AIAA Atmospheric Flight Mechanics Conference
10 - 13 August 2009, Chicago, Illinois

AIAA 2009-6317

Handling Qualities Evaluation at the USAF Test Pilot School

William R. Gray, III*
USAF Test Pilot School, 220 S. Wolfe Ave, Edwards AFB, CA, 93524, USA

A contradiction lies at the core of handling qualities evaluations: The most complex subsystem of the handling qualities loop is both the least understood and most familiar system. This subsystem is, of course, the human body. A test pilot's lack of understanding of the mechanisms involved in his cognition and motor control is only exceeded by the confidence he shows in his ability to report on them. Nevertheless, satisfactory handling qualities are vital for safe mission accomplishment so the USAF Test Pilot School (TPS) must endeavor to prepare its graduates to rationally conduct these evaluations. In the last five years, the USAF TPS has undergone a steady transformation in the education of its students for these tasks. This change has stemmed partly from changes in the foundational assumptions of how pilots control, partly from TPS research into objective measurements of pilot inceptor motion, and partly from the recognition that traditional pilot modeling approaches serve only as useful analogies that have little in common with the actual mechanisms of human cognition and motor control. This paper will review the history of handling qualities evaluations at the USAF TPS and describe how shifts in the foundations of the School's understanding of pilot control have encouraged significant adjustments to the aircraft handling qualities curriculum. These adjustments are helping students better understand how their minds and bodies work and are helping them learn how to create and execute flight test techniques that more efficiently and accurately predict aircraft capability.

I. Introduction

Aircraft handling qualities (HQ) testing has been, and remains, dependent upon pilot opinion. From the early days of flight test when the pilot dictated change based solely on intuition, to modern times when the phrase "the most important data is the pilot's comments" is universally preached, the indispensability of pilot opinion has always been part of the assumed wisdom of HQ testing. In spite of valiant efforts to mathematically model pilots, the flight test community has essentially left unquestioned the pilot's role as an almost impenetrable black box upon whom the engineer must rely for subjective data. Test pilots, thanks to the high baseline self-confidence that seems to be a prerequisite for the role, are more than happy to leave this methodology unchallenged. Twenty years ago at the USAF Test Pilot School (TPS), HQ testing was an afterthought limited to a few hours of classroom, a flight or two, and a single flight test technique (FTT). About fifteen years ago, the USAF TPS began the process of academically examining the methods and assumptions of HQ testing. HQ testing now encompasses many hours of instruction and simulation, at least three simulator sessions, and five dedicated demonstration and data sorties. New FTTs born from a new appreciation of the complexity of tracking strategies promise to make test pilots and test engineers much more aware of pilots' "inner loops" and much more thorough and scientific evaluators.

The purpose of this paper is to tell the story of USAF TPS HQ curriculum changes over the last fifteen years, sketch out a roadmap for the future of the HQ curriculum, and summarize the challenges and opportunities that remain. It is a whirlwind tour of a turbulent time full of conflict and discovery, written in the fervent hope that the aircraft handling qualities community at large might benefit from the knowledge gained in an environment focused on the day-to-day instruction of curious, intelligent, and demanding students. First, we will examine the history of HQ testing and education at the USAF TPS. From there, we go on to examining our limited understanding of the pilot as a subsystem and how we might improve that

*Chief Test Pilot, USAF Test Pilot School, 220 S. Wolfe Ave, Edwards AFB, CA, 93524, AIAA Senior Member.

understanding. The goal of understanding the pilot is, of course, to understand the closed-loop pilot/aircraft system and this subject will be our next focus of attention. Finally, we will examine the various HQ flight test techniques that the USAF TPS is using, with extra attention given to the Cooper-Harper Rating Scale on this fortieth anniversary of its first publication.

II. A Short History of Handling Qualities Testing at the USAF TPS

The recent history of handling qualities testing at the USAF Test Pilot School can be roughly divided into three "eras." The first era was the Era of Benign Neglect, when aside from some non-testable discussion of pilot "gain," pilot induced oscillations (PIOs), and the Cooper-Harper Rating (CHR) Scale,[1] students were left with the distinct impression that HQ flight test was all about ensuring that open-loop flying qualities (hereafter referred to as "flying qualities" or "FQ") met specifications. The second era, the Era of HQDT, began with any extremely aggressive and necessary change to bring closed-loop handling qualities to the front-and-center of the aircraft flying qualities curriculum. Given the scope of the change, the high degree of confidence placed on a particular methodology (HQDT—"Handling Qualities During Tracking"), and the opinion-centric nature of HQ testing, conflict was assured. The USAF TPS is now entering a third era.

A. Traditional Techniques and the Cooper-Harper Rating Scale

The Cooper-Harper Rating Scale is, far and away, the most important and common handling qualities FTT in existence. This scale, essentially a human factors survey created in the early 1960's specifically for the needs of airplane handling qualities evaluation, is the final product of years of lessons-learned in the employment of earlier rating scales. It enjoyed almost immediate success and is now used almost universally in aircraft handling qualities testing. When the author attended the USAF TPS from Summer 1991 to Summer 1992, the only flight test technique practiced in the curriculum for evaluating aircraft handling qualities was the CH rating.

The 1991 curriculum for the USAF Test Pilot School set aside just three hours for an "Operational Handling/HQDT" lecture. The courseware has been lost to history; there was no textbook provided so the only classroom materials were acetate slides. All that remains is the outline of the course, "Course discusses the basic principles of operational handling and handling qualities during tracking (HQDT), and teaches the correct use of the Cooper-Harper Rating scale. Terminology and task design are emphasized."[2] The author recalls a total of about one-half hour dedicated to practicing use of the CHR scale in several different aircraft. HQDT was only discussed in passing and no attempt was made to demonstrate this technique in flight or in a simulator. (TPS did not have any simulator capability at the time.) Pilot induced oscillations (PIO) were discussed in depth during lectures on the prediction of aircraft HQ, but aside from noting their occurrence during difficult tracking tasks on demonstration sorties, there was no attempt to show how to test for PIO tendencies or severity.

When the history of PIO and flight test is taken into account, the paucity of instruction on in-flight methods at TPS to find, describe, and prevent PIO was surprising. This may have stemmed from the hesitancy of the the pilot instructors—based largely on their own operational experience—to see PIOs as anything but the result of poor or inexperienced flying, it may have come from the lack of a commonly accepted flight test technique, or perhaps it came from the inability of the rapidly changing and inexperienced staff to devise or incorporate new material. (At the time the flight instruction staff and most of the academic instructors were active-duty military officers on a two-year rotation.) Even the acronym "PIO" was a point of contention. It is difficult to instruct a subject when even the name of the subject isn't agreed-upon.

B. When PIO was King; the "HQDT Decade"

On a beautiful Saturday afternoon in the Spring of 1992, a single column of smoke rose from the main runway at Edwards AFB. The YF-22 had just landed hard with the gear up and caught fire. The pilot, a highly experienced and respected test pilot, was safe but the aircraft was a complete loss. A book could be written on the ensuing conflict over the cause of the accident, but in the end the proximate cause was a severe pitch PIO encountered during a low approach. This event was seen as predictable and unsurprising (in retrospect, of course) and it was probably a major catalyst for the USAF TPS to reexamine their PIO training approach and curriculum.

At the time, the only significant AFFTC flight test technique for rooting out PIO was HQDT. Documentation of this technique in the AFFTC publication "Handling Qualities Testing" confidently states,

> HQDT is perhaps the most important handling qualities test technique at your disposal. At present, it is unmatched as a means of evaluating high bandwidth handling qualities. As flying qualities test pilots and engineers, you will use HQDT test techniques in simulator evaluations and in flight test evaluations.[3]

These words were published in February 2002, almost a decade after the YF-22 mishap, but were in reference to a technique that had been in sporadic use around the AFFTC since the 1970s. In the mid-1990's, with the recent accident and an "unmatched" test technique at its disposal, TPS set about completing a major reorganization of its flying and handling qualities curriculum. The new curriculum would step away from simple flying qualities specification compliance and CHR evaluations and fully embrace the HQDT assessment methodology. Again from "Handling Qualities Testing:"

> **Phase 1 Low Bandwidth Handling Qualities Testing** During Phase 1 handling qualities testing we conduct an evaluation of low gain, or low bandwidth handling qualities at a safe, up-and-away flight conditions...
> **Phase 2 High Bandwidth Handling Qualities Testing** Phase 2 handling qualities testing is an evaluation of high gain, or high bandwidth handling qualities. The purpose of Phase 2 high bandwidth testing is to force the pilot to minimize his compensation and increase his gain, or bandwidth. This is intended to emulate the way less skilled or less experienced pilots might fly, or the way any pilot might fly whose level of stress, excitement, anxiety, or fear exceeds a certain threshold. Hence, Phase 2 high bandwidth test is often referred to as handling qualities stress testing...
> **Phase 3 Operational Handling Qualities Testing** Phase 2 testing is an operational evaluation of the airplane handling qualities. During Phase 3 testing, pilots evaluate handling qualities while performing the various tasks that make up the design mission of the airplane.[3]

It is hard to understate the importance or comprehensiveness of the mid-1990's curriculum change. Handling qualities were literally brought front-and-center and every other aspect of flying qualities testing was relegated to a supporting role. Some of the most important curriculum shifts that occurred with this change were

- The first course taught in the flying qualities curriculum was "Handling Qualities," discussing the main concepts of open- and closed-loop systems and the HQDT flight test technique. From there the curriculum continued with a more traditional approach (equations of motion, statics, dynamics, etc.).
- Simulator sessions were added in a simple but effective system to instruct students on the HQDT FTT, PIO, the PIO rating (PIOR) scale, and CHRs.
- The first flying qualities demonstration flight focused almost exclusively on handling qualities evaluations; HQDT, CHR, and PIOR.
- Open-loop flying qualities test techniques were deemphasized along with basic flying qualities specification compliance testing.
- All PIOs were very conservatively defined as being a "loss of control," requiring a CHR of "10" and correction.[3]
- Student data flights were shifted to place higher emphasis on HQDT and CHR as opposed to open-loop FQ specification compliance.

The intent of this curriculum change was without fault. PIOs had been responsible for numerous major flight test incidents, including aircraft losses and fatalities, and the emphasis on flying qualities in the curriculum missed the point that, for operational employment, flying qualities are just one part of the system necessary to produce satisfactory handling qualities and successful mission accomplishment. It seemed then that the curriculum finally placed the emphasis where it belonged. Graduates would understand the "most important handling qualities test technique" and be fully prepared to find and eliminate all PIOs. The new curriculum was deployed in 1995.

