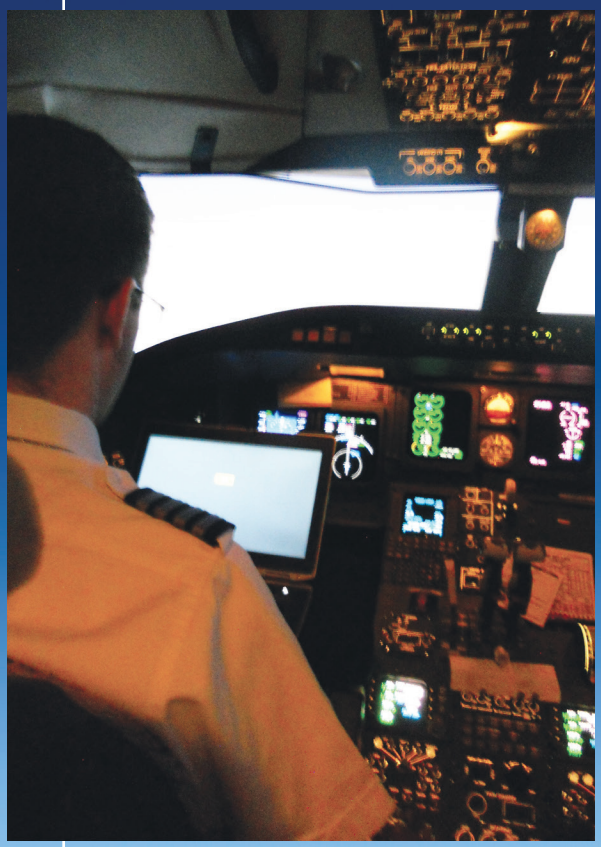


Simulator Experiment on Fatigue in Multi-Segment Operations

WSU-RAA Fatigue Study

Washington State University

RAA
Regional Airline Association

WSU–RAA Fatigue Study

Simulator Experiment on Fatigue in Multi-Segment Operations

Scientific Steering Committee:

Hans P.A. Van Dongen, PhD, WSU principal investigator

J. Lynn Caldwell, PhD, US Navy senior research psychologist

Bob Frisch, Air Wisconsin vice president of flight operations

Ryan Gibson, Air Wisconsin regional chief pilot

Paul Preidecker, Air Wisconsin chief flight instructor

Richard Swindell, Air Wisconsin MEC chair

Matthew Hintze, Air Wisconsin MEC representative

Peter F. Demitry, MD, MPH, ALPA consultant

Scott Foose, RAA senior vice president of operations and safety

Consensus Report

2015 February 25

Prepared by:

Hans P.A. Van Dongen, PhD
Principal Investigator

Sleep and Performance Research Center
Washington State University Spokane

EXECUTIVE SUMMARY

At the time of this research project, passenger-carrying domestic flights in U.S. commercial aviation were governed by Title 14 Code of Federal Regulations (CFR) Part 121. CFR Part 121 restricted pilot duty time to a maximum of 16 hours and flight time to a maximum of 8 hours, regardless of the number of flight segments flown. In 2009, the Federal Aviation Administration (FAA) convened a "Flight and Duty Time Limitations and Rest Requirements" Aviation Rulemaking Committee (Flight Time ARC) to make recommendations on new flight and duty time limits. In 2011, at the conclusion of the Flight Time ARC and following a public comment period, the FAA made a final ruling on new flight and duty time regulations, based in part on the extant science of fatigue. The new regulations, included in CFR Part 117, took effect on January 4, 2014.

Under CFR Part 117, duty duration will be limited as a function of duty start time and as a function of the number of flight segments in the duty period. The relevance of duty start time is supported by established fatigue science. However, the science pertaining to number of segments is far less comprehensive. Washington State University (WSU) and the Regional Airline Association (RAA) launched the WSU-RAA Fatigue Study to help fill this science gap. As part of a multi-phase project, the phase described here consisted of a flight simulator-based experiment. In this phase, fatigue experienced by pilots was compared between a duty day with multiple take-offs and landings versus a duty day of equal duration with a single take-off and landing. The data and findings of this experiment are presented here.

Twenty-four active-duty regional airline pilots (12 Captains and 12 First Officers; age range 24–49) participated in the study. Each crew of two pilots flew a high-fidelity, full-flight (motion) Bombardier CRJ-200 flight simulator during two consecutive 9-hour duty days each starting at 05:15. In randomized order, one duty day contained five flight segments and the other duty day contained one flight segment. The simulator flights were the same for all pilot crews and were flown like normally operated, passenger-carrying flights. The experimental conditions were operationally realistic to the highest degree possible within a simulator facility.

During each of the two duty days, 10 fatigue test bouts were administered. In the multi-segment duty day, the fatigue test bouts occurred immediately after top of climb (while on autopilot) and immediately after landing. In the single-segment duty day, the fatigue test bouts occurred at times of day matching those in the multi-segment duty day. All test bouts contained the Samn-Perelli (SP) Subjective Fatigue Checkcard, a 10-minute Psychomotor Vigilance Test (PVT), and the Karolinska Sleepiness Scale (KSS). After every flight segment, a simulator instructor rated the pilots' flight performance in terms of airspeed, altitude, course control and air traffic control (ATC) communication using a standard flight evaluation form. Throughout the study, the pilots wore a wrist activity monitor to measure their sleep times.

The study results showed a greater build-up of fatigue in the multi-segment duty day than in the single-segment duty day, both objectively assessed (PVT) and subjectively reported (SP and KSS). In relation to studies of simulated night shifts and nighttime sleep deprivation, the fatiguing effect of the multiple take-offs and landings was comparatively modest. The increased

fatigue in the multi-segment duty day was not associated with lower ratings of the pilots' flight performance by the simulator instructor.

To understand the operational applicability and generalizability of these key findings, a number of study limitations should be considered. The study was conducted over nine-hour duty periods commencing at 05:15, on days two and three of a four-day pairing. Although pilots participating in the study were instructed to be ready to respond to any emergencies that might occur, no particularly abnormal circumstances or emergencies were included in the flight scenarios. Additionally, to avoid confounds from variables not within the scope and intent of the study, no situations such as adverse weather, high density air traffic, flight delays, or other potentially complicating factors were included.

While these study limitations are to be considered in the applicability of the results to real-world operations, an important strength of the study design was that the increased fatigue in the multi-segment day could be attributed specifically to the additional take-offs and landings associated with the multiple segments, as other factors that could systematically affect fatigue were held constant or experimentally and statistically controlled for. As such, the study findings represent a significant and effective first step into the scientific area of fatigue in multi-segment operations.

INTRODUCTION

Fatigue is widely considered to be a risk factor in flight performance and safety (e.g., Caldwell, 2005). Fatigue is determined principally by time awake (sleep/wake homeostasis) and time of day (circadian rhythm) (Daan et al., 1984; Dijk et al., 1994; Van Dongen & Dinges, 2000; Åkerstedt, 2003; Satterfield & Van Dongen, 2013), and thus by duration and timing of duty hours (see Figure 1). Following recommendations provided by a "Flight and Duty Time Limitations and Rest Requirements" Aviation Rulemaking Committee (ARC), the Federal Aviation Administration (FAA) recently adopted these principles in its final rule "Flightcrew Member Duty and Rest Requirements" (Federal Aviation Administration, 2011). An additional variable, number of segments flown in a duty period – a correlate of workload – was also included in the new rule. However, there is a dearth of systematic scientific research supporting the inclusion of this variable (Van Dongen et al., 2011a).

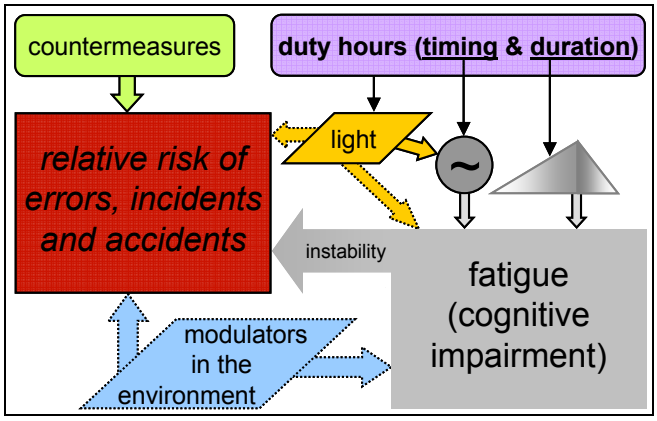


Figure 1: Fatigue risk is determined by interaction of the operational environment with the neurobiology of sleep and sleepiness. Colored elements: operational environment (solid: enduring influences; dotted: transient influences). Gray elements: sleep neurobiology. Triangle: homeostatic pressure for sleep (increasing during wakefulness, decreasing during sleep). Circle: circadian rhythm (oscillating). Adapted from Satterfield & Van Dongen (2013).

The Regional Airline Association (RAA) partnered with Washington State University (WSU), Air Wisconsin Airlines Corporation (AWAC), the Air Line Pilots Association (ALPA), and CAE Inc. to begin to fill the science gap in the FAA's rule making regarding the effect of the number of flight segments in a duty period. The RAA sponsored the WSU-RAA Fatigue Study project, which included a carefully controlled, full-flight simulator experiment described in this report. In the experiment, objectively and subjectively assessed fatigue during a multi-segment duty day was compared with that during a single-segment duty day of equal duration. AWAC and CAE Inc., respectively, provided pilot and simulator resources for the study.

A Scientific Steering Committee (SSC) with broad representation was constituted to oversee the experiment. The SSC consisted of members from the scientific community (one from WSU and one from an independent institution); AWAC airline representatives; AWAC pilots through their representation nationally (ALPA) and locally (ALPA Master Executive Council; MEC); and a representative of the RAA. The members of the SSC are identified on the title page of this report. Employing consensus decision-making, the SSC oversaw the design of the study, discussed the

data analysis approach and the interpretation of the results, and reviewed and approved the present report prior to its release.

