

# A RATING SCALE FOR SUBJECTIVE ASSESSMENT OF SIMULATOR FIDELITY<sup>1</sup>

Philip Perfect<sup>2</sup>,  
Emma Timson, Mark D. White,  
Gareth D. Padfield  
The University of Liverpool  
Liverpool, U.K

Robert Erdos,  
Arthur W. Gubbels  
Flight Research Laboratory,  
National Research Council,  
Ottawa, Canada

Andrew C. Berryman  
Consultant Test Pilot, UK

## ABSTRACT

A new quality rating scale for the subjective assessment of simulation fidelity is introduced 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 and use of the new rating scale in an evaluation of the fidelity of the Liverpool HELIFLIGHT-R research simulator, in conjunction with the Flight Research Laboratory's ASRA in-flight simulator, is described.

## NOTATION

ACAH Attitude Command, Attitude Hold  
ASRA Advanced Systems Research Aircraft  
ATD Aircrew Training Device  
FRL Flight Research Laboratory  
HQs Handling Qualities  
HQR Handling Qualities Rating  
IAI Israel Aircraft Industries  
IWG International Working Group  
MTE Mission Task Element  
NRC National Research Council of Canada  
RAeS Royal Aeronautical Society  
RCAH Rate Command, Attitude Hold  
SFR Simulator Fidelity Rating  
ToT Transfer of Training  
UCE Useable Cue Environment  
UoL University of Liverpool  
VCRs Visual Cue Ratings

## INTRODUCTION

The evaluation of the fidelity of a simulation device for flight training is typically based upon the ability to meet a series of quantitative requirements contained within simulator qualification documents such as JAR-FSTD H [1]. 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 fixed, 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 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 fill this gap, the qualification documents 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”. 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 [2] (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 ill-defined, and is most certainly open to interpretation by the operator and qualifying body.

It is suggested that the existing requirement for the subjective aspect of simulator qualification is unsatisfactory and should be improved. It is proposed that an effective method by which the process may be

<sup>1</sup> Presented at the 37th European Rotorcraft Forum, Gallarate, Italy, 2011

<sup>2</sup> Corresponding author, Email: p.perfect@liv.ac.uk; Tel: +44 (0) 151 7946844

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 Simulator 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. Also, 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 support and 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 [3] and [4]. This stage of the fidelity evaluation would be completed after the 'predicted fidelity' analysis examined each individual component of the simulator.

