

Heart Rate Measures of Flight Test and Evaluation

Malcolm A. Bonner
AFOTEC Det 5/OL-KM
Keesler Air Force Base, MS

Glenn F. Wilson
Air Force Research Laboratory
Wright Patterson Air Force Base, OH

One of the goals of aircraft test and evaluation is to determine whether the crew can operate a new system safely and effectively. Because flying is a complex task, several measures are required to derive the best evaluation. This article describes the use of heart rate to augment the typical performance and subjective measures used in test and evaluation. Heart rate can be nonintrusively collected and provides additional information to the test team. Example data illustrate the nature of the results provided by heart rate during the test and evaluation of a transport aircraft. Comparison with subjective workload estimates shows discrepancies that provide valuable insights into the crews' responses to the demands of the test missions. Heart rate should be considered as an additional measure in the test and evaluation tool kit.

The test and evaluation of new and modified aircraft is required to determine whether crew members can operate these systems. This is especially true of modern aircraft and aircraft upgrades that may replace crew members with automation. The purpose of the testing is to determine whether the planned crew complement can fly the aircraft safely and effectively and accomplish their mission. The test and evaluation environment provides unique problems for the test team. Due to the nature of the testing environment, optimal experimental design is typically not possible. The number of available crew members is limited, and the test design usually does not permit complete coverage of all of the desired conditions.

Requests for reprints should be sent to Glenn F. Wilson, AFRL/HECP, 2255 H Street, Wright-Patterson Air Force Base, OH 45433-7022. E-mail: glenn.Wilson@wpafb.af.mil

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2001		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Heart Rate Measures of Flight Test and Evaluation				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Wright-Patterson AFB, OH 45433-7022				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 15	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Because the flight tests are expensive, the number of flights is constrained and must be shared with groups who are testing other aircraft systems and capabilities.

The most widely used human factors tool is the subjective report. These measures are relatively easy to collect, have face validity, and enjoy crew acceptance. However, subjective reports may not provide a complete picture of the cognitive demands placed on the operators (Hankins & Wilson, 1998). Subjective reports may be susceptible to memory lapses and bias (Eggemeier & Wilson, 1991). Performance data may also be recorded. The nature and number of performance data that can be collected depends on the system or aircraft being tested. Although performance data may be available from the aircraft bus, access may be limited. Psychophysiological measures have been used in flight test and evaluation. These measures are continuously available and relatively easy to implement. Due to the complex nature of flying, it is unreasonable to expect any one measure to provide a complete assessment of the crew members' functional state. This is why a battery of measures is recommended, with psychophysiological data included when possible.

In the context of flight test and evaluation, psychophysiological measures are typically used to monitor the mental workload of crew members during test missions (Wilson, 2002). The missions are designed to provide an environment that will permit the determination of whether the aircraft or new aircraft system meets the stated specifications. Cognitive workload is always a concern, and psychophysiological measures have been used to determine the level of workload crew members experience. As has been the case with flight research in general, heart rate (HR) has experienced the most widespread application in flight test and evaluation. Reviews can be found in Roscoe (1992), Wilson and Eggemeier (1991), and Wilson (2001). Using HR, Roscoe (1979) demonstrated that Harrier ski-jump takeoffs were no more difficult than conventional short takeoffs from a runway. Roscoe (1975) evaluated the effects on pilots of steep gradient approaches that were to be used for noise abatement. His results demonstrate that the steeper approaches do not involve higher pilot workload than the customary approaches. Rokicki (1987) used HR collected during test flights as a debriefing aid. The data were examined to locate periods of high HR. These periods were then used to identify segments of the test flights for more in-depth analysis.

The test and evaluation environment presents several unique problems for recording psychophysiological data. The missions may be long—from 1 to 12 hr—during which crew members may move about the aircraft. This requires ambulatory recording equipment and knowledge about when they are moving to separate those segments from periods when they are seated. Other types of movement are also required, such as looking for air traffic during landing. Although it may be possible to identify specific events such as switch closures, communication, and aircraft turns, the usual approach in analysis is to use broader categories of performance. For example, the events surrounding aircraft landing are included in one epoch. All of the specific actions are included in the landing seg-

ment. This is done for several reasons. The evaluation may be focusing on aircraft handling during landing and not on the specific actions that take place during the landing. Also, the total number of data and the short time permitted for analysis preclude extensive analysis.

This article describes the procedures used in flight test and evaluation and shows typical psychophysiological data from these tests. Electrocardiographic data were collected from pilots, copilots, and loadmasters during the test and evaluation of a new transport aircraft. The tests included aircraft handling, which was partially tested with a series of touch and go landings as well as other aspects of normal flight. Landing at short airfields was also tested because this was a required capability for the aircraft. The testing also included long-duration flights. Data from these tests provide examples of the sorts of findings that one can expect from transport aircraft test and evaluation.