METHODS

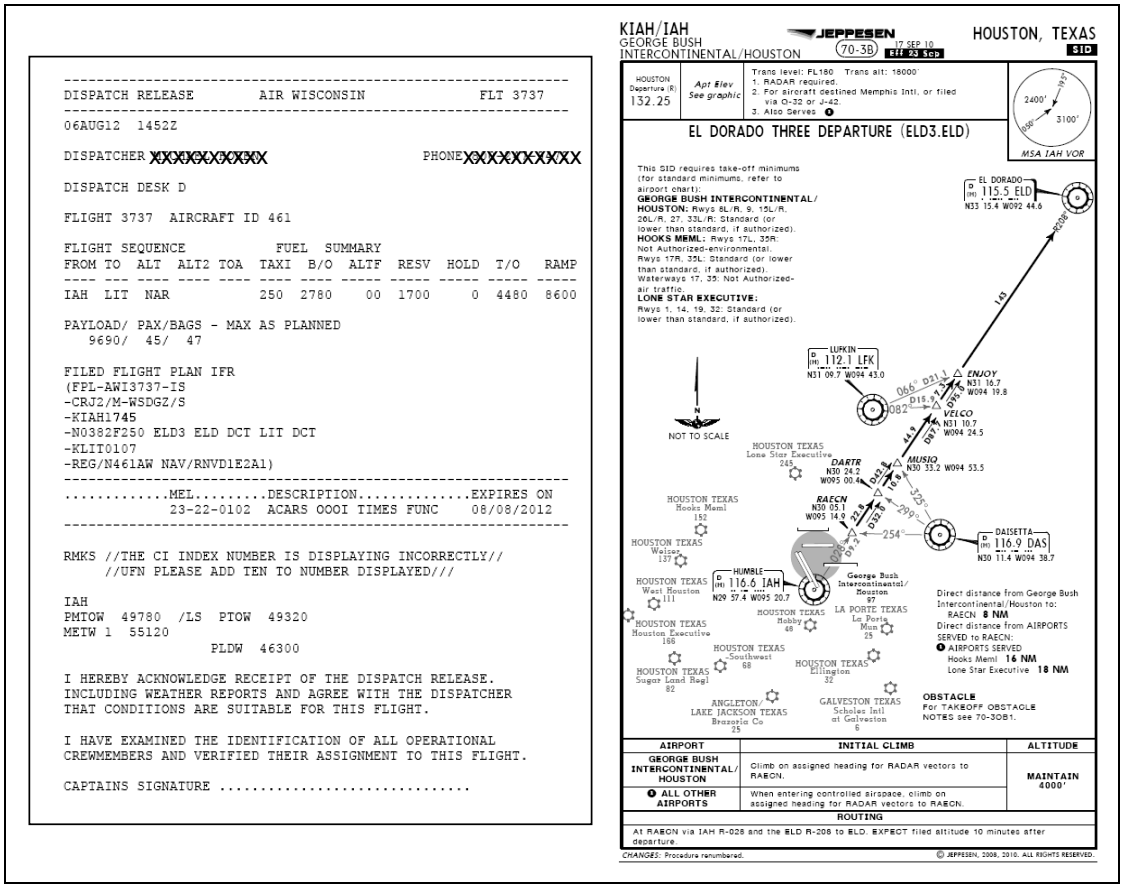
Experimental Design

The full-flight simulator experiment took place at the US Airways Training Center (USATC) in Charlotte, NC. AWAC pilots flew a CAE high-fidelity, moving-base, full-flight simulator of a regional jet – the Bombardier CRJ-200 – flown by the airline (see Figure 2). The objective of the study was to compare fatigue in a duty day with multiple take-offs and landings versus fatigue in a duty day with only one take-off and landing, holding everything else constant as much as possible. Therefore, under the control of a simulator operator – a flight instructor who did not intervene with regard to the pilots' flight performance – two different flight schedules were simulated. One schedule was a multi-segment flight day in which pilots flew from St. Louis, MO (STL) to Springfield, IL (SPI) to Dallas, TX (DFW) to Corpus Christi, TX (CRP) to Houston, TX (IAH) to Little Rock, AR (LIT) in a single duty day. The other schedule was a matching single-segment flight day in which pilots flew from Miami, FL (MIA) to Seattle, WA (SEA). The specific airports in these schedules were selected to represent class B airspace typical of large airline hubs and smaller satellite airports, and were deemed unlikely to be familiar to the pilots participating in the study.

Figure 2: High-fidelity, moving-base Bombardier CRJ-200 flight simulator at the USATC used during the study.

The experimental procedures were scripted in detail so as to provide the same experience for all participating pilot crews and a high degree of standardization for fatigue measurements, yet were as realistic as possible to resemble normal duty days in the real world. Dispatch releases for each of the flight segments and navigation charts for each of the airports (to the level of detail that they were represented in the flight simulator) were generated specifically for the study (see Figure 3). For both the multi-segment day and the single-segment day, scheduled (first) block-

out was at 06:00 and scheduled (last) block-in was at 14:00. With 45 minutes preparation time beforehand and 15 minutes wrap-up time afterwards, each schedule constituted a 9-hour duty period. This duty duration was selected because it contained an 8-hour flight in the single-segment day, which by rule is the maximum currently allowable flight time for unaugmented crews.



```
-----------------------------------------------------------
DISPATCH RELEASE        AIR WISCONSIN              FLT 3737
-----------------------------------------------------------
06AUG12  1452Z

DISPATCHER XXXXXXXXXX                      PHONEXXXXXXXXXX

DISPATCH DESK D

FLIGHT 3737  AIRCRAFT ID 461

FLIGHT SEQUENCE           FUEL  SUMMARY
FROM TO  ALT  ALT2 TOA  TAXI  B/O  ALTF  RESV  HOLD  T/O   RAMP
---- --- ---- ---- ---- ---- ---- ----- ----- ----- ----- -----
IAH  LIT  NAR           250  2780    00  1700     0  4480  8600

PAYLOAD/ PAX/BAGS - MAX AS PLANNED
   9690/  45/  47

FILED FLIGHT PLAN IFR
(FPL-AWI3737-IS
-CRJ2/M-WSDGZ/S
-KIAH1745
-N0382F250 ELD3 ELD DCT LIT DCT
-KLIT0107
-REG/N461AW NAV/RNVD1E2A1)
-----------------------------------------------------------
............MEL.........DESCRIPTION.............EXPIRES ON
            23-22-0102  ACARS OOOI TIMES FUNC    08/08/2012
-----------------------------------------------------------

RMKS //THE CI INDEX NUMBER IS DISPLAYING INCORRECTLY//
     //UFN PLEASE ADD TEN TO NUMBER DISPLAYED///

IAH
PMTOW  49780  /LS  PTOW  49320
METW 1  55120
                 PLDW  46300

I HEREBY ACKNOWLEDGE RECEIPT OF THE DISPATCH RELEASE.
INCLUDING WEATHER REPORTS AND AGREE WITH THE DISPATCHER
THAT CONDITIONS ARE SUITABLE FOR THIS FLIGHT.

I HAVE EXAMINED THE IDENTIFICATION OF ALL OPERATIONAL
CREWMEMBERS AND VERIFIED THEIR ASSIGNMENT TO THIS FLIGHT.

CAPTAINS SIGNATURE ..............................
```


Figure 3: Example dispatch release cover page (left, de-identified) and example Jeppesen airport chart (right) as used in the study.

For the study, pilots were scheduled for a pairing of 4 consecutive days, following at least one rest day. On the first day, a crew pair – one Captain (CA) and one First Officer (FO) – deadheaded to Charlotte, NC (CLT) and reported for a briefing in the crew hotel at 17:00. On the second and third days, the pilots were scheduled to fly the simulator in the USATC. On both of these days, they took a shuttle from the crew hotel to the USATC at 05:00, and the start of their duty period was at 05:15. On the fourth day, the pilots were scheduled to deadhead back to their domicile or other pre-arranged destination (although several arranged to deadhead back home on the third day after completion of the experiment).

During the briefing on the first day of the pairing, WSU research assistants explained the study procedures to the CA and the FO. Both pilots practiced a 10-minute Psychomotor Vigilance Test (PVT). They also practiced filling out the Samn-Perelli (SP) Subjective Fatigue Checkcard and the Karolinska Sleepiness Scale (KSS) subjective sleepiness questionnaire. (See “Fatigue

Testing" below for further details.) In addition, the pilots started a paper-and-pencil sleep/wake/duty log, and recorded their schedule over the past two days. They began wearing an Actiwatch-2 wrist activity monitor (Mini-Mitter Company, Inc.) to measure estimated sleep times during the study (see Figure 4). The pilots kept the log and wore the wrist activity monitor through to the end of the third day of the pairing.

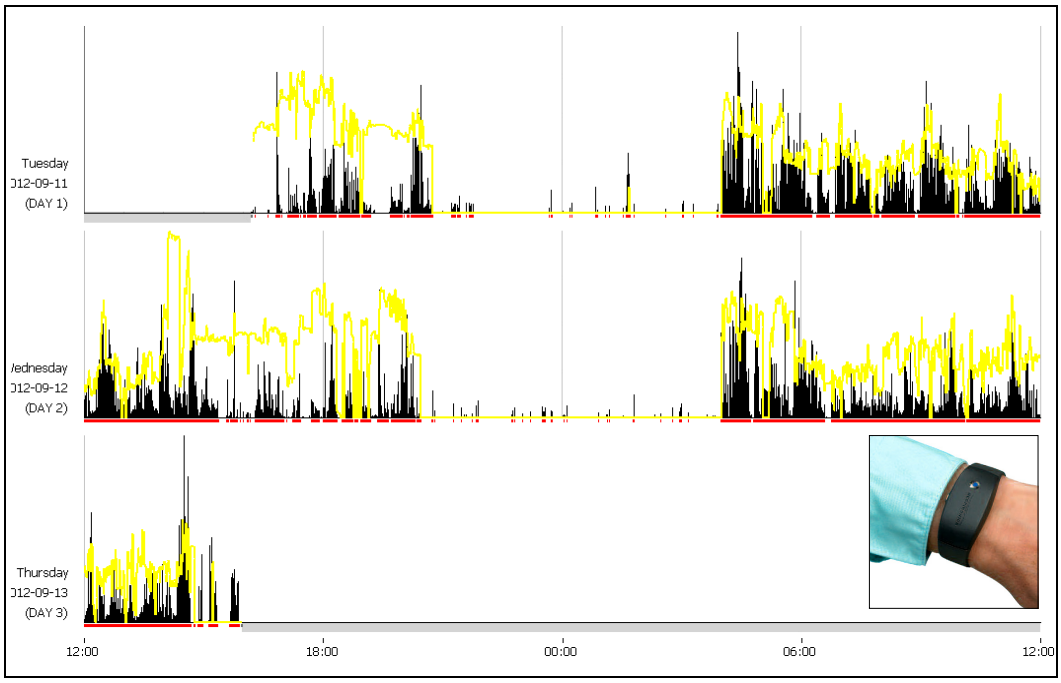


Figure 4: Sample record from a wrist activity monitor worn by one of the pilots during the study. The record shows data from the first (top), second (middle) and third (bottom) 24-hour periods in the pairing (data are plotted from noon to noon). Black peaks (underlined in red) indicate periods of high activity. Yellow curves display light exposure. Sleep periods can be clearly recognized by the almost complete absence of activity and significantly reduced light exposure. Gray-marked sections indicate periods when the wrist activity monitor was not worn (before the briefing on the first day and after the end of duty on the third day). Inset on bottom right: photograph of the Actiwatch-2 wrist activity monitor used in the study.

For the simulator flight days (the second and third days in the pairing), half the pilot sample was assigned randomly to the multi-segment duty period on the second day (first simulator session) and the single-segment duty period on the third day (second simulator session); the other half had the order reversed. The pilots were told that the simulator flights would be flown like real passenger-carrying flights and under the provisions of CFR Part 121. They were asked to report fit for duty as per normal duty requirements (they were not given any specific instructions about how to manage their sleep). They came to the simulator dressed in uniform and carrying their flight bags just as they would during normal passenger-carrying flights.

On each of the two simulator flight days, after arriving at the USATC, the CA and the FO met with a WSU research assistant in the AWAC briefing room to review expectations for the day. They were asked to keep their cell phones turned off in the simulator, to give their best effort and refrain from talking during performance testing, to adhere to time limits for breaks during the single-segment day, etc. Although the flight scenarios contained no unexpected procedures or

emergencies, the pilots were instructed to be ready to respond to any emergencies that might occur, just like in the real world. They were also told who was to fly which segment during the duty period. For the multi-segment day, the CA always flew segments 1 (STL–SPI), 3 (DFW–CRP) and 5 (IAH–LIT); the FO always flew segments 2 (SPI–DFW) and 4 (CRP–IAH). For the single-segment day, either the CA or the FO flew the entire segment (MIA–SEA) as determined by randomization (counterbalanced).

The pilots then obtained a breakfast bag and a dispatch release for their (first) flight of the day. They went to the simulator and prepared it as if readying a real airplane for a real flight. They conducted a simulated walk-around (using a marked path around the flight simulator), settled into the cockpit, and conducted their first-flight-of-day checks. Throughout the duty period they interacted with air traffic control (ATC), where the simulator operator served as the air traffic controller. Cued by ATC at the scheduled time, they began taxiing and took off. While en route, they made typical use of the autopilot. Airspace-specific ATC background chatter (recorded prior to the study) was played back over the pilots' headsets or overhead speakers to add to the realism of the flight. At the end of the flight segment, the pilots conducted a routine descent and landed the airplane, again interacting with ATC, and taxied from the runway to a designated gate. All flights were conducted in accordance with FAA instrument flight rules (IFR).