C. Turmoil and Transition

The deployment of the new curriculum was smooth enough, but gaining acceptance with TPS instructor pilots (IPs) was another thing entirely. Even the simplest and most obvious changes will cause some conflict as the pilot community is taught a new way, but HQDT was a significant change in a completely new direction. T-38, F-15, and F-16 IPs were charged with teaching the Phase 2 high bandwidth HQDT technique but, for the most part, either did not grasp the academic concepts behind the technique, did not "get" the technique itself, or lacked the ability to properly instruct this particularly difficult FTT. While a few instructors were quite confident of the technique, many instructors felt that the technique was oversold and not particularly effective for its intended purpose. After a few classes of frustrated instructor pilots and confused students (imagine having an instructor tell you, "this technique is crap!") the TPS instituted a requirement that IPs must successfully complete the student curriculum Phase 2 HQDT simulator sessions with a qualified handling qualities engineer. This didn't seem to help much. Unfortunately, the conflict became trench warfare between a majority of IPs convinced that the technique was too difficult to teach, invalid, unnecessary, or all three and handling qualities engineers utterly convinced of the validity of the technique and its primary importance in a flight test team's tool set. The conflict was so apparent that within two years of Phase 2 HQDT's introduction into the curriculum, the TPS Commandant directed the staff to conduct a research project evaluating the technique. The authors summarized the staff's complaints in the following list:

a) "Phase 2 [HQDT] makes all airplanes look bad.

b) The technique is not suitable for heavy aircraft.

c) The technique "breaks" airplanes (over-g).

d) Operational Pilots do not fly that way.

e) Results are inconsistent and pilot dependent.

f) Hard FTT to accomplish correctly, and difficult to teach."[4]

The study satisfactorily answered all of the concerns on the list except for items a) and f). The limited nature of the study (only four pilots flew in the study and two of them were the authors) made the conclusions difficult to accept for those with doubts. The completion of the study only served to put the status of the methodology on hold, which in the high-workload environment of the USAF TPS is equivalent to maintaining the status quo. The technique and the curriculum change remained in place and unchanged. Yet the conflict continued, essentially unabated, and the flight test technique continued its run as a methodology occasionally used only at the Air Force Flight Test Center.

As more classes and instructors flowed through the school, the complaints about Phase 2 HQDT began to settle into three essential elements. First, the technique worked "too well." It could find PIO in just about any aircraft. This was just a less pessimistic statement than item a) in the 2000 study. Second, When it did find PIO, it gave no information on the level of effort required to create the PIO and it didn't answer the question "is this PIO a problem?" The methodology was built on an assumption that *all* PIOs were an unsatisfactory deficiency, so the threshold of effort to cause the PIO was essentially immaterial. This conflict (Does the PIO matter or not?) originated with test pilots comparing their Phase 2 HQDT results with their operational experiences learning to fly the same aircraft type. Aircraft with which the FTT showed a definitive PIO problem had been flown for decades in the same conditions and the PIO only showed itself with new students learning the task. Students and instructors unsurprisingly asked, "How could TPS endorse a technique that calls a known satisfactory aircraft unsatisfactory?" Third, the technique was extraordinarily difficult to instruct—very few students seemed to be able to do it correctly even if given quality instruction. At first this was attributed to IP inexperience, but after a few years it became evident to experienced IPs that they just couldn't seem to figure out a way to get most students to properly execute the technique. It became clear that time and experience would not produce acceptance of the technique. There is no question that Phase 2 HQDT will, when properly executed, find PIO tendencies, so in that respect it is a "successful" FTT. Why then, was achieving acceptance so difficult?

The fundamental difficulty with the technique may well have been a matter of terminology. In the terminology of the HQDT regimen, PIO is defined so strictly that the Flying Qualities Testing textbook definitively states "You will discover during your training at the Test Pilot School, that a genuinely Level 1 or Level 2 airplane cannot be made to PIO."[3] Yet Phase 2 HQDT was showing PIO in almost every application and in aircraft with long histories of satisfactory, even excellent, handling qualities! The 2000 staff research project demonstrated this nicely, with PIO identified by at least one of the four pilots in over half of the test conditions.[4] In the HQDT way of seeing pilot-aircraft instability, all PIOs are bad. But this philosophy is counter to the experience of most pilots, who have encountered PIO tendencies throughout their flying career and with almost no exception safely handled them. Close formation is a particularly good example of a task where all pilots start by PIOing until they finally learn to do the task. It is a testament to the Phase 2 HQDT technique that, when properly conducted, it will allow a skilled test pilot to identify the very PIO modes that students encounter every day when learning to fly formation. But telling a test pilot that an aircraft with decades of operational use, that they have flown in very tough real-world conditions, is deficient for that task is not a good way to sell your technique!

In the end then, it became clear that the problem with the technique may have had much more to do with its underlying assumptions and definitions than with its effectiveness for finding PIO tendency. The disconnect between pilots' perceptions of PIOs (not uncommon, rarely hazardous, often not worth correcting) and the HQDT developers' perceptions of PIOs (often hazardous, never acceptable) made a difficult technique even more difficult to accept. But these were problems worth solving and the conflict has created much discussion about PIOs, handling qualities FTTs, pilot tracking strategies, and seen movement to a more realistic perception of pilot in-the-loop instabilities.

The conflict raised larger issues about the USAF TPS HQ curriculum. The theoretical models behind Phase 2 HQDT were quite complex so it was necessary to provide a very low-level academic treatment of them and rely on FTT training to prepare pilots for handling qualities evaluations. It is possible that more in-depth education would have made the FTT instruction easier, but there was no time for additional academic treatment. So what is more important, the ability of pilots to accomplish a given FTT or their understanding of that FTT? How much information is too much? The difficulties encountered with HQDT raised important questions about the curriculum as a whole.

1. *Are Training to Standard Flight Test Techniques the Correct Educational Approach?*

The fundamental mission of the Test Pilot School is to educate its charges in the necessary engineering disciplines, scientific theories, and developmental test and evaluation methodologies so that they may enter the world of flight test prepared to contribute and rapidly move into leadership positions. Training is generally limited to only what is necessary to allow a smooth and safe educational experience. For instance, students are trained in the operation of curriculum aircraft and day-to-day flying operations at Edwards AFB. They then use the curriculum aircraft, under the tutelage of their instructors, to experience selected characteristics that they learned about in academics. Their training supports their education, not the other way around. Contrary to this approach, many flight test techniques are seen as discrete maneuvers that the student test pilots must be trained to accomplish. But in reality, a flight test technique is nothing more than an operation to efficiently and effectively extract the data necessary to confirm that an aircraft or subsystem is capable of performing as designed. Each airplane and each requirement combine to demand a unique test technique. There are accepted practices and common methodologies, but if a test team is selecting test techniques from a reference book (like a flight test cookbook), they are increasing their chances of failure by failing to take their particular aircraft and specifications into account.

To continue the cooking analogy, consider the difference between a "chef" and a "cook." A chef can take available ingredients and, using a comprehensive knowledge of ingredients and techniques, create a delicious meal. A cook takes a procedure, or a recipe, then gathers the ingredients and prepares them in accordance with prepared instructions. Cooks can make excellent meals but are limited by their training. Chefs are limited only by their knowledge and their ingredients. If the Test Pilot School concentrates on, or arguably even delves into, training for specific FTTs without providing sufficient education to provide their students with a full understanding, the TPS is creating skillful but extremely limited "flight test cooks." Graduates of the USAF TPS must have sufficient knowledge to create flight test techniques for whatever mix of airplane and requirements they are given.

2. *How Much is Too Much?*

In spite of the USAF Test Pilot School's lofty goals, there is a limit to what may be successfully taught in a year. A full understanding of aircraft handling qualities design, prediction, and evaluation would require more time than is available for the entire course. Clearly, priorities must be set and the best base of knowledge that can be provided in the time available must be identified. Handling qualities evaluation requires some understanding of the entire closed-loop system, starting with the aircraft from forces and moments, to control systems and displays, and finishing with the human system from perception to computation to action. To date, the USAF TPS has spent many hours on the systems external to the pilot and very little (aside from analogous linear approximations) on the human pilot. Given the rapidly advancing state of cognitive science, the science of the human mind, it is time to redress this imbalance.

III. On Evaluating the Pilot as a Subsystem

Writing courseware on the pilot as a subsystem will be a difficult task. There are no cognitive science or cognitive engineering specialists at TPS, much less at Edwards AFB. (Do "cognitive engineers" even exist?) Yet a cursory examination of books and papers on cognitive science and motor control reveal that specialists in these subjects know much more about the inner workings of the human mind than most flight testers can even imagine. Injecting some of this knowledge into the flight test community can only help!

A. The Importance of Test Pilot Self-Observation

It is important to note in passing that pilots, in regard to their hands-on flying skills, have much more in common with children on bicycles than with cognitive scientists. While observing student test pilots attempting new tasks, it is quickly apparent that at the start of their TPS education they have very little awareness of the motions they make to control their aircraft. Like children on bicycles, their subconscious control is so automatic that they have virtually no appreciation of the methods their subconscious mind uses to create the result they consciously desire. Test pilots should be able to go far beyond observing their performance; they should be able to provide some information on the operation of their subconscious feedback loops.

B. Is There "Gain" in a Pilot's Brain?

For many years engineers have sought to model pilots by using the mathematics of feedback theory. From the simplest pilot models, created by feeding back a time-delayed error signal multiplied by a gain, to models that attempt to take into account the pilot's structural dynamics, learning, lead and lag control, and so on, these models have produced useful results. The first engineers that modeled pilot response by using the mathematics of linear feedback theory may have thought that they were modeling the true operation of the human mind. But as scientific knowledge of the human mind advances it has become abundantly clear that these models must be thought of as only informative analogies.

It is extraordinarily important that analogies be taught as such. When discussing the pilot in a closed-loop system, "pilot gain" is an analogy. An important and powerful one, but still an analogy. Most test pilots speak confidently of changing their "gain," but few have an understanding of the mathematical origin of the term, much less what's really going on while they are controlling their aircraft. This is an unsatisfactory state. The relative simplicity of linear pilot modeling and the way that a few terms from these efforts (i.e. gain and bandwidth) have gone adrift from their original meaning and come to be used as shorthand for a subject the flight test community poorly understands, points to the need to look outside of the flight test community for a more correct representation of how pilots control airplanes.

C. Human Cognition and Motor Control

At the USAF Test Pilot School, aircraft handling qualities is currently taught without any background science on human cognition and motor control. The situation is analogous to teaching aerodynamics without considering fluid mechanics. But just as aerodynamics is a small subset of fluid mechanics, aircraft handling qualities is a small subset of human motor control. It would seem that much could be gained by teaching the basics of motor control before moving to aircraft handling qualities as a special (and more complex) case. Cognitive science might be thought of as the "mother science" of aircraft HQ; after all, an airplane is just one more tool that the mind/body is called upon to control.