METHODS

Del Mar 463 microcassette recorders (Del Mar Medical Systems, Irving, CA) gathered a minimally intrusive recording of aircrew electrocardiographic (ECG) data. The recorders provide 26 hr of time-indexed data on three recording channels. A Del Mar 563 Holter analysis system digitized the ECG data. A Workload Assessment Monitor detected R waves from the ECG and recorded the interbeat intervals. Missing and extra beats were detected and corrected. Three ECG channels were recorded using ConMed Ultratrace adult ECG electrodes. The electrodes were placed at the right manubrial border of the sternum and at the left anterior axillary line of the sixth rib. A ground electrode was placed on the lower right rib. All of the electrodes were placed directly over bone to reduce muscle artifacts. The three ECG channels provided redundancy in case of electrode failure or artifact problems. The electrodes were applied approximately 3 to 4 hr prior to takeoff.

A technician ensured that the ECG recorder was turned on, its clock synchronized, and the recorder was operating properly. The electrode wires were routed out the neck of the undershirt and around the flight suit collar, and the recorder was placed in the upper flight suit pocket. Loadmaster recorders were encased in a small aluminum protector to prevent damage during loading and offloading. Baseline recordings were gathered during mission brief and/or mission planning. Once digitized with the Del Mar 563 Holter analysis system, reduced data were electronically transmitted to Air Force Research Laboratory for analysis.

During the missions, human factors personnel recorded times at the beginning and end of specific mission segments. They noted periods when the pilots were out of their seats. They also noted significant events that might potentially affect HR or workload along with their times of occurrence.

There were two phases of the test and evaluation. The first was the basic airland portion, and the second was the tactical airland part. The basic airland missions involved transporting cargo from one airstrip to another with a high cruise altitude. Mission durations, show time to final landing, were typically 7 to 9 hr with 4 to 6 hr of actual flying time. The longest missions were 12 hr. The basic airland mission segments were mission planning, preflight, taxi, takeoff, climb, cruise, descent, approach, landing, taxi/shutdown, go-around, loading, offloading, and reconfiguration. A series of touch and goes also took place on several missions.

The tactical airland missions involved transporting cargo at low-level altitudes to an assault landing strip with simulated threats along the route. The mission durations were typically around 7.5 hr with 3.5 hr of actual flying time. The tactical airland mission segments were mission planning, preflight, taxi, takeoff, low-level segment, entry maneuvers to the assault runway, landing, takeoff from the assault runway, taxi, and shutdown.

The Modified Cooper-Harper Workload Rating Scale was used to record subjective workload ratings from each segment from each crew position (pilot, copilot, and loadmaster). It is a 10-point scale designed to rate perceived workload. The 10 points are grouped into four categories. The acceptable category includes ratings of 1 to 3, high mental workload is 4 to 6, major deficiencies includes ratings of 7 to 9, and mandatory system redesign is a rating of 10. The numbered rating scale was taped to the airplane console to be visible to both pilots, and all aircrew members were trained in the proper use of the scale. Immediately following each selected mission segment, the observer verbally prompted the aircrew for a response, and all the operators verbally gave their rating. In-flight workload ratings were obtained on a noninterference basis, and safety always took priority over data collection.

Because of the highly variable nature of the flight tests, inferential statistics were not used to analyze the data. Only five pilots, five copilots, and seven loadmasters flew the aircraft during both phases. Some pilots flew as copilot on at least one mission. The pairings of pilots and copilots were not systematically varied, and the number of missions each loadmaster flew was not consistent. Furthermore, the actual number of times that the various maneuvers or segments were executed varied from flight to flight. All of these factors contributed to the total variance in the data that made interpretation of inferential statistical results extremely difficult if not impossible.

RESULTS

In the basic airland phase, HRs and subjective workload estimates were recorded from five pilots, three copilots, and seven loadmasters. Data were recorded from

