Chapter 5 – NVG IN-FLIGHT RESEARCH

Gregory Craig, Todd Macuda, Sion Jennings
National Research Council of Canada
Ottawa, Ontario
CANADA

greg.craig@nrc-cnrc.gc.ca

5.1 FLIGHT TESTING NIGHT VISION GOGGLES

The following section presents experimental flight test methods and approaches for evaluating NVGs. In the previous section, the techniques were geared toward assessing the impact of the goggles on operational flight tasks to determine if there are restrictions or limitations on how/when the goggles are used in flight. Typically, the tests outlined in the previous section would be done to determine the acceptability of the NVGs and the aircraft for conducting future NVG operations or research. In contrast, the current section presents methods which can allow the researcher to compare in detail goggle effectiveness, compare between different goggle systems, examine effects of different types of aircraft, cultural or environmental lighting, and different weather effects, for example. The following sections describe pilot/subject selection, the different categories of NVG flight testing that might occur, objective and subjective performance measurements and some of the considerations in using each type of measurement.

5.1.1 Pilot Selection

The current section examines the basic issues associated with identifying the pilot population for the NVG research to be conducted. It is important to select pilots that are representative of the final user population of the NVG or the symbology. For example, when examining the ease of use of symbology, the researcher must consider both the pilot who has hundreds of hours of NVG flying and the pilot who has only just qualified on NVG flying. Pilot currency should also be considered in the selection of research subjects to ensure that the pilots have flown NVGs recently, or can be given additional practice flying with the goggles in the case of a non-current pilot. While qualified test pilots often provide invaluable feedback on the aspects of the system being tested, they will invariably compensate for the deficiencies in the system in ways that might be difficult for non-test pilots to match. If the time and resources are available one may wish to test both an expert and a group representative of the final users. From the test pilots, one may gain a better sense of the underlying system problems, while the end users will provide a sense of potential training issues and issues that may arise only in the operational environment.

5.1.2 Basic Categories of Flight Testing

The following sections detail the different broad categories of research flight testing including operational flight tests, developmental flight tests and certification flight tests.

5.1.2.1 Evaluating the Operational Context

As flight tests are designed, it is important to approach the flight test with a considered understanding of the operational environment. By utilizing representative pilot input as above, evaluators will be able to examine NVGs from an operational context rather than a strict “bench” science approach. This will be useful in understanding system usability in real missions. For example, NVG performance will change as function of operational context (e.g. urban to non-urban lighting conditions, changes in terrain, etc.). Flight test scientists and evaluation pilots will need to consider the impact of a dynamic flight environment on NVG function when assessing new and mature technologies in various flight test environments.

For example, in keeping with laboratory methods as described above, flight tests will need to vary environmental context (e.g. lighting, terrain features, weather) to ensure that testing is representative of operational requirements. In addition, when testing specific NVG capabilities, flight tests should consider the Concept of Operations (CONOPs) in developing studies. It is important to note that any methodological approach will have to mesh applied operational science with theoretical approaches towards studying the visual system in order to establish a full understanding of how NVGs influence pilot vision and flight performance. Finally the authors would like to stress that all NVG test scientists should be well versed with operational implementation of NVGs and should be participating in flight tests and flying operations on various airborne platforms where possible.

The approach toward operational flight testing is typically quasi-experimental utilizing domain expertise from flight test scientists and evaluation pilots as described above. Evaluation of NVGs involves incorporation of laboratory and flight test engineering evaluations (e.g. see summary in Chapter 4). Once data is available from these well controlled environments it will be necessary to flight test NVGs in a real operational environment (Rash et al. 1996; 2000). For example, NRC recently conducted an operational evaluation for the Ontario Ministry of Natural Resources to examine goggle performance in highly incompatible lighting conditions (Jennings et al., 2006). OMNR wanted to examine the feasibility of conducting operations around forest fires at night to perform personnel extractions and scan for hot spots. These are not typical applications for NVG systems and so NRC conducted a preliminary examination of the viability of these operations. The preliminary flight test involved flying around/near controlled burns, ranging in size from $1m^2$ to $5m^2$. Based on the preliminary flight test, an initial set of strategies for flying around the fire were developed and disclosed to OMNR along with a recommendation for additional phased testing around actual forest fires to fine tune the strategies.

