# A RATING SCALE FOR THE SUBJECTIVE ASSESSMENT

# OF SIMULATION FIDELITY

Philip Perfect,

Emma Timson, Mark D. White,

Gareth D. Padfield

The University of Liverpool

Liverpool, UK

Robert Erdos,

Arthur W. Gubbels

Flight Research Laboratory,

National Research Council, Ottawa, Canada

**ABSTRACT**

**A new rating scale for capturing pilot subjective assessment of simulation fidelity is described in this paper. The scale has been developed through a series of flight and simulation trials using six test pilots from a variety of backgrounds, and is based on the methodology utilised with the Cooper-Harper Handling Qualities Rating scale and the concepts of transfer of training, comparative task performance and task strategy adaptation. The development of the new rating scale has been undertaken using the University of Liverpool's HELIFLIGHT-R research simulator, in conjunction with the Canadian Flight Research Laboratory’s ASRA in-flight simulator. The utility of the scale applied to locating fidelity boundaries for quantitative metrics is illustrated for an inter-axis coupling criterion.**

**INTRODUCTION**

**The evaluation of the fidelity of a simulation device for flight training typically includes a series of quantitative requirements contained within simulator qualification documents such as JAR-FSTD H [1] or FAA AC120-63 [2]. These quantitative requirements examine the response or behaviour of the individual elements of a simulation device – the visual system, the motion platform (if so equipped), the flight dynamics model etc. – to a set of predetermined inputs. The results of these tests are typically termed “engineering fidelity” and only partially serve to characterise the utility of a simulator. The implicit assumption in tests of engineering fidelity is that a strong quantitative match of simulator component systems with the flight vehicle will assure a high degree of simulator utility. Experience has shown that this assumption is not always valid, and that tests of engineering fidelity are insufficient to guarantee a sufficiently accurate simulation. Hence, the qualification standards require a piloted, subjective assessment of the simulation in addition to the quantitative elements. These subjective tests “*verify the fitness of the simulator in relation to training, checking and testing tasks*” [3]. However, the guidance provided in the qualification documents regarding the approach taken to subjective evaluations is very limited. Paragraph 4.2.4 of Section 2 of JAR-STD 1H [3] (one of the predecessors to JAR-FSTD H) states:**

***“When evaluating Functions and Subjective Tests, the fidelity of simulation required for the highest Level of Qualification should be very close to the aircraft. However, for the lower Levels of Qualification the degree of fidelity may be reduced in accordance with the criteria contained (within the document)”***

**This requirement is poorly defined, and open to interpretation by the operator and qualifying body. It is suggested that this existing requirement for the subjective aspect of simulator qualification is unsatisfactory and should be improved. It is therefore proposed that an effective method by which the process may be improved is through the incorporation of a repeatable, prescriptive rating scale for the subjective assessment of fidelity into the overall qualification process.**

**In this paper, a new Simulation Fidelity Rating (SFR) scale is introduced that may be used to complement and augment the existing simulator evaluation processes of JAR-FSTD H and other applicable simulator qualification processes. It is proposed that the SFR scale may be used as part of a fidelity evaluation methodology based on the use of engineering metrics for both the *prediction* of the fidelity of the individual simulator components (flight model, motion platform, visual system etc.) and the assessment of the *perceptual* fidelity of the integrated simulation system, as experienced by the pilot (**Figure 1**). Within this methodology, the SFR scale would augment the numerical analysis of the pilot’s behaviour and vehicle response in the ‘perceptual fidelity’ component of a simulator fidelity assessment, as discussed in [4] and [5]. As proposed in [5], this stage of the fidelity evaluation would be completed after the ‘predicted fidelity’ analysis, examining each individual component of the simulator, was completed satisfactorily.**

**An additional benefit of the SFR scale beyond direct evaluation of training devices is the ability to locate boundaries between levels of fidelity for quantitative metrics, such as the compensation/adaptation metrics based on control attack.**



Figure 1: Methodology for the assessment of simulator fidelity [5]

**The paper begins with a brief review of previous attempts at the development of fidelity rating scales. This is followed by a description of the process and methodology under which the new scale has been developed, followed by a discussion of the concepts that have been selected to form the basis of the new scale. The process by which the new scale may be used in an evaluation of a training simulator is described. Two case studies are provided. The first looks at an evaluation of a research simulator as a training device, while the second demonstrates the use of the SFR scale in the identification of level boundaries for fidelity metrics. The paper is brought to a close with some concluding remarks, including prospective future developments and consideration of wider applications for the new scale.**

**REVIEW OF FIDELITY RATING SCALE CONCEPTS**

**An early fidelity rating scale was developed during a NASA programme to validate the General Purpose Airborne Simulator (GPAS) [6]. The scale (**Table 1**) was configured to be similar in format and consideration to the Cooper-Harper Handling Qualities Rating (HQR) Scale [7]. While the scale was initially developed to support evaluation of the suitability of the GPAS for Handling Qualities (HQs) experiments, a read-across to training effectiveness could be made, where ‘results directly applicable to actual vehicle’ would correspond to ‘training completely effective’ and so forth. Indeed, a desirable characteristic of a new SFR scale would be its applicability to a wide range of simulation uses beyond pure training tasks, which may include HQ developments, aircraft development etc.**

Table 1: Early NASA Simulator Fidelity Rating Scale [6]

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Rating** | **Adjective** | **Description** |
| **Satisfactory** | **1** | **Excellent** | **Virtually no discrepancies. Simulator results directly applicable to actual vehicle with high degree of confidence.** |
| **2** | **Good** | **Very minor discrepancies. Simulator results in most areas would be applicable to actual vehicle with confidence.** |
| **3** | **Fair** | **Simulator is representative of actual vehicle. Simulator trends could be applied to actual vehicle.** |
| **Un-satisfactory** | **4** | **Poor** | **Simulator needs work. Simulator would need some improvement before applying results directly to actual vehicle.** |
| **5** | **Bad** | **Simulator not representative. Results obtained here should be considered as unreliable.** |
| **6** | **Very Bad** | **Possible simulator malfunction. Gross discrepancies prevent comparison from being attempted.** |

**This scale was used to quantify and improve the modelled dynamics of the GPAS over a number of flights, being applied both to individual characteristics (such as Dutch Roll damping) and also the overall system. However, the scale did not subsequently see more widespread adoption in the simulation community.**

**A discussion on the use of subjective rating in the evaluation of simulation devices was compiled by the US Air Force in 1982 [8]. Here, the effectiveness of a wide variety of schemes for subjective evaluation was considered. Two examples are discussed below, beginning with rating of the task cues provided by a simulator (**Table 2**).**