1. Some Intriguing Results from Cognitive Science

Several hours spent with a single source, *The MIT Encyclopedia of the Cognitive Sciences*,[5] produced some intriguing information relative to aircraft handling. First, in the area of motor control, there are multiple hypotheses about how commands are sent to muscles. The "equilibrium-point hypothesis" holds that muscles are not commanded to produce a force. Muscles are commanded to achieve or maintain an equilibrium position. This hypothesis nicely explains things like muscle response to external perturbation; the muscle automatically compensates so the control system in the nervous system is minimally involved. Does it also help to explain why holding a control stick in one place is easier than holding a constant force on a fixed stick?[a] Another hypothesis holds that there are only a limited number of commands that can be sent to a muscle, but they can be sent simultaneously in any combination giving a very wide range of capability. Both of these ideas have experimental support; it seems that they could both play a part in motor control. Second, there is evidence that the brain is able to represent the location of an object in both body-centered space and in body-independent space. This is important because the mind has to plan a trajectory to accomplish something like picking up a ball or rolling to 30 degrees of bank. Third, researchers have recently shown that we humans quickly incorporate tools into our body; our mind treats tools like an extension of our selves.[6] A science blogger described this effect nicely, "In martial arts classes, students are often taught to treat weapons as extensions of their own body. But this is more than just a metaphor. It turns out that when we use tools—not just swords and spears, but toothbrushes and rakes as well—our brain treats them as temporary body parts."[7] Pilots often talk about "strapping on the jet" and of how flying brings a sense of freedom. The aircraft seems to become an extension of their will, almost a part of their body. It seems that this romantic notion may, in fact, reflect the truth. It should be entirely unsurprising that cognitive scientists investigating the very basics of motor control should speak directly to the complex subject of aircraft handling qualities, yet somehow it is surprising. Perhaps it is time to look beyond aircraft handling qualities for answers on how pilots control aircraft.

2. Kids on Bikes

When it comes to controlling his aircraft, a Thunderbird pilot flying in formation has more in common with a five year-old riding a bicycle than he can imagine. Most handling qualities engineers and pilots have learned to ride a bicycle, usually at an early age. The process of controlling a bicycle is quite complex. Forces on the handlebars produce counter-intuitive results and they have more to do with balance than steering. Under way, balancing the bicycle has less to do with leaning and more to do with turning the front wheel with the handlebars. Anyone who has taught someone to ride a bicycle knows that there are two phases with a distinct and obvious phase change. The first phase is characterized with lots of handlebar movement, rapid loss of control, and lots of frustration. During this phase, some improvement can be seen but the pupil has the distinct sense that she's "never going to get it." The phase change is literally an "AHA!" moment, where the new rider suddenly realizes that she can do it, the wobbling rapidly decreases and the second phase, typified by rapid improvement, begins. There is nothing special about this pattern of learning, it occurs in any human learning a new control task, from walking on a tightrope, to wrist-flipping food in a saute pan, to driving a car. There is also nothing special about flying aircraft, as the many control tasks demanded of the pilot are subject to the same two-phase learning pattern.

Learning motor control is largely a process of allowing your brain to create an internal model. Once this model is created, your autonomous system can "take over" from your conscious system. You make the switch from thinking "move this to do that" to "do that." Imamizu *et al.*[8] provided physical evidence for such a model being created in the brain, specifically in the cerebellum. The phase transition so common to learning complex motor tasks is likely the point at which the cerebellar model becomes accurate enough that the autonomous system can take over the computation of motions, leaving the conscious mind free to treat the modeled object, be it a bicycle or airplane, as just another part of the body.

[a]The author has experienced this repeatedly with students attempting similar tasks in the T-38 and F-16. The F-16 stick senses only force and has a very small range of motion but the T-38 has a large range of motion with similar forces. Students are much better at holding a constant control input with the T-38 stick than with the F-16 stick.

3. *Bringing Cognitive Science into the Cockpit*

In 1997, the National Academy of Sciences published *Aviation Safety and Pilot Control*.[9] The committee members identified three categories of aircraft-pilot coupling (APC),[b] Category I, II, and III were described in turn:

- relatively benign, initial or early encounters that occur when the pilot is learning to adapt to the effective aircraft dynamics
- severe, potentially dangerous oscillations stemming from a combination of extreme task demands, which require very high gain in the PVS [pilot-vehicle system], and deficiencies in the effective aircraft dynamics, such as excessive time lag
- severe, potentially dangerous oscillations occasioned by pilot commands that are usually motivated by task demands and are large enough to cause a major nonlinear change (flying qualities cliff) in the effective aircraft dynamics[9]

These categories may have roots that follow from what cognitive science has learned about motor control.

- Category I, APC in the learning phase, can be seen as a consequence of cognitive control. Without a good cerebellar model, the pilot must make cognitive decisions about control inputs. This results in a much longer time delay and more likelihood of PIO and other handling mistakes. These APC issues may disappear entirely or become inconsequential once the motor task is learned.
- Category II, APC after successful cerebellar modeling, can be seen as a result of human autonomous control limitations. Even with the best possible cerebellar model, autonomous control can still be lost due to limitations of human motor control. Aircraft must be designed to remain within the limitations of human controllers.
- Category III, APC due to a nonlinear change, can be seen as the result of rapidly and unexpectedly having the cerebellar model fail. Pilots maneuver aggressively and confidently provided their internal model is sufficient for automatic control. If that model is suddenly wrong, the pilot will either recognize that the system has changed and revert to slow and incompetent cognitive control or fail to recognize the change and just allow his autonomous system to make things worse.

Cognitive science has much to bring to the topic of aircraft handling qualities, both in helping engineers craft better aircraft and in helping flight test professionals safely test these aircraft and identify the cause of handling qualities faults.

4. *Bringing Cognitive Science to the USAF TPS*

The TPS has just started examining cognitive science in preparation for bringing it into the curriculum. The current flying qualities curriculum starts with the aircraft as a physical system, and builds from the equations of motion to the consequences of aircraft dynamics to testing aircraft flying qualities. Perhaps handling qualities education must take the same pattern, starting with instructing the students on the basics of cognition and motor control then continuing into the specifics of controlling aircraft. Surely a better understanding of how their bodies—including their brains—work will make pilots better able to observe themselves “in the loop” and provide higher quality information on the capabilities and limitations of the pilot/aircraft system.

IV. On Evaluating the Closed-Loop Pilot/Aircraft System

It is abundantly clear that understanding the aircraft portion of the closed-loop system is mandatory for understanding the entire system. Open loop flying qualities, including generalized characteristics that are likely to result in acceptable closed-loop performance, are well understood relative to the pilot portion of the system. The level of understanding is so much greater for the aircraft than the pilot that this paper will not go beyond asserting that a modern test pilot school must impart some of this understanding to its students as a minimum baseline for understanding closed-loop handling qualities evaluation.

[b]PIO is a subset of APC. The committee members preferred the term “pilot-involved oscillations,” yet another way to create the acronym “PIO.”

If pilot modeling to-date has provided nothing more than an informative analogy of pilot control, how can the TPS move forward to instruct its students in the actual mechanisms of pilot cognition and motor control? The first step is to back away from any particular hypothesis of how the mind works and observe without assumption of cause. How does a pilot's performance change with effort? What does a pilot pay attention to, both consciously and unconsciously? What can we say about PIO that fits pilots' actual PIO experiences? In other words, how do we define "PIO?" How do we get pilots to reduce their tolerance for error and/or increase their aggressiveness so we can see when and how the closed-loop system breaks down? Is there such thing as a "nuisance PIO?" If so, how do we differentiate between a nuisance PIO and a dangerous PIO potential? How do we measure pilot inputs without resorting to assumptions about the source of those inputs? What do we do with industry standard handling qualities FTTs like the Cooper-Harper Rating scale?

In order to advance the discipline of airplane handling qualities evaluation, the flight test community may need to return to observation and let those observations create data in hopes that the data might be used to truly understand how a pilot's inputs become outputs. If "gain" is a multiplier in a linear feedback system then there is no gain in the pilot's brain, and the handling qualities community must break from this analogy if it is to truly understand the pilot/aircraft closed-loop system. Design standards must not be abandoned but the path to improving those standards may lie with an entirely different understanding of the pilot part of the closed-loop system.

A. The Point of Diminishing Returns

The concept of "diminishing returns" finds application in almost every human endeavor from economics to education. It will be no surprise, then, that it has application in aircraft handling qualities. Put most simply, a pilot does better with increased effort until increasing effort produces no improvement or even degrades results. This can be easily illustrated by plotting achieved tracking performance against desired tracking performance as depicted in Figure 1. The in-flight data necessary to support this hypothesis was obtained as part of a TPS student test management project called "BAT DART."[10] The plot depicts several important hypotheses. First, for a given task (including system, task, and environment) there is a limit to a human's ability; an unattainable performance. Second, when the task is easy the pilot may choose to do less well on the tracking task so that more time may be spent on other tasks. If there are additional tasks (whether other tracking tasks, systems operation, etc.) then the pilot will allow their tracking performance to degrade within the requirement so he can attend to other tasks.[c] Third, as the task becomes more difficult, the pilot is forced to concentrate more and more on the task until a point of "single focus" is reached where the pilot has no excess capacity and can barely achieved the desired tracking performance. Finally, the task becomes so difficult that the pilot cannot achieve the desired tracking performance. At this point, the closed-loop system breaks down into a PIO or the pilot simply runs out of sufficient control authority to do any better. Depending on the severity of the PIO, the achieved performance begins to degrade. With a catastrophic PIO, obviously the performance degrades quite rapidly!

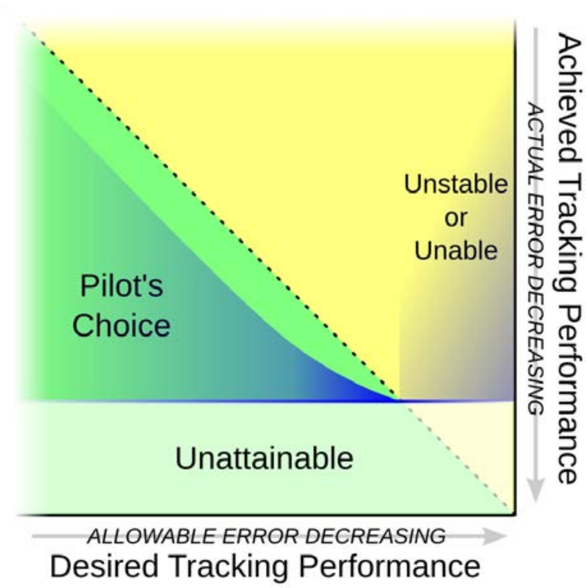


Figure 1: Achieved vs Desired Tracking Performance. (Blue shading corresponds to the ranges of expected pilot tracking performance depending on the task and how the closed-loop system breaks down at high pilot gain.)

Use of this chart has helped TPS student test pilots and engineers understand how they and their aircraft are limited as a complete system. This conceptual framework helps TPS students observe, describe, and predict their performance.

[c]This was clearly shown in another student test management project, "AT BAT."[11]

B. Just What Does a Pilot Want, Anyway?

Flying qualities engineers have always appreciated that pilots are constantly shifting their focus of control. At one moment, the pilot makes an input to control altitude. The next moment, the pilot makes a throttle motion to check an airspeed change. And the next, he turns a dial to set a new frequency. Aside from making aircraft control an unconscious act, the most important thing a pilot learns in training is how to appropriately shift attention across the many items that she must monitor and/or control. When the author was an instructor pilot in the mid-1980's the most common reason for eliminating a student pilot from training was "task saturation," meaning the student simply couldn't learn to manage the required tasks in the time alloted. Task saturation manifested itself in many ways, including poor aircraft control during complex procedural tasks, inability to accomplish procedural tasks during complex maneuvers, and so on. In the end, the student "ran out of brain bytes"[d] and critical tasks were left undone.