42 missions that lasted from 4 to 12 hr. Four- to 8-hr-long missions were typical. More than 50 takeoffs and landings were performed with more than 230 touch-and-go maneuvers executed. In all, in excess of 4.96 million heartbeats were recorded from these flights. Figure 1 shows an example of the HR data from one flight from takeoff to final landing. The mission lasted over 4 hr. Note the increases in HR during takeoff, touch and goes, and final landing for the pilot-in-command (PIC) of the aircraft during these maneuvers. The relatively large increases in HR during these events are noteworthy. The pilots were seated and controlling the aircraft, which did not require a great deal of physical effort, yet the increase in HR was in the range of 20 beats per minute (bpm). These large increases are characteristically found during flight and are much larger than the HR increases found in response to laboratory task performance. This is one of the hallmarks of HRs recorded during flight. Even though the peak HR for the different landings varied, there is no trend toward adaptation over the duration of the flight. Each landing is reflected in notable HR increases in the PIC. Although not shown in this figure, other landings were associated with smaller HR increases. Also of importance is the period when the pilots were out of their seats and walking about the aircraft. This activity brought about a large increase in HR and is marked with arrows in the figure. This underlines the necessity for recording mission events so that such episodes can be located and removed from consideration because the increased HR is almost totally due to the physical activity and not cognitive endeavors.

It is well known that the PIC exhibits a higher HR than the non-pilot-in-command (non-PIC). This is especially true during high workload situations such as landing (see Figure 1). Note that the pilot who is actually in command of the aircraft during the touch and go has a higher HR than the second pilot, who is not responsible for the landing. When a pilot is non-PIC during a touch and go, there are generally only small increases in HR. These small increases probably reflect the increased cognitive activity required of the non-PIC, who is engaged in systems work.

Figure 2 depicts group mean HRs for PIC and non-PIC over 2-min periods for several flight segments. Note that the PIC HRs are consistently higher and more volatile with a greater range than those of the non-PIC. This demonstrates the effects of cognitive activity associated with piloting. If the increased HRs seen during landing, for example, were due to fear or other emotions, then one would expect to see the same changes in the non-PIC, who is sitting next to the PIC and may suffer the same fate in case of mishap. It is possible that the increased HRs from the PICs were due to increased physical activity associated with flying. However, Wilson (this issue, "An Analysis of Mental Workload") recorded arm movements during flight and reported very low correlation with the pilots' HRs. Electromyographic (EMG) data were also collected and were not highly correlated with HR or electrodermal activity. Physical activity and EMG data were not

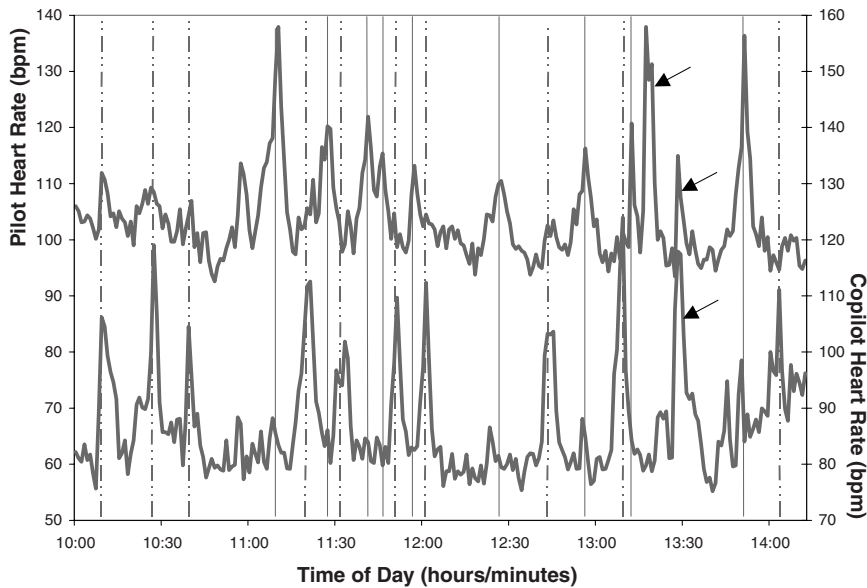


FIGURE 1 Pilot and copilot HRs for one mission. Takeoff, touch and goes, and final landing are marked with vertical lines. The solid lines are events when the pilot was PIC, and the dashed lines indicate when the copilot was PIC. The leftmost line corresponds to the initial takeoff and the rightmost line indicates final landing. The means for every 60 sec are displayed with the pilot's data at the top. The arrows indicate times when the pilot or copilot was out of their seat and moving about the aircraft. The values on the left y axis represent values for the pilot's data and those on the right are for the copilot's data. The time of day is indicated on the x axis.

recorded during the test and evaluation flights; however, the flight results of Wilson strongly suggest that physical activity did not cause the PIC HR increases reported here.