Table 2: Example of Subjective Rating for Task Cues [8]

|  |  |
| --- | --- |
| **Rating** | **Description** |
| **S1** | **The cues provided by the Aircrew Training Device (ATD) are Sufficiently similar to the aircraft cues required to train the task/subtask** |
| **S2** | **The cues provided by the ATD are Sufficient to support training, but, if improved, would improve/enhance training.** |
| **NS** | **The cues provided by the ATD are Not Sufficiently similar to the cues required to train the task/subtask.** |

**The suggestion is made that the evaluation of device effectiveness is ultimately a binary decision – the cues are either sufficient or not sufficient. The extension to this basic decision in** Table 2 **allows the identification of areas where an improvement in the cueing would enable an improvement in the training effectiveness to be made. Schemes with four or five choices (e.g. identical; highly similar; moderately similar; slightly similar or not at all similar cues) were also discussed in [8]. While these offer more flexibility in the evaluation process and can provide additional insight into the severity of deficiencies in the simulation, ultimately the sufficient/insufficient boundary must be placed between two of the ratings, restoring the binary nature of the evaluation. This is a similar structure to that used in the established Deck Interface Pilot Effort Scale (DIPES) [9], used to determine safe conditions for helicopters operating from ships at sea. DIPES is a five point scale, but any rating of three or lower represents a safe condition, while a higher rating represents a potentially unsafe condition.**

**A second type of rating considered in [8] relates to the training capability offered by a device. In this rating system, the evaluator is asked to consider the effectiveness of the training received relative to that provided in the live aircraft. In this scale, a basic interval rating between 1 and 7 is awarded, with the upper and lower boundaries as described in** Table 3**. The intermediate points were left undefined, with the evaluator instructed to consider these intermediate points as being equally spaced between the two boundaries.**

Table 3: Training Capability Rating Scale [8]

|  |  |
| --- | --- |
| **Rating** | **Description** |
| **7** | **Training provided by the ATD for this task is equivalent or superior to training provided in the aircraft.** |
| **1** | **Training provided by the ATD for this task is in no way similar to training provided in the aircraft; no positive training can be achieved for this task** |

**More recently, a scale was developed by Israel Aircraft Industries (IAI) to support the development of a Bell 206 simulator [10]. This rating system used a six point scale to assess the individual aspects of a pilot’s interaction with the aircraft/simulator. The rating points were given simple descriptors, such as ‘identical’, ‘very similar’ etc., while the aspects of the simulation that were considered included**

* **Primary response magnitude**
* **Control sensitivity**
* **Control position**
* **Secondary response direction**
* **Etc.**

**Thus, this rating system allows the evaluating pilot to specify the precise aspects of the simulation that are deficient, but does not offer a consideration of the overall, integrated simulation.**

**While each of these scales has seen positive applications, no single rating scale structure has yet emerged as a standard or gained acceptance by certification agencies for the formal evaluation of simulator fidelity. This brief review demonstrates that there are many different methods that can be applied to fidelity assessment, but as yet, none has gained precedence over the others. Each of the rating methods discussed above brings its own benefits. The IAI scale and the task cue ratings of** Table 2 **allow the engineer to assess the individual elements of the simulator and determine which require remedial action. The requirement here is for the pilot to introspect in order to identify the individual aspects of the simulation that are deficient – whilst engaging with a task at the overall simulation level. The assessment of the overall simulation effectiveness in** Table 1 **and** Table 3**, or the use of basic suitable/unsuitable decision points in** Table 2**, allows the simulator operator and qualification agency to understand whether the device is acceptable or not. With these approaches, the evaluating pilot is required to make judgements regarding the overall suitability of the simulator.**

**RATING SCALE DEVELOPMENT**

**The lack of an internationally accepted fidelity rating system led to a new initiative. The research has been conducted by the University of Liverpool (UoL) in collaboration with the Flight Research Laboratory (FRL) of the National Research Council of Canada (NRC), and supported by the US Army International Technology Center (UK). The objective of the research has been to address the gaps in the evaluation of the overall fidelity of simulators in the existing qualification processes.**

**Initial development process**

**As the structure of the HQR scale [7] has gained international acceptance and worldwide adoption over a forty-year period, it forms a natural starting point for the development of a new scale, especially so as many of the principles of HQ engineering (task performance, compensation, etc.) must also be considered during a fidelity evaluation. An examination of HQ practices led to three rules that were to be adhered to during the SFR scale development:**

1. Qualifiers of simulation functional fidelity (e.g. quality is low, medium, high; or poor, fair, good, very good, excellent) need to be defined in such a way that a common, unambiguous understanding can be developed between the community of pilots, simulation and test engineers and regulatory bodies. Reference cases (scenarios, manoeuvres, configurations) for these qualifiers should also be defined.
2. Functional fidelity relates to fitness for purpose, and the purpose needs to be clearly defined within the experimental design, e.g. by using Mission Task Elements (MTEs) [11] with appropriate performance standards. For example, if the simulator is to be used to practice ship landings in rough weather (i.e. skill development), then task performance criteria, simulated environmental conditions and the consequent specific training needs should be detailed.
3. The pilot has to rate an entity that is experienced at the cognitive level – the quality of the ‘illusion’ of flight created by a suitable combination of visual, vestibular, auditory etc. cues and the mathematical flight model. This is very important and has, to a large extent, bedevilled the development and application of other rating scales that attempt to quantify pilot perception, e.g. the Usable Cue Environment (UCE) in which Visual Cue Ratings (VCRs) [11] feature. During the early development phases of the UCE system, test pilots were asked (or attempted) to rate the quality of the cues that they used to fly a task. This assumed that the pilots were able to introspect on how their perception system worked with the available cues. In practice the perceptual system largely works at the subliminal level so a pilot’s ability to reflect on what and when they used different cues, and describe the process, is very limited [12].

