spot information to match their procedures.

# South to North Transition

# South to North Transition

- Collected at the end of each run in the post-run questionnaires
- 5 Questions
  - 5.1 Rate procedures for S→N transition
  - 5.2 Pushback advisory impact on S→N transition
  - 5.4 Information presented during S→ N transition was easy to understand
  - 5.5 Information available in correct location during S→N transition
  - 5.6 Needed information was available during S→N transition
- Used a rating scale of 1 (poor) to 5 (good)

# South to North Transition: Results



# South to North Transition: Summary

- Overall, the scores for the South to North flow transition were high

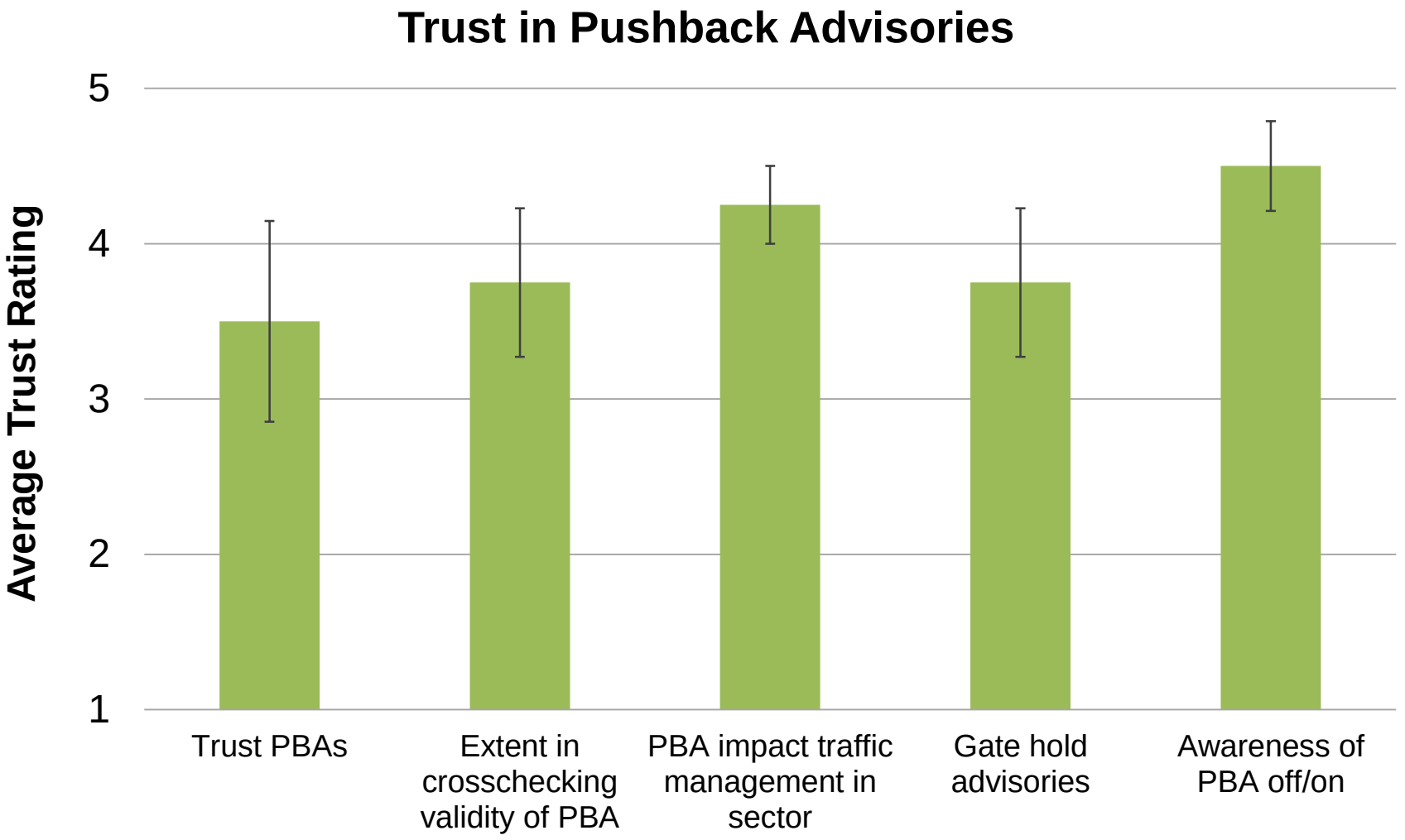
- The pushback advisories were not giving good times during the transition, which likely resulted in the lower rating for the impact pushback advisories had on the transition.  Improvements should be seen with the new scheduler.
- Or we may have to turn off the pushback advisories during transition between flows. This should be done automatically