Figure 3 depicts the mean subjective estimates of mental workload. The subjective reports were restricted to a small portion of the overall range of 0 to 10. The mean subjective estimates ranged from about 2.75 to 4.0 on a 0-to-10-point scale. These ratings are confined to the lowest workload category, acceptable. The rating of 4 is the lowest rating of the next highest workload category, high workload. The mean HRs covered an approximately 11-bpm range. The mean subjective ratings for the simulated emergencies and go-arounds were the highest, with the ratings for the rest of the segments at about the same level—in the range of 3.0. Overall, the mean HRs across all missions ranged from about 84 bpm to 95 bpm. The highest HRs were associated with takeoff, touch and go, go-around, and final landing. The lowest HRs during flight were associated with the climb-out following takeoff, cruise, simulated emergencies, and the descent

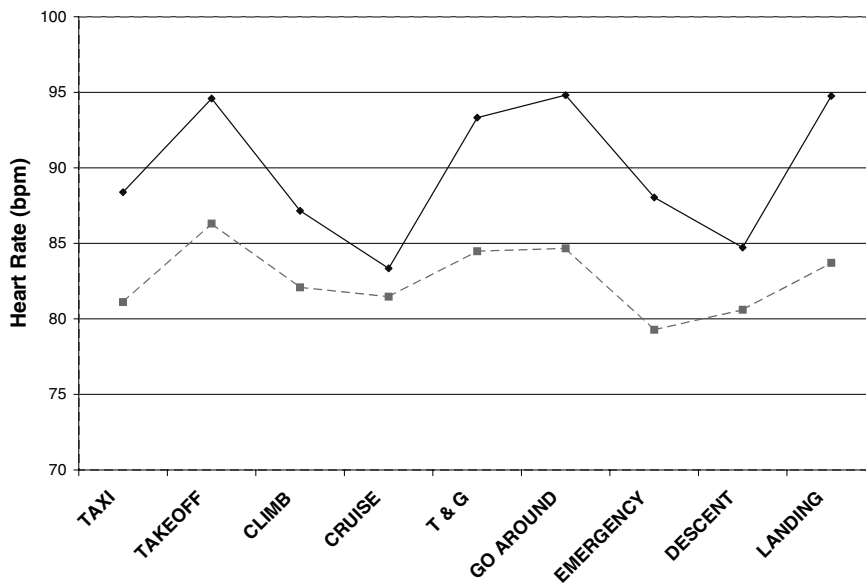


FIGURE 2 Heart rate for the PIC, top curve, and non-PIC, lower curve. Each point represents means for 2 min during each of the labeled events.

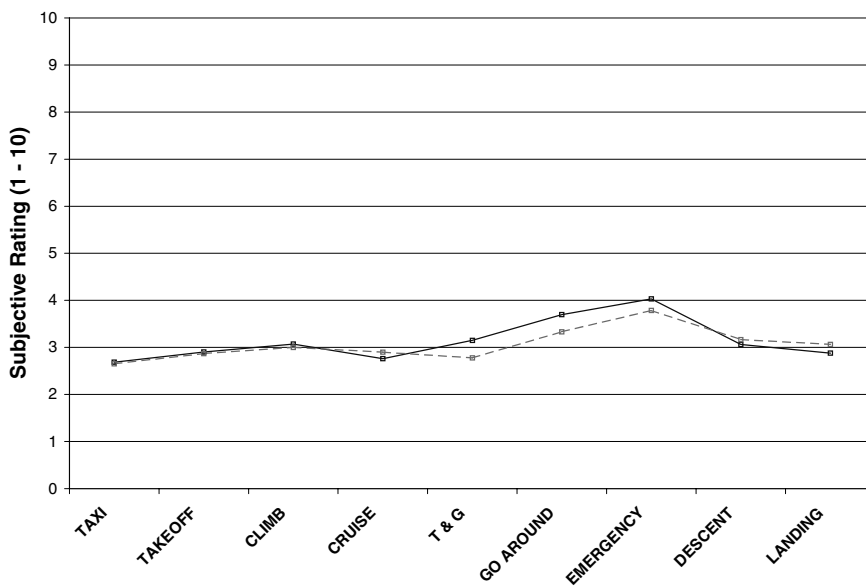


FIGURE 3 Mean subjective workload ratings for PIC, solid line, and non-PIC, dashed line.

to final landing. The discrepancy between the HR and subjective workload shows that the two measures are responding to different aspects of the workload that pilots experience.

The loadmaster's job primarily involved loading and unloading the aircraft cargo and reconfiguring the cargo area to match the expected type of load. The loadmaster had to check the cargo during flight, but most of the flight time was spent relaxed. HR and subjective workload estimates were collected from the loadmaster. As expected, the loading and reconfiguration duties brought on increased HRs, whereas the segments during the flights were associated with lower HRs (Figure 4). There were isolated periods of increased HRs during the flights when the loadmasters were moving about the aircraft. The subjective ratings were consistent with the loadmasters' duties: higher during the loading and lower during the actual flights.