5.1.2.2 Development Tests

As suggested by the name, development tests examine new equipment, new concepts of operation for the equipment, or new symbology. Typically the research will compare the new system or concept against the current standard to determine if the new concept is worth pursuing. Specifically, does the new system increase the safety or the capability of the NVG operation(s) enough to warrant the cost involved in changing the system. For example, one may be tasked to compare a new version of image intensifiers to determine if it is worth upgrading to the new tubes. A small increase in the dynamic range or acuity of the goggles may not be worth the increased cost of the new tubes. Significant increases in the field of view (without resolution or acuity loss), visual acuity through the goggles, or performance under a wide range of illumination conditions may increase capability sufficiently to warrant adopting new goggle technology. Significant decreases in the weight of the tubes or optics without appreciable degradations to goggle performance may also merit replacing the NVGs.

Simple subjective evaluation flights may be adequate to address the question of improvements in goggle performance, particularly when there are obvious performance differences between the old and new technologies. However, conducting tests under a wide variety of illumination/weather conditions may not be feasible if there are time constraints involved. In such cases, laboratory or simulation evaluations may compliment the flight test(s). In some cases, the performance difference(s) between old and new goggles or symbology sets needs to be quantified. When this is the case, a structured series of laboratory and flight tests will be needed, using the flight tests suggested by AETE (see summary in Chapter 4). The operational requirements of the organization may serve to expand or narrow the range of tests performed. Similarly, an initial series of laboratory tests may be useful in pointing to aspects of the system that need to be examined further in flight. Likewise, the flight test may reveal aspects of the goggles or the symbology that need to be addressed in further detail in the lab (where the tests can be performed at lower cost and lower risk).

As an example of a developmental flight test assessment of symbology for NVGs, QinetiQ and NRC conducted flight tests of new NVG symbology in 2002 – 2005 (Craig, Jennings & Thorndycraft, 2003;

Craig & Thorndycraft, 2004; Jennings, Craig, Carignan, Ellis & Thorndycraft, 2005; Thorndycraft, Jennings & Craig, 2005). In each of these flight tests, the manoeuvres selected for the symbology evaluation were suited to the maturity of the symbology being tested. In the earlier tests, the symbology was being tested to determine how it would function on the aircraft, which called for concise and controlled flight test tasks such as is the case with racetrack and the ADS-33 manoeuvres (Craig et al., 1993). The racetrack manoeuvre consisted of a number of elements (lateral hover taxi, hovering, bob-up, etc.) which the pilot performed serially. The overall score for the racetrack was determined by the speed and accuracy of the pilot's performance of the manoeuvre. As the design of the symbology and the underlying aircraft data systems improved, the percentage of ADS-33 tasks decreased and flight tasks representative of operational missions played a larger part in the evaluation (Thorndycraft et al., 2005). This incremental approach is an important aspect to developing any new system for flight test. Generally developmental tests would start with basic elements of the final mission(s) to determine the basic usability of the system. As the system improves, the complexity of the flight tasks can increase to represent more realistic or operational flight tasks.

Another reason for the gradual (staged) approach to the tests is one of safety. Many of the more recent flight tests (Thorndycraft et al., 2005) involved low-level navigation and world referenced symbology. As such, the safety pilot flew in daylight VMC, while the evaluator flew the NVG system with neutral density filters and a mask to simulate night flight. This allowed for an increased safety margin compared to having both the safety pilot and the evaluation pilot under goggles. The testing arrangement also allows the safety pilot to act as an additional observer and to note performance deficiencies that might not be visible under goggles.