In the multi-segment day, the pilots flew a total of five segments. They utilized the time available during the turns between the flight segments to use the restroom and to get food and drink (sandwiches, salads, granola bars, apples, water bottles, and quarters for coffee and soda machines were provided). A few went outside to smoke. One pilot of each pair also picked up the next flight release in the AWAC briefing room, while the other conducted a walk-around of the simulator. The pilots then did their normal flight checks. Based on data provided by regional airlines prior to the study, which showed that five-segment duty periods include one airplane swap on average, a minimum equipment list (MEL) associated airplane swap was put in after the second segment (after landing at DFW). The pilots removed their belongings from the airplane, re-entered the simulator as a new airplane (with different aircraft number and log), and conducted acceptance checks like they would normally do after an airplane swap.

In the single-segment day, there were simulator off-motion opportunities to allow pilots to go to "the back of the plane." These opportunities were scheduled at the same times as when the turns were scheduled in the multi-segment day. The pilots could opt to use the off-motion opportunities, one pilot at a time (while the other continued to fly the simulator), to quickly use the lavatory (a nearby restroom in the USATC) and to get food and drink (sandwiches, salads, granola bars, apples, water bottles, and quarters for coffee and soda machines were provided). Pilots could not go outside to smoke. Normally, the CRJ-200 does not carry enough fuel for the 8-hour flight of the single-segment day, but for the study fuel was added artificially by the simulator operator during the flight to circumvent this practical limitation.

At the end of each of the two simulator flight days, after the last block-in of the day at around 14:00, the pilots gathered their belongings and exited the simulator. On the first simulator flight day, they then took the USATC shuttle back to the crew hotel. On the second simulator flight day (the third day in the pairing), they filled out a feedback form and then took the USATC shuttle to the crew hotel or to the airport.

Fatigue Testing

Real-time fatigue testing during the study was done with a 10-minute Psychomotor Vigilance Test (PVT) (Lim & Dinges, 2008). The PVT is a simple reaction time task, with stimuli occurring randomly at intervals of 2 to 10 seconds. It was administered on a calibrated laptop (Pulsar Informatics Inc.) and performed by both pilots individually at the same time while the simulator was flown using the autopilot (in cruise flight) or parked on a taxiway (after landing) (see Figure 5). The primary outcome variable extracted from the reaction time data was the total number of lapses of attention ("PVT lapses"), defined as the number of reaction times ≥ 500 milliseconds, which was selected a priori as the primary objective measure of fatigue for the study. Secondary outcome variables extracted were the mean reaction time ("PVT mean RT") and the number of false starts ("PVT false starts").

Figure 5: A pilot taking the PVT during a simulator flight. The photograph was taken during a dry run of the study, showing a pilot who did not participate in the actual study. Used with permission from AWAC and from the pilot shown.

Fatigue testing was also done with two paper-and-pencil subjective measures: the Samn-Perelli (SP) Subjective Fatigue Checkcard (Samn & Perelli, 1982)[*] and the Karolinska Sleepiness Scale (KSS) (Åkerstedt & Gillberg, 1990). On the SP scale, pilots rated whether they felt better than,

[*] A simplified version called the Crew Status Check is available and in use as well (Samn & Perelli, 1982). It is a Likert-type scale that resembles the KSS and ranges from 1 (fully alert) to 7 (completely exhausted).

same as, or worse than 10 given statements related to fatigue (see Figure 6, left). Scores on the SP scale can range from 0 (severe fatigue) to 20 (unusually wide awake). On the KSS, which is a Likert-type scale, pilots rated their sleepiness from 1 (extremely alert) to 9 (extremely sleepy) (see Figure 6, right).

STATEMENT	BETTER THAN	SAME AS	WORSE THAN
1. Very Lively			
2. Extremely Tired			
3. Quite Fresh			
4. Slightly Pooped			
5. Extremely Peppy			
6. Somewhat Fresh			
7. Petered Out			
8. Very Refreshed			
9. Fairly Well Pooped			
10. Ready to Drop			

1. ☐ extremely alert
2. ☐
3. ☐ alert
4. ☐
5. ☐ neither alert nor sleepy
6. ☐
7. ☐ sleepy-but no difficulty remaining awake
8. ☐
9. ☐ extremely sleepy-fighting sleep

Figure 6: SP subjective fatigue questionnaire (left) and KSS subjective sleepiness questionnaire (right).

Each fatigue test bout was comprised of the SP scale, the 10-minute PVT, and the KSS. For each flight during the multi-segment day, pilots completed fatigue testing right after reaching cruise speed while on autopilot and right after landing while parked on a taxiway before taxiing to the gate. During the single-segment day, pilots completed fatigue testing at the same times as during the multi-segment day, with nine tests en route while flying level and on autopilot, and one right after landing while parked on a taxiway before taxiing to the gate. A WSU research assistant was present in the back of the simulator to administer the fatigue test bouts at the predetermined times.

Each pilot's flight performance was evaluated by the simulator operator after every flight segment, using a standard flight evaluation form (see Appendix 1). The evaluation yielded scores for the pre-flight, gate departure, take-off, climb, cruise, descent, approach, landing, and gate arrival phases of flight, as well as an overall (i.e., calculated average) flight performance score. All scores were on a scale from 0 to 5, with 5 representing best flight performance.

Two simulator operators and two WSU research assistants took turns in staffing the simulator sessions of the study. Each pilot crew worked with the same simulator operator and the same WSU research assistant throughout the experiment. The simulator operators and WSU research assistants were assigned to the study runs in counterbalanced fashion.

Statistical analyses employed mixed-effects analysis of variance (ANOVA), with a random effect over subjects on the intercept to account for inter-individual differences (Van Dongen et al., 2004b). Statistical analyses were performed using SAS for Windows (version 9.2).

Research Subjects

The number of pilots to be studied, which was 24, was determined in advance based on a statistical power calculation (see Appendix 2). The research volunteers were 12 active-duty CAs and 12 active-duty FOs, randomly recruited from among the AWAC lineholder pilots certified to fly the Bombardier CRJ-200. One lineholder pilot scheduled for the study became unavailable and was replaced by a reserve. All 24 pilots completed the study. The pilots were 33.2 years old on average (range 24–49), and the sample included two women. The CAs had 9,688 hours of total flight experience and 5,979 hours of CRJ-200 flight experience on average; the FOs had 2,829 hours of total flight experience and 1,475 hours of CRJ-200 flight experience on average (see Figure 7).

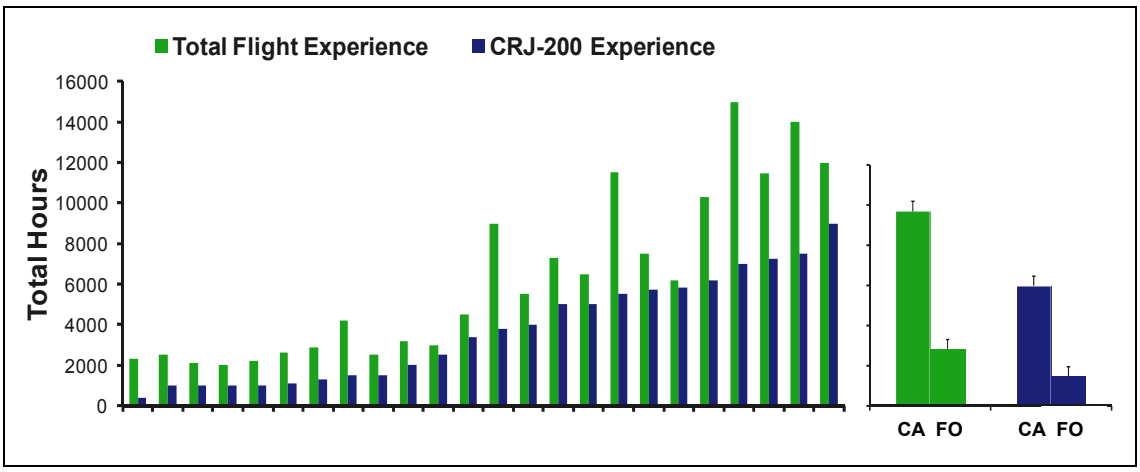


Figure 7: Breakdown of flight experience by participating pilot. The pilots are ordered from the least to the most hours of flight experience on the Bombardier CRJ-200 (left). Mean flight experience for the Captains (CA) and First Officers (FO) is also shown (right). Black whiskers: standard errors.

The participating pilots were domiciled at DCA (3 pilots), LGA (8 pilots), ORF (6 pilots), and PHL (7 pilots). They flew to CLT on the day of the briefing from a variety of locations in the Eastern, Central and Mountain time zones: Burlington, IA; Easton, PA; Fairfield, OH; Fort Lauderdale, FL; Fort Worth, TX; Fuquay, NC; Galena, IL; Gilcrest, CO; Lexington, KY; Melbourne, FL; Miami, FL; New York, NY; Norfolk, VA; Orlando, FL (2 pilots); Raleigh, NC (2 pilots); Rockford, IL; Salt Lake City, UT; South Hadley, MA; St. Louis, MO; Virginia Beach, VA; Virginia Beach, VA; Washington, DC.

The study was approved by the WSU Institutional Review Board (IRB). All pilots gave written, informed consent. The data collected from the individual pilots have been kept confidential and were not shared with AWAC, ALPA, RAA, the SSC, or any other parties besides the WSU research team.

RESULTS

Experimental Procedures

Figure 8 depicts the timeline of the flight schedules, as actually flown in the simulator, during the multi-segment and single-segment days of the study. Since the 24 pilots were studied in crew pairs (CA and FO), there were 12 study runs. Table 1 shows the block-out times, block-in times and flight durations of each of the segments in the multi-segment and single-segment days. The standard deviations (SDs) over study runs are all in the order of a few minutes, demonstrating high consistency in flight schedules across the study runs. Also, the first block-out and last block-in of the multi-segment day occurred at nearly the same time of day as the block-out and block-in times of the single-segment day, and the flight duration in the single-segment day was almost exactly 8 hours as planned. Table 2 shows the times and durations of the off-motion opportunities during the single-segment day, which exhibited a reasonable level of consistency across the study runs as well. Figure 9 displays the start times of the fatigue test bouts administered to the pilots during the multi-segment day versus the single-segment day, demonstrating a high degree of matching.