# Trust

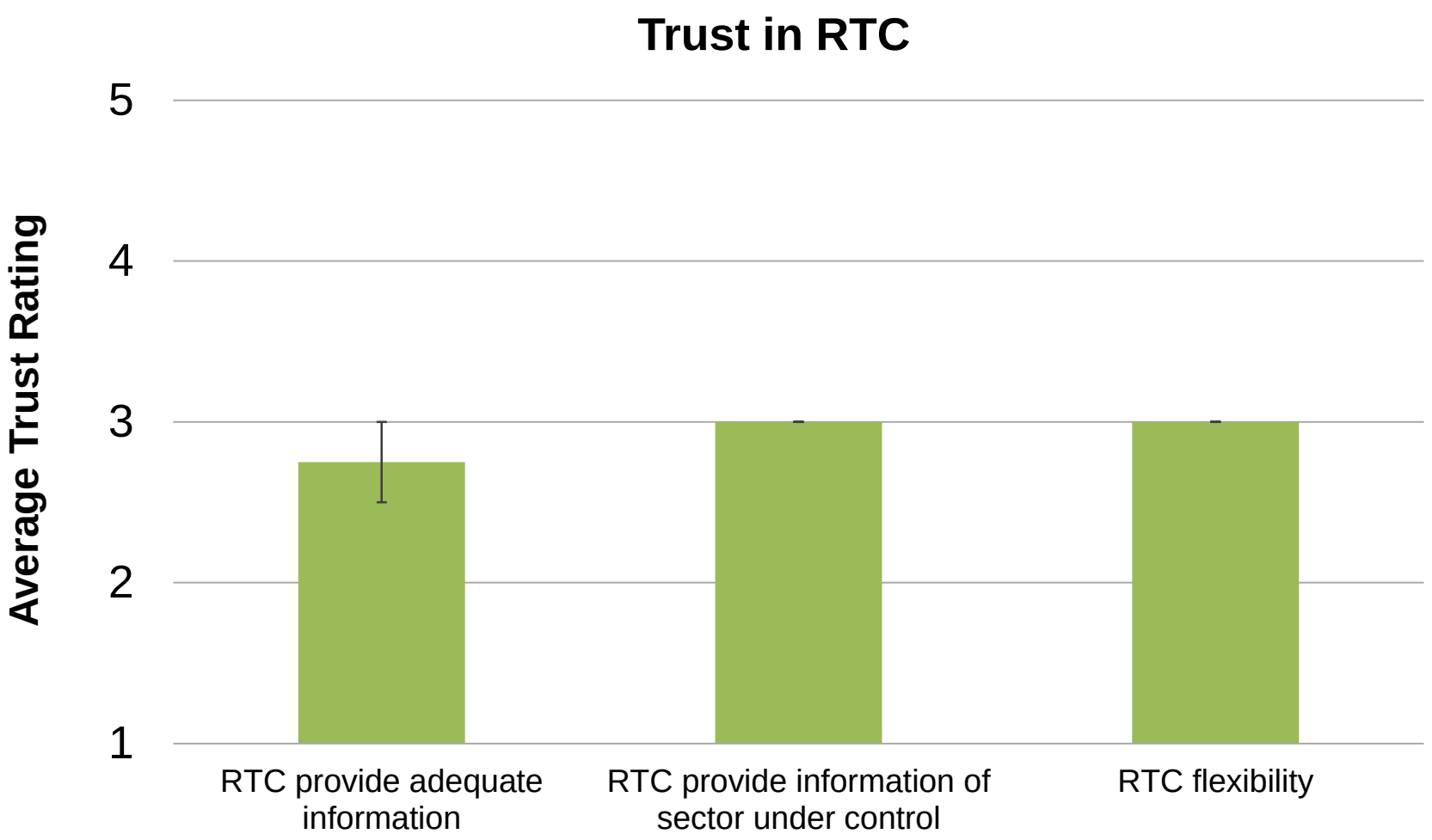
# Trust - RTC

- Collected at the end of the week in the post-study questionnaire
- 8 Questions
  - 1.1 Trust that PBA provided adequate times
  - 1.2 Extent to which you had to crosscheck the validity of PBA
  - 1.3 RTC provide adequate information to manage operations in sector
  - 1.4 RTC provide enough info to keep you aware of sector
  - 1.5 PBA impact ability to manage traffic in sector
  - 1.6 Gate hold advisories impact ability to manage traffic in sector
  - 1.7 RTC give you flexibility to complete your task
  - 1.8 Awareness of advisories when turned off or on
- Used a rating scale of 1 (low trust) to 5 (high trust)
- Data analysis based on four subjects (no RM)
- Note: Results are low power and not statistically significant