5.1.2.3 Qualification and Certification

Qualification and certification testing examines the functioning of the NVG system to ensure that the goggle is working in the correct manner. The most basic of these tests is done during the standard pre-flight checks and during routine goggle maintenance checks, as described in Chapter 2. These are the basic tests performed on goggles that are already in service. Even for new equipment, the basic tests of interest (e.g. checking on goggle resolution and gain) can be done with basic test equipment such as the Hoffman ANV-126. The performance of the goggles with respect to incompatible lights and halo can easily be examined in the lab. A variety of light sources (compatible and incompatible) can be examined with the goggles to determine the relative size and intensities of the halo in response to the different lights.

5.1.3 Objective Flight Test Measures

The following sections will examine the different categories of objective measurements that can be used to assess performance of the pilot with the goggles. There are three basic ways to quantitatively measure performance of a flight test manoeuvre:

- Precision
  - Measuring RMS error in position, height, attitude, rates, etc.
  - Measuring frequency and or magnitude of the exceedance of tolerances.
  - Measuring accuracy of response to a secondary task.
- Speed
  - Time to react to a new stimulus or situation; reflects on the interpretability of the information (NVG image or symbology).
  - Time to complete as a measure of the level aggressiveness that can be attained with the system.
  - Measuring time to respond to a secondary task.

- Control inputs
  - Measuring the number of control reversals.
  - Measuring the magnitude of the control movements.
  - Frequency content of the movements.
  - Movement rates.

There are pros and cons associated with each of these kinds of measurements. The first issue is that the methods noted above measure the performance of the whole system. There must be a thorough examination of the methodology and the actual test conditions to ensure that any differences that arise are due to the experimental manipulation.

The next issue is that the goggles will affect the pilot's visual cues and so the performance will typically be denoted in terms of precision and aggressiveness. Indeed, both the precision and the aggressiveness need to be quantified as different pilots will employ a different mixture of aggression versus precision in completing the task. This source of statistical variation can be minimized by instructing the pilots to perform the manoeuvre as precisely or as aggressively as possible. While this approach represents a departure from the traditional ADS-33 methodology such that handling qualities ratings are no longer appropriate, it does make statistical analysis more viable. Alternatively, a composite score can be developed to take both the speed and precision of task performance into account. The "racetrack" manoeuvre developed by QinetiQ encouraged pilots to perform the manoeuvre quickly, but added time penalties for losses in precision so that the time to complete the course was representative of both the level of aggressiveness and the level of precision that was achieved (Thorndycraft, 2002).

One difficulty with measuring performance objectively is to find a measurement that captures performance differences that arise from the NVG test conditions. This is particularly true when developing new manoeuvres for testing. With the ADS-33 manoeuvres (see Section 5.1.4.2) there are predetermined performance criteria which can be examined objectively, in concert with handling qualities ratings (Aeronautical Design Standard, 2000). For example, with the ADS-33 precision hover task, one can measure the position error to determine and if there are any biases in position in each axis. Similarly, for the side-step and the acceleration-deceleration manoeuvres, one can measure the aggressiveness of the manoeuvre via the attitudes (roll and pitch, respectively). For the pirouette the aggressiveness can be measured via the time to complete the manoeuvre. In each of these tasks, the performance will be affected by the quality of the information provided by the NVG imagery or symbology. For example, if goggle A has a higher resolution and higher contrast image than goggle B, the increased cuing available in goggle A should allow the pilot to perform the task more precisely/aggressively.

5.1.4 Subjective Flight Test Measures

Traditionally, many of the methods for evaluating aircraft or aircraft systems were based on the pilot's subjective rating of the system. These evaluations ranged from basic verbal protocols to structured questionnaires used to derive a handling qualities rating of the pilot/aircraft system. The following sections deal primarily with the standardized handling qualities evaluation procedures, including a discussion of the ADS-33 usable cue environment (UCE) and visual cue ratings (VCR), as well as some traditional flight test workload assessment techniques.