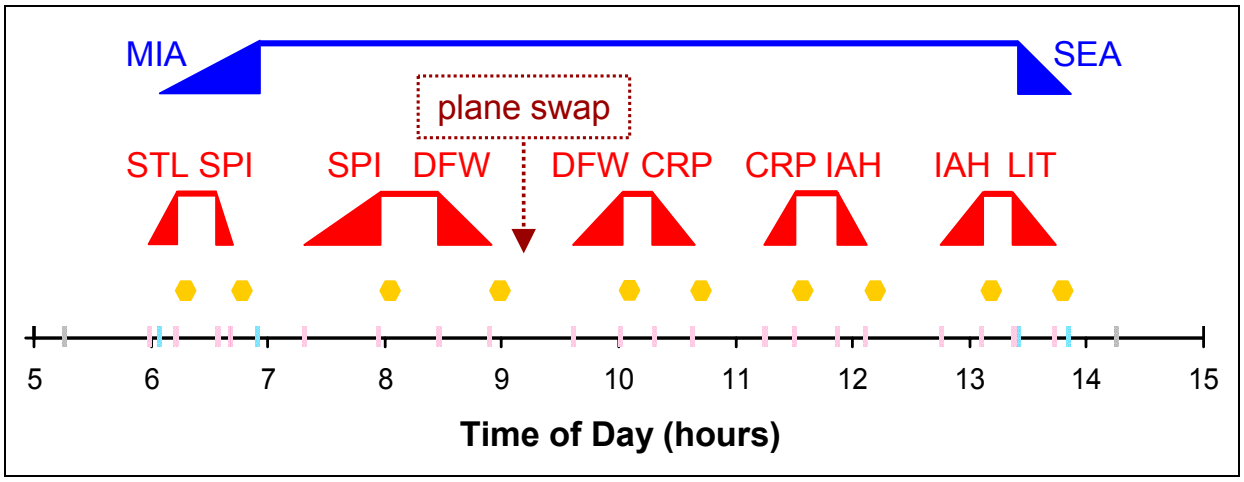


Figure 8: Study timeline as the study was executed. The black line represents the timeline from 05:00 until 15:00. The beginning (05:15) and end (14:15) of the 9-hour duty period for both the multi-segment and single-segment days are indicated by gray tick marks. The times of take-off, top of climb, top of descent and landing for each of the segments in the multi-segment day, averaged over the pilot crews, are indicated by shaded red tick marks. The take-off (red upward triangle), cruise (red horizontal bar) and landing (red downward triangle) phases of flight are again indicated above the timeline; the respective airports are shown in red as well. An airplane swap occurred between the second and third segments. The times of take-off, top of climb, top of descent and landing for the single-segment day, averaged over the pilot crews, are indicated by shaded blue tick marks. The take-off (blue upward triangle), cruise (blue horizontal bar) and landing (blue downward triangle) phases of flight are again indicated on top; the respective airports are shown in blue as well. The yellow markers indicate the times of the fatigue test bouts in both the multi-segment and single-segment days.

Table 1: Means and standard deviations (SD) over study runs for block-out times, block-in times, and flight durations of each of the segments in the multi-segment duty day and of the single-segment duty day.

Segment	*Block-out time ± SD (in minutes)*	*Block-in time ± SD (in minutes)*	*Duration (in minutes) ± SD (in minutes)*
1	5:59 ± 2	6:54 ± 3	55 ± 2
2	7:18 ± 4	9:07 ± 3	110 ± 4
3	9:37 ± 2	10:52 ± 2	75 ± 2
4	11:15 ± 2	12:20 ± 3	66 ± 3
5	12:45 ± 3	13:57 ± 3	72 ± 3
Single-segment	6:04 ± 8	14:04 ± 8	479 ± 4

Table 2: Means and standard deviations (SD) over study runs for simulator off-motion opportunities during the single-segment duty day.

Off-motion opportunity	*Start time ± SD (in minutes)*	*Duration (in minutes) ± SD (in minutes)*
1[a]	7:00 ± 7	12 ± 8
2	9:13 ± 7	21 ± 6
3	10:53 ± 7	18 ± 4
4[a]	12:21 ± 2	14 ± 6

[a] In two of the 12 study runs, pilots chose to skip the first off-motion opportunity. In one of these two study runs, pilots also chose to skip the fourth off-motion opportunity. In these instances, the simulator stayed on motion and the flight continued normally.

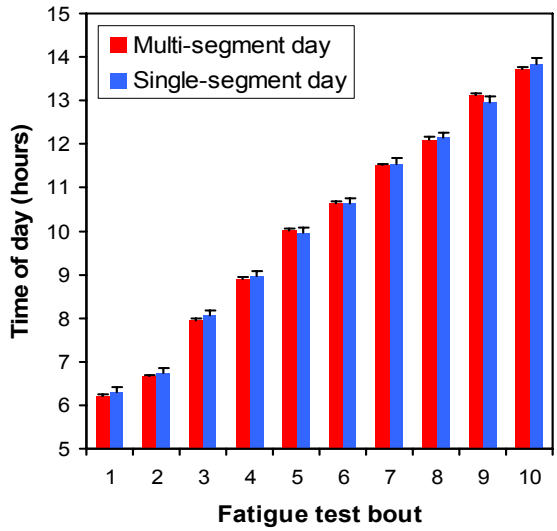


Figure 9: Side-by-side comparison of start times of the fatigue test bouts administered to the pilots in the multi-segment and single-segment days (cf. Figure 8). Black whiskers: standard deviation over pilots.

A few minor glitches occurred during the execution of the study, two of which are noteworthy. During the multi-segment day of the second study run, a fire alarm went off and the USATC had to be evacuated for about 30 minutes. After the event, the simulator operator advanced the flight scenario to get back on the flight schedule timeline, and the fatigue test bout scheduled for 12:55 was performed 12 minutes late. During the single-segment day of the eighth study run, there was a simulator malfunction at 13:08, which prevented normal completion of the flight segment. The

pilots continued the study as if still flying and performed the two remaining fatigue test bouts as scheduled, but were unable to actually land the simulator and block in at SEA. These events did not result in any outliers in the study data.

The participating pilots varied in the amount of sleep they obtained, as assessed with the wrist activity monitors, before each of the two flight simulator sessions (see Figure 10). Before the first simulator session (on the second day of the pairing), between the briefing on the first day of the pairing and the beginning of the first simulator session, they slept 6.1 ± 1.3 hours (mean ± SD).[†] Before the second simulator session (on the third day of the pairing), between the end of the first simulator session and the beginning of the second simulator session, they slept 7.4 ± 1.2 hours, with 8 pilots taking naps in addition to nighttime sleep (without the naps, sleep duration was 7.1 ± 1.5 hours). The difference in sleep duration before the first and second simulator sessions was statistically significant ($F_{1,23}$=15.8, P<0.001). To interpret this finding, it is important to consider that the order of the multi-segment and single-segment flight schedules was randomized over the two simulator sessions. Sleep duration before flying the multi-segment flight schedule was 6.8 ± 1.3 hours. Sleep duration before flying the single-segment schedule was 6.7 ± 1.5 hours. The difference in sleep duration before the multi-segment and single-segment days was not statistically significant ($F_{1,23}$=0.05, P=0.82). Self-reported sleep durations in the paper-and-pencil sleep/wake/duty log correlated with the data from the wrist activity monitors (r=0.64, $F_{1,22}$=34.8, P<0.001).

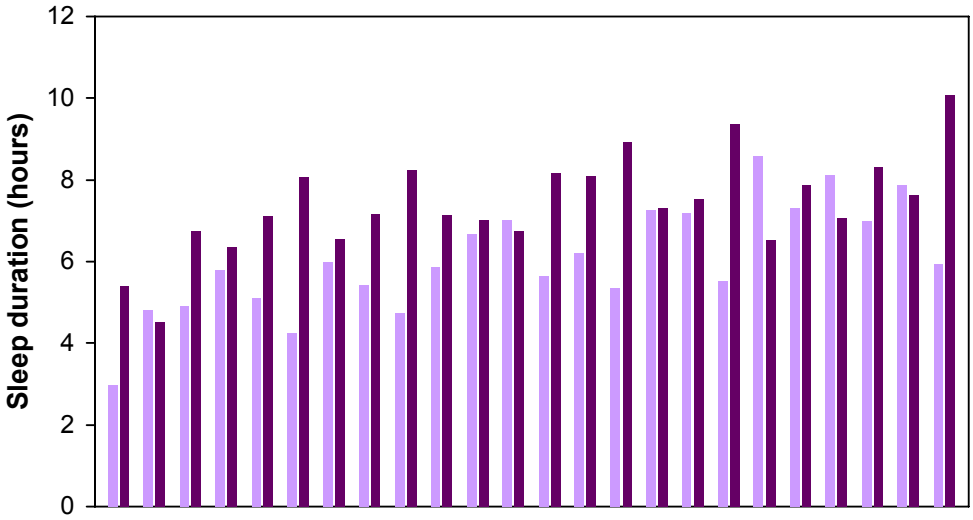


Figure 10: Sleep duration, as measured by wrist activity monitor, preceding the two simulator study days. Each pair of bars represents an individual pilot; from left to right, the pilots are ordered by average sleep duration over the two periods. The light purple bars represent sleep preceding the first simulator session (since the briefing the day before); the dark purple bars represent sleep preceding the second simulator session (since the end of the first simulator session the day before, and including any naps).

[†] Four pilots reported taking a nap during the first day of the pairing, but this occurred before the briefing and thus before the wrist activity monitor was worn. These naps could therefore not be included in the wrist activity monitor-based estimates of sleep duration before the first simulator session.

On the first flight simulator day, the pilots woke up at 04:12 ± 22 minutes (mean ± SD). On the second flight simulator day, they woke up at 04:17 ± 20 minutes. Wake-up time was not significantly different between the first and second simulator days ($F_{1,23}$=1.16, P=0.29). Wake-up time was also not significantly different between the multi-segment and single-segment days ($F_{1,23}$=0.08, P=0.78).

For the day before the briefing (i.e., before the wrist activity monitors were worn), self-reported sleep duration was 7.7 ± 2.3 hours, with 4 pilots reporting naps in addition to nighttime sleep. For two days before the briefing, self-reported sleep duration was 8.3 ± 1.8 hours, with 6 pilots reporting naps in addition to nighttime sleep.

Fatigue Measurements

Statistical analyses of the data from the fatigue test bouts focused on head-to-head, within-pilot comparisons across test bouts between the multi-segment and single-segment days. The primary analysis involved mixed-effects ANOVA of condition (multi-segment versus single-segment) by test bout (1 through 10). Secondary analyses included covariates for pilot age, total flight experience, CRJ-200 flight experience, prior sleep duration, seat (CA or FO), pilot flying or monitoring, order of conditions, simulator session (first or second), and simulator session by condition interaction. Results from secondary analyses are only reported here if the covariate was statistically significant.

Figure 11 shows the results for PVT lapses – the primary objective measure of fatigue in the study – as well as PVT mean RT and PVT false starts. For PVT lapses, there was a main effect of condition ($F_{1,437}$=8.7, P=0.003) indicating greater fatigue in the multi-segment day compared to the single-segment day. Although the difference in PVT lapses between the multi-segment and single-segment days appeared to increase across the test bouts, the interaction of condition by test bout was not statistically significant ($F_{9,437}$=0.9, P=0.52), nor was the main effect of test bout ($F_{9,437}$=1.1, P=0.38). Thus, fatigue as measured with PVT lapses was greater overall in the multi-segment day. Secondary analyses revealed an effect of simulator session ($F_{1,436}$=11.8, P<0.001) and a significant simulator session by condition interaction ($F_{2,436}$=6.7, P=0.001). Due to the randomization of the multi-segment and single-segment flight schedules over the two simulator sessions, when the effects of simulator session and simulator session by condition interaction were controlled for, the main effect of condition remained significant ($F_{1,436}$=8.9, P=0.003).

For PVT mean RT, there was a main effect of condition ($F_{1,437}$=15.4, P<0.001), a main effect of test bout ($F_{9,437}$=2.1, P=0.031), and an interaction of condition by test bout ($F_{9,437}$=2.6, P=0.007). These findings indicate greater fatigue in the multi-segment day compared to the single-segment day, with the difference increasing across the test bouts (see Figure 11, right). Secondary analyses revealed that when pilots were monitoring, they had greater mean RT than when they were flying ($F_{1,436}$=6.8, P=0.010), but the difference was only 5 ± 2 milliseconds (mean ± standard error) and did not affect the condition and test bout main effects and interaction. PVT mean RT was correlated with PVT lapses (r=0.66, $F_{1,436}$=408.0, P<0.001).