**Using these principles, a series of ‘straw-man’ SFR scales was drawn up at UoL and FRL, and their advantages and disadvantages examined during an exploratory trial with one test pilot, using the UoL HELIFLIGHT-R full-motion flight simulator [13]. The most promising elements of each of the ‘straw-man’ scales were combined to give a preliminary baseline SFR scale.**

**SFR scale refinement**

**As the initial piloted simulation trial continued with the now baseline SFR scale, the terminology, definition of the levels of fidelity and indeed the overall structure of the scale evolved. This evolutionary process continued over a further two simulation trials at UoL, involving three additional test pilots.**

**As these developmental trials focused purely on evaluations using a single simulation device, a means of introducing a fidelity comparison was required in order to be able to exercise the new rating scale over a wide range of test cases. The pilots were familiarised with a baseline flight model configuration, which was used to represent the ‘simulation’ side of the fidelity assessment. The configuration of the model was then altered, for example by introducing additional inter-axis coupling between the pitch and roll axes, and this modified configuration was used to represent the ‘flight’ side of the fidelity assessment.**

**The simulation trials matured the rating scale structure, which was for the first time employed during a flight test with the NRC’s Bell 412 Advanced Systems Research Aircraft (ASRA) airborne simulator [14] in April 2011. Two of the test pilots involved in the SFR scale development took part in the flight tests, allowing the scale to be used for real flight vs simulation comparisons. The unique airborne simulation capabilities of the ASRA were also employed to replicate the configuration changes adopted at UoL, a process that helped to expand the matrix of evaluation points with which to exercise the SFR scale.**

**At the conclusion of the flight tests, further modifications to the SFR scale were made. A further piloted simulation trial was conducted at UoL, involving two additional test pilots. While extensive test data was gathered during this trial, no further developments of the scale were found to be necessary.**

**Workshop at AHS Forum**

**The development of the SFR scale involved a relatively small number of research staff from UoL and FRL. To broaden involvement in the development of the SFR scale to the wider rotorcraft, simulation and HQs communities, a workshop was organised in conjunction with the 67th Annual Forum of the American Helicopter Society in May 2011. A total of 48 people attended the workshop, which ran over a two-day period. During the workshop, presentations were given on a wide range of subjects pertaining to simulation fidelity, including the UoL/FRL SFR scale and related quantitative metrics development, fidelity assessment research in the US Army, CAE and the University of Iowa, and the use of (and issues pertaining to) flight simulation in aircraft development programmes. The workshop also included sessions where various aspects of a fidelity rating scale were discussed, including ‘what should the pilot be rating?’, ‘what makes a good simulation fidelity test pilot?’ and ‘what test methods are appropriate?’**

**THE SIMULATION FIDELITY RATING SCALE**

The proposed SFR **scale employs a number of key concepts that are considered fundamental to the utility of a simulation device. They are as follows:**

* **Transfer of Training (ToT) – the degree to which behaviours learned in a simulator are appropriate to flight.**
* **Comparative Task Performance (CTP) – comparison of the precision with which a task is completed in flight and simulator.**
* **Task Strategy Adaptation (TSA) – the degree to which the pilot is required to modify their behaviours when transferring from simulator to flight and *vice versa*.**

**The relationship between task performance and strategy adaptation is similar to that between performance and compensation in a handling qualities evaluation. In the HQR scale, the expectation is that the pilot’s perception of deteriorating performance will stimulate higher levels of compensation, indicative of worsening HQ’s. While this correlation can be expected in measuring HQs, in the context of fidelity assessment task performance and adaptation will not necessarily change in correlation with each other, but will instead depend on the nature of the fidelity deficiencies present in a simulator.**

**A matrix presenting all possible combinations of comparative performance and task strategy adaptation was constructed (**Figure 2**); this was used to form the basic structure of the SFR scale (**Figure 3**).**

|  |  |
| --- | --- |
|  | **Comparative Performance** |
| **Equivalent** | **Similar** | **Not Similar** |
| **Strategy Adaptation** | **Negligible** | **LEVEL 1** ***Full Transfer of Training*****(SFR 1-2)** |  |  |
| **Minimal** |
| **Moderate** |  | **LEVEL 2*****Limited Transfer of Training*****(SFR 3-6)** |  |
| **Considerable** |
| **Excessive** |  |  | **LEVEL 3*****Negative Transfer of Training*****(SFR 7-9)** |

Figure 2: Fidelity Matrix



*Figure 3: Simulation Fidelity Rating Scale*

**Each of the ratings SFR=1 to SFR=9 corresponds to a region in the fidelity matrix. The SFR=10 rating indicates a simulation that is entirely inappropriate for the purpose, and so comparisons with flight cannot be made. As with the HQR scale, boundaries have been defined between the potential combinations of comparative performance and adaptation, reflecting value judgements on levels of fidelity. As the SFR worsens through each level, it can be seen from** Figure 3 **that the individual comparative performance and adaptation measures may not degrade in a progressive manner. However, the intention is that the overall ‘experience’ of the simulation fidelity degrades progressively as the SFR worsens.**

**SFR Scale Terminology**

**The SFR scale has been designed to evaluate a simulator on a task-by-task basis. Consequently, where fidelity defines *fitness for purpose*, a collection of ratings for various MTE’s would define the boundaries of positive training transfer for a given simulator. This is similar to the approach being adopted by an International Working Group (IWG) led by the Royal Aeronautical Society (RAeS) which has been revising and updating the existing training simulator certification standards; in the new proposals ([15],[16],[17]), the required complexity for each of the simulation components is based on the tasks that will be trained.**

**The first definition that must be made prior to the commencement of fidelity assessment with the SFR scale is that of the purpose of the simulator. The purpose describes the range of tasks to be flown using the simulator, and hence the scope of the SFR evaluations. Each task identified in this step would be assessed on an individual basis, the results for each task then being used to create a “usage envelope”, within which the simulator may be effectively (and safely) employed.**

**In the context of a training simulator, the definition of the levels of fidelity has been made relative to the transfer of training (ToT) that occurs when a pilot transitions between the simulator and the aircraft. It should be noted that in assigning the level of fidelity, the evaluating pilot is being asked to make a subjective judgement on the degree of ToT that is likely to take place. This is in contrast to the work of Bürki-Cohen *et al*, where quantitative metrics for the measurement of ToT have been developed [18]. For the SFR scale, the definition of the four levels of fidelity depends on the objective of the training. Three types of simulation training have been defined – *skills acquisition*, *skills development* and *skills assessment*, where acquisition and development would correspond to the processes of initial training and recurrent training respectively. For *skills acquisition*, the levels have been defined as follows:**

* **Level 1 fidelity: Simulation training is sufficient to allow operational performance to be attained with minimal pilot adaptation. There is complete ToT from the simulator to the aircraft in this task.**
* **Level 2 fidelity: Additional training in the aircraft would be required in order to achieve an operational level of performance. There is limited positive ToT from the simulator to the aircraft in this task, but no negative ToT.**
* **Level 3 fidelity: Negative ToT occurs (i.e. the pilot learns an inappropriate technique), and the simulator is not suitable for training to fly the aircraft in this task.**

**Similarly, the levels for *skills development* have been defined as:**

* **Level 1 fidelity: Simulation training is sufficient to restore previous performance capabilities.**
* **Level 2 fidelity: Simulation training provides limited improved performance capability. Additional training is required.**
* **Level 3 fidelity: No positive ToT occurs. The simulator is unsuitable for training.**