5.1.4.1 Handling Quality Evaluation

Cooper and Harper (1969) defined flying qualities as "Those qualities or characteristics of an aircraft that govern the ease and precision with which a pilot is able to perform the tasks required in support of an aircraft role." A great number of engineers have occupied themselves with answering the questions of what makes good flying qualities and how to rigorously assess them (see also Harper and Cooper, 1986).

The engineers have developed a thorough methodology, discussed in Hodgkinson (1998) and Padfield (1995). In these texts it is quite clear that in the evaluation of flying qualities, engineers do not restrict themselves to looking at the aerodynamics of an aircraft. Instead, they have a more inclusive view that includes characteristics of the pilot, the aircraft, the displays, the controls and the intended mission (or task) in their investigations. The pilot is one of the key elements in the aircraft control loop and in fact, basic texts on flight tests stress that "the pilot is part of a system that is intended to accomplish a mission" (Hodgkinson 1998).

Within the flying qualities field, researchers have recognized for a long time that the visual environment influences handling qualities. Handling qualities researchers have attempted to develop methods to quantify the visual environment through the use of a usable cue environment (UCE) methodology. This methodology has not been rigorously evaluated, but it has found useful application in the assessment of the interaction between control system augmentation and the visual environment. In ADS-33 it is used to define levels of control augmentation to compensate for degraded visual cueing. This approach suggests that the pilot's visual environment can influence his control strategy. It also highlights the fact that night vision devices, cockpit displays and lighting form a key part of the pilot's visual environment and factors such as display sensitivity, display dynamics and display time delay affect control. There is also a realization that the external visual environment affects control. Changes to the visual environment may alter the presence of visual cues used by the pilot.

"The same aircraft might have perfectly good handling qualities for nap-of-the-earth operations in the day environment, but degrade severely at night; obviously the visual cues available to the pilot play a fundamental role in the perception of flying qualities…. e.g. the quality of the vision aids, and what the symbology should do, becomes part of the same flying qualities problem as what goes into the control system." (Padfield, 1995).

There are many complex interactions between the visual environment, the visual aid (i.e. the NVGs), the task, the pilot, the control system and the aircraft. This combination is complex, non-linear and intricate making it inherently difficult to analyze. Effective analysis of any display must include more than modelling, simulation and laboratory tests. Pilot-in-the-loop testing is essential to uncover the full nature of the complex interactions among the pilot, aircraft, control and display systems.

There are often cases where different test conditions result in the same Handling Qualities Rating (HQR). HQR is not meant as a stand alone and requires description of why the score was given. Also cases occur where the pilot is aware of differences in workload, but the differences are not sufficiently large to warrant a different HQR. While this is a deficiency in the Cooper-Harper ratings scale, the scale was not intended to resolve fine differences in visual cueing systems such as NVGs.

5.1.4.2 ADS-33 Manoeuvres

In order to evaluate goggles or other NVG systems using ADS-33 one needs to ensure that everything except the experimental conditions are held relatively constant. The ADS-33 methodology tests the whole system and, as such, any factor that could affect performance is incorporated into the HQR. While the aircraft performance remains essentially constant, a significant amount of test time may be spent awaiting consistent environmental conditions (e.g. low winds, similar weather/lighting conditions). However, having consistent environmental conditions is necessary to isolate the effects of the test conditions (the goggles or the symbology) so that one can see only the effect of the goggles on the whole pilot-in-the-loop system.