For PVT false starts, which tend to increase as a result of fatigue in tandem with PVT lapses (Doran et al., 2001), there was a main effect of condition ($F_{1,437}$=12.8, P<0.001) indicating

greater fatigue in the multi-segment day compared to the single-segment day (see Figure 11, bottom). The interaction of condition by test bout was not statistically significant ($F_{9,437}$=0.8, P=0.59), nor was the main effect of test bout ($F_{9,437}$=1.2, P=0.30). Secondary analyses revealed a significant effect of prior sleep duration as measured by wrist activity monitor ($F_{1,436}$=6.8, P=0.010), where an hour of additional prior sleep led to a reduction of 0.17 ± 0.07 false starts (mean ± standard error). Controlling for prior sleep duration did not affect the condition and test bout main effects and interaction. PVT false starts were somewhat correlated with PVT lapses (r=0.27, $F_{1,436}$=17.6, P<0.001).

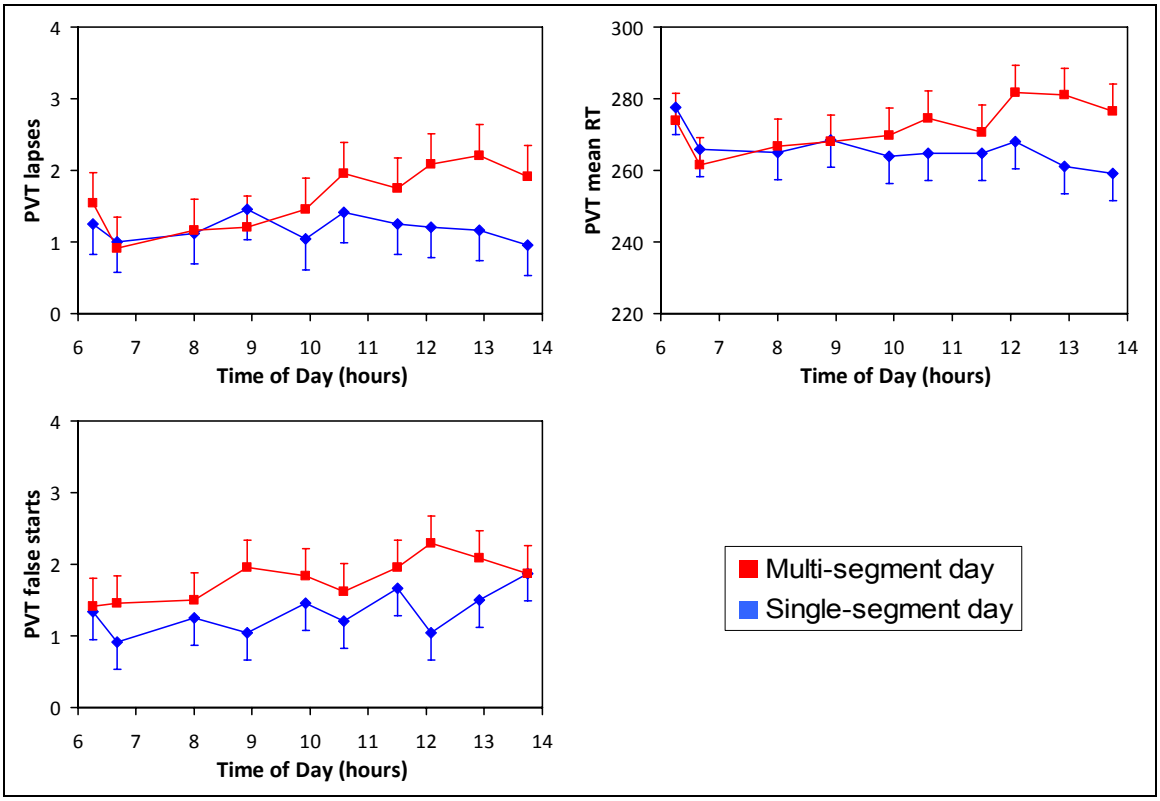


Figure 11: PVT outcome variables plotted over fatigue test bouts in the multi-segment and single-segment duty days. The outcome variable "PVT lapses" (top left) was selected a priori as the primary objective measure of fatigue for the study. Whiskers: standard errors from mixed-effects ANOVA.

Figure 12 shows the results for SP subjective fatigue and KSS subjective sleepiness. For SP subjective fatigue, there was a main effect of condition ($F_{1,437}$=20.1, P<0.001) and a main effect of test bout ($F_{9,437}$=9.1, P<0.001), but no significant interaction of condition by test bout ($F_{9,437}$=0.7, P=0.71). These findings indicate greater subjective fatigue in the multi-segment day compared to the single-segment day, and a build-up of subjective fatigue across the test bouts regardless of condition (see Figure 12, left). Secondary analyses revealed an effect of simulator session ($F_{1,436}$=13.2, P<0.001) and a significant simulator session by condition interaction ($F_{2,436}$=7.0, P=0.001). Due to the randomization of the multi-segment and single-segment flight schedules over the two simulator sessions, when the effects of simulator session and simulator session by condition interaction were controlled for, the main effect of condition remained significant ($F_{1,436}$=20.7, P<0.001) as did the main effect of test bout ($F_{9,436}$=9.3, P<0.001). The

pilots reported greater subjective fatigue when they were flying than when they were monitoring ($F_{1,436}$=4.0, P=0.047), the difference being 0.5 ± 0.2 units (mean ± standard error). This did not affect the condition and test bout main effects and interaction. There was also a significant effect of prior sleep duration as measured by wrist activity monitor ($F_{1,436}$=12.5, P<0.001), where an hour of additional prior sleep led to an improvement in subjective fatigue of 0.4 ± 0.1 units (mean ± standard error). Controlling for prior sleep duration did not affect the condition and test bout main effects and interaction. SP subjective fatigue scores were not significantly correlated with PVT lapses ($F_{1,436}$=3.2, P=0.073).

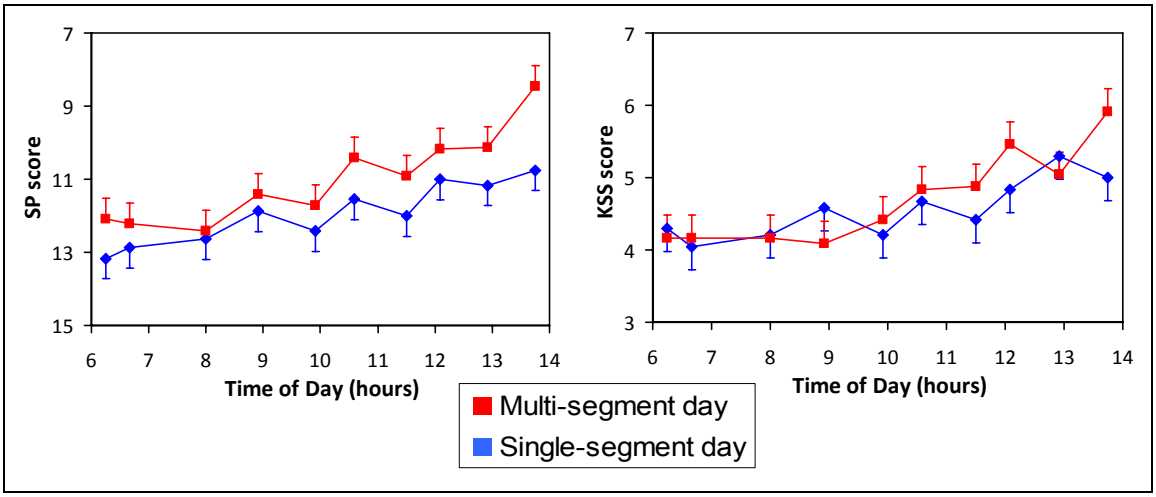


Figure 12: Subjective fatigue on the Samn-Perelli (SP) Subjective Fatigue Checkcard and subjective sleepiness on the Karolinska Sleepiness Scale (KSS), plotted over fatigue test bouts in the multi-segment and single-segment duty days. Note the reversal of the axis for SP score; upwards corresponds to greater fatigue/sleepiness in both panels. Whiskers: standard errors from mixed-effects ANOVA.

For KSS subjective sleepiness, there was a main effect of condition ($F_{1,437}$=17.0, P<0.001) and a main effect of test bout ($F_{9,437}$=7.9, P<0.001), but no significant interaction of condition by test bout ($F_{9,437}$=1.3, P=0.21). These findings indicate greater subjective sleepiness in the multi-segment day compared to the single-segment day, and a build-up of subjective sleepiness across the test bouts regardless of condition (see Figure 12, right). Secondary analyses revealed an effect of simulator session ($F_{1,436}$=10.6, P<0.001) and a significant simulator session by condition interaction ($F_{2,436}$=6.4, P=0.002). Due to the randomization of the multi-segment and single-segment flight schedules over the two simulator sessions, when the effects of simulator session and simulator session by condition interaction were controlled for, the main effect of condition remained significant ($F_{1,436}$=17.3, P<0.001) as did the main effect of test bout ($F_{9,436}$=8.1, P<0.001). There was a significant effect of prior sleep duration as measured by wrist activity monitor ($F_{1,436}$=36.7, P<0.001), where an hour of additional prior sleep led to a reduction in subjective sleepiness of 0.3 ± 0.1 units (mean ± standard error). Controlling for prior sleep duration did not affect the condition and test bout main effects and interaction. KSS subjective sleepiness scores were correlated with SP subjective fatigue scores (r=0.62, $F_{1,436}$=316.4, P<0.001). KSS subjective sleepiness scores were somewhat correlated with PVT lapses (r=0.24, $F_{1,436}$=16.3, P<0.001).

Statistical analyses of the data from the pilots' flight performance evaluations by means of a standard flight evaluation form (see Appendix 1) involved mixed-effects ANOVA over flight segments, with a planned contrast (*t* test) comparing the five segments in the multi-segment day to the one segment in the single-segment day. Secondary analyses included covariates for pilot age, total flight experience, CRJ-200 flight experience, prior sleep duration, seat (CA or FO), pilot flying or monitoring, order of conditions, simulator session (first or second), and simulator session by condition interaction. Results from secondary analyses are only reported here if the covariate was statistically significant, with the exception of pilot flying versus pilot monitoring, which is of particular interest in the context of flight performance. Recall that for the multi-segment day, the CA flew segments 1, 3 and 5, whereas the FO flew segments 2 and 4; for the single-segment day, either the CA or the FO flew the entire segment while the other was monitoring as determined by randomization.

Figure 13 shows the results of the evaluations for overall flight performance. There was a significant effect of condition (t_{115}=2.6, P=0.010), with generally better pilot flight performance in the multi-segment day than in the single-segment day. The difference was 0.15 ± 0.6 units (mean ± standard error). Secondary analyses revealed a trend for generally better ratings of flight performance when pilots were monitoring compared to when they were flying ($F_{1,114}$=3.5, P=0.064), the difference being only 0.08 ± 0.04 units (mean ± standard error) and not affecting the effect of condition.

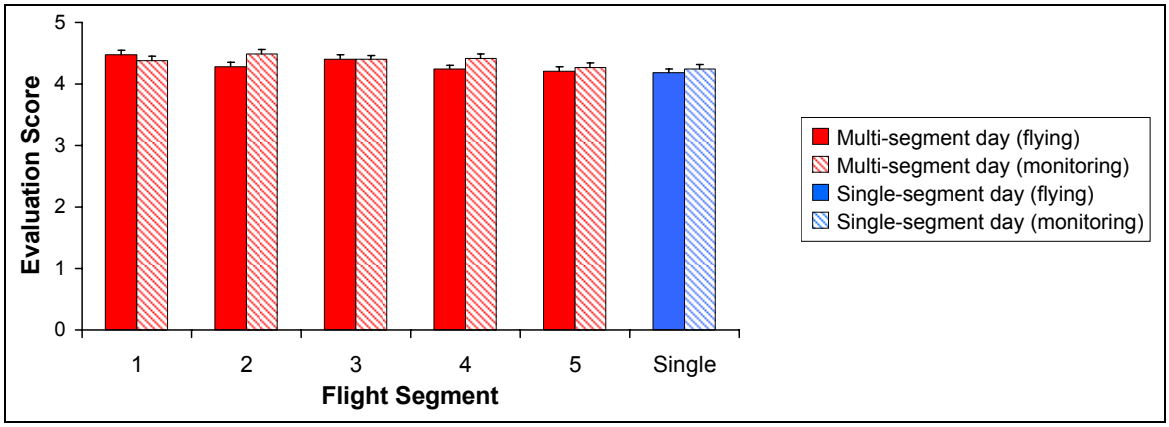


Figure 13: Overall flight performance evaluation scores in each of the segments of the multi-segment and single-segment duty days, differentiating when pilots were flying from when they were monitoring. Means (bars) and standard errors (whiskers) were derived from full-factorial mixed-effects ANOVA of segment by pilot flying/monitoring.