A pilot's conscious desires (set that airspeed, maintain this altitude, etc.) are complex and constantly changing. But these obvious tasks also draw our attention, as outside observers, from where their unconscious attention may lie. The pilot's unconscious mind is constantly monitoring and responding to countless inputs, largely unknown to the pilot, that make habituated control possible. Like a child on a bicycle, the pilot is constantly making unobserved and unappreciated control inputs. What does a pilot want? Consciously, the pilot wants to be the best pilot he can by accomplishing all assigned tasks while skillfully and efficiently handling the aircraft. Unconsciously, the pilot's brain is simultaneously closing and integrating multiple control loops to produce the conscious intent of the pilot. The unconscious part of the pilot makes conscious action possible; all are part of the whole and an appreciation of handling qualities requires an appreciation of this whole. That is for the future. For now, we must begin to seriously take apart the problem.

1. To Maintain a Condition: Point Tracking

Prior to 2004 it was assumed that when the pilot was actively controlling the aircraft the pilot was trying to maintain something. That "something" might be pitch attitude, pitch rate, minimizing an error from a moving target, controlling altitude above the ground during landing, or any number of things. In 2004, the author gave this generalized type of tracking the name "point tracking."[12] All pilot models prior to 2004 assumed that the pilot was engaged in point tracking. But in spite of the many levels of complexity within point tracking, a simple story can show that it is not all-inclusive of pilot control strategies (or any person engaged in any control task, for that matter).

2. To Avoid a Condition: Boundary-Avoidance Tracking

Imagine riding a bicycle on a nicely paved country path. Now imagine taking that path and elevating it several stories above the ground, and just for fun, imagine that the ground is littered with nasty spikes. This scenario sets up an analogy routinely used to help pilots understand the concept of "gain." As you, the evil path-maker, make the path narrower and narrower the rider will increase her gain to stay on the path and prevent certain doom. At some point the path will become so narrow and the rider's gain so high that she will create a closed-loop oscillation with the bicycle (a rider-induced oscillation?) and her doom will be sealed.

This is a fine story and seems to illustrate pilot gain quite nicely. But it is missing one very important consideration—or two in this case! Should the rider approach the side of the path, the rider must maneuver to avoid careening into the void. This maneuver, made in relation to the edge of the path, has nothing to do with maintaining a condition, or "point tracking." Controlling the bicycle to avoid the boundary is another type of tracking altogether, "boundary-avoidance tracking," and in this special case it is more specifically called "boundary-escape tracking." The paper that first identified this type of tracking, *Boundary-Escape Tracking: A New Conception of Hazardous PIO*,[12] showed how opposing boundaries (like the two sides of the dangerous path) could create hazardous PIOs. At the 2004 Society of Experimental Test Pilots Symposium where the paper was presented, numerous test pilots excitedly described PIOs to the author that they had been involved in and how boundary avoidance correctly explained what their intent was during the PIO. In the intervening years, boundary avoidance has been recognized as a cause of PIOs by Mitchell and Klyde[13] and the Federal Aviation Administration.[14]

[d] The concept of "brain bytes" as an analogy for a limit to mental processing power is ubiquitous in the US Air Force pilot community. Every military pilot has experienced saturation of their mental faculties. A proper training program ensures that all pilots learn to perceive when they run out of brain bytes.

C. Rethinking PIO

PIOs are a constant source of contention in the flight test community.[e] Ample evidence exists for this in the many definitions of the term, several interpretations of the "I" in "PIO" ("induced," "in-the-loop," "involved," etc.), and numerous methods of classification. There is complete agreement, though, that *some* PIOs are sufficiently hazardous that the flight test community must make the effort to understand them and prevent them when necessary. Understanding and preventing PIO has concentrated primarily on finding ways to predict them and test for them, not on truly *understanding* them. Predicting PIOs using the linear pilot analogy has been helpful but it has not helped the community understand them. Several flight test techniques have been created to find them but these haven't helped our understanding much, either. Assuming that our very incomplete understanding of the human mind will prevent us from a complete understanding of PIO in the foreseeable future, perhaps it is time to just observe them and, using these observations, create a taxonomy of PIOs. A proper taxonomy, or categorization, of PIOs might help us better create flight test techniques that tell us not only whether or not an aircraft/pilot combination can create unintentional oscillations, but also whether or not it matters.[f]

Proper characterization and assessment of PIOs is of critical importance to the flight test community. There is a vast difference between the perception of PIOs by pilots and by non-flying program managers. Non-test pilots see PIOs as indicative of poor piloting and therefore probably not worth worrying about after they are eliminated through sufficient training or practice. Program managers tend to see PIO only as the cause of incidents like the YF-22 mishap. Engineers run the gamut. Given that PIOs are a natural phenomenon and nature doesn't care what we think about them, it is time to drop preconceived notions and let nature show the way.

1. Snakes on a Plane

For a new pilot student at the USAF Test Pilot School, appreciating PIO first requires unlearning their operationally-derived notions of PIO. PIO, to an operational pilot, is something that happens while you are learning to fly an airplane and is much more about the pilot's inexperience or incompetence than it is about a problem with the aircraft. Some of this perception is due to the work of the aircraft designers and test teams and some is due to the simple fact that PIO *is* a common symptom of a pilot learning to fly a complex high-gain task. At the TPS, there is a school of thought that all PIOs are "bad" and require correction to the aircraft design. These two perceptions sit at opposite ends of a broad spectrum.

PIOs are a little like snakes; some are deadly, most are harmless, and to someone unfamiliar with them the safest reaction is fear. Aircraft developers have done such a good job eliminating the deadliest PIOs that the vast majority of operational pilots only experience the harmless ones and routinely learn to handle them. Aircraft testers have been bitten so many times by deadly PIOs that they tend to see any PIO as a significant risk. Neither approach, habituated more than learned, is correct. The correct approach for aircraft developers is to learn to tell the difference so they can identify and eliminate hazardous PIOs while reporting inconsequential PIOs to the users instead of wasting time and money correcting them.

Creating a taxonomy of PIO should help make these distinctions. When is a PIO clearly hazardous? When is a PIO clearly a nuisance? How can they be categorized relative to the task, the environment, and the aircraft? Is it possible to make a specification that doesn't require a test pilot's opinion? How do you test for PIO and PIO hazard? These are all questions that the USAF TPS is trying to answer, starting with simply observing PIO across the entire spectrum through research and through students' eyes.

2. We've Seen Something Like this Before!

The snake analogy is instructive but, aside from a Hollywood script, a bit separated from aircraft. An analogy from the world of manned flight can provide some important lessons on how to handle the intersection of

[e] Even the volume of this paper spent discussing PIOs is a statement about the contentiousness of the issue. If PIOs occupy such a small part of the flight envelope, why do they get all the attention? The best answer for this lies in their place on the spectrum of pilot control. PIOs are one of the most important ways that the pilot/aircraft control loop breaks down. They mark the point where safety, not precision, limits the operational handling qualities of the aircraft. After all if you don't understand how a system fails you don't understand the system.

[f] The creation of a scientifically valid PIO taxonomy is, at the time of this paper's presentation, just the start of an idea. The USAF TPS will be examining the many forms of PIO classification created over the years in order to create a provisional taxonomy and, hopefully, conduct research to fine-tune the taxonomy into a useful tool for categorizing and testing for PIO.

our human weaknesses with highly capable machinery. Humans did not evolve in aerial maneuvers or for aerial maneuvers. For millions of years our ancestors lived in a one-g world where "down" was always down and "up" was always up. Our sensor systems (eyes, inner ear, somatosensory system, etc.) evolved in an environment where the only long-term source of acceleration was Earth's gravity. Our brains evolved to control our bodies in this environment, so if our environment is changed in a way that creates a disconnect between our many ways of perceiving motion, we are left without the necessary tools to compensate and can only learn to adjust. Nothing makes this clearer than spatial disorientation. Learning to safely fly through clouds or at night without a visible horizon is truly about learning to fly in relation to instruments in spite of the confusing signals sent from the many "sensors" scattered through the body. Indeed, trusting "seat of the pants" feel in instrument conditions is often a death sentence. Pilots even give these feelings names, such as "the leans," referring to the intense sense that the aircraft is in a turn (even upside down) in spite of the instruments clearly showing that the wings are level. Training and experience do not make "the leans" go away, they only make them easier to handle.

Spatial disorientation (SD) is defined in AF Manual 11-217V1, *Instrument Flight Rules,*[15] as "an incorrect perception of one's linear and angular position and motion relative to the plane of the earth's surface." It is a significant cause of fatal aircraft accidents, and with a 90% fatality rate for SD-related accidents it is certainly much more significant than PIO! Like PIO, SD is never really a "good thing" (although it is unavoidable thanks to our human instincts and limitations; occasional SD is a necessary condition for manned instrument flight), but it can occasionally be a very bad thing indeed.

Spatial disorientation has numerous causes and effects but the flight medicine community has created an important categorization of them that we can borrow from when considering PIOs. SDs are divided into three "types" depending upon the pilot's perception of them, not upon their cause or effect. The three types are I) Unrecognized, II) Recognized, and III) Incapacitating. These categories are quite useful for describing different PIOs so we shall examine them from a SD and a PIO standpoint. Keep in mind that SD is not optional—all pilots experience it—so the key to safe manned flight in instrument conditions is *not* eliminating SD, it is eliminating unrecognized or incapacitating SD.

1. Type I: Unrecognized

   (a) Unrecognized SD: "...unrecognized spatial disorientation; the pilot is unaware that anything is wrong and controls the aircraft in response to the false sensations of attitude and motion."[15] This is, by far, the cause of most fatal SD accidents. Many aircrew, including a friend of the author's in an F-117 accident, have died completely unaware that their aircraft attitude was not what they thought. Not all Type I SD events result in an accident, though. Any time a pilot looks up from a non-flying task to find that the aircraft attitude doesn't match his expectations, he has experienced Type I SD.

   (b) Unrecognized PIO: As with SD, the unrecognized PIO can be extraordinarily dangerous. Unlike SD, there is no way a pilot will not notice that something is wrong! Typically, the pilot misinterprets the oscillation of the aircraft as a flight control problem and automatically responds by attempting to stop the oscillation with the flight controls. In actuality, it is the pilot's flight control inputs that started the PIO and his intensified control inputs that make it worse. These types of PIOs typically have one of three end points; either 1) the PIO grows until structural failure, departure from controlled flight, or impact with the ground, 2) the pilot recognizes the PIO and gets out of the loop to stop it, or 3) another task forces the pilot out of the loop. If the PIO is sustained until stop-to-stop control inputs are being made, but the resulting oscillations are insufficient to cause a crash, the pilot is likely to recognize that his inputs are the problem and abandon his attempts at controlif a Type I PIO becomes a Type II PIO, recovery becomes far more likely.