There is a set repertoire of manoeuvres to choose from and thus care must be taken to choose a manoeuvre that may highlight the differences between the conditions under test. For example, to examine different field-of view characteristics (e.g. normal NVG field of view vs. panoramic NVGs), a manoeuvre like the

ADS-33 pirouette is a good candidate because it requires the pilot to frequently scan between the center marker and the longitudinal axis performance markers. In examining field of view, performance differences will be more likely to occur if the pilot has to make large and frequent head movements with the smaller field of view compared to the larger field of view. The ADS-33 precision hover is also a useful task in this regard, as the pilot's ability to detect drift off of the optical axis is impaired by smaller fields of view. The pilots will typically report failures or difficulties to achieve desired performance in the longitudinal axis of the hover task with small fields of view. Similarly, objective performance measures would tend to show larger RMS errors in the longitudinal position when the pilots perform the precision hover task with a small field of view.

5.1.4.3 ADS-33 Usable Cue Environment for Evaluating Pilot Visual Systems[1]

To document improvements in the visual display that the pilot uses to control the aircraft, a controlled and measured methodology is required. If this methodology is standardized throughout the research community, then the experimental results can be shared. The usefulness of this data set is enhanced if the experiment approaches the operational context in terms of task difficulty, cueing environment and flight duration. Notwithstanding this, external factors such as stress, adverse weather and ancillary tasks are sometimes difficult to control and simulate and often must be accounted for during the operational trials.

This use of flight test manoeuvres for handling qualities evaluations works well when evaluating cross-coupling, power margins or response types; unfortunately, the evaluation of a helmet mounted display also adds to the mixture the requirement to determine the visual cue environment of the pilot. The following factors as addressed above have a bearing on the pilot's ability to conduct the task:

- The amount of texture and contrast;
- The colour content;
- The value of the visual acuity;
- The size of the field of view;
- Field of regard;
- Binocular overlap;
- The presence and type of symbology;
- The image time lag;
- The quality of the image fusion; and
- The weather and lighting conditions.

All these factors are grouped together in what is termed Useable Cueing Environment (UCE). The UCE is evaluated by providing Visual Cue Ratings (VCR) during specific manoeuvres and focuses on the pilot's ability to perceive horizontal and vertical translation and attitude cues. According to ADS-33, low ratings (1-2) mean that the pilot can make aggressive/precise corrections with confidence and precision. The UCE boundary was defined using an aircraft that met the ADS 33 rate response type requirements and received a level 1 mean pilot rating during the hover, landing, pirouette, sidestep, acceleration-deceleration and bob up flight test manoeuvres. Using goggles, the field of view and opacity was varied until the pilot could no longer achieve level 1 ratings during the manoeuvres detailed above. Level 2 and 3 ratings were used to determine the UCE 2 and 3 areas respectively.

One must however remember that the main use of the UCE is as a guideline to aircraft manufacturers to determine the response type needed to complete the chosen mission task elements, given a chosen visual

[1] The following discussion of UCE is taken from Gubbels et al. (2002) and used with permission of the authors.

display (be it head-down displays, head up displays, goggles, or helmet mounted displays). In other words a scout helicopter required to conduct near earth night missions using NVG without symbology (i.e. most likely in UCE 2 conditions) would require an attitude command/attitude hold plus a rate command/ direction hold control system (Aeronautical Design Standard, 2000).

There are several limitations to the current VCR/UCE methodology for evaluating upgrades to a helmet mounted display. The first of these limitations deals with the usefulness of the UCE methodology as an engineering tool while the last two of these limitations deal with the usefulness of the results as applied to an operational application. Firstly, the use of a three-point scale with very few descriptors leads to poor repeatability. Secondly, since the cueing environment is the critical factor in the evaluation of an NVG, the external cueing must be carefully chosen. The use of cones to define manoeuvres may lead to false results if they are automatically applied to the operational context where such well-defined cues are lacking. Thirdly, the variation in aircraft flight characteristics adds to the scatter when comparing fielded equipment on different aircraft. Similarly, results obtained using rate response controls will be different from results obtained using the all-up systems of the aircraft (such as attitude command).