For the distinct phases of each segment (pre-flight, gate departure, take-off, climb, cruise, descent, approach, landing, and gate arrival), there were significant effects of condition for gate departure (t_{115}=2.5, P=0.016), climb (t_{112}=2.9, P=0.004) and cruise (t_{112}=4.3, P<0.001), with better flight performance in the multi-segment day than in the single-segment day in each case. Figure 14 shows an omnibus analysis of flight performance in the distinct phases of each of the different flight segments. Contrasts yielded significant differences between the five segments of the multi-segment day and the one segment of the single-segment day only for the climb (t_{1197}=2.7, P=0.008) and for the cruise phase (t_{1197}=3.4, P<0.001).

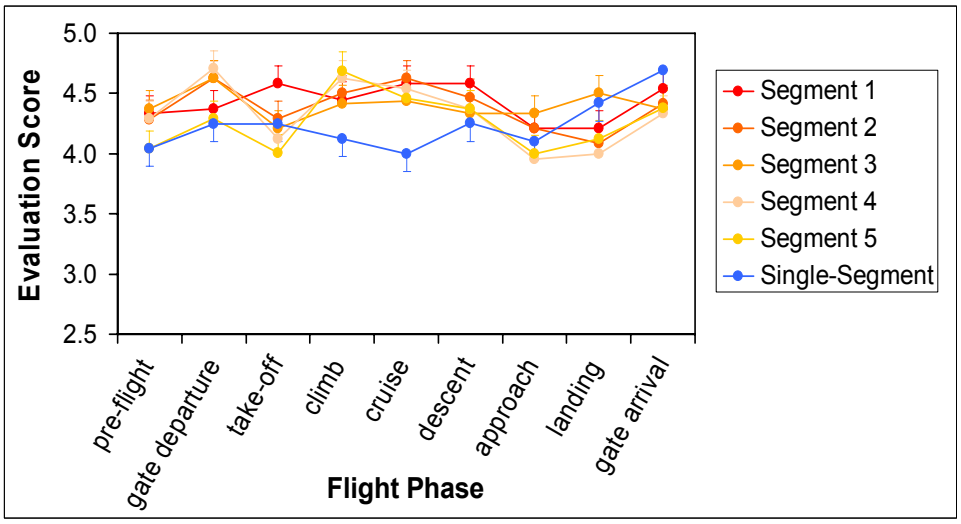


Figure 14: Flight performance evaluation scores across the distinct phases of flight in each of the segments of the multi-segment and single-segment duty days. Note that the axis for evaluation score is truncated at 2.5 to better show the small differences between the curves. Means (circles) and standard errors (whiskers) were derived from full-factorial mixed-effects ANOVA of segment by phase of flight.

DISCUSSION

The flight simulator experiment of the WSU-RAA Fatigue Study involved 24 active-duty pilots each flying two 9-hour duty days in the full-flight simulator with 10 fatigue test bouts per duty day. The data set comprised a total of 432 duty hours, 144 take-offs and landings, and 480 fatigue test bouts (including 4,800 minutes of PVT performance). In this regard, and in the degree of realism and operational detail, the experiment stands out among aviation fatigue studies in simulated operational settings (cf. Caldwell et al., 2000, 2004; Thomas et al., 2006; Elmenhorst et al., 2009). Scientists, airline management, labor representation, trade organization, and pilots worked together to begin to fill a gap in fatigue science with regard to the fatiguing effect of workload, in particular workload associated with take-offs and landings in multi-segment operations.

Study Findings

The flight simulator experiment yielded consistent evidence of a greater build-up of fatigue, both objectively measured (see Figure 11) and subjectively reported (see Figure 12), in a five-segment duty day as compared to a single-segment duty day of the same duration and start time. The experiment was designed such that the additional fatigue in the multi-segment day could be attributed to the additional take-offs and landings associated with the multiple segments, as other factors that could systematically affect fatigue were purposely held constant or experimentally and statistically controlled for. Thus, the study findings indicated that the workload associated with multiple take-offs and landings resulted in increased fatigue over the duty day. There was no evidence that the increased fatigue in the multi-segment schedule also had adverse effects on flight performance measures selected in this study (see Figures 13 and 14), although whether this

was due to a true absence of adverse effects on flight performance or insufficient sensitivity of the flight performance measures used remains to be determined.

Figure 15 shows that based on the objective measure of fatigue in the study, PVT lapses, the magnitude of the fatiguing effect of multiple take-offs and landings was modest relative to that seen in laboratory studies of simulated night shifts and nighttime sleep deprivation with similar fatigue testing regimes (Van Dongen et al., 2011b, 2013). As expected, the number of PVT lapses observed in the single-segment day was similar to that seen in a laboratory study of simulated day shifts. In comparison, the number of PVT lapses in the multi-segment day was similar to that seen in the early night hours of simulated night shifts, but stayed below the number of PVT lapses in the late night / early morning hours of simulated night shifts.[‡] Further, the number of PVT lapses in the multi-segment day was considerably smaller than that seen during a night of sleep deprivation.

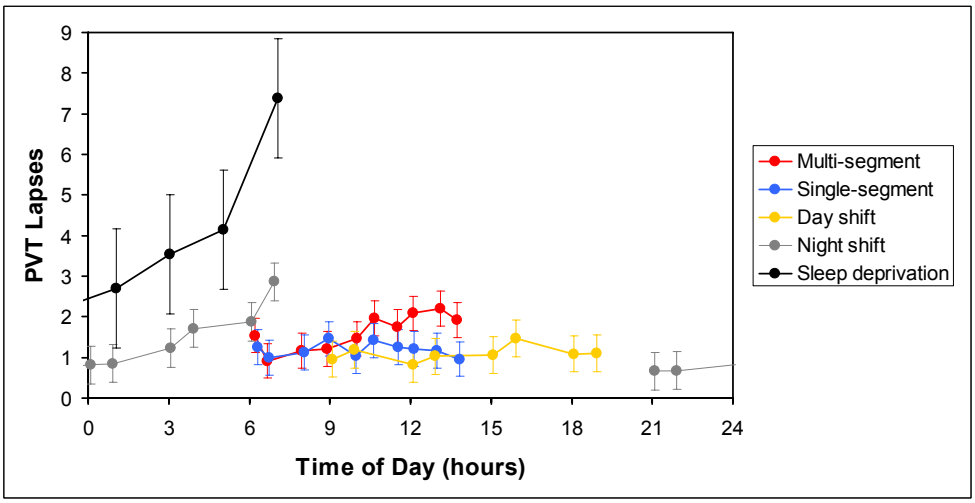


Figure 15: Comparison of fatigue in the multi-segment and single-segment days of the present study with fatigue in a study of simulated day shifts and night shifts and a study of sleep deprivation. Means (circles) and standard errors (whiskers) are plotted against time of day. Note that when the curve for night shifts reaches midnight (24:00) on the right, it continues at midnight (00:00) on the left. The multi-segment and single-segment day curves are the same as shown in Figure 11 (top left). The shift schedule curves are from 14 healthy subjects in the day shift condition and 13 healthy subjects in the night shift condition of a laboratory study as shown in Van Dongen et al. (2011b; Figure 3, upper left panel, first duty cycle). The sleep deprivation curve is from 13 healthy individuals during the first 24 hours of a 62-hour total sleep deprivation period in a laboratory study (Van Dongen et al., 2013).

The temporal profiles of SP subjective fatigue and KSS subjective sleepiness resembled those of the PVT outcome variables (cf. Figures 11 and 12). Nonetheless, the observed correlations between the subjective measures and PVT lapses were low to negligible, which warrants further discussion. The source of the apparent discrepancy is inter-individual variability in vulnerability to fatigue, which has repeatedly been found to be incongruent between objective and subjective measures in other studies (Leproult et al., 2003; Van Dongen et al., 2004a; Van Dongen et al., 2011c). Since inter-individual variability was accounted for in the statistical analyses of the study

[‡] The simulated night shift experiment allowed for 10 hours of sleep opportunity per day (Van Dongen et al., 2011b), which is more than what is typically encountered in night shift operations and as such tempered the level of fatigue observed in that study.

data, the incongruence of inter-individual variability between objective and subjective measures tempered the correlations between them. Additionally, the pilots reported greater subjective fatigue as measured on the SP when they were flying than when they were monitoring, but this effect was not found in the objective fatigue measurements (nor on the KSS). This phenomenon further reduced the correlation between SP subjective fatigue and PVT lapses (to the point of non-significance). As such, the low correlations between the subjective measures of fatigue and sleepiness on the one hand and PVT lapses on the other hand do not negate the high correspondence in the overall temporal profiles of the outcome variables of the study, as seen in Figures 11 and 12.

Experimental Control

The design and execution of the study and the analysis and interpretation of the data were discussed extensively, and decided upon by consensus, in two face-to-face meetings of the SSC. Also, before the study was begun, the research team conducted two dry runs at the USATC to verify and fine-tune the experimental procedures.

The pilots were between 24 and 49 years old, and 8% were women. They were certified to fly the CRJ-200 and covered a range of CRJ-200 and total flight experience (see Figure 7). They were randomly selected for recruitment into the study. All were domiciled on the East Coast of the U.S., which is also where the experiment was conducted.

Important study design features and experimental controls provided a solid foundation for head-to-head comparison of fatigue from multiple take-offs and landings versus only one take-off and landing. The pilots were randomly assigned to fly the multi-segment schedule on the first flight simulator day and the single-segment schedule on the second flight simulator day or the other way around. They were also randomly assigned to be either flying or monitoring during the single-segment day. The multi-segment duty day and the single-segment duty day had the same duty duration (9 hours) and the same early morning start time (05:15).

The pilots flew the multi-segment and single-segment schedules in the CRJ-200 flight simulator used by their airline (see Figure 2). They flew to airports that were a priori deemed unlikely to be familiar to them, so as to avoid study confounds from differences in how the airports were implemented in the simulator versus their real-world counterparts as well as any differences among pilots in airport familiarity. (One pilot reported to have prior, non-recent familiarity with some of the airports as he had lived and flown in Texas.) The airports selected for the study represented class B airspace, and all take-offs and landings took place in IFR meteorological conditions so as to help bring out the workload associated with these critical phases of flight (after which fatigue testing took place immediately).

The experimental procedures were detailed in standard operating procedures and minute-by-minute study protocols in order to provide consistency across study runs in the pilots' experiences and standardization of circumstances during fatigue testing. All pilots were briefed at the same time (17:00) on the day before the first simulator flight day, and at the start of their duty period (05:15) on both simulator flight days. The pilots had ample opportunity for sleep before each of the two simulator flight days, and the commute between the crew hotel and the

simulator site was short (less than 15 minutes). Sleep was measured with wrist activity monitors (see Figures 4 and 10) and sleep/wake/duty logs. Prior sleep was accounted for in secondary statistical analyses of the fatigue measurements, and the results of these analyses invariably corroborated the primary findings.