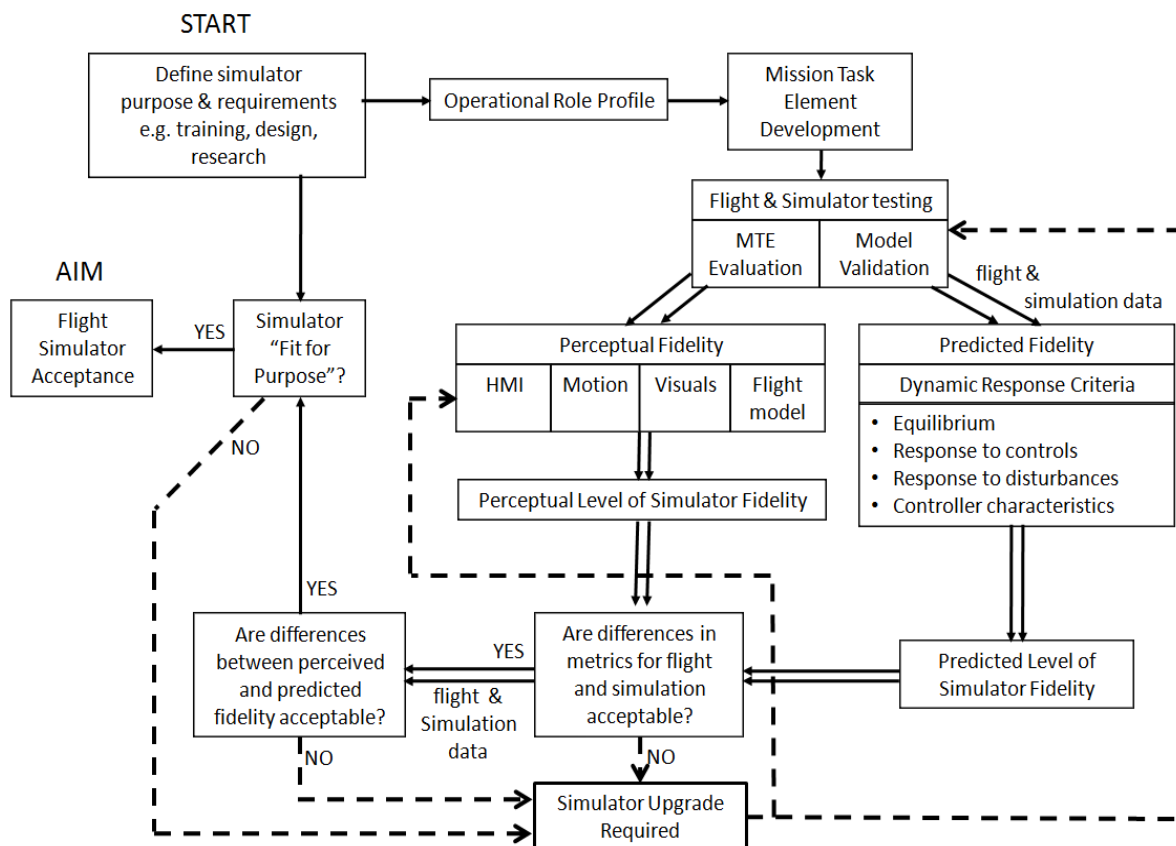


Figure 1: Methodology for the assessment of simulator fidelity

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 it is envisaged the new scale may be used in an evaluation of a training simulator is described. A selection of results from the developmental trials is presented to demonstrate the function of the scale before the paper is brought to a close with some concluding remarks, including future developments and consideration of wider application for the new scale.

## REVIEW OF EXISTING FIDELITY RATING SCALE CONCEPTS

An early fidelity rating scale was developed during a NASA programme to validate the General Purpose Airborne Simulator [5]. The scale (Table 1) was configured to be similar in format and considerations to the Cooper-Harper Handling Qualities Rating (HQR) Scale [6]. While it should be noted that the scale was initially developed to support evaluation of the suitability of the Airborne Simulator 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 HQs, aircraft development etc.

Table 1: Early Simulator Fidelity Rating Scale [5]

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.

More recently a discussion on the use of subjective rating in the evaluation of simulation devices was compiled by the US Air Force [7]. Here, the effectiveness of a wide variety of schemes for subjective evaluation was considered. Two examples are discussed below.

The first relates to the rating of the cues provided by a simulator, and is summarised in Table 2. 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 were also discussed in [7]. 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.

Table 2: Example of Subjective Rating for Task Cues [7]

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.

A second type of rating considered in [7] is 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 [7]

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

To bring the story of the fidelity rating up to date, a six-point scale was developed by Israel Aircraft Industries (IAI) to support the development of a Bell 206 simulator [8]. 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
- Controls sensitivity
- Controls position
- Secondary response direction
- Etc.

Thus, this rating system allows the evaluating pilot to specify the precise areas where the simulation is deficient, but does not offer a consideration of the overall, integrated simulation.

While no single rating scale has yet gained formal acceptance for the evaluation of simulator fidelity, this review demonstrates that there are many different methods that can be applied to fidelity assessment, and again, no method 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 immediately see whether the device is acceptable or not, but require the evaluating pilot to make judgements on a high level regarding the overall suitability of the simulator.

## RATING SCALE DEVELOPMENT

The lack of an internationally-accepted fidelity rating system led to a new initiative, launched 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), to develop a new Simulator Fidelity Rating (SFR) scale to address the subjective evaluation gaps in the existing simulator qualification processes.