2. Type II: Recognized

   (a) Recognized SD: "...recognized disorientation; the pilot is aware that something is wrong..."[15] It is safe to say that there are no IFR-qualified pilots that have not experienced Type II SD. Learning to safely fly in spite of being disoriented is an important part of learning to be an instrument pilot.

   (b) Recognized PIO: Recognized PIOs are a part of flying aircraft. Indeed, learning to perceive that your attempts at controlling the aircraft are causing oscillations is an important part of many flying tasks requiring tight control. Many fledgling test pilots are surprised by the emphasis given PIOs in the flight test community because in the "real world" they are just something you learn to stop or control. (PIO can actually be a very good indicator to a pilot that he is trying too hard!) In-and-of-themselves, recognized PIOs are not hazardous. They certainly reduce tracking accuracy, but as long as the pilot can accept more error, reduce stick motion, or allow the aircraft to return to trim they do not present much safety risk. On the other hand, recognized PIO can severely limit a pilot's ability to accomplish a task. If the pilot cannot fly accurately enough to, say, land then the PIO can create a dangerous lack of accuracy.

3. Type III: Incapacitating

   (a) Incapacitating SD: "...incapacitating spatial disorientation; the pilot knows something is wrong, but the physiological or emotional responses to the disorientation are so great that the pilot is unable to recover the aircraft."[15] We do not have anything approaching complete control over our bodies (including our minds). If SD is severe enough, instinctive or trained reflexes can override conscious control so that a pilot's cognitive awareness of disorientation can be coupled with an inability to stop an incorrect or even dangerous response. The response is coming from the disoriented part of the brain but is so strong that the pilot cannot stop it. This type of disorientation is often perceived as a control system failure discovered after recognizing SD. This type of disorientation carries a high risk of a fatal crash but at least the pilot is aware that something is wrong and might have the ability to eject or eventually regain control. Note that incapacitating SD probably starts as unrecognized SD—the SD is recognized late enough that the physical cues become so strong that the instinctive response cannot be easily overridden. Pilots experience a mild form of this type of SD during spin testing. Following a multiple-turn spin they find that they can't seem to stop trying to roll in the opposite direction of the spin.

   (b) Incapacitating PIO: When can a pilot be in a recognized PIO, yet be unable to recover? There are PIOs in which each oscillation creates a motion that must be corrected and for which the trained "muscle-memory" correction creates an opposing motion that must be corrected... and so on. The pilot may recognize the oscillation as a PIO but he cannot relinquish control to let the aircraft recover because he cannot allow any more deviation. This is the "rock and a hard place" PIO typical of boundary-escape tracking; the pilot knows that he has to stop controlling so aggressively in order to stop the PIO but cannot because allowing the aircraft to continue it's current motion is too hazardous. The only way a pilot can stop this type of PIO may be to "pick his poison." Imagine a PIO in the flare where the aircraft oscillates between averting ground impact and avoiding a pitch-up into departure; if the pilot cannot go around he may be faced with no good choice, but at least one survivable one (provided he has an ejection seat).

The PIO/SD analogy may be more than just an analogy. Both result from a disconnect between conscious perception, unconscious perception, and reality. In both cases, pilots find themselves in the most dangerous situations when their conscious and unconscious perceptions match each other but do not match reality. But test pilots and engineers must not avoid PIOs—they must seek them out. Safely identifying PIOs requires the use of methodologies that prevent their unrecognized inception. Flight testers must also determine if an unrecognized PIO will give the operational pilot time to discern his errors of perception.

V. Flight Test Techniques

Flight test is fundamentally about safety and effectiveness. A savvy reader may ask, "Isn't safety just a subset of effectiveness?" Sure, but it is so important and so easy to lose sight of with time and money constraints that safety must be foremost in the test team's minds. Just as safe and effective test programs examine structural strength, flutter, stalls, departure, and so many other "edges of the envelope," they must examine the full spectrum of closed-loop handling qualities. Are there aircraft dynamics that make required tasks too difficult, or even impossible? Are there non-linearities that will break the unconscious control loops and cause a PIO? Do any PIOs occur in a way that could leave a pilot unaware that she is in a PIO? These are complex questions that test teams must seek to answer. In the world of flight test, there are few quantitative test techniques that are effective at answering these questions. Unfortunately, there isn't even widespread agreement as to what the questions should be or what the test team should be looking for. The most common handling qualities FTT, the Cooper-Harper Rating, is suitable for assessing utility but can leave dangerous high-gain instabilities completely unnoticed.

The USAF TPS, as an accredited academic institution, is not interested in just teaching the start-of-the-art, such as it is. It is vital that TPS improve the science of handling qualities evaluation, treating FTTs not as fixed maneuvers that must be trained, but as *techniques* that may be used to teach its students about aircraft, themselves, how the two interact, and how to objectively measure and observe that interaction. TPS has adopted a three-part strategy to examine aircraft handling qualities; 1) characterize the aircraft with open-loop FTTs, 2) characterize the full spectrum of closed-loop interaction including the implications of any PIOs encountered, and 3) characterize the mission-relevant handling qualities.

A. Getting Within a Reasonable Range of Characteristics: Open Loop FTTs

Many talented engineers have labored to answer the question, "What does a good-handling aircraft act like?" Designers strive to place aircraft inside the resulting specifications and it is important that test teams first labor toward ensuring that the aircraft acts as designed and at least behaves reasonably. It is easy to get caught up in handling qualities evaluation and forget that the "aircraft" part of the closed-loop system is the only part that can be reasonably adjusted. (In spite of the common management solution to fixing an HQ problem, "Can't we just fix it by training the pilots?") To adjust the aircraft, the aircraft must be sufficiently understood. To have any hope of understanding the human part of the closed-loop system, we must understand the aircraft part. TPS handling qualities FTTs and demonstrations are almost always preceded by examining the open-loop aircraft dynamics so that students can begin to appreciate how these characteristics affect their closed-loop performance.

B. Examining the Spectrum of Pilot Input: Workload Buildup

Safe and effective flight test is largely made possible by the concept of "buildup." Whether checking the structural load capacity, flutter margins, or stability at maximum dynamic pressure, test teams almost always use a buildup process. They build up in loads, Mach, dynamic pressure, etc., by checking interim conditions to ensure that the aircraft is matching predictions before moving to the next condition and ultimately the edge of the envelope. The buildup process has saved many aircraft, pilots, and programs.

1. The Workload Buildup FTT

The workload buildup (WLB) FTT was created at the USAF TPS during boundary avoidance theory (BAT) research.[16] During a WLB FTT, the test pilot is given a succession of tracking tasks where the goal is to maintain a given condition completely within clearly defined limits. Figure 2 illustrates how boundaries are used to drive desired tracking performance so achieved performance may be measured. At each pilot gain buildup step the test pilot conducts the tracking task by doing whatever it takes to remain within the designated limits. This test technique may be used to find the most important elements of the pilot tracking performance spectrum for a given task, including

- the best achievable performance,
- the point of highest tracking workload—where "single focus" is necessary, and
- how the system breaks down at high gain due to PIO or insufficient control authority.

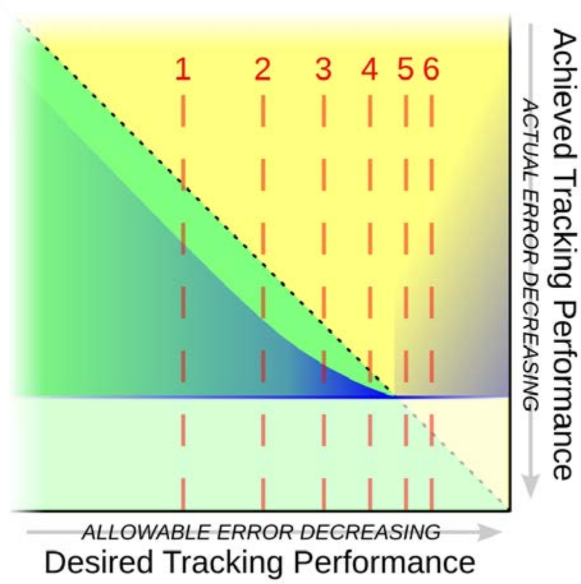
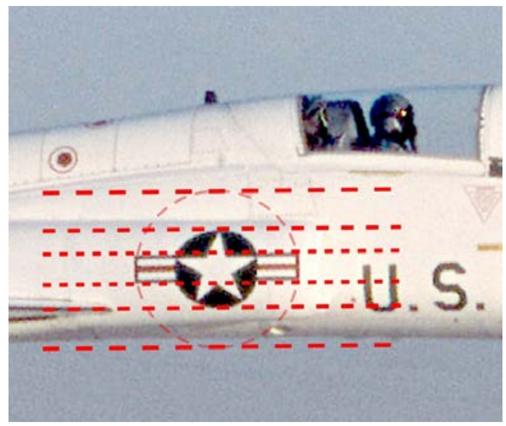


Figure 2: Using Boundaries to Drive Pilot Gain

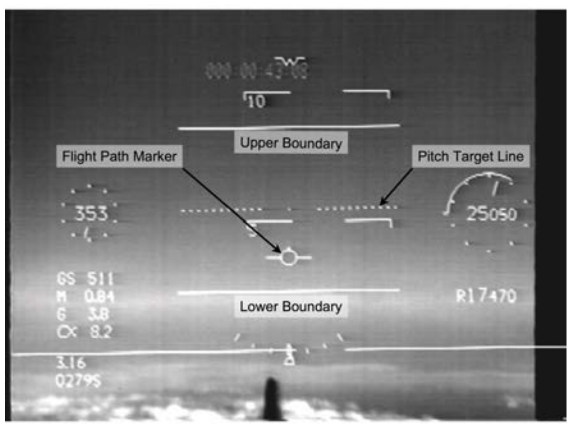
The WLB FTT is currently used by TPS for in-flight handling qualities demonstrations in the T-38, F-16, variable stability NF-16D VISTA, and KC-135R. Several years of use have shown that the technique is easy to teach and easy to use although sometimes it is difficult to get student test pilots to "role-play" as if the artificial boundaries are real. But once they understand the importance of role-playing, the results from student-to-student are surprisingly consistent. One of the most interesting characteristics of the technique is its usefulness for handling qualities research conducted without the benefit of experienced test pilots. Operational pilots and non-pilots can successfully execute the technique simply by doing their best to remain within the set boundaries at each step.[g] CHR evaluations and HQDT FTTs require education and training for the test subjects, the WLB FTT does not.

The WLB FTT does have its disadvantages, chief of which is the amount of time it takes to step through the boundary settings. It shares this disadvantage with buildup processes in general, so just as in other forms of buildup it is probably not suited for situations where a buildup process isn't required for safety or efficacy. The FTT also requires the presentation of boundaries and the ability to change the boundaries in real time. This is relatively easy to do if flight test displays (especially HUDs) are available. If they are not, as is the case with most of the USAF TPS in-flight demonstrations, there are other ways to set boundaries.