During the tactical airlift phase, data were collected from four pilots and five copilots. Data from 17 missions were evaluated. The main purpose of this phase of testing was to test the aircraft during landings on short, narrow, assault runways. Assault landing zones are much smaller than normal runways. They are typically 3,000 feet long and 60 feet wide and may be unimproved (dirt). This more difficult maneuver produces increased mental workload on the crews. Dur-

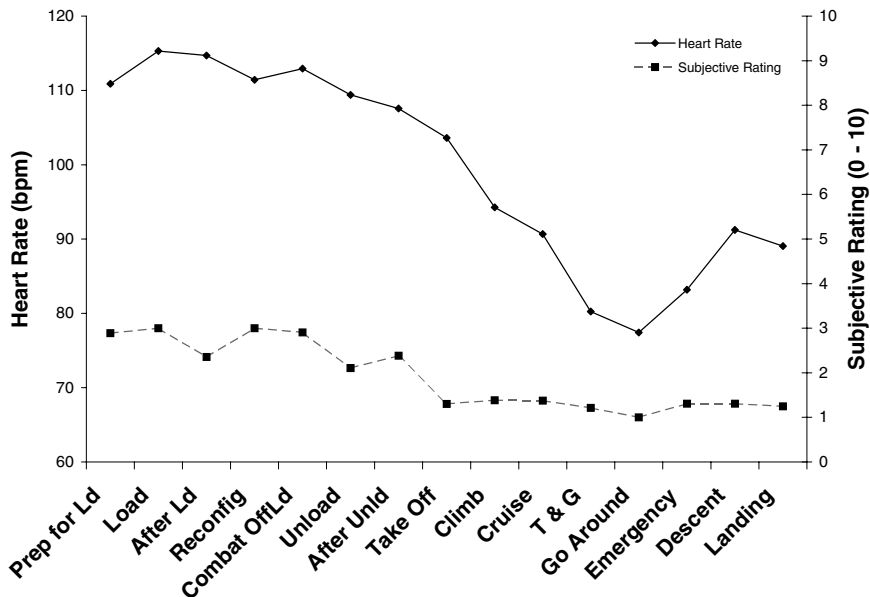


FIGURE 4 Loadmaster mean HRs and subjective workload ratings. The HRs, solid line, are means for 2-min periods at each of the labeled mission events.

ing testing, two types of runways were used. The first was implemented by painting markers on a normal (prepared) runway that simulated the boundaries of an actual assault runway. The second was an actual assault (unprepared) runway that was narrower and shorter than a typical runway; its surface was packed earth rather than the typical prepared surface of normal runways. The HRs associated with landing and taking off on both of these runways were higher than those associated with standard takeoffs and landings on prepared, normal-length runways (Figure 5). Furthermore, landing and taking off on the unprepared runways produced the highest HRs. These higher HRs were produced by the higher workload levels associated with this more difficult maneuver. The HRs for assault takeoffs from the prepared surface assault runways were about 10 bpm higher than those while taking off from the standard runways. Takeoffs from the unprepared assault runways produced mean HRs about 30 bpm higher than those from a standard runway. Landing on the prepared surface assault runways was associated with an approximately 25-bpm increase in HR over that found while landing on a standard runway. The largest increase in HR, 45 bpm, occurred when landing on the unprepared assault runway. These large increases highlight the high demands placed on pilots when taking off and landing on shorter, narrower runways.

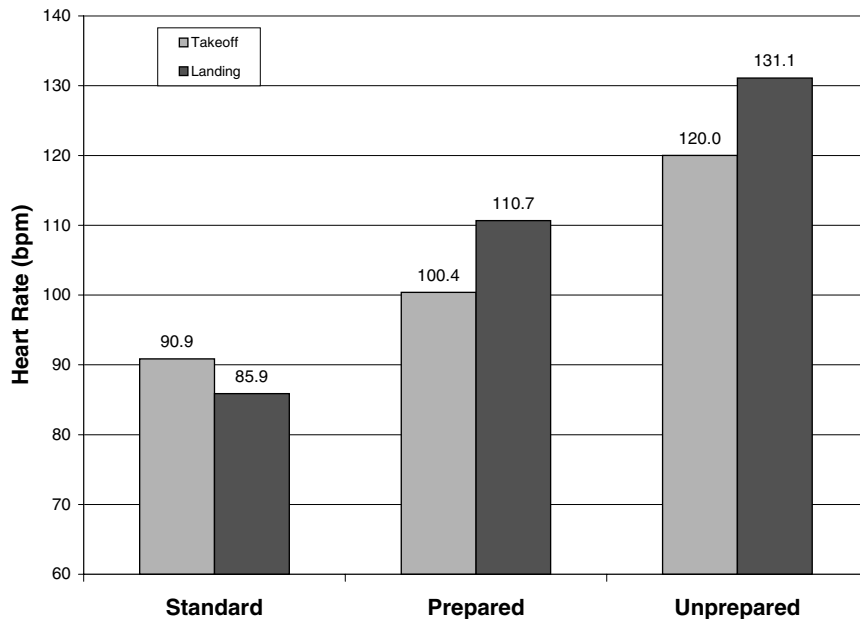


FIGURE 5 Heart rate means for PICs during takeoffs and landings at standard, prepared assault, and unprepared assault runways.