**And the levels for *skills assessment* are defined as:**

* **Level 1 fidelity: Simulation is sufficient to comprehensively demonstrate skills associated with qualified performance.**
* **Level 2 fidelity: Performance in the simulator demonstrates limited elements of the required skills.**
* **Level 3 fidelity: Performance in the simulator does not serve to demonstrate the required skills.**

**In each of these cases, a Level 4 fidelity rating indicates that it is not even possible to complete the task using the simulator.**

**The task may be defined as the training manoeuvre/procedure, accompanied by a set of performance requirements and environmental conditions. In an HQR evaluation, a Mission Task Element (MTE) specification consists of the target manoeuvre profile alongside a set of ‘desired’, and ‘adequate’ performance tolerances for each element of the manoeuvre profile (height, airspeed, heading etc.), where the achievement of a certain category of performance assists the pilot with determining the level of handling qualities of the aircraft. The same style of task definition is adopted for an SFR evaluation. The comparison of the achieved level of performance between flight and simulator assists the evaluating pilot with the judgement of comparative performance. The three levels of comparative performance have been defined as follows:**

* **Equivalent performance: The same level of task performance (desired, adequate etc.) is achieved for all defined parameters in simulator and flight. Any variations in performance are small.**
* **Similar performance: There are no large single variations in task performance, or, there are no combinations of multiple moderate variations across the defined parameters.**
* **Not similar performance: Any large single variation in task performance, or multiple moderate variations, will put the comparison of performance into this category.**

**Definition of ‘moderate’ and ‘large’ variations has proven to be a complex process. Initially, the test pilots were instructed to consider these as being a deviation from desired to adequate, or adequate to beyond adequate for a moderate variation, and from desired to beyond adequate and *vice versa* for a large variation. However, this proved to be too restrictive, with the pilots commenting that with certain test configurations, desired performance may be just achievable on one side of the flight-simulator comparison, but marginally unachievable on the other, forcing the pilot to degrade the fidelity rating to Level 2 despite a very small change in the actual task experience. In the more recent simulation trials, the pilots have been allowed a greater degree of flexibility in making decisions regarding whether a deviation is small, moderate or large. This approach allows the pilots to ensure that they rate the simulation in the level that they consider to be appropriate, rather than being driven by the task performance. The use of a fidelity rating questionnaire (Appendix A) helps to ensure a group of evaluating pilots applies consistent interpretations to these judgements.**

**A second area where the pilots are asked to make a qualitative distinction is for strategy adaptation. This is intended to capture all aspects of a pilot’s behaviour, and would include:**

* **Control strategy – differences in the size, shape and frequency of the applied control inputs.**
* **Cueing – differences in the way in which task cues are presented to the pilot.**
* **Workload – including differences in the physical effort of moving the controls; scanning of the available task cues; and the mental work associated with interpreting cues and determining the required control inputs.**
* **Vehicle response – differences in the perceived response of the vehicle.**

**Any other aspects of the task, other than the achieved level of performance, that are perceived to be different between the simulation and flight test should also be included within the level of adaptation required. Five levels of strategy adaptation are defined – negligible, minimal, moderate, considerable and excessive. These terms have deliberately been selected to be familiar in name and meaning to pilots who have used the HQR scale and have thus rated compensation/workload during a task. There are, however, differences in the interpretation of some of the terms when used in the SFR scale:**

* **The shift from minimal to moderate adaptation signifies the Level 1/Level 2 boundary, as is the case with minimal to moderate compensation in the HQR scale. However, minimal adaptation additionally features as a Level 2 fidelity rating when found in combination with ‘similar’ performance.**
* **The boundary between Level 2 and Level 3 HQRs occurs between compensation levels ‘extensive’ and ‘maximum tolerable’. Both of these terms were considered to be representative of insufficient simulation fidelity, and so have been replaced by a single adaptation level – ‘excessive’, which exists only in the Level 3 fidelity region.**

Due to these inherent complexities in assessing the level of adaptation and comparative performance, and analogous to the use of the HQR scale, satisfactory performance during simulation fidelity assessment may be limited to trained practitioners only. In order to ensure reliable SFRs, it is necessary for the evaluating pilot to possess a strong awareness of training effectiveness, training requirements and the processes of skills acquisition and development.

**A final aspect of SFR terminology is the term ‘fidelity’ itself. In the common vernacular, a full-flight simulator may be referred to as a ‘high fidelity’ device, while a part-task trainer may be ‘medium fidelity’ and a procedures trainer a ‘low fidelity’ device. In the context of the SFR, however, these labels are inappropriate. Instead, the intention is for ‘fidelity’ to be reflective of the suitability of the simulation device for the role it is performing. In this sense, all of the above devices can be ‘high fidelity’, as long as they provide the appropriate degree of transfer of training for the tasks for which they are employed. Using these concepts, the definition of ‘fidelity’ according to Heffley is used - *“****the simulator's ability to induce the pilot trainee to output those behaviours known to be essential to control and operation of the actual aircraft in performance of a specific task"* [19].

**Use of the rating scale**

**In the context of a training simulator evaluation as part of a certification process, the missions and scenarios for which the simulator is expected to be used need to be broken down into a series of small sections that are representative of individual training tasks – for example, these could be engine start, takeoff, hover etc. For each of these tasks, the expected profile and the allowable deviation away from the profile must be specified; these will form the basis of the comparative task performance section of the fidelity evaluation.**

**The evaluating pilots would be expected to be proficient and current at flying each of the tasks on the aircraft, and thus to be able to carry that experience across onto the simulator during the evaluation process. While this evaluation method is consistent with that used currently in training simulator evaluations, it is not always the same as that in which the trainees experience the simulator – where the simulator may be used to provide initial training prior to the first experience on the aircraft. Thus, an alternative evaluation method would be for the evaluating pilot to fly the tasks in the simulator, and then to repeat the tasks in the aircraft and award the SFRs following this flight. An essential aspect of either evaluation method, however, must be that the time period between the flight and simulator experiences be short, and uncontaminated with other aircraft or simulator types; thus, the memory of the first system remains reasonably fresh when the second system is flown.**

**A further consideration here is the duration of the simulator evaluation process. One of the outcomes of the trials at UoL was that the pilots became acclimatised to the fidelity deficiencies of the simulator after a period of exposure (a process distinct from the initial adaptation used in the fidelity assessment), and thus became less sensitive to those deficiencies as further tasks were evaluated. The ideal assessment process may therefore be for short periods of simulator evaluation, followed by periods of re-familiarisation in the aircraft.**