(a) T-38 Lateral Displacement Target



(b) T-38 Formation Target



(c) NF-16D Pitch WLB FTT HUD

Figure 3: Some Workload Buildup FTT Boundaries Used at the USAF TPS

Figure 3 depicts some boundaries used for workload buildup FTTs at the USAF TPS. Figure 3a shows how the underside of a T-38 is used for lateral capture tasks with imposed boundaries. This setup has shown good predictive capability for lateral-directional tracking during offset landings. Figure 3b shows the boundaries used for formation pitch (relative altitude) tracking. The actual position of the test aircraft is further aft than shown, so the wingtip can be used along with the depicted boundaries. Finally, Figure 3c depicts the VISTA NF-16D HUD for pitch WLB FTTs. The pilot is tasked with keeping the flight path marker between the boundaries. All of these methods are easily understood by students and very effective for

[g]Techniques for motivating operational pilots to remain within the boundaries have ranged from a steak dinner award for the pilot that does the best to intimating that participants will be publicly rank-ordered following the test. Competition is a powerful motivator for most pilots!

showing them how their tracking changes with increasing task difficulty. Once they have learned to role-play and observe the consequences of increasing gain, learning more complex handling qualities FTTs, such as Phase 2 HQDT, becomes significantly easier.

2. Handling Qualities Stress Testing FTT

Phase 2 HQDT, also known as handling qualities stress testing (HQST) is a second option for closed-loop buildup and PIO assessment. HQST can be accomplished quickly and safely if the test pilot has the right experience and expertise. On the other hand, HQST does a poor job of defining when the aircraft becomes susceptible to PIO and leaves PIO identification, classification, and prevention solely in the test pilot's hands. Test safety is also left to the test pilot because a continuous increase in pilot gain is required instead of an incremental buildup.

The USAF TPS has deemphasized HQST but it retains an important place in the curriculum. If a buildup process in pilot gain is not seen as necessary then this FTT can be used to rapidly check for the presence of PIO at high pilot gain.

C. Can we Identify the Good, the Bad, and the Ugly PIO?

Once an aircraft has shown itself to be capable of producing PIO the implications of the PIO must be ascertained. It is clear that there are a wide variety of PIOs, from nuisances that only occur early in training, to hazardous oscillations that will endanger even the most skilled pilots. We have discussed the importance of determining the taxonomy of PIO. Until we have this, test teams must have some interim guidance on what to look out for when they have identified a PIO.

1. What is a "Good" PIO?

The concept of a necessary or even acceptable PIO tendency is anathema to some handling qualities engineers. Yet many aircraft, and certainly most modern fighter aircraft, can easily be driven into PIO by pilots using unnecessarily high gain inputs. So the existence of PIO tendency doesn't necessarily mean that an aircraft's flying qualities must be corrected. Is it possible to go even further and assert that some PIO tendency may be *necessary* for some handling qualities capabilities? Students at the USAF TPS have long asserted that aircraft designed to the strict standards of some PIO prevention criteria are excessively sluggish—a high price to pay for some missions!

Outside of the world of airplanes, operator-induced oscillations (OIO?) are common and accepted in the learning phase. Almost everyone standing on a Segway for the first time creates a rapid and often unstable back-and-forth oscillation with the center of rotation roughly at the device's handles. Children learning to ride bicycles oscillate quite a bit until they get past the initial learning stage. Skilled users are not immune to the same oscillations if they get their gains too high or encounter hazardous boundaries. Would a Segway or bicycle that wasn't susceptible to OIO be maneuverable enough for the sidewalk or the street?

2. The Switching-Induced Simulated PIO FTT

With two proven techniques for PIO identification, Phase 2 HQDT and the workload buildup FTT, the TPS was in need of a technique to determine the criticality of an identified PIO. An "intentional" PIO may be created by using a series of constant amplitude step inputs, called a switching-induced simulated PIO (SISPIO). The pilot starts with a small displacement from a desired point tracking condition, then makes a small step input to return toward the center. The input is held until the center is crossed then the input is reversed to the other side of trim. The frequency of the pseudo square-wave input is entirely determined by the rate at which the aircraft returns to center and the pilot makes no effort to compensate. The data from this FTT is very simple. Does the oscillation grow when driven by a constant amplitude simulated PIO input? If the answer is "no" then a unrecognized PIO is less likely to become dangerous. If the answer is "yes" then the growing oscillations would force a frightened pilot to increase the size of his inputs and the PIO would probably rapidly diverge.

The SISPIO FTT may also be an effective way to ensure that there are no lurking non-linearities. If the test pilot is being as aggressive as possible while switching from one polarity of control position to the other, then the flight control system is likely to exhibit nonlinearities either in the data or in the pilot's perceptions.

Much work remains to perfect this technique. Does the size of the pilot's inputs matter? If it does, should they be increased in a build-up fashion? How much growth in the aircraft response is too much? At the "safe" end of the spectrum, the technique is quite effective—if the SISPIO oscillations immediately stabilize, as they do with T-38 and F-16 fixed gunsight tracking of a target, any PIO tendency seen in normal aircraft operation has no history of causing safety problems. At the other end of the spectrum, if the oscillations rapidly grow as they do with the T-38 pitch axis in close formation at high speed, operational experience has shown that these PIOs can be problematic. (In this case causing occasional over-g during T-38 pilot training.)

D. Characterizing Pilot Inputs

While measuring a pilot's inputs to the control system is simple, analyzing those inputs is currently something of an art. There are only two pilot output parameters that can be measured; control position and applied control force. These parameters are related through control dynamics, with the control position typically resulting from control force. Modern fly-by-wire aircraft make this relationship relatively simple because there is no aerodynamic feedback to the stick through the control system.

Handling qualities data is most often based on control force. Thanks to controller damping (an important condition, it seems, for good handling qualities) the pilot's force inputs are filtered through the control dynamics into the control position, so measuring position provides a smaller range of frequencies to analyze—and less information. Pilots perceive stick force and position and their assessment of workload is based upon both. Yet the equilibrium-point hypothesis of motor control suggests that muscles are commanded to a specific position, not to a specific force, so control force may be more of a by-product of control than a measurement of intent. Nevertheless, characterizing a pilot's control inputs probably cannot start by ignoring either control position or force.

1. Don't Say "Bandwidth" in a Pilot Bar

The concept of "gain" is central to handling qualities evaluation and universally used by flight test engineers and test pilots. Its origins were in linear pilot modeling where it referred to a numerical value in the approximation of the pilot's feedback. But is it universally understood? When two engineers are discussing gain do they envision the same thing, or has the word just become a shorter way of saying "level of effort and/or intolerance for error?"

When attempting to measure pilot gain, it is important to recognize that there probably isn't any parameter in a pilot's brain that could be directly translated as "gain." Neural cognition and muscular motor control bare little resemblance to the linear pilot analogy, so when measuring pilot inputs and attempting to characterize them it is probably not helpful to refer to any given parameter as "gain." Nevertheless, this is a standard approach.

One of the most common methods of measuring pilot gain is through bandwidth analysis. By taking the pilot's control force inputs and plotting them as a power-spectral density (PSD) against frequency one can measure the frequency bandwidth of the pilot's inputs. Bandwidth analysis assumes that as the pilot's gain increases the bandwidth increases. But this assumption is only true in a limited way and for limited conditions; it does not take into account other sources of pilot effort (like stick position) and it filters out information from the original pilot input trace by running the data through a Fourier transform. Telling a pilot that his gain has increased because you have measured that the rate and/or rapidity of his stick movements has increased strikes most pilots as blindingly obvious. Although an engineer can use changes in the PSD to get a feel for changes in pilot gain, the data gathered is remarkably bereft of useful information. How did the pilot's gain change? Did the pilot change tracking strategy? Many questions are left unanswered. Is bandwidth analysis an oversimplification of filtered data?

The most problematic feature of bandwidth analysis is that it is seems to fail when PIO is encountered—the very time when correct measurement of pilot gain is most critical. When a pilot enters a full-blown PIO the PSD of her inputs collapses around a single frequency, the PIO frequency. She will report being at very high gain, yet the most common method of measuring her gain has completely broken down.

Several years ago the author, in collaboration with Mr. Dave Vanhoy (at the time the Flying Qualities Branch Master Instructor, now the School's Technical Director), developed an outline for an alternative method of measuring pilot gain. The method abandons frequency-domain analysis in favor of parameterizing pilot inputs and measuring objective correlates to pilots' subjective assessments of their gain.[16]

2. How often, and how hard, is the pilot working? "Duty Cycle" and "Aggressiveness"

Measuring the amount of effort that a pilot is putting into controlling an aircraft—his "gain"—is not a simple task. It is often estimated through rating scales, where the pilot describes effort in terms of workload and compensation, but numerical measurement has historically been something of a black art. More importantly to the USAF TPS, current numerical techniques, such as PSD analysis, serve little to help pilots understand how they fly aircraft. To alleviate any possible confusion that the following methods have anything to do with frequency analysis or pilot gain as the mathematical entity K_p, the term "inceptor workload" is used to describe how workload is measured as a combination of the independent variables "duty cycle" and "aggressiveness."

Duty Cycle When a pilot is involved in a tracking task, even a difficult one, he is unlikely to be constantly moving the controls. He will occasionally stop changing the position of the control to allow the aircraft to respond by itself; sometimes to let the aircraft finish the job, sometimes to allow his input time to take effect, and sometimes because an input simply isn't necessary. "Duty cycle"[h] is nothing more than the percentage of time the pilot is changing his input on the stick, whether through force or position. (In data analysis, the inceptor is assumed steady when its rate of motion is below a threshold determined by observation of the data.) In terms of workload, it is clear that as duty cycle is increased, pilot inceptor workload is increased as well. Duty cycle fully describes the time the pilot spends with the inceptor held nearly motionless, so there is nothing more to be gained out of examining the inceptor down-time. All that remains is to start the process of characterizing how the pilot moves the inceptor.

Aggressiveness When the pilot is moving the inceptor, the movement can be characterized any number of ways. The characterization the USAF TPS needed was a simple first approximation that captured the effort the pilot is putting into the motion. There is no attempt in this method to explain why the motions are occurring when they are or why they have the characteristics they have. Several methods for calculating average aggressiveness have been tried, 1) the root-mean-squared average of the stick displacement rate, 2) the root-mean-squared average of the stick force rate, and 3) the total work the pilot is applying to the stick (displacement change times the force applied). The third method has proven the most useful, not surprisingly since with proper calculation it becomes a measurement of the "work" of the pilot's inputs. Power (work per unit time) can be found by dividing the total work by the time taken for the work.