Fatigue testing occurred at the same times of day during the single-segment day as during the multi-segment day (see Figures 8 and 9). Fatigue was measured while the simulator was on autopilot (after take-off) or parked on a taxiway (after landing), under controlled circumstances with distractions removed and supervised by a WSU research assistant (see Figure 5). Pilots practiced the PVT, as well as the SP subjective fatigue and KSS subjective sleepiness measures (see Figure 6), during the briefing on the day before the simulator flight days.

The number of lapses of attention on the PVT was selected a priori as the primary objective measure of fatigue. The PVT is considered a gold-standard task for measuring fatigue (Lim & Dinges, 2008). It exhibits no practice effect across repeated test bouts (Van Dongen et al., 2003), and displays minimal aptitude differences among healthy individuals (Dorrian et al., 2005). The PVT is highly sensitive to fatigue because it captures the primary effect of fatigue on brain functioning, that is, wake state instability (Doran et al., 2001; Jackson & Van Dongen, 2011). Wake state instability is believed to be the main reason for fatigue's contribution to risk of errors, incidents and accidents (Van Dongen & Hursh, 2010; Satterfield & Van Dongen, 2013).

Significant steps were taken to maintain a high level of operational realism in the study. The flight schedules were executed like real passenger-carrying flights. The pilots were to report fit for duty as per normal duty requirements, and they came to the simulator dressed in uniform and carrying their flight bags. They prepared the simulator like a real airplane, and conducted walk-arounds of the simulator as they would for an actual airplane. A MEL-associated airplane swap occurred after the second segment in the multi-segment duty day. Although the flight scenarios contained no unexpected procedures or emergencies, the pilots were instructed to be ready to respond to any emergencies that might occur, just like in the real world. The pilots were given realistic dispatch releases and Jeppesen charts prepared specifically for the study (see Figure 3). There was airspace-specific ATC background chatter during flights, and there were realistic ATC interactions with the simulator operator serving as the air traffic controller.

While on the ground between flights, the pilots had to use the available time to obtain the next dispatch release, eat and drink, use the restroom, and prepare for the next flight, just like they would do at real airports. During the single-segment day, pilots were restricted to brief simulator off-motion opportunities to go "to the back of the plane" to eat and drink and use the lavatory. The off-motion opportunities in the single-segment day occurred at the same times as the turns in the multi-segment day.

As is bound to happen in full-flight simulator studies, a few minor glitches and malfunctions occurred during the execution of the study. These events were reviewed in detail by the SSC, and were not deemed to have impacted on fatigue testing or to have caused any outliers in the data.

Limitations

The study was simulator-based in order to achieve a high degree of consistent operational integrity during multi-segment and single-segment flight schedules, standardization of experiences among the different flight crews, and safe opportunities for uninterrupted fatigue testing while in simulated flight. This would not have been possible in a field study. The main strength of the study as designed and conducted was that the additional fatigue in the multi-segment schedule could be attributed specifically to the workload from multiple take-offs and landings. This was also its chief limitation, as it required the exclusion of other factors that may or may not be involved in differentiating fatigue between multi-segment and single-segment flight operations. The study did not examine the interaction of fatigue from multiple take-offs and landings with earlier or later duty start times (circadian effects), longer duty days, or more than two consecutive days on duty (cumulative fatigue effects). Furthermore, no adverse weather events, no high density air traffic, no delays, and no other complicating factors were included in the flight scenarios. The operational applicability and generalizability of the study findings must be considered in the context of these limitations.

To quantify absolute levels of fatigue in multi-segment operations or to make direct comparisons with different commercial aviation operations (e.g., long-haul flights), other operational differences potentially affecting fatigue besides multiple take-offs and landings need to be considered (see Figure 1). In addition to the number and duration of segments flown, regional airline operations generally differ from other types of commercial aviation in terms of aircraft models, and sometimes differ in terms of airports and routes. There may also be differences in exposure to weather and temperature, airplane swaps, and delays. The fatiguing effects of these factors, and their possible interactions, have not been systematically studied scientifically in this or other published experiments.

The relevance of such factors and conditions notwithstanding, it would have been difficult to interpret the study findings and translate them to the airline operational environment if the study had been designed to attempt to measure the fatigue associated with these variables (if any) rather than the specific effect of multiple take-offs and landings. It would have required the disentanglement and quantification of too many variables for which fatigue science is limited or non-existent. It is for this reason that we limited this study to examining the effect of flying multiple segments as an initial start into this scientific area.

The increased fatigue levels observed during the multi-segment day did not appear to have adverse consequences for the set of pilot flight performance indices rated during the study by a simulator instructor in this study. Flight performance during the climb and cruise phases of flight was actually rated less favorably for the single-segment duty day than for the multi-segment duty day, although the reason for this difference is not clear. Both the climb and cruise phases were longer in the single-segment day than in any of the segments of the multi-segment day, thus providing greater opportunity for any flight performance changes to be exposed, but whether this was a contributing factor remains unknown.

Although the pilots participating in the study were instructed to be ready to respond to any emergencies that might occur, no particularly abnormal circumstances or emergencies were

included in the flight scenarios. In the FAA's duty regulations, such unknowns are accounted for by including conservative margins in the duty and flight time restrictions (and by requiring augmentation on longer flights).

A practical limitation of the present study was that the duty days were limited in duration to 9 hours. This duty duration was selected because it allowed an 8-hour flight in the single-segment day, which is currently the maximum allowable flight time for unaugmented crews. This in turn restricted the number of segments in the multi-segment day to five, as otherwise the duration of the turns would have become too short. Five-segment duty days are common in regional airline operations. Nonetheless, this practical limitation prohibited exploring the fatiguing effect of multiple take-offs and landings across duty durations closer to the limits given by the FAA's new final rule (Table B in Federal Aviation Administration, 2011).

Another issue related to the duration of the duty day was that the 8-hour flight it allowed in the single-segment day exceeds the CRJ-200 airplane's fuel capacity. The SSC debated this issue and decided to allow for fuel to be added artificially by the simulator operator during the cruise phase of the single-segment flight schedule, seeing that aircraft range was not a factor of interest in the study and had no bearing on fatigue levels per se. Taking aircraft range into account nonetheless would have required shortening of the duty day in the single-segment flight schedule and, to preserve equivalence for fatigue testing, of the multi-segment flight schedule as well (and thus also reducing the number of segments). An alternative solution would have been to simulate a different airplane for the single-segment flight schedule, but that would have introduced a number of other confounds in head-to-head comparisons of fatigue between the multi-segment and single-segment duty days.

Finally, at the end of the second simulator flight day (the third day in the pairing), the pilots filled out a feedback form. Their comments provided further insight into the study and its strengths and limitations, and are summarized in Table 3.

Table 3: Summary of pilots' comments on the feedback form they filled out at the completion of their participation.

"The study would be more realistic with… ...irregular operations (weather, delays, maintenance, etc.)." ...different personalities in gate and ramp agents (more real-world interactions)." ...hot airplanes." ...weather during the walk-around." ...longer or varied sits during turns (sometimes several hours)." ...varied start of duty." ...longer duty days (10−16 hours)." ...more aircraft swaps." ...longer pairing (4+ days of flying)." ...flying in the real world."
"We are regional pilots and are accustomed to short legs and aren't conditioned for long flights, so that was hard in itself because it was a shift from the norm."
"I feel more tired towards the end of cruise before the descent begins than right after take-off or right after landing, since those are usually busy and give my mind some stimulation."
"As close to a line operation as possible." "Good detailed orientation." "Professional and organized study."

Conclusion

The flight simulator experiment of the WSU-RAA Fatigue Study involved 24 active-duty pilots each flying two 9-hour duty days in the full-flight simulator with 10 fatigue test bouts per duty day. The data set comprised a total of 432 duty hours, 144 take-offs and landings, and 480 fatigue test bouts (including 4,800 minutes of PVT performance). In this regard, and in the degree of realism and operational detail, the experiment stands out among aviation fatigue studies in simulated operational settings (cf. Caldwell et al., 2000, 2004; Thomas et al., 2006; Elmenhorst et al., 2009). Scientists, airline management, labor representation, trade organization, and pilots worked together to begin to fill a gap in fatigue science with regard to the fatiguing effect of workload, in particular as associated with take-offs and landings in multi-segment operations.

With the new US flight and duty time regulations of CFR Part 117 taking effect on January 4, 2014, duty duration in passenger-carrying domestic flights in commercial aviation will be limited as a function of duty start time and as a function of the number of flight segments in the duty period. The carefully controlled simulator study on fatigue associated with flying multiple segments as described in this report is the first to provide systematic evidence vis-à-vis number of flight segments as a factor relevant to duty duration. Specifically, the study showed greater build-up of fatigue in a five-segment duty day than in a single-segment duty day of the same duration and start time. While study limitations must be consider with regard to the applicability of the study results to real-world operations, an important strength of the study design is that the increased fatigue in the multi-segment day could be attributed specifically to the additional take-offs and landings associated with the multiple segments. The study findings therefore represent an important first step into the scientific area of fatigue in multi-segment operations.

ACKNOWLEDGMENTS

We gratefully acknowledge the many individuals whose concerted efforts made it possible to conduct this study. They were too numerous to list them all by name, but several deserve special mention. The individuals identified on the title page of this report constituted the Scientific Steering Committee. Twenty-four dedicated Air Wisconsin pilots served as volunteer research participants. Ryan Gibson, Matthew Hintze, Paul Preidecker and Sandy Pierson were instrumentally involved as part of a study execution committee. Jason Weirick and Judd Brinkman served as the simulator operators. Kimberly Honn and Brieann Satterfield were the WSU research assistants charged with running the study.

Air Wisconsin Airlines Corporation made substantial resources and personnel available for conducting the study. CAE provided access to and technical support for the Bombardier CRJ-200 flight simulator at the US Airways Training Center in Charlotte, NC. The Air Line Pilots Association and its Air Wisconsin-based Master Executive Council expressed their essential support for the study. The Regional Airline Association provided logistical support and research funding for Washington State University.

REFERENCES

Åkerstedt T (2003). Shift work and disturbed sleep/wakefulness. Occupational Medicine 53: 89-94.

Åkerstedt T, Gillberg M (1990). Subjective and objective sleepiness in the active individual. International Journal of Neuroscience 52: 29-37.

Caldwell JA (2005). Fatigue in aviation. Travel Medicine and Infectious Disease 3: 85-96.

Caldwell JA, Caldwell JL, Smith JK, Brown DL (2004). Modafinil's effects on simulator performance and mood in pilots during 37 h without sleep. Aviation, Space, and Environmental Medicine 75: 777-784.

Caldwell JA, Smythe NK, Leduc PA, Caldwell JL (2000). Efficacy of Dexedrine for maintaining aviator performance during 64 hours of sustained wakefulness: a simulator study. Aviation, Space, and Environmental Medicine 71: 7-18.

Daan S, Beersma DGM, Borbély AA (1984). Timing of human sleep: recovery process gated by a circadian pacemaker. American Journal of Physiology 246: R161-R178.

Dijk DJ, Czeisler CA (1994). Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. Neuroscience Letters 166: 63-68.

Doran SM, Van Dongen HPA, Dinges DF (2001). Sustained attention performance during sleep deprivation: evidence of state instability. Archives of Italian Biology 139: 253-267.

Dorrian J, Rogers NL, Dinges DF (2005). Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: Kushida CA, ed. Sleep deprivation: clinical issues, pharmacology, and sleep loss effects. Marcel Dekker, New York, pp. 39-70.

Elmenhorst EM, Vejvoda M, Maass H, Wenzel J, Plath G, Schubert E, Basner M (2009). Pilot workload during approaches: comparison of simulated standard and noise-abatement profiles. Aviation, Space, and Environmental Medicine 80: 364-370.

Federal Aviation Administration (2011). Final rule: flightcrew member duty and rest requirements. 14 CFR, Parts 117, 119, and 121.