**During an evaluation, the pilot would fly the training tasks individually, and provide an SFR based upon each one. It is recommended that the evaluating pilot performs a number of repeats of each task. While the initial experience of a new system is important in quantifying the magnitude of fidelity deficiencies, and forming an impression of the differences that exist, continued exposure allows the pilot to determine the nature of the deficiency. During the simulation trials in which the SFR scale was developed, the test pilots were asked to award an SFR on the first run with a new configuration, and then performed three additional repeats of the task. At the end of this repeat phase, the pilot could revise their SFR if necessary. Continued repetition of tasks should, however, be avoided, as the evaluator may begin to lose memory of the original reference – the flight test.**

**For any fidelity evaluation, but especially in the case of ratings in Levels 2 or 3, the justifications for the ratings are critical. This is particularly the case given that the SFR scale, as with the HQR scale, is ordinal, so that the interval between individual ratings is not uniform across the scale. Hence, the pilot’s narrative supporting the rating explains the specific deficiencies that exist and the reasons why an SFR=3 was given, rather than SFR=2, for example. This assists the simulator engineer in determining the areas of the system that must be upgraded if fidelity is to be improved. During the UoL trials, each pilot was asked to complete a questionnaire following each fidelity evaluation; the questionnaire documenting the areas where task performance changed and adaptation was considered to have taken place.**

**Following the evaluations with each of the individual training tasks, the fidelity of the simulator in its overall role can be considered. In the event that different levels of fidelity are determined for different tasks, a breakdown of the utility of the simulator may be made – for those tasks for which Level 1 fidelity ratings were awarded, the simulator can be used with no additional training, while for those tasks awarded Level 2 SFRs, the simulator may still be used, but in the knowledge that the trainee will require additional training on the aircraft prior to reaching operational proficiency. The narrative substantiating the SFR should help determine the specific aircraft training requirements. Finally, for any tasks for which a Level 3 SFR has been awarded, the simulator should not be used, as it will impart incorrect behaviours to trainees.**

**CASE STUDY 1: USE OF THE SFR SCALE IN THE ASSESSMENT OF THE TRAINING BENEFIT OF A SIMULATOR**

**The two test pilots who participated in the flight tests on the Bell 412 ASRA awarded SFRs for the comparison of the HELIFLIGHT-R simulator [13] (featuring the FLIGHTLAB [20] simulation of the Bell 412 (F-B412)) against flight. Conforming with the developed practice, the SFRs were awarded during simulation trials (i.e. the flight testing had been completed first). The results are presented in** Figure 4 **for a selection of Mission Task Elements (MTEs). In the tests, the ASRA and HELIFLIGHT-R were configured to offer a variety of different response types (attitude command, attitude hold – ACAH; rate command, attitude hold – RCAH; and unaugmented – Bare). The MTEs were chosen from the US Army HQs standard, ADS-33E-PRF [11]. While these MTEs are primarily designed to highlight handling qualities, and specifically in military operations, they form a useful starting point for the exposure of fidelity deficiencies, especially in a research simulator such as HELIFLIGHT-R, where training is not the main purpose. The HQRs awarded by the pilots for each configuration in flight and the simulator are presented in** Figure 5**.**



Figure 4: Sample of SFR Results



Figure 5: Handling Qualities Ratings

**As HELIFLIGHT-R is a research simulator, it features a generic crew station, and is therefore not representative of the Bell 412 in terms of the layout and functionality of the instrumentation, controls and other cockpit elements. The pilots were instructed to consider the role of the simulator during the evaluations as being training for the vehicle handling only, rather than the fully interactive flight and mission management role that would be experienced in flight. The ASRA, as an airborne simulator, can function in a very similar manner to this role. However, if any crew station aspects influenced the SFR the pilot was required to report these.**

**Of the two pilots, Pilot A was very experienced with the ASRA and had not flown the HELIFLIGHT-R simulator prior to the SFR evaluations. Pilot B, on the other hand, had considerable prior experience with the HELIFLIGHT-R simulator and F-B412 model, but only a small amount of time in the Bell 412 helicopter. In addition, the gap between the most recent Bell 412 flight experience and the simulator evaluations was different for the two pilots – for Pilot A the gap was short (less than one week), while for Pilot B a period of approximately two months separated the two trials. This gap would normally be expected to reduce the validity of the awarded SFRs, and it can indeed be seen in** Figure 4 **that Pilot B has typically been less critical of the simulator than Pilot A. The time interval between flight and simulator tests might mean that Pilot B would only be able to consider the most significant differences, whereas the more recent experiences of Pilot A might allow more subtle, but potentially important, differences to also be considered.**

**As detailed in** Figure 4**, the results from the two pilots generally agree, with the same SFRs being awarded for the ACAH precision hover and accel-decel, and the RCAH lateral reposition MTE. In addition, the results for the RCAH accel-decel were a single rating point apart (and within the same level of fidelity). Greater differences were recorded for the bare airframe flown in the precision hover (although the ratings remained in the same fidelity level), and the RCAH precision hover and pirouette. Note the differences in the flight-simulator HQR comparisons (**Figure 5**) in these cases. For example, while Pilot B recorded the same HQR in both flight and simulator for the RCAH precision hover, Pilot A awarded a Level 1 HQR in the simulator but a Level 2 HQR in flight. These results indicate that the HQR can be an effective tool in the assessment of the utility of a simulator. However, there are other tasks (such as the Accel-Decel in the ACAH configuration) where the HQRs are similar and a poor SFR was awarded. The HQR by itself is insufficient to capture all possible scenarios of fidelity deficiency, emphasising the importance of a dedicated rating scale for the assessment of simulator fidelity.**

**In the cases where different SFRs were awarded by the two pilots, Pilot B considered the simulator to be entirely representative of the flight experience, while Pilot A felt the simulator to be lacking in certain areas. Pilot A reported that the unaugmented flight dynamics were somewhat easier to control in the simulator than was the case in the ASRA. This was noted as being particularly true in heave, where the ASRA exhibits a highly under-damped collective governor dynamic that is very easy to excite during precision tasks. This dynamic was not as evident in the simulator, causing a reduction in apparent workload across all axes. Although Pilot B also reported the difference in the heave and torque dynamics, he did not feel that it warranted degradation of the simulator into Level 2 fidelity for the precision tasks.**