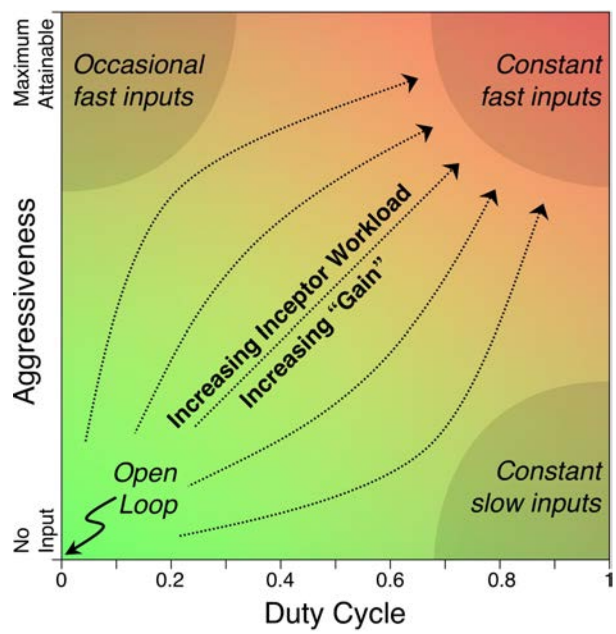


Figure 4: Achieved vs Desired Tracking Performance

The Two-Dimensional Picture of Pilot Inceptor Workload With pilot inceptor workload divided into two theoretically independent factors, it is convenient to plot them on a two-dimensional chart, with duty cycle on the x axis and aggressiveness on the y axis. This chart is depicted as Figure 4 and includes several important generalizations based upon the author's experience using the workload buildup FTT as an instructional tool and data from research conducted after the first concept of this measurement technique was created. What can this figure tell us? First, as the measured workload for a given task moves away from the origin, it can be said that the pilot's workload is increasing. Hypothetically, this should correspond to increasing pilot gain. Second, the proportion of duty cycle to aggressiveness tells us something about the aircraft. For instance, when duty cycle is low and aggressiveness is high (the upper left-hand part of the figure) the pilot's inputs are consistent with lead compensation, where an input is

[h]The term "duty cycle" has a strict technical meaning that does not quite align with the way it is being used here. Over time, the name for this parameter will likely be changed to "duty factor." Nevertheless, the meaning remains the same!

made and the system allowed to respond (often in order to minimize the effects of a response delay). Third, the upper right-hand corner, where duty cycle is 1.0 and aggressiveness is maximized, corresponds to the worst possible PIO—stop-to-stop at maximum effort.[i] Thus the pilot inceptor workload may be used to show the change in pilot inceptor workload (or "gain" in the shorthand sense) and be used to compare the workload between different pilots and different attempts at the same task. Comparing inceptor workload across aircraft and tasks will require a method to normalize aggressiveness—an effort not yet attempted.

Scientific measurement and observation of the closed-loop pilot/aircraft system can answer many questions about that system's stability and safety. But open-loop FTTs and gain buildup testing cannot answer the question, "Do the aircraft handling qualities enable the pilot complete his mission?" For that question, the only significant FTT is the Cooper-Harper handling qualities evaluation.

E. The Cooper-Harper Rating Scale

The presentation of this paper will nearly coincide with the 40th anniversary of the Cooper-Harper Rating Scale.[1] There is no more important or more widely used FTT in aircraft handling qualities testing. But instead of providing hard data the scale is, in essence, a survey. The resulting data is one-dimensional, placing the handling qualities of the airplane for the given task on a scale from 1 ("Excellent/Highly desirable") to 10 ("Control will be lost during some portion of required operation"). For readers not yet intimately familiar with the scale, it is depicted in Figure 5.

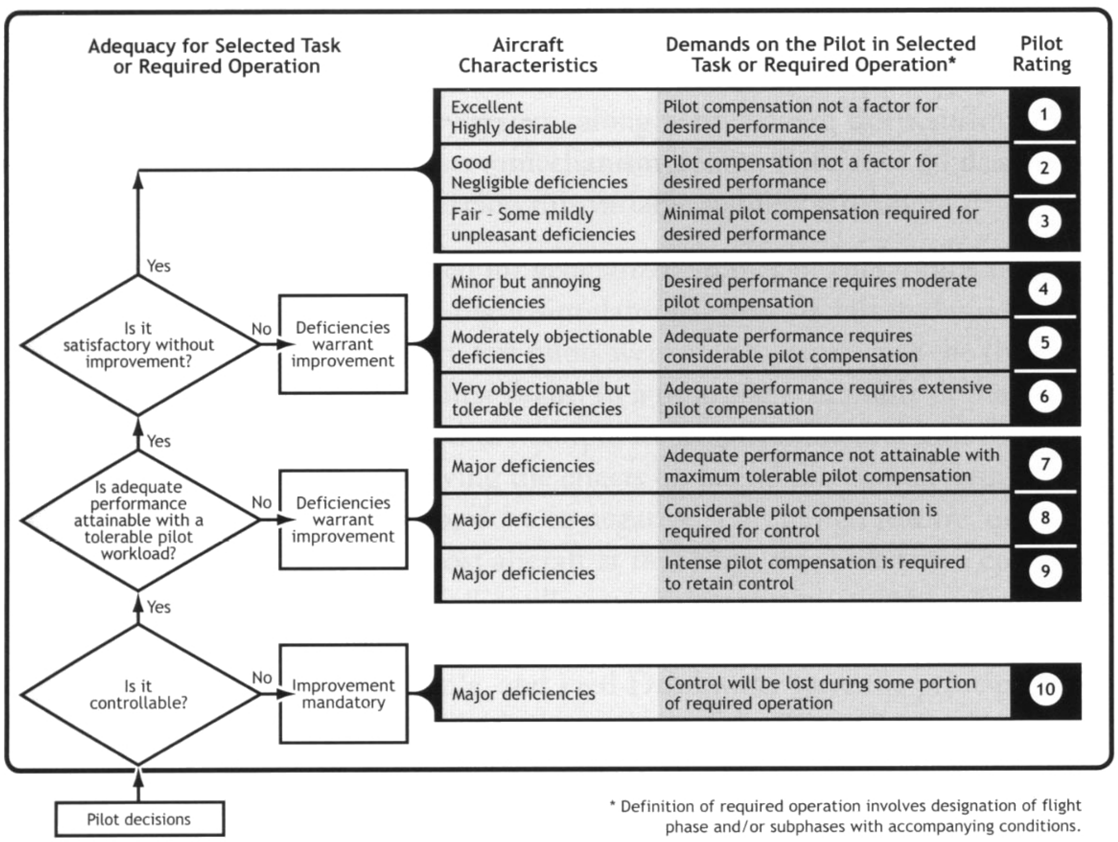


Figure 5: The Cooper-Harper Rating Scale

[i]Holding the inceptor at the physical stop is counted as moving the inceptor for the purpose of duty cycle, and the inceptor rate at impact is used to calculate aggressiveness; after all, the pilot only stopped at that point because of the physical limit!

Acceptance of the scale in the flight test community has been essentially universal, yet some find room for improvement. Payne and Harris[17] find four difficulties with the CHR scale.

1. "...the Cooper-Harper Scale can be criticized for lacking diagnosticity. Aircraft handling qualities are complex, but this complexity is not reflected in the scale." They go on to restate the importance of pilot comments but also point out (correctly) that pilot comments are often neglected—the numerical ratings are given more importance than the comments. The scale may be more about eliciting comments in a controlled manner than about determining a numerical rating. The numerical rating is important, but it is more of a summary than a complete result.

2. "This lack of diagnosticity is related to a lack of content validity. For any scale with psychometric properties to have content validity, the items in it must evaluate all major aspects of the domain to be assessed."[j] A single-dimensional scale cannot hope achieve content validity over a complex subject. The content validity (such as it is) for the CHR scale is obtained through pilot comments; the pilot is responsible for evaluating all major aspects of the handling qualities task, not just providing a numerical rating. The education and training of the evaluation pilot is a critical factor for achieving valid results.

3. "There are also some questions about the reliability of the Cooper-Harper Scale. Although the scale is widely used, there have been few studies that have explicitly evaluated the intrarater and interrater reliability of the scale." Payne and Harris specifically cite the results of work by Wilson and Riley[18] showing high levels of variability for tests conducted at their facility. There is much evidence that sufficient training and discipline can reduce variability but variability remains a problem even with an expert test team.

4. "...the Cooper-Harper Scale may be criticized as lacking in sensitivity." This criticism is not significantly different from the first, but the authors concede that increasing sensitivity increases pilot workload and the time for each evaluation, and that the extra data might not be worth the cost of increased sensitivity.

The strength of the Cooper-Harper Rating scale is not likely to be found in its diagnosticity, content validity, reliability, or sensitivity. Its strength is in its *efficiency.* With proper expertise and training, a test team can use the scale to create and execute mission-relevant evaluations in minimum time and with minimum risk. But the CHR scale is deceptively simple. It looks quite easy to use but if the user lacks the necessary expertise, it is unlikely to produce useful data. The USAF TPS is committed to graduating pilots and engineers that can step into a test team ready to create and execute high-quality CHR evaluations.

1. The Importance of CHR Education and Training

Proper education and training in the use of the Cooper-Harper Rating Scale is vital for the FTT to produce useful data. TPS student education starts with airplane flying qualities and the basics of handling qualities. Students are given basic training in the scale's use in a simulator[19] then apply it during several in-flight handling qualities demonstrations. Ideally, students graduate understanding the following key issues on the use of the scale.

THE SCALE DOESN'T RATE AIRPLANES, IT RATES AN AIRPLANE FOR A GIVEN TASK. It is not uncommon to hear test pilots compare the CHRs of different aircraft. But noting that airplane X has a better CHR in close formation than airplane Y without confirming the conditions of the task makes little sense. Cooper-Harper ratings require the strict definition of a task, the desired and adequate performance required of the pilot/aircraft system in the task, the flight environment in which the task must be performed, and the expected expertise of the pilot. Demonstrating this to USAF TPS students is an important part of their CHR education. It is easy to make a perfectly acceptable aircraft look bad (or a bad aircraft look good) when improperly using the CHR scale.

- Task definition is vital; it must conform to the required performance of the aircraft in operational use. Test pilots with applicable operation experience must be involved in CHR planning so the task

[j] "Content validity" is a technical term that refers to the degree to which a measurement methodology encompasses the target subject. A test without content validity would leave important questions unasked.

definition is neither too difficult nor too easy, either in the definition of the desired and adequate criteria or the definition of the maneuvers to be performed. For instance, in a close formation task a poorly defined set of maneuvers for the lead aircraft can lead to inconsistencies in ratings caused by different lead aircraft pilots being more or less aggressive in their maneuvering.

- The environment must be realistic. Conducting offset landing tasks in a strong crosswind will add scatter to ratings and probably shift the median rating unfavorably. On the other hand, these tests might be necessary, so the test team would be gathering ratings in different conditions to ensure the aircraft meets requirements in all conditions.

- The expertise of the pilot must be considered. CHRs tend to improve with pilot practice; do the first ratings count or the last ratings? The test team must know the answer to this question before they start testing. Along these lines, students are discouraged from conducting CHR FTTs during one-flight qualitative aircraft evaluations; it is extremely unlikely that they will have sufficient familiarity with the aircraft or the mission to produce a credible rating.

THE RATER MUST UNDERSTANDING THE MEANING OF THE WORDS IN THE SCALE. Without a clear understanding of the meanings of terms like "tolerable pilot workload," "extensive pilot compensation," or "control will be lost" a test pilot is unlikely to produce reliable CHRs.