On the other hand, the subjective ratings of these landings were only slightly higher than those given to the standard landings and takeoffs (Figure 6). The HR and subjective data for the copilots (non-PIC) during the assault landings on the unprepared runways did not show the high HRs exhibited by the pilot (PIC), whose subjective ratings were also low. The pilot's mean HRs for takeoffs and landings on the unprepared runway were 85.6 and 87.1 bpm, respectively. He also showed very little difference in his subjective workload ratings for the takeoffs and landings (3.3 and 3.1, respectively). This suggests that the higher pilot HRs were driven by the increased cognitive demands of the landings and not a fear response regarding safety. The copilot faced the same danger but did not exhibit the large increases in HR.

An unplanned event produced data that provide information regarding the upper limits of HR response during very high cognitive demands while flying. Because the aircraft has to operate during all types of weather, flying in winter conditions was tested. This included flying to snow-covered airfields to test the aircraft. During one mission a full-stop landing on an ice-covered runway was attempted. However, due to the existing weather conditions (strong crosswinds), the decision was made to change the full-stop landing to a touch-and-go landing. Had the weather degraded further, the approach could not have

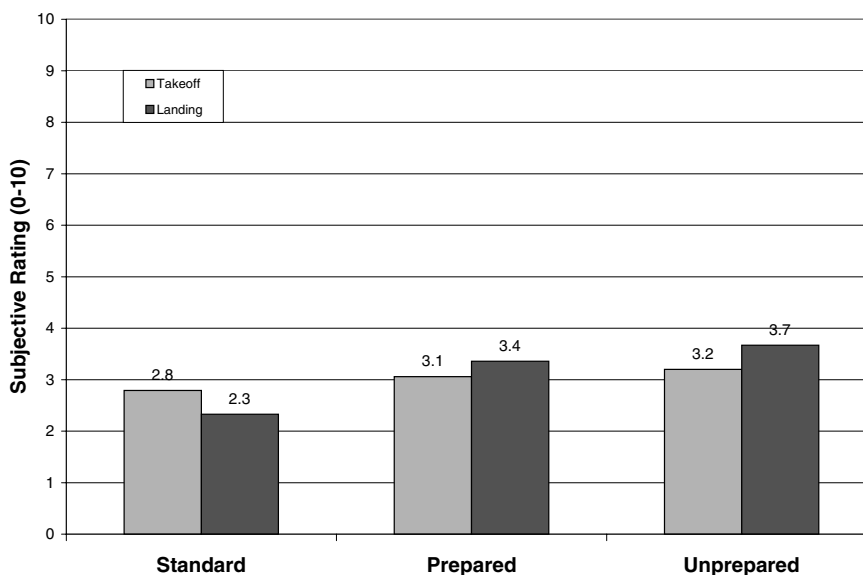


FIGURE 6 Mean subjective workload ratings from PICs in response to takeoffs and landings at standard, prepared assault, and unprepared assault runways.

been attempted. This produced very high pilot (PIC) HRs. Figure 7 shows the interbeat interval plot of the pilot's cardiac response to the two touch-and-go landings (attempted full-stop landings) under these conditions. This resulted in the highest HRs recorded during the testing. The pilot's (PIC) HR peaked at about 175 bpm during each of the two touch and goes. The mean heart rate for the 2 min surrounding these touch and goes was about 140 bpm for the first touch and go and approximately 158 bpm during the second. His HR during the time between the two touch and goes decreased to only about 120 bpm. The pilot's (non-PIC) mean HR during the final landing at the home base was 99 bpm. The pilot rated his workload at 10 on both touch and goes and at 2 for the initial takeoff and final landing. He rated a go-around just prior to the first touch and go as 5 while his mean HR was 111 bpm. During the two touch and goes, the copilot's (non-PIC) mean HR was 86.2 and 86.5 bpm. The copilot's subjective workload ratings were 5 and 6 for the first and second touch and goes. He rated the final landing as a 2, and his mean HR was 94.8 bpm as he was PIC. These data show the pilot's and copilot's responses to extreme situations. Moreover, the differences between the PIC's and non-PIC's HR responses highlight the relationship between the cognitive demands of each crew position and HR responses. In this situation the HR and subjective workload ratings were highly correlated.