**This difference in the pilots’ ratings are likely to be related to the different backgrounds and lengths of time between flight and simulator tests; the greater in-flight experience of Pilot A helping him to identify simulation deficiencies with conviction. In addition to this, there may be other factors contributing to different perceptions of the flight-simulator comparison. These additional contributions include the accuracy of the simulated atmospheric disturbances and any effect that may result from different piloting techniques. This can be seen in** Figure 5**, which shows that Pilot A consistently awarded better HQRs for the simulated aircraft than Pilot B. This was largely a result of Pilot A adopting a predominantly ‘low gain’ control strategy, taking advantage of the stability of the RCAH and ACAH configurations and the relative lack of simulated atmospheric disturbances. In contrast, Pilot B adopted stronger loop closure with a ‘high gain’ control strategy where he attempted to minimise positional errors at all stages, leading to the dynamic modes of the model being excited, and giving the perception of a higher workload [22].**

**Turning back to the SFR results, Pilot A generally found that the simulator provided a degree of training benefit, although this varied depending on the task and vehicle configuration. However, in no case did Pilot A feel that the simulator provided full ToT for flight in ASRA. Pilot B, on the other hand, offered a more positive evaluation of the simulator, with full ToT from simulator to flight judged to occur for two of the MTEs in the RCAH configuration.**

**Both pilots felt that the simulator was most deficient for the Accel-Decel (in the ACAH configuration) and the Lateral Reposition (in the RCAH configuration). For both manoeuvres, this was reported to be due to restricted visual cueing of longitudinal translational rate. As part of the assessment process, the pilots were asked to award Visual Cue Ratings (VCRs) [11] for each MTE to allow the Useable Cue Environment (UCE) to be determined, and therefore provide an indication of the quality of the visual cues being presented in flight and in the simulator. This approach is a variation on the formal UCE process in ADS-33E-PRF [21] where the VCRs of three pilots flying a number of MTEs are consolidated.**

**Longitudinal translational rate VCRs of 1.5 (aggressive and precise corrections can be made with confidence; achievable precision is good) in flight and 3 (only limited corrections can be made with confidence; achievable precision is fair) in the simulator were awarded by Pilot B for the Accel-Decel task. The discrepancy was judged to be a result of a difference in the vertical field of view (FoV) available directly ahead of the pilot. It should be noted that, as part of a HQ investigation, the Test Guide for ADS-33E-PRF [21] requires that cockpit field of view limits are not considered in a VCR assessment. As the purpose of the current assessments was to identify differences between flight test and simulation, the pilots were instructed to consider all aspects of the visual scene in their VCRs.**

**In the ASRA, the vertical FoV is sufficient directly ahead of the pilot to maintain sight of all required cues for each of the MTEs. In HELIFLIGHT-R, the instrument panel is somewhat higher, and this restricts the vertical FoV by approximately 5 degrees in the forward direction (**Figure 6**). The effect of this difference is to reduce the number of task cues visible to the pilot during the deceleration phase of the Accel-Decel MTE, particularly during the latter stages when the nose-up pitch attitude may reach +30°. In the Lateral Reposition MTE, the pilots’ view of the final hover point cues was restricted in the simulator, but not in flight.**



Figure 6: Fields of View - ASRA and HELIFLIGHT-R

**The SFR results are in line with the previously reported predicted and perceptual fidelity metric analyses of the four manoeuvres [4], which showed a greater degree of control activity in flight; e.g. Figure 5, showing the cut-off frequency [5] in flight and simulator lateral reposition MTEs with the RCAH configuration for Pilot B. The cut-off frequency is a measure of the rate at which the pilot is applying corrective control inputs, and is calculated as the frequency at which 70.7% of the total energy contained within the pilot’s control activity has accumulated [23]. In flight, Pilot B reported a minimal level of compensation with the desired performance objectives easily being met. A HQR=3 was awarded. In contrast, while desired performance could still be achieved in the simulator, the pilot felt that there was an overall reduction in the level of precision, accompanied by a higher level of compensation, particularly so in the longitudinal axis where the above-mentioned visual cueing differences impacted. These differences, including the requirement to utilise the available visual cues differently, led Pilot B to award SFR=6 for this case. This measured difference reinforces the restricted utility of the simulator for training to complete the manoeuvres in flight.**

****

Figure 7: Quantitative Analysis of Lateral Reposition Task: Cut-Off Frequency

**CASE STUDY 2: USE OF THE SFR SCALE TO DEFINE FIDELITY BOUNDARIES**

In addition to the utility of the SFR scale in the direct assessment of the fidelity of an overall simulation, the scale may be used to support the development of new metrics for the quantification of fidelity. A series of ‘predictive’ metrics for flight model fidelity based upon ADS-33E-PRF handling qualities criteria, and ‘perceptual’ metrics for the assessment of overall simulation fidelity have been proposed, and their sensitivity to variations in fidelity demonstrated [4], [5]. However, insufficient data exist to define the location of boundaries between the different levels of fidelity for these metrics.

During the trials supporting the development of the SFR scale, data were gathered to begin the process of locating boundaries for a small selection of the metrics. As described earlier, one of the testing processes was to modify the simulation model with specific deficiencies, and to use these to award SFR ratings against a baseline simulation. An example was the addition of inter-axis coupling between the pitch and roll axes, and *vice versa*. This effect was created by feeding a proportion of the pilot’s longitudinal control input into the lateral control channel, and a proportion of the pilot’s lateral control input into the longitudinal control channel [22].

A second example of these tests has been the introduction of progressively larger time delays between the pilot’s control inputs and their actions being applied to the rotor cyclic and collective controls [24].

The result of increasing the time delay, or proportion of inter-axis coupling, was to degrade the handling qualities of the flight model in a controlled manner. As the HQs of the modified model were degraded, the perceived difference between the baseline and modified models increased, leading to the award of progressively more degraded SFRs. Through correlation of the degradation in SFR across level boundaries with the values of the various predictive and perceptual metrics, the change in the metrics at the boundary crossing points can be determined. With sufficient data, accurate mapping of the fidelity boundaries may be achieved.

While data gathering is ongoing, the results so far are promising, and highlight the utility of the SFR scale within the process of creating quantitative fidelity requirements.

In the case of the additional inter-axis coupling, a series of tests were performed with 20%, 40% and 50% of the longitudinal control input being added to the lateral control channel (and *vice versa*) of the RCAH-configured F-B412. In the model these additional dynamics correspond to inter-axis roll-from-pitch coupling ratios of approximately 0.4, 0.8 and 1.0 respectively (Figure 8), where the ratio is defined as

*Δφpk* is the largest roll perturbation and *Δθ4* the final pitch attitude change following a four second longitudinal cyclic step control input. The inter-axis coupling present in the baseline RCAH F-B412 is minimal (coupling ratio < 0.03).