IT'S AN ORDINAL, NOT NUMERICAL, SCALE It is likely that much confusion may have been avoided if Bob Harper and George Cooper had chosen non-numeric ratings vice the 1-10 ratings chosen. An alphanumeric rating scale, say S1, S2, S3 ("S" for satisfactory), M1, M2, M3 ("M" for marginal), U1, U2, U3 ("U" for unsatisfactory), and I ("I" for impossible) would have discouraged inappropriate attempts at statistical analysis of CHRs. The relationship of the ratings is in no way linear—one may only say that they represent an ordering. Mitchell and Aponso illustrated this nicely, showing how the spacing between CHRs is neither evenly distributed nor constant.[20] Students at the USAF TPS are *strongly* discouraged from giving decimal ratings—even the commonly accepted half-ratings—because these ratings imply a confidence that simply cannot exist given the highly subjective nature of nearly every element of a CHR. Overattention to the exact rating also drives attention from the most important result of a CHR evaluation—the pilot's comments!

THE CHR SCALE OCCUPIES ONE DIMENSION, NOT TWO. USAF TPS students frequently try to simplify their in-flight CHR data acquisition by skipping the scale altogether and just answering two questions:

- What was the performance? (i.e. Desired/Adequate/Neither)

- What was the pilot's level of compensation? (i.e. None/Minimal/Moderate/Considerable/Extensive/ Maximum Tolerable/Intense)

These questions imply that the scale has two dimensions, performance and compensation, but this perception is incorrect. The CHR scale has but one dimension. The problem with these students' imaginative methodology can be seen in a simple question, "What rating corresponds to intense compensation achieving desired performance?" Using the scale as intended would immediately result in a Level III rating (CHR7-CHR9) because the pilot workload should probably not be assessed as "tolerable." One may also imagine a lazy pilot, unwilling to work hard enough for adequate criteria, reporting that adequate performance was not achieved but compensation was minimal. These examples provide insight into one of the most important reasons for proper test pilot preparation; *skilled CHR evaluators adjust their performance and workload during the task so they complete the task with a performance/compensation mix that will fall neatly into the scale.* This explains why experienced CH raters often forgo carefully reading through the scale. They know the scale and experiment during the test to find where the aircraft, in the task, fits on the scale. Experienced handling qualities test pilots essentially choose the rating while they fly the test. This is probably anathema to many CHR practitioners but the simplicity of the scale allows for this technique and the structure of the scale demands it.

RATINGS IMPROVE WITH ATTEMPTS; THIS MUST BE TAKEN INTO ACCOUNT. That ratings improve with practice is immediately obvious to TPS students and can be a source of confusion. Some students seem disappointed with this, because it puts into doubt the objectivity they assumed that the scale provides. The CHR method is, of course, subjective but the nature of the questionnaire ensures that the evaluator take into consideration the most important elements of handling qualities evaluation and sets a standard language and methodology so that the subjective assessments might be reasonably compared against a standard. Students learn that their ratings will improve with practice, and they learn that pilots or non-pilots flying difficult or unfamiliar tasks (like close formation or landing) will almost always rate the aircraft quite poorly while their more experienced classmates are giving Level I (CHR1-CHR3) ratings. This just brings home the point that CHR ratings are about the pilot, the airplane, and the task. Students are encouraged to report in their comments if they perceive that they can improve with practice or if they believe that their lack of experience resulted in a rating that does not reflect what could be expected from a properly trained pilot.

A PIO during a Cooper-Harper evaluation requires careful consideration and comments. It is not uncommon for students to encounter pilot-induced oscillations during a CH evaluation. These PIOs may be classified in three general categories.

1. An unstable PIO that requires abandoning the task altogether to maintain aircraft control. These PIOs clearly require an "uncontrollable" (CHR10) rating, not because the aircraft is uncontrollable in a larger sense—after all, the pilot had control of the aircraft prior to the task—but because the aircraft is uncontrollable in the task. It is also likely that this PIO will cause problems in other parts of the envelope.

2. An unstable PIO that requires abandoning some level of performance to maintain aircraft control. This type of PIO is often seen in close formation flying, where a student that attempts to meet the desired criteria is driven to PIO but, by reducing gain and accepting adequate performance, can stop the PIO and continue the maneuver. In these cases, students are encouraged to treat the PIO as a deficiency that increased their level of compensation. These PIOs are less likely to be a safety risk but must be evaluated for that possibility.

3. An incipient or minor PIO that does not prevent them from achieving desired performance. These oscillations may not even prevent a Level I rating, but they must be reported in the pilot's comments to ensure that the tendency receives the necessary attention. There are too many examples of aircraft that did well during CHR tests only to fall victim to PIO. Any observed PIO tendency must be investigated to ensure that it is either inconsequential or corrected.

A Cooper-Harper evaluation leaves important questions unanswered. Cooper-Harper evaluations are an important part of the flight tester's handling qualities evaluation repertoire, but they can leave important questions unanswered. If the CH evaluations never challenged the pilots enough to stress the closed-loop system, they can leave dangerous PIOs completely hidden. They are excellent for tracking performance specification compliance assessment but they are not suitable for handling qualities stress testing, thus they can leave the question "is it safe?" unanswered and hidden behind a positive "is it effective?" result.

2. Is the CHR the Siné Qua Non *of Handling Qualities Evaluation?*

The Cooper-Harper rating scale is so pervasive in aircraft handling qualities testing that it is tempting to believe that it cannot be improved upon. Whether or not this is true almost doesn't matter. It is so pervasive and so ingrained in handling qualities evaluation that it may not be possible to propagate changes in the methodology even if they are clearly improvements. The method is essentially a survey that provides useful data when properly conducted yet says almost nothing about the inner workings of the closed-loop system. Perhaps the scale will remain the primary method of conducting HQ evaluations until we can model the human controller well enough to make closed-loop handling qualities FTTs as pilot-neutral as open-loop handling qualities FTTs.

George Cooper and Bob Harper's handling qualities rating scale is an interim solution awaiting our better understanding of human cognition and motor control. Flight test professionals must not let a useful approximation hinder efforts toward understanding the true nature of ourselves and our interactions with our machines. The USAF Test Pilot School hopes to be at the leading edge of incorporating improved understanding of the human mind and body into the evaluation of aircraft handling qualities.

VI. The Future of Handling Qualities Evaluation at the USAF TPS

The USAF Test Pilot School is leveraging several important recent changes to improve handling qualities instruction and evaluation. First, TPS has the strongest motivation to understand handling qualities and improve evaluation techniques; its students form the front line for many critical flight test efforts both in the USAF and out. Second, the recent addition of civilian flight test professionals, both flight test engineers and experienced test pilots, has created a base of expertise with the longevity to carefully incorporate significant curriculum changes. Third, the recent accreditation of the school as a Master of Science Degree-granting institution has encouraged the creation of an academic framework that can provide stability in the rapidly fluctuating world of US Air Force systems acquisitions and constant military leadership turnover. Fourth, TPS has world-class simulation facilities including the variable-stability NF-16D VISTA and an indigenously-created reconfigurable simulator. Finally, the School has a long tradition of conducting handling qualities research projects as part of its students' academic experience. To provide the best possible instruction, the staff of the USAF TPS must have a strong foundation of knowledge. Creating this foundation is an ongoing process that will continue aggressively and indefinitely.

VII. Acknowledgments

The author appreciates the outstanding staff and extraordinarily talented students of the USAF TPS School. The former demonstrates day-after-day their commitment to excellence in the conduct and instruction of all facets of flight test and the latter provides a constantly-renewing source of curiosity, excitement, and fresh ideas.

References

[1]Cooper, G. E. and Harper, Jr., R. P., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," Tech. Rep. NASA TN D-5153, Cornell Aeronautical Laboratory, 1969.

[2]US Air Force Test Pilot School, *Flying Qualities Curriculum*, 1991.

[3]US Air Force Flight Test Center, *Flying Qualities Testing*, February 2002.

[4]Prosser, K. E. and Thurling, A. A., "Handling Qualities Stress Testing," *44th Society of Experimental Test Pilots Symposium Proceedings*, 2000.

[5]Wilson, R. A. and Keil, F. C., editors, *The MIT Encyclopedia of the Cognitive Sciences*, The MIT Press, 2001.

[6]Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A. C., and Farne, A., "Tool-use induces morphological updating of the body schema," *Current Biology*, Vol. 19, 2009, pp. 1157.

[7]Yong, E., "Brain treats tools as temporary body parts," http://scienceblogs.com/notrocketscience/2009/06/brain_treats_tools_as_temporary_body_parts.php, June 2009.

[8]Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Putz, B., Yoshioka, T., and Kawato, M., "Human cerebellar activity reflecting an acquired internal model of a new tool," *Nature*, Vol. 403, 2000, pp. 192–195.

[9]Committee on the Effects of Aircraft-Pilot Coupling on Flight Safety, *Aviation Safety and Pilot Control*, National Academy Press, 1997.

[10]Dotter, J. D., "An Investigation into Pilot Performance and PIOs as Influenced by Boundaries," Tech. rep., United States Air Force Flight Test Center, Edwards AFB, CA, 2006.

[11]Blake, R., Lee, C. H., Keithley, C., Baccheschi, N., and Cooley, G., "Limited Investigation of Auxiliary Tasking during Boundary Avoidance Tracking (AT BAT)," will be published as AFFTC-TIM-08-09.

[12]Gray, III, W. R., "Boundary-Escape Tracking: A New Conception of Hazardous PIO," *48th Society of Experimental Test Pilots Symposium Proceedings*, 2004.

[13]Mitchell, D. G. and Klyde, D. H., "Recommended Practices for Exposing Pilot Induced Oscillations or Tendencies in the Development Process," Tech. rep., Systems Technology, Inc., 2004.

[14]Arnold, E., "PIO Testing of Transport Category Aircraft," Society of Experimental Test Pilots European Symposium, June 2009, conference presentation.

[15]U.S. Air Force, *Instrument Flight Procedures*, af manual 11-217v1 ed.

[16]Gray, III, W. R., "A Boundary Avoidance Tracking Flight Test Technique for Performance and Workload Assessment," *51st Society of Experimental Test Pilots Symposium Proceedings*, 2007.

[17]Payne, K. H. and Harris, D., "A Psychometric Approach to the Development of a Multidimensional Scale to Assess Aircraft Handling Qualities," *The International Journal of Aviation Psychology*, Vol. 10, 2000, pp. 343–362.

[18]Riley, D. R. and Wilson, D. J., "Cooper-Harper Pilot Rating Variability," *AIAA Atmospheric Flight Mechanics Conference*, 1989.

[19]Gray, III, W. R. and Kemper, J. C., "Creating an Aircraft Handling Qualities Simulator for the USAF Test Pilot School," *U.S. Air Force T&E Days*, 2009.

[20]Mitchell, D. G. and Aponso, B. L., "Reassessment and extensions of pilot ratings with new data," *AIAA Atmospheric Flight Mechanics Conference*, 1990.