# Trust in Pushback Advisories (RTC)



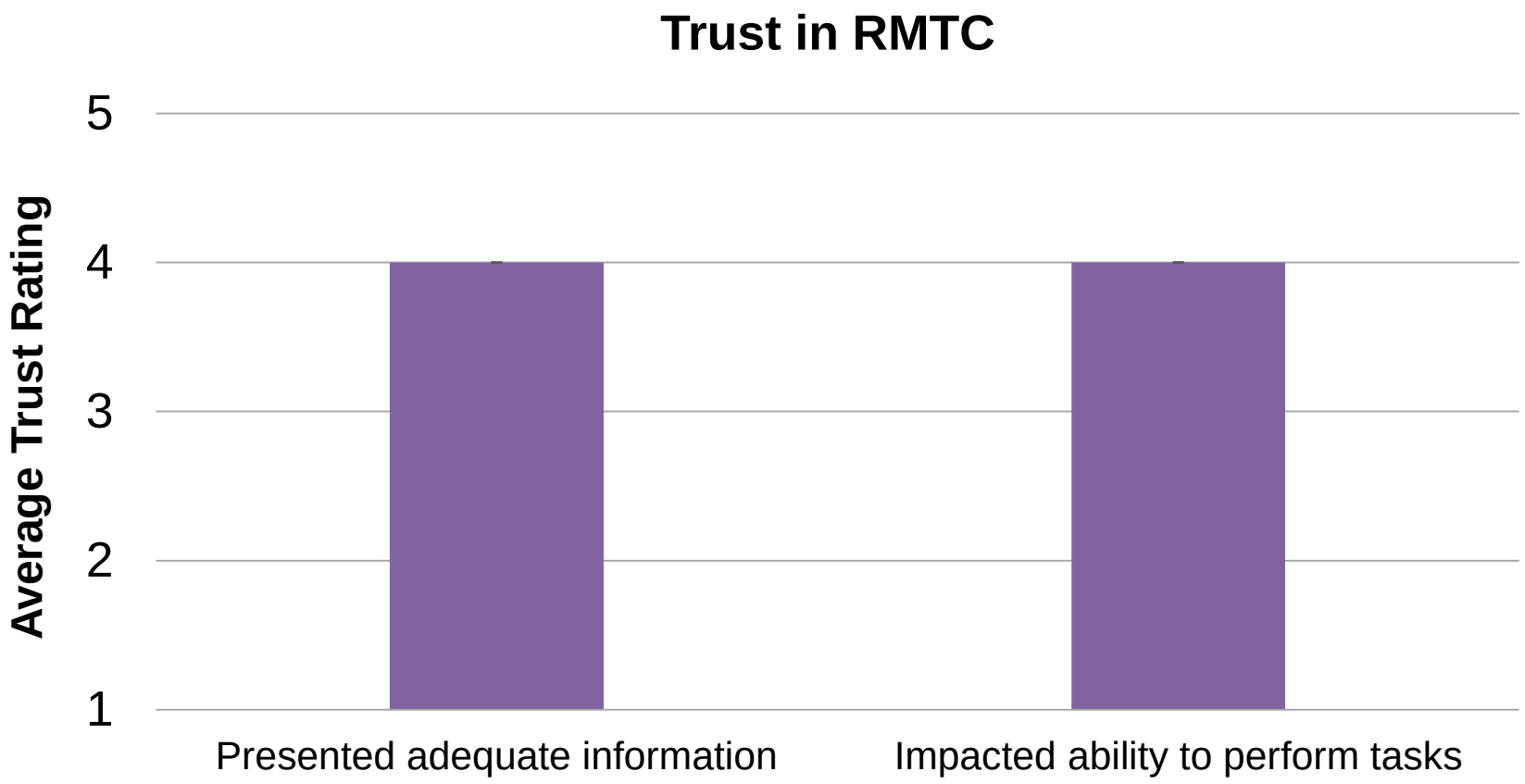
# Trust in the RTC



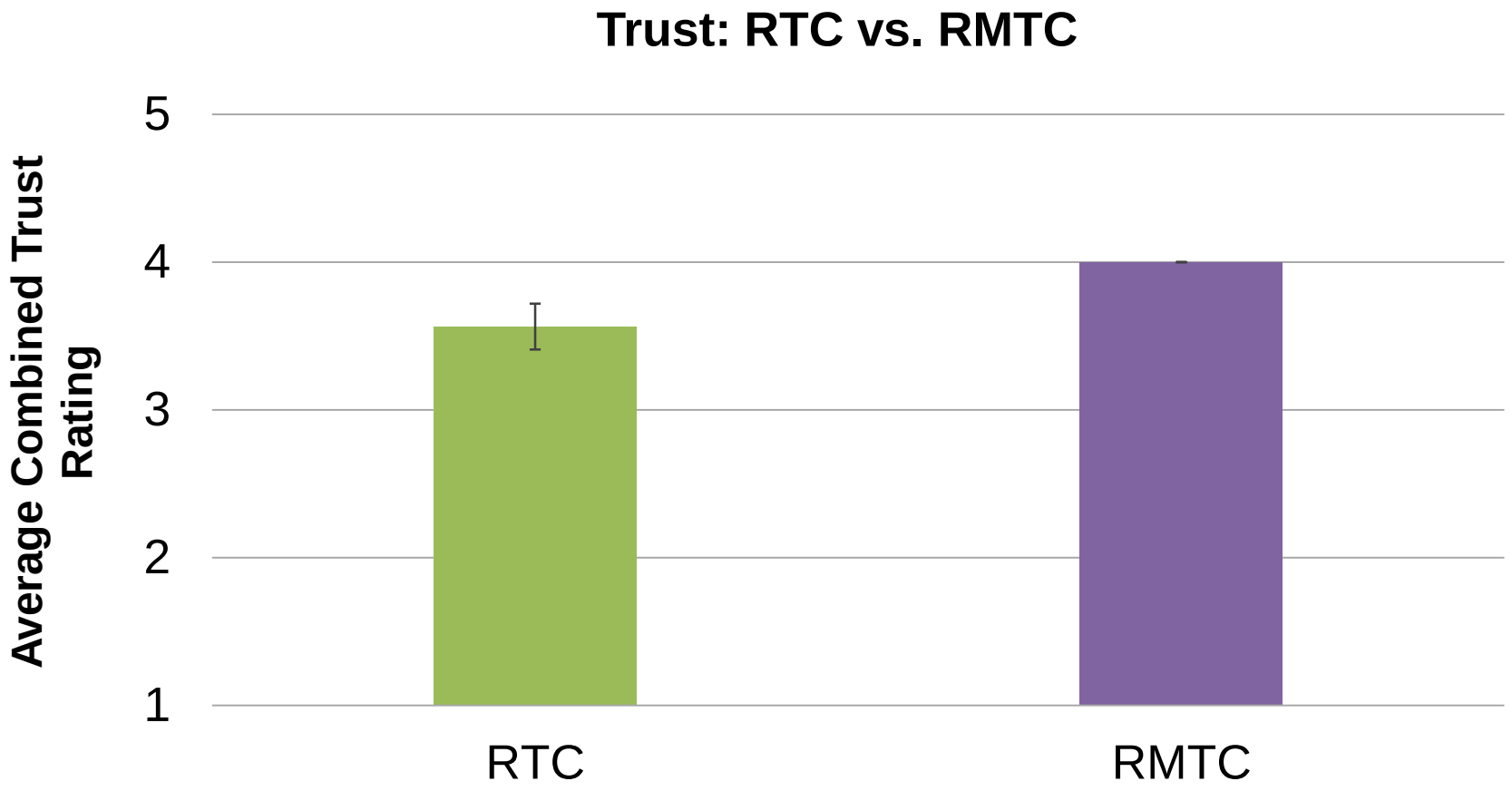
# Trust - RMTC

- Collected at the end of the week in the post-study questionnaire
- Questions
  - 1.1 Did the RMTC provide you with adequate information to manage operations?
  - 1.2 How did the RMTC display impact your ability to perform your ramp manager tasks?
- Used a rating scale of 1 (low trust) to 5 (high trust)
- Data analysis based RM only (one data point per question)
- Note: Results are low power and not statistically significant

# Trust - RMTC


Trust in RMTC
Average Trust Rating
5
4
3
2
1
Presented adequate information
Impacted ability to perform tasks