Jackson ML, Van Dongen HPA (2011). Cognitive effects of sleepiness. In: Thorpy MJ, Billiard M, eds. Sleepiness: causes, consequences and treatment. Cambridge University Press, Cambridge, pp. 72-81.

Leproult R, Colecchia EF, Berardi AM, Stickgold R, Kosslyn SM, Van Cauter E (2003). Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. American Journal of Physiology – Regulatory, Integrative and Comparative Physiology 284: R280-R290.

Lim J, Dinges DF (2008). Sleep deprivation and vigilant attention. Annals of the New York Academy of Sciences 1129: 305-322.

Samn WS, Perelli LP (1982). Estimating aircrew fatigue: a technique with application to airlift operations. Brooks Air Force Base Report SAM-TR-82-21.

Satterfield BC, Van Dongen HPA (2013). Occupational fatigue, underlying sleep and circadian mechanisms, and approaches to fatigue risk management. Fatigue: Biomedicine, Health and Behavior 1: 118-136.

Thomas MJW, Petrilli RM, Lamond N, Dawson D, Roach GD (2006). Australian long haul fatigue study. In: Proceedings of the 59th Annual International Air Safety Seminar. Flight Safety Foundation.

Van Dongen HPA, Baynard MD, Maislin G, Dinges DF (2004a). Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep 27: 423-433.

Van Dongen HPA, Belenky G, Krueger JM (2011a). Investigating the temporal dynamics and underlying mechanisms of cognitive fatigue. In: Ackerman PL, ed. Cognitive fatigue: multidisciplinary perspectives on current research and future applications. American Psychological Association, Washington, DC, pp. 127-147.

Van Dongen HPA, Belenky G, Vila BJ (2011b). The efficacy of a restart break for recycling with optimal performance depends critically on circadian timing. Sleep 34: 917-929.

Van Dongen HPA, Caldwell JA, Caldwell JL (2011c). Individual differences in cognitive vulnerability to fatigue in the laboratory and in the workplace. Progress in Brain Research 190: 145-153.

Van Dongen HPA, Dinges DF (2000). Circadian rhythms in fatigue, alertness, and performance. In: Kryger MH, Roth T, Dement WC, eds. Principles and practice of sleep medicine, 3rd ed. WB Saunders, Philadelphia, pp. 391-399.

Van Dongen HPA, Dinges DF (2007). Individual differences in response to sleep deprivation. Final Technical Report, NIH R01 HL70154.

Van Dongen HPA, Hursh SR (2010). Fatigue, performance, errors, and accidents. In: Kryger MH, Roth T, Dement WC, eds. Principles and practice of sleep medicine, 5th ed. Elsevier Saunders, St. Louis, pp. 753-759.

Van Dongen HPA, Maislin G, Dinges DF (2004b). Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: importance and techniques. Aviation, Space, and Environmental Medicine 75: A147-A154.

Van Dongen HPA, Maislin G, Mullington JM, Dinges DF (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 26: 117-126.

Van Dongen HPA, Whitney P, Hinson JM, Jackson ML (2013). Feedback blunting: a novel mechanism underlying deficits in real-world decision making due to sleep loss. Sleep Science 6: S42-S43.

APPENDIX 1: FLIGHT EVALUATION FORM

Each pilot's flight performance was evaluated by the simulator operator (a flight instructor) after every flight segment, using a standard AWAC flight evaluation form shown below (Figure A1).

AIR WISCONSIN AIRLINES CORPORATION — **PILOT TRAINING FORMS**

PREFLIGHT *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Crew reviews all necessary flight information including the appropriate weather, flight release, flight plan, AeroData, NOTAMs, W&B, and MEL
- ☐ Review A/C MX Logbook
- ☐ All MX irregularities have been resolved in accordance with Air Wisconsin's approved MX procedures and complies with MEL procedures
- ☐ Crew performs appropriate exterior/interior preflight duties
- ☐ Crew performs a detailed review of the ATC clearance, filed route, FMS programming and review, and departure briefing
- ☐ Complies with Safety and Acceptance Checklists, if required
- ☐ Performs required system tests
- ☐ Reviews/discusses performance calculations from ACARS and/or AeroData
- ☐ Complies with Before Start, Start, and After Start Checklists
- ☐ Monitors engine parameters during engine start procedures
- ☐ Establishes communication with ground/ramp using accepted procedures
- ☐ Other _____

GATE DEPARTURE *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Complies with taxi clearances
- ☐ Controls taxi speed consistent with safe operations
- ☐ Complies with hold lines and other surface markings
- ☐ Complies with Before Takeoff Checklist
- ☐ Complies with ramp/ATC communications
- ☐ Other _____

TAKEOFF *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Complies with Takeoff Checklist
- ☐ Maintains runway centerline alignment
- ☐ Uses appropriate crosswind control techniques
- ☐ Sets thrust for conditions/performance requirements
- ☐ Sets thrust to calculated flex thrust or full thrust settings
- ☐ Uses radar, if applicable
- ☐ Uses approved call-outs
- ☐ Adheres to appropriate V-speeds during takeoff
- ☐ Complies with recommended rate and degree of initial rotation
- ☐ Complies with initial climb speed profile
- ☐ Uses FD and AP in accordance with approved procedures
- ☐ Performs gear and flap retraction schedules per profiles/procedures
- ☐ Complies with After Takeoff Checklist
- ☐ Complies with ATC clearances/instructions
- ☐ Other _____

CLIMB *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Complies with climb profile/area departure
- ☐ Maintains airspeed control
- ☐ Maintains navigational tracking/heading control
- ☐ Uses proper thrust management techniques
- ☐ Uses radar, if applicable
- ☐ Manages automation in accordance with approved procedures
- ☐ Maintains awareness of climb performance limitations
- ☐ Complies with ATC clearances/instructions
- ☐ Other _____

CRUISE *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Maintains airspeed control
- ☐ Maintains navigational tracking/heading control
- ☐ Uses radar, if applicable
- ☐ Complies with turbulence procedures, if applicable
- ☐ Monitors fuel used compared to fuel planning
- ☐ Maintains awareness of maximum performance ceilings
- ☐ Maintains coordination with cabin crew
- ☐ Complies with ATC clearances/instructions
- ☐ Other _____

DESCENT *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Maintains awareness of descent planning requirements
- ☐ Complies with crossing restriction requirements
- ☐ Maintains navigational tracking/heading control
- ☐ Uses radar, if applicable
- ☐ Maintains awareness of V_{MO}/M_{MO} speeds and other speed restrictions
- ☐ Manages automation in accordance with approved procedures
- ☐ Sets and cross-checks altimeters using appropriate references
- ☐ Complies with In Range procedures and checklist
- ☐ Complies with ATC clearance and instructions
- ☐ Other _____

APPROACH *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Approach briefings (chart and operational)
- ☐ Reviews/discusses destination weather
- ☐ Reviews/discusses performance from ACARS and/or AeroData
- ☐ Complies with Approach Checklist
- ☐ Maintains navigational tracking/heading control
- ☐ Uses radar, if applicable
- ☐ Sets V-speeds according to landing weight
- ☐ Selects flap and gear in accordance with approved profiles
- ☐ Manages automation in accordance with approved procedures
- ☐ Complies with approach procedures
- ☐ Maintains stabilized approach criteria in the landing configuration
- ☐ Uses approved call-outs and maintains crew coordination
- ☐ Transitions to visual segment, if applicable
- ☐ Complies with go-around/missed approach procedures, if applicable
- ☐ Complies with ATC clearances/instructions
- ☐ Other _____

LANDING *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Complies with Landing Checklist
- ☐ Maintains runway centerline alignment
- ☐ Uses appropriate crosswind control techniques
- ☐ Maintains stabilized descent rates to touchdown
- ☐ Performs touchdown in the TDZ
- ☐ Uses thrust reverse and braking consistent with conditions
- ☐ Coordinates transfer of control, if applicable
- ☐ Complies with ATC clearances/instructions
- ☐ Other _____

AFTER LANDING / ARRIVAL *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Complies with After Landing Checklist
- ☐ Complies with ATC/ramp communications
- ☐ Complies with post-flight inspections
- ☐ Completes appropriate logbook entries, including discrepancies and complies with MEL procedures
- ☐ Other _____

ABNORMALS *(circle one)* 5 4 3 2 1

- ☐ Adverse WX ☐ Airline Operational ☐ Ground / Ramp
- ☐ ATC ☐ Cabin ☐ Dispatch / Paperwork
- ☐ Airport Conditions ☐ Aircraft Malfunctions ☐ Manuals / Charts
- ☐ Other Environmental ☐ Ground MX
- ☐ Complies with Emergency/Abnormal Checklist and procedures
- ☐ Maintains crew coordination
- ☐ Communicates with ATC, Dispatch, and Maintenance as required
- ☐ Other _____

GRADING SCALE

- 5 .... Excellent – No Errors
- 4 .... Standard – Trapped Errors
- 3 .... Standard with Debrief – Un-trapped Errors, No Safety Issues
- 2 .... Marginal – Un-trapped Errors, Safety Issues
- 1 .... Unsatisfactory

CL-65 OBSERVATION JOB AID — Page 2 of 2 — REV DATE: 22 FEB 12

22 FEB 12

Figure A1: Second page of the standard flight evaluation form filled out by the simulator operator for each pilot after every flight segment. (The first page of the form, containing only demographics and identifiers, was not used.)

APPENDIX 2: STATISTICAL POWER CALCULATION

To determine the number of pilots to be studied, a statistical power calculation was performed in advance. We made use of the only data set available at the time regarding the effects of workload on performance in the context of fatigue, from an NIH-funded laboratory study described in Van Dongen & Dinges (2007). In that study, 21 healthy subjects were exposed to sleep deprivation in a moderate workload condition (30 minutes of intensive testing every 2 hours) and again in a high workload condition (60 minutes of intensive testing every 2 hours) in randomized, counterbalanced order. The study revealed a modest but significant increase in the number of lapses on the PVT in the high workload condition (first 10 minutes of a 20-minute PVT) relative to the moderate workload condition (10-minute PVT). This effect did not interact significantly with prior time awake nor with time of day.

By design, the methodological structure of the present study was similar to the earlier workload study. In the present study, we set out to compare – within subjects – average lapses in repeated administrations of the 10-minute PVT in a high workload condition (five take-offs and landings) versus a moderate workload condition (one take-off and landing with the same total duty duration) to statistically test the effect of workload on fatigue in a simulated aviation environment. For the power calculation, we postulated that in order for the five-fold workload increase in the high workload condition to be operationally relevant, it should have at least the same observed magnitude as the two-fold increased workload effect in the earlier study.

The earlier study observations were 10.04 ± 1.96 lapses (mean ± standard error) in the moderate workload condition and 15.17 ± 1.99 lapses (mean ± standard error) in the high workload condition. From this we derived a workload difference of 5.12 ± 9.05 (mean ± conservative estimate of pooled SD), which implied an effect size of at least 0.57. It follows that in order to have at least 80% statistical power with a one-sided type I error threshold of 0.05 in the present within-subject study, we needed to have a sample of at least 21 pilots (nQuery Advisor 7.0). Given that pilots operate in pairs (Captains and First Officers), we thus had to study a total of 22 pilots (11 crew pairs). To allow for modest attrition, the target sample size for the study was set to 24 (12 crew pairs).

Regional Airline Association
2025 M Street, NW, Suite 800
Washington, DC 20036
Tel: (202) 367-1170 Fax: (202) 367-2170
raa@raa.org www.raa.org

Washington State University Sleep and Performance Research Center
P.O. Box 1495
Spokane, WA 99210-1495
Tel: (509) 358-7755 Fax: (509) 358-7810
hvd@wsu.edu www.wsu.edu/sprc