5.1.4.4 The UCE Rating Scale is a 3-Point Subjective Scale

UCEs are evaluated by having the pilot, given a specific set of equipment, fly specific manoeuvres to Degraded Visual Environment (DVE) criteria and provide a visual cue rating. Unlike the HQR scale, the VCR does not have a decision tree to help the evaluator. Though the scale has 5 ratings, it only provides descriptors for 3. The descriptors that are given are fairly vague and subject to wide interpretation. Compounded by the fact that the test pilot community as a whole has limited experience with these scales, the repeatability of test conditions and results is poor. The three descriptions, for ratings of one, three and five, respectively, are detailed below (note: X refers to attitude, horizontal translation rate or vertical translation rate cues):

- Good X Cues: Can make aggressive and precise X corrections with confidence and precision is good.
- Fair X Cues: Can make limited X corrections with confidence and precision is only fair.
- Poor X Cues: Only small and gentle corrections in X are possible, and consistent precision is not attainable.

These ratings are given for attitude, horizontal translation rate and vertical translation rate cues and are then plotted on Figure 5-1 to determine the UCE. Note that the poorer of the horizontal or vertical translation rates is used for the vertical axis and that heading is implied in the attitude rating.

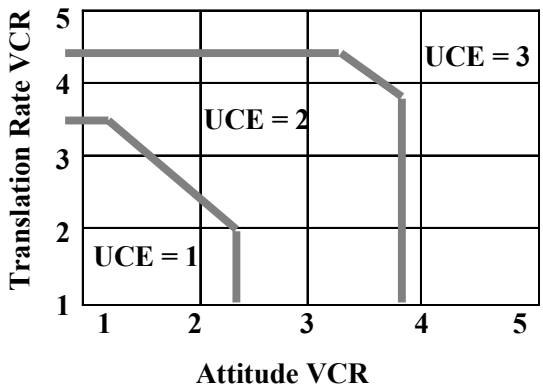

Figure 5-1: UCE Determination.

To be used as an engineering tool the VCR scale will have to be enhanced. A brief description of the factors influencing the perception of cues should be provided as background reading. This should also include which elements are to be considered when making the evaluation, including looking at issues involving the standard aircraft instruments, additional symbology and field of regard in addition to evaluating the goggles. To increase repeatability, the VCR scale will also have to incorporate better word descriptors for all five ratings. The possibility of developing a decision tree, similar in concept to the Cooper Harper rating scale, should be investigated. Questions such as ability to perceive a given amount of deviation (in terms of distance and/or rate) as well as percentage of control input and speed of input could be used to help the pilot derive a meaningful VCR. However, as the rating gets more detailed the time to go through the process will also increase and in the end, a compromise between the time to give the VCRs and the repeatability of the results will have to be made. The other option is to move away from the traditional visual cue ratings and develop a new and/or more simplified scale.

5.1.4.5 Use of Cones in the Evaluation

ADS-33 flight test manoeuvres require the use of cones/lines to define adequate and desired criteria. The unfortunate consequence of this is that the pilot, using an NVG, now has far superior point style cues than his operational community counter-part. For instance during an unmask-mask manoeuvre the pilot will perform a vertical climb and descent generally in reference to a series of nearby trees (at least one in the forward field of view to achieve lateral positioning and one to the side to achieve fore and aft positioning). These sometimes blend in with the other trees and in some cueing environments the pilot has great difficulty in detecting even moderate drift rates. On the other hand, even with frequent head movements the pilot can generally determine his position with respect to a line of cones, quickly estimate the relative position and input the corrective action. It is not suggested that the use of cones for flight-test manoeuvres be eliminated; however care must be taken in interpreting the results from an ADS-33 test environment with respect to operational flight conditions. Therefore any full-scale evaluation of an NVG must incorporate both ADS-33 flight test manoeuvres and operational type manoeuvres. To enhance comparison of results between different test agencies, a series of operational type manoeuvres should be defined that do not rely on cues such as cones.