Figure 8: Predicted Inter-Axis Coupling Handling Qualities (non-tracking tasks)

Figure 9 shows the SFRs awarded by Pilot B for the modified F-B412 while performing an Accel-Decel task, while Figure 10 shows the HQRs awarded by the same pilot for each configuration. The pilot did not report an adverse effect on his ability to perform the task with the introduction of the 20% coupling, meaning a requirement for negligible adaptation to maintain consistent task performance. Figure 10 shows that there was a small degradation in the HQR between the baseline model and this configuration. As the artificial coupling strength was increased from 20% to 40%, the pilot noted a moderate adaptation of his task strategy compared to the baseline simulation, accompanied by a reduction in the precision with which the task could be performed. These characteristics led the pilot to award an SFR=5 for this model – Level 2 fidelity. Although the additional cross-coupling resulted in an increase in the required compensation, the pilot was still able to achieve the desired level of performance, leading again to the award of HQR=4. This result once again highlights the insufficiency of the HQR scale for the detection of fidelity deficiencies. Finally, when the coupling strength was increased to 50%, the pilot needed to apply an excessive level of adaptation in order to perform the task, leading to an SFR=7, in the Level 3 region. Simultaneously, the awarded HQR also degraded into the Level 3 region, with HQR=7 being awarded for this configuration.



Figure 9: Degradation in SFRs with Increased Artificial Cross-Coupling – Accel-Decel Task



Figure 10: Degradation in HQRs with Increased Artificial Cross-Coupling - Accel-Decel Task

While this is a very limited data set, and intermediate points would be required to refine the results, the trend in the degradation of SFRs seen in Figure 9 is clear. Using this data, it could be hypothesised that the Level 1-2 fidelity boundary will lie at a roll-from-pitch coupling ratio error of approximately 0.5 between the flight test aircraft and the simulation model. Likewise, the Level 2-3 fidelity boundary will lie at a coupling ratio error of approximately 1.0.

It should be noted that these results apply to one particular aircraft configuration which does not exhibit inter-axis coupling in its baseline form. It is likely that the results would vary depending on the nature of the MTEs being performed and the baseline HQs of the test vehicle (generally excellent HQs could expose smaller simulation deficiencies; conversely, generally poor HQs could make simulation deficiencies more critical). Therefore, a much more extensive range of tests, encompassing different baseline vehicle dynamics, a greater resolution of test points and a range of pilots and flying tasks would be required to define the boundaries for this single predictive fidelity metric more accurately.

**DISCUSSION**

The case studies have illustrated the use of the SFR scale in two distinct applications at UoL/FRL. In both cases, the scale was shown to be an effective tool. However, important lessons for the future use of the scale can be drawn from each example. The first study highlighted the importance of conducting the flight and simulation tests within a short period of time, ensuring fresh recall of flight characteristics when conducting the simulation evaluations. The results also indicated benefit from use of pilots with considerable experience of the vehicle, and the importance of matching the prevailing flight test atmospheric conditions in the simulator. Differences in the HQRs awarded in flight and simulator generally correlated well with the SFRs, but there were occasions when similar HQRs were accompanied by poor SFRs. While the HQR and associated narrative provides useful insight into the behaviour of the vehicle in flight and simulator, it is not by itself sufficient to capture all fidelity deficiencies.

In the second case study, Figure 9 presents a picture of degradation in the simulator fidelity as the difference between ‘flight’ and ‘simulation’ increased. However, this result was achieved with a very small data sample, and additional tests would be required to confirm the accuracy of the predicted fidelity boundaries. The HQRs awarded during these tests degrade in a generally similar manner to the SFRs; again, conditions are observed where the HQR fails to detect a significant change in fidelity.

Guidance notes for the effective use of the SFR scale, covering these and other topics, are currently under development, and will be made available to facilitate the employment of the scale. The guidance notes will outline the range of applications of the SFR scale with case studies for each application. Information for test engineers on the definition of fidelity assessment tasks; test planning; pilot briefing/de-briefing and anomaly detection will assist new users in the generation of valid SFRs. Notes on pilot selection and training will be provided. Information for the evaluation pilots will also be included within the guidance notes. The process of completing the fidelity rating questionnaire and awarding an SFR will be highlighted within the case studies. The parameters for consideration during the rating procedure will be detailed with examples, and the definitions of the qualitative descriptors will be discussed.

**CONCLUDING REMARKS**

**This paper has presented the evolution of a new rating scale for the subjective assessment of flight simulator fidelity – the SFR Scale, its development, and its use in the evaluation of training effectiveness and the identification of level boundaries for predictive fidelity metrics.**

**Existing simulator certification processes are considered to be lacking in the rigour of the subjective assessments of the fidelity of the overall, integrated simulation experience. However, previous attempts at development of a fidelity rating system have not led to widely accepted practice.**

**The handling qualities rating scale structure has proved a firm foundation on which to build a new SFR scale. The use of comparative task performance and task strategy adaptation as the two dimensions of the fidelity rating allows the evaluating pilot to provide feedback on aspects of the simulation that are directly experienced, and limit the degree of introspection required to award the rating. The degree of transfer of training makes a suitable operant definition of “fidelity”, and can be used as the differentiator of the three levels of fidelity – full transfer signifies Level 1; partial transfer signifies Level 2 and negative (or adverse) transfer signifies Level 3 fidelity.**

**The use of the SFR scale in an evaluation of the HELIFLIGHT-R simulator produced a number of important outcomes related to the award of robust fidelity ratings:**

1. **The level of experience of the evaluating pilot in the test aircraft, and the length of time between the flight and simulator tests, are likely to have a significant impact on the accuracy of the SFR results.**
2. **A good match of the flight test conditions must be achieved in the simulator for accurate ratings.**
3. **Use of the SFR scale should be limited to trained practitioners to ensure authentic ratings.**

**Case Study 1 presented results for an evaluation of the fidelity of a research simulator for performance of a series of ADS-33E-PRF MTEs. In the majority of cases, the SFRs awarded were consistent between the evaluating pilots. In the cases where the ratings differed significantly, it is believed that the greater time interval between flight and simulator experiences of the less critical pilot contributed to a less valid rating, highlighting the importance of recent flight experience in the award of accurate fidelity ratings. The awarded SFRs were shown to correlate with quantitative analysis of the differences between the flight and simulator task performance and control activity.**

**Although the primary focus of the SFR scale is in the direct assessment of simulator utility, the use of the scale in the development of quantitative fidelity metrics has been demonstrated. Case Study 2 demonstrated the process whereby subjective fidelity ratings can be employed to correlate quantitative differences with their impact on pilot training. The results of this study showed a strong match between the magnitude of an artificially-introduced cross-coupling (between pitch and roll) and simulation fidelity. These results enabled an estimate to be made of the allowable cross-coupling error between flight and simulation for Level 1 and Level 2 fidelity. However, the case study constitutes a single pilot flying a single baseline vehicle configuration in a single task. Work is ongoing to expand the available dataset of test results so as to improve the validity of the calculated fidelity level boundaries.**