### *Initial development process*

As the HQR scale [6] 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) [9] with appropriate performance standards. For example, if the simulator is to be used to practice ship landings in rough weather, then task performance criteria 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) [9] 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 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 [10].

Using these principles, a series of ‘straw-man’ SFR scales were drawn up at UoL and FRL, and their advantages and disadvantages were examined during an exploratory trial with one test pilot, using the UoL HELIFLIGHT-R full-motion flight simulator [11]. 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 combined 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 test 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 cross-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 [12] in April 2011. Two of the test pilots who had already been involved in the SFR scale development took part in the flight tests. This flight trial allowed the scale to be used for flight vs simulation comparisons, and the unique airborne simulation capabilities of the ASRA were 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. Results from the flight and simulation configuration change experiments are reported by Timson *et al* [13], and so will not be discussed in detail in this paper.

At the conclusion of the flight tests, the final (to date) modifications to the SFR scale had been made. A further piloted simulation trial was held at UoL, involving two additional test pilots. No further developments of the scale were found to be necessary during or following this trial.

#### *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 67<sup>th</sup> Annual Forum of the American Helicopter Society. 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 a series of discussion sessions where various aspects of a fidelity rating scale were debated, including ‘what should the

pilot be rating?’, ‘what makes a simulation pilot?’ and ‘what test methods are appropriate?’ (Ref [14]).

## **THE SIMULATOR FIDELITY RATING SCALE**

The proposed SFR scale employs a number of key concepts that are considered to be fundamental to the utility, and determination of the efficacy, 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 – comparison of the precision with which a task is completed in flight and simulator.
- Task strategy adaptation – the degree to which the pilot is required to modify their behaviours when transferring from simulator to flight.

The relationship between task performance and strategy adaptation is similar to that between performance and compensation in the HQR scale. 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 to obtain in measuring HQ’s, in the context of fidelity assessment task performance and adaptation will not necessarily change in correlation with each other, depending on the nature of the fidelity deficiencies present in a simulator.

A matrix combining all possible combinations of comparative performance and task strategy adaptation was constructed (Figure 2), and it is this which forms the basis of the SFR scale (Figure 3) in its present form.

		Comparative Performance		
		Equivalent	Similar	Not Similar
Strategy Adaptation	Negligible	LEVEL 1 (SFR 1-2)		LEVEL 3 (SFR 7-9)
	Minimal			
	Moderate	LEVEL 2 (SFR 3-6)		
	Considerable			
	Excessive			