# Trust – RTC vs. RMTC



# Trust: Summary

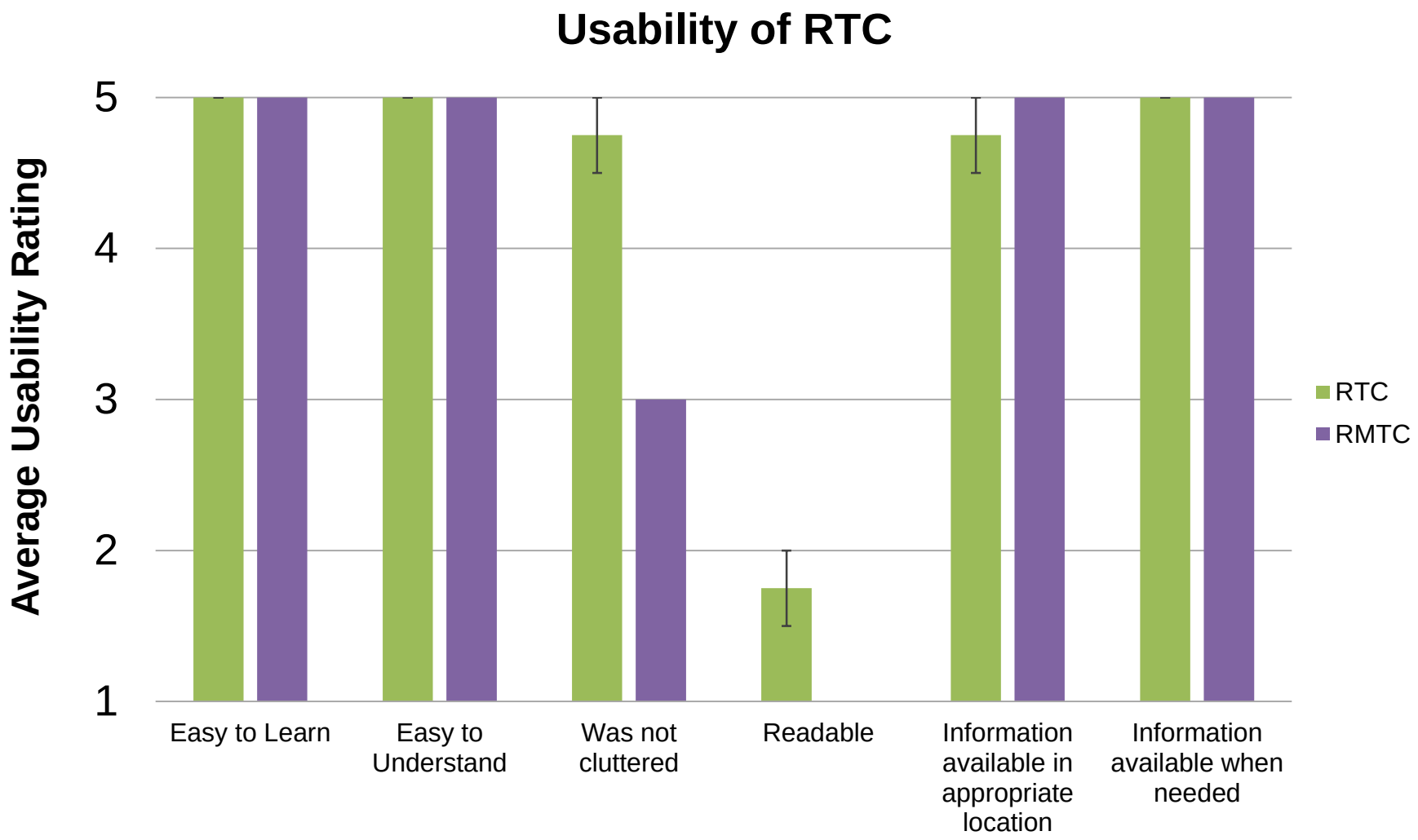
- Overall, trust scores were relatively high
- No major difference between RTC and RMTC trust levels
- Enabled RCs and RM to perform their tasks
- Participants commented that the RTC display was missing some key information like arrival gate numbers and aircraft types, but also commented that for controllers who didn't know the CLT airspace, the available information on the RTC was easy to follow

# Usability

# Usability – RTC vs. RMTC

- Collected at the end of the week the post-study questionnaires
- 6 Questions
  - 1.1 The features were easy to learn
  - 1.2 The features were easy to understand
  - 1.3 The RMTC display was not cluttered
  - 1.4 The RMTC display was readable
  - 1.5 The information was available in an appropriate location
  - 1.6 The information was available to me when I needed it
- Used a rating scale of 1 (poor usability) to 5 (good usability)
- Data analysis for RTC based on four subjects (no RM)
- Data analysis for RMTC based RM only (one data point per question)
- Note: Results are low power and not statistically significant

# Usability – RTC vs. RMTC



# Usability: Summary

- Overall, usability scores were high
- RM was concerned about the clutter when more arrivals are on the map
- Biggest concern for both RMTC and RTC was the readability of the font sizes. RM was also concerned that the 27” RMTC display was too small for performing RM tasks and readability.
- This has been fixed in the subsequent versions of RTC