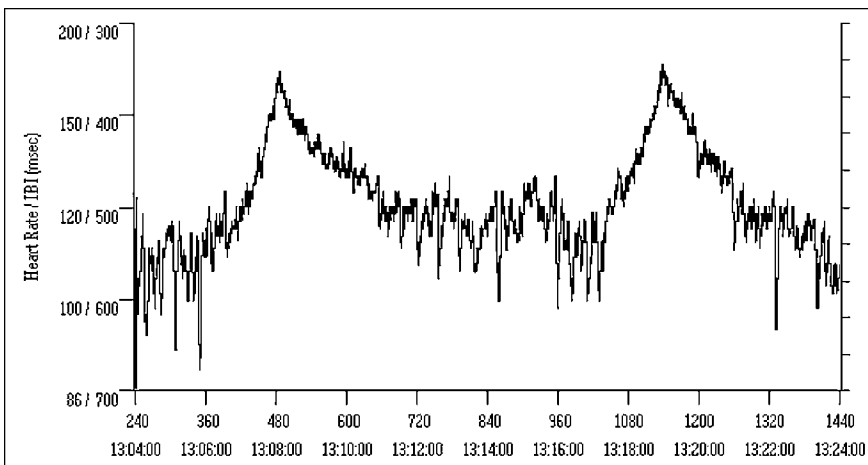


FIGURE 7 Interbeat intervals recorded from PIC during two touch-and-go landings on an icy runway. The two peaks in the interbeat intervals corresponded with touchdown. Note the very high heart rates at touchdown and the high heart rates during the time between touch and goes.

DISCUSSION

HR is another tool available to the test team to provide insight into the demands placed on the aircrew. Because of the complexity of flying, the additional information provided by HR is useful in understanding the effects of the activity on the crew. HR can augment subjective and performance data and enhance our understanding of the effects of complex task performance on operators. It can serve as a means to confirm subjective workload ratings. Among its strong points is the ability to continuously monitor operator state while in flight on a noninterference basis. Because crew members quickly adapt to wearing electrodes these do not interfere with job performance. Because the data are continuously recorded, responses to unexpected events are available for analysis. The touch and goes on the icy runway illustrate this point. If the crew had not been instrumented, the extreme HRs in response to this high workload situation would have been lost.

Finding the HR increases when the PIC was landing the aircraft confirms the validity of the data. Several studies have reported increased HRs from the PIC (Hart & Hauser, 1988; Roscoe, 1978; Wilson, 1992). Widely reported phenomena such as this can help determine whether data are showing expected results. If unexpected patterns are found, the data can be used to focus attention on those segments of the test missions.

The small HR increases in response to simulated emergencies has been reported previously. Wilson, Skelly, and Purvis (1989) reported increases of up to 50% in HR in response to actual in-flight emergencies whereas simulator emergencies showed no change in HR. This might be caused by the rote nature of crew members' responses to emergencies. The required responses to most aircraft emergencies are highly practiced so that the crew members can quickly and effectively respond to these situations. It may be that the simulated emergencies elicited these rote responses and did not require a great deal of cognitive activity, resulting in small increases in HR. The actual emergencies required more processing to confirm that they had in fact occurred, which required further cognitive processing, resulting in the higher HRs. There is an emotional component to actual in-flight emergencies that is responsible for some component of the cardiac response.

HR is an objective measure; however, the interpretation of the data may involve subjective elements. Because there are no set criteria to determine whether an operator's state has entered a dangerous range, subjective interpretation is required. There are no agreed-on thresholds that can determine whether a state such as mental workload overload has been reached. However, one can make useful decisions in conjunction with subjective and performance data. Precautions must be taken to ensure that changes in HR are not the result of artifacts such as moving about the aircraft. An example of this is shown in Figure 1. The

HR increases while the pilot and copilot were moving about the cockpit were as large as the increases during touch-and-go maneuvers. Although it is remarkable that touch and goes are associated with such large increases in HR, one must be aware of the circumstances surrounding any changes in HR. This is easily accomplished. Observers are part of the test team, and they can easily note when crew members move about the aircraft.