*Figure 2: Fidelity Matrix*

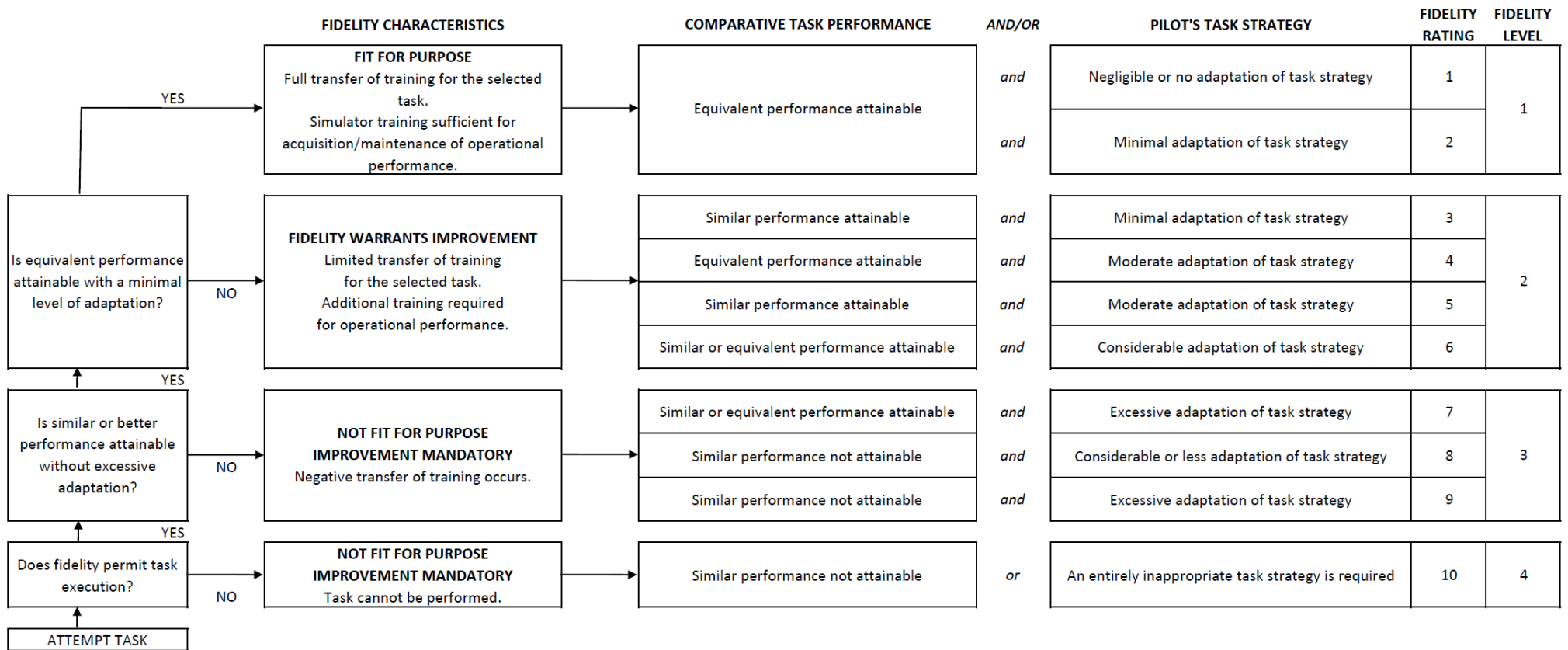


Figure 3: Simulation Fidelity Rating Scale

Each of the ratings SFR=1 to SFR=9 corresponds to a region in the fidelity matrix. 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 in 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 simulator 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, rather than to comprehensively assess an overall system. 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) of 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 allows the range of tasks that will be flown using the simulator to be defined, and hence the scope of the SFR evaluations. Each task identified in the previous step would be assessed on an individual basis, the results then being combined to produce an overall evaluation of the simulator.

In the context of a training simulator (the scenario used in the development of the SFR scale), 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 takes place. This is in contrast to the work of Bürki-Cohen *et al*, where metrics for the measurement of ToT have been developed [18]. The four 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 operational performance in the aircraft. There is limited positive ToT from the simulator to the aircraft in this task.
- Level 3 fidelity: Negative ToT occurs, meaning that the simulator is not suitable for training to fly the aircraft in this task.
- Level 4 fidelity: The task cannot be performed in the simulator. The simulator in no way prepares the pilot for this task in the aircraft.

The task may be defined as the training manoeuvre/procedure, accompanied by a set of performance requirements. In a 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 (e.g. 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, where the comparison of the achieved desired/adequate/beyond adequate 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.
- 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, they were still able to achieved desired task performance in all parameters, but had degraded from a point where desired performance was easily achieved to being just capable of achieving desired

performance, leading to the pilots wanting to degrade the fidelity rating to Level 2 (as with a HQR=4 – desired performance but with moderate compensation). In the more recent simulation trials, the pilots have been allowed a greater degree of flexibility, to make their own decisions regarding whether a deviation is small, moderate or large. While 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, a question remains over the consistency with which ratings are returned if each pilot is applying their own interpretation. A second area where the pilots have been asked to make a qualitative distinction is in strategy adaptation. This is intended to capture all aspects of a pilot's behaviour, including, but not limited to:

#### Control Inputs:

- Input size
- Input direction
- Input shaping
- Input frequency

#### Perceived use of cues:

- Visual
- Vestibular
- Proprioceptive
- Auditory

Any other aspects of the task, beyond the achieved level of performance, that are perceived to be different between the simulation and flight test runs 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 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 at simulator fidelity assessment may be limited to trained practitioners only.