5.1.4.6 Requirement for a Level 1 ADS 33 Rate Response Type

The VCR and UCE are based on the evaluations being conducted using an aircraft that has met the ADS 33 rate response type requirements and received a level 1 mean pilot handling qualities rating during a specified set of flight test manoeuvres. As defined in ADS-33, a level 1 aircraft must meet all of the standards for that level, which are intended to ensure that handling qualities do not a limitation on the capacity to perform the intended missions. There can be a multitude of criteria required to achieve level 1 for the various roles/missions for which the aircraft is intended. However, even if the aircraft meets level 1 for the tests performed, it could fail to meet level 1 handling qualities on dimensions/tasks that were not tested. Few aircraft meet the all of the specifications for level 1 and not all aircraft have rate-damped systems. Results obtained with a sub-level 1 (i.e. level 2 or greater) aircraft may lead to the wrong conclusions in determining the response type required for a given Mission Task Element (ADS-33). For instance, a bordering UCE 2 rating given using a sub-level 1 aircraft may in fact be UCE 1. For this reason (and others such as field of regard and cockpit instrument layout), results obtained on a specific aircraft type may not be transferable to others.

5.1.4.7 Selection of Response Type for Evaluation

The investigation or development of the fielded system might be aimed at an aircraft with much higher levels of augmentation such as translational velocity command. In this case what is the transferability of results obtained using Rate Command? The evaluation of an NVG using the VCRs and a rate-damped aircraft is more of an exercise used to benchmark the system and determine the response type required.

Past research has shown that system development performed with a rate-damped system helps define the improvements required by magnifying the deficiencies. In other words, if you evaluate a particular NVG system, each small incremental improvement will show up as a larger variation in handling quality ratings if Rate Command is used as the basis of evaluation in lieu of Attitude Command. However, along the same lines as the requirement to test the NVG in an operational context to remove the influence of the increased point cueing provided by the cones, the final system handling quality assessments of the fielded equipment should be evaluated using the intended aircraft control systems.

5.1.4.8 Workload or Pilot Compensation

In the planning stages of a goggle evaluation, the concept of pilot workload is often raised as a means of evaluating the goggles. For example, does the pilot's workload increase when using goggle A vs. goggle B? While the concept is intuitively compelling, in practice measuring workload has proven to be a complicated proposition. Workload is a multidimensional representation of the information (sensory/perceptual) available to the pilot, the amount of processing (cognitive activity) required to make sense of the information and whatever output is required of the pilot (motor or control activity). The tools available to measure this process range from simple subjective rating scales (e.g. rate your workload from 1 – 10) to more complicated subjective rating scales (e.g. NASA TLX Hart & Staveland, 1988, Modified Cooper-Harper Workload Scale Wierville & Casali, 1983) to objective measures of pilot activity (DIMSS, physiological measures as above). For small subject samples, none of the measures works particularly well due to the variability of interpretations between subjects. For example, the modified Cooper-Harper workload scale often is treated as a simple 10-point workload rating, while other evaluators use the decision tree to arrive at a workload rating. The variability associated with the set of subjective questionnaires might be alleviated to a certain extent by pre-training. Pre-training may diminish inter-subject variability enough to allow workload ratings to distinguish between the workload associated with using the different goggles (or symbology sets). However, typically the subjective rating scales either cannot resolve differences in workload between test conditions or can only resolve obvious differences in workload.

Similarly, standard handling qualities evaluations can be used to determine the level of pilot compensation required to attain a given level of performance. Within each level of performance (desired, adequate and sub adequate) the pilot may need to employ different levels of compensation or strategies to deal with the limitations of the system. As indicated in Section 5.1.4.2, when all other aspects of the system (helicopter, weather, winds, lighting) are consistent, the remainder of the pilot compensation should be due to the test conditions. In general terms, the level of pilot compensation can be regarded as an element of the pilot's workload.