**While this paper presents a mature version of the SFR scale that has been successfully used by a range of test pilots across flight and simulation, the development and verification process is not yet complete. In particular, the scale has been used exclusively for evaluations of a full-flight rotorcraft research simulator. It is envisaged that the SFR scale will have a much wider utility, encompassing fixed-wing simulation, part-task and procedural trainers. Indeed, beyond flight simulation, the SFR scale could potentially be deployed for the evaluation of medical simulation, virtual realities and so forth.**

**ACKNOWLEDGEMENTS**

The research reported in this paper was funded by the UK EPSRC through grant EP/G002932/1 and the US Army International Technology Center (UK) (reference W911NF-11-1-0002). The work also contributes to the University’s Virtual Engineering Centre, supported by the European Regional Development Fund. The contributions of test pilots Andrew Berryman (British Airways), and Lt. Cdr. Lee Evans (RN), Lt. Cdr. John Holder (RN), Lt. Christopher Knowles (RN) and Flt. Lt. Russ Cripps (RAF) of the UK Rotary Wing Test and Evaluation Squadron to the development, test and improvement of the SFR scale are gratefully acknowledged. The authors would like to thank all at the NRC-FRL, especially test pilot Stephan Carignan, for their contributions to the flight testing phase of this work.

**REFERENCES**

1. anon., “JAR-FSTD H, Helicopter Flight Simulation Training Devices”, Joint Aviation Authorities, May 2008.
2. anon., “FAA Advisory Circular AC120-63, Helicopter Simulator Qualification”, Federal Aviation Authority, November 1994.
3. anon., “JAR-STD 1H, Helicopter Flight Simulators”, Joint Aviation Authorities, April 2001.
4. White, M.D., Perfect, P., Padfield, G.D., Gubbels, A.W. and Berryman, A.C., “Progress in the development of unified fidelity metrics for rotorcraft flight simulators”, 66th Annual Forum of the American Helicopter Society, Phoenix, AZ, USA, May 2010.
5. Perfect, P., White, M.D., Padfield, G.D., Gubbels, A.W. and Berryman, A.C., “Integrating Predicted and Perceived Fidelity for Flight Simulators”, 36th European Rotorcraft Forum, Paris, France, September 2010 (Accepted for publication in the The Aeronautical Journal).
6. Szalai, K.J., “Validation of a General Purpose Airborne Simulator for Simulation of Large Transport Aircraft Handling Qualities”, NASA TN-D-6431, October 1971.
7. Cooper, G. E. and Harper, Jr. R. P., “The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities”, NASA TN-D-5153, April 1969.
8. Hagin, W.V., Osborne, S.R., Hockenberger, R.L., Smith, J.P. and Gray, T.H., “Operational Test and Evaluation Handbook for Aircrew Training Devices: Operational Effectiveness Evaluation”, AFHRL-TR-81-44(ii), February 1982.
9. Advani, S.K. and Wilkinson, C.H., “Dynamic Interface Modelling and Simulation – A Unique Challenge”, The Challenge of Realistic Rotorcraft Simulation, RAeS Conference, London, November 2001.
10. Zivan, L. and Tischler, M.B., “Development of a Full Flight Envelope Helicopter Simulation Using System Identification”, Journal of the American Helicopter Society, Vol. 55, pp. 22003 1-15.
11. anon., “ADS-33E-PRF, Handling Qualities Requirements for Military Rotorcraft”, US Army, March 2000
12. Key, D.L., Blanken, C.L. and Hoh, R.H., “Some lessons learned in three years with ADS-33C”, in ‘Piloting Vertical Flight Aircraft’, a conference on flying qualities and human factors, NASA CP 3220, 1993.
13. White, M.D., Perfect, P., Padfield, G.D., Gubbels, A.W. and Berryman, A.C., “Acceptance testing of a rotorcraft flight simulator for research and teaching”, Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering DOI: 10.1177/0954410012439816, March 2012
14. Gubbels, A.W. and Ellis, D.K., “NRC Bell 412 ASRA FBW Systems Description in ATA100 Format”, Institute for Aerospace Research, National Research Council Canada, Report LTR-FR-163, April 2000.
15. anon., “Manual of Criteria for the Qualification of Flight Simulation Training Devices, Volume 2 – Helicopters (Draft)”, ICAO 9625, Volume 2, 3rd Edition, 2011.
16. Jennings, M., “Introduction to, and Overview of IWG on Helicopter FSTDs”, The Challenges for Flight Simulation – The Next Steps, RAeS Conference, London, November 2010.
17. Phillips, M., “The H-IWG Training Matrix: Unlocking the Mystery”, The Challenges for Flight Simulation – The Next Steps, RAeS Conference, London, November 2010.
18. Go, T. H., Bürki-Cohen, J., Chung, W. W., Schroeder, J., Saillant, G., Jacobs, S., and Longridge, T., “The Effects of Enhanced Hexapod Motion on Airline Pilot Recurrent Training and Evaluation”, AIAA Modeling and Simulation Technologies Conference, AIAA, 2003.
19. Heffley, R. K., et al, 'Determination of motion and Visual System Requirements for Flight Training Simulators', Systems Technology Inc., Technical Report No. 546, August 1981.
20. DuVal, R.W., “A Real-Time Multi-Body Dynamics Architecture for Rotorcraft Simulation”, The Challenge of Realistic Rotorcraft Simulation, RAeS Conference, London, November 2001.
21. Blanken, C.L., Hoh, R.H., Mitchell, D.G. and Key, D.L., “Test Guide for ADS-33E-PRF”, Special Report AMR-AF-08-07, US Army AMRDEC, July 2008.
22. Timson, E., Perfect, P., White, M.D., Padfield, G.D. and Erdos, R., “Pilot Sensitivity to Flight Model Dynamics in Flight Simulation”, 37th European Rotorcraft Forum, Gallarate, Italy, September 2011.
23. Tischler, M.B. and Remple, R.K., “Aircraft and Rotorcraft System Identification: Engineering Methods with Flight Test Examples”, AIAA Education Series, Reston, Virginia, 2006.
24. Timson, E., Perfect, P., White, M.D., Padfield, G.D., Erdos, R. and Gubbels, A.W., “Subjective Fidelity Assessment of Rotorcraft Flight Training Simulators”, 68th Annual Forum of the American Helicopter Society, Fort Worth, TX, USA, May 2012.