The discrepancies between HR and the subjective workload ratings provide useful data. These measures are obviously sensitive to different aspects of flying and the crew members' responses to them. In the current data the subjective ratings are for the most part quite uniform and are restricted to a small portion of the lower range of the available 0-to-10 rating scale. The HR shows a wider range of responses that at times seems to better fit the workload demands of the flight segments. One interesting example is the differences in the HR and subjective responses to the simulated emergencies. The HR shows little change, whereas the mean subjective ratings for these events are the highest of all segments. It is possible that the subjective ratings are related to the assumed mental demands and uniqueness of the emergency situation, whereas the HR may be sensitive to the actual demands of the simulated emergency, which required only routine procedures. Another example that produced opposite results is the responses to the final landings. The HRs increased, whereas the subjective ratings were essentially the same as for most of the other, lower rated segments. The lower subjective ratings for the final landings may reflect the perceived routine nature of landing, whereas the HR data show the actual cognitive demands required to land the aircraft. However, when the mental workload was extremely high, during the touch and goes on the icy runway, for example, the HR and subjective ratings are highly correlated—175 bpm and a rating of 10. Subjective and performance data from laboratory experiments are known to dissociate under circumstances of low or high mental workload (Eggemeier & Wilson, 1991; Yeh & Wickens, 1988). Hankins and Wilson (1998) also reported dissociation between subjective ratings of flight segment difficulty and psychophysiological measures.

The relationship between HR and subjective workload ratings provides useful data that can provide insights into the workload that crew members experience during flight. HR data can also provide quick looks at the data that can be used to readily identify high stress areas for further study. If the HR can quickly be made available following each flight, it can also provide the aircrew with immediate feedback on their workload. Artifacts can be quickly filtered out prior to presentation to the crew. As pointed out earlier, Rokicki (1987) successfully used this strategy in the test and evaluation environment. HR can help identify the balance or sharing of workload duties among crew members in the cockpit during a particular phase of flight. This may be useful in improving crew and cockpit resource management. HR or other psychophysiological measures could

assist in monitoring the workload of crew members during specific flight tasks. These data could then help determine whether an imbalance of work existed among the crew. If imbalances were observed, then one could investigate ways of reassigning tasks to lower the workload of the overburdened crew members.

Overall, HR can add valuable information to the test and evaluation community. When used as a component of a larger battery of measures, it adds value to the testing. Currently available hardware and software have greatly improved the field use of HR as an applied measure. Further developments and increased use will no doubt see more widespread use of HR in a wide range of test and evaluation environments.

ACKNOWLEDGMENTS

The authors wish to thank Major Greg King, Captain Reginald Kabban, Second Lieutenant Chris McClernon, and Major Cynthia Brown for their assistance in the study design and data collection. They also wish to thank Dennis Allen, Jared Lambert, and George Reis for their assistance with data analysis.

REFERENCES

- Eggemeier, F. T., & Wilson, G. F. (1991). Performance and subjective measures of workload in multi-task environments. In D. Damos (Ed.), *Multiple-task performance* (pp. 217–278). London: Taylor & Francis.
- Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, 69, 360–367.
- Hart, S. G., & Hauser, J. R. (1987). In-flight application of three pilot workload measurement techniques. *Aviation, Space, and Environmental Medicine*, 58, 402–410.
- Rokicki, S. M. (1987). Heart rate averages as workload/fatigue indicators during OT&E. *Proceedings of the Thirty-First Annual Meeting of the Human Factors Society*, 2, 784–785.
- Roscoe, A. H. (1975). Heart rate monitoring of pilots during steep-gradient approaches. *Aviation, Space, and Environmental Medicine*, 46, 1410–1413.
- Roscoe, A. H. (1978). Stress and workload in pilots. *Aviation, Space, and Environmental Medicine*, 49, 630–636.
- Roscoe, A. H. (1979). Handling qualities, workload and heart rate. In B. O. Hartman & R. W. McKenzie (Eds.), *Survey of methods to assess workload*. Paris: AGARD.
- Roscoe, A. H. (1992). Assessing pilot workload: Why measure heart rate, HRV, and respiration? *Biological Psychology*, 34, 259–288.
- Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology*, 34, 163–178.
- Wilson, G. F. (2001). In-flight psychophysiological monitoring. In F. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory monitoring* (pp. 435–454). Seattle, WA: Hogrefe & Huber.

- Wilson, G. F. (2002). Psychophysiological test methods and procedures. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of human factors testing and evaluation* (2nd ed., pp. 127–156). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wilson, G. F., & Eggemeier, F. T. (1991). Physiological measures of workload in multi-task environments. In D. Damos (Ed.), *Multiple-task performance* (pp. 329–360). London: Taylor & Francis.
- Wilson, G. F., Skelly, J., & Purvis, B. (1989). Reactions to emergency situations in actual and simulated flight. In *AGARD Conference Proceedings: Human behavior in high stress situations in aerospace operations* (pp. 9–13). Paris: AGARD.

Manuscript first received May 2001