A final aspect of SFR terminology is that of 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.

#### *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 must 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 (based on that for the aircraft on which the simulator is modelled), 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 pilot, or 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 will frequently 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 flight in the aircraft. An essential aspect of either evaluation method, however, must be that the time period between the flight and simulator experiences should be short, and uncontaminated with



other aircraft or simulator types, such that the memory of the first system remains 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 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. Repeating test points should be discouraged, as this allows the evaluating pilot to modify their strategies in response to differences between the simulator and the aircraft, which may thus not be reflected in the final SFR. Instead, the SFR should be based largely upon the first attempt at a specific test point, forcing the pilot to adopt the strategies learned during the familiarisation period, and therefore driving the assessment of adaptation and performance to be based on the first, immediate impression of the differences that may exist.

For any fidelity evaluation, but especially in the case of ratings in Levels 2 or 3, the justification behind the rating given by the evaluating pilot is critical. The identification of the specific simulator deficiencies allows the simulator engineer to determine 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 evaluation of each of the individual training tasks, the fidelity of the simulator in its overall role can be considered. In the likely 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 zero additional training – ‘zero flight time’, while for those tasks awarded Level 2 SFRs, the simulator may still be used, but in the knowledge that the trainee will still 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 onto trainees.

## FIDELITY RATING SCALE RESULTS

The two test pilots who participated in the flight tests on the Bell 412 ASRA awarded SFRs for the comparison of flight against a FLIGHTLAB [19] simulation of the Bell 412 running on the UoL HELIFLIGHT-R simulator [11]. The comparisons were made during simulation trials (i.e. the flight testing had been completed first). The results, with the ASRA and simulation model configured to offer a variety of different response types (attitude command, attitude hold – ACAH; rate command, attitude hold – RCAH; and unaugmented – Bare) are presented below (Table 4) for a selection of Mission Task Elements (MTEs) chosen from the US Army HQs design specification, ADS-33E-PRF [9]. While these MTEs are primarily designed to highlight handling deficiencies in an aircraft, 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.

Table 4: Sample of SFR Results

		SFRs	
		Pilot A	Pilot B
Precision Hover	ACAH	3	3
	RCAH	4	2
	Bare	5	3
Pirouette	RCAH	5	2
Lateral Reposition	RCAH	6	6
Accel-Decel	ACAH	6	6
	RCAH	4	3

As HELIFLIGHT-R is a research simulator, it features a generic crew station, and is therefore not fully 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 vehicle management role that would be experienced in flight. The ASRA, as an airborne simulator, can function in a very similar manner to this role, with the Safety Pilot managing the flight.

Of the two pilots, Pilot B had considerable prior experience with the HELIFLIGHT-R simulator, but only a small amount with the Bell 412. Pilot A, on the other hand, was very experienced with the ASRA, but had not flown the HELIFLIGHT-R simulator prior to the SFR evaluations. 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.

As detailed in Table 4, the results from the two pilots 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 in the unaugmented precision hover (although the ratings remained in the same fidelity Level), and the RCAH precision hover and pirouette, where 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, particularly 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 due to the strongly cross-coupled nature of the unaugmented vehicle. 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 may 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 increased confidence. The impact on the fidelity results of varying levels of pilot experience should be further evaluated during the continuing development of the SFR scale. In addition, differences may have been caused by different atmospheric conditions on the days of the flight tests. All of the simulation runs were performed in constant, steady wind conditions appropriate to the flight tests. However, in flight these winds would have been continuously varying, would have contained a turbulence component, and may have had

a gusting element. A difference in the perceived level of disturbances during flight between the two pilots would result in differences in the flight-simulator comparison. This would be especially so with the unaugmented aircraft, which lacks the stabilisation systems of the other configurations, hence requiring the pilot to work harder to suppress disturbances. This highlights the importance of achieving an accurate match of the flight test conditions during the simulator evaluation process.

Turning to the SFR results themselves, 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 two of the MTEs in the RCAH configuration being considered to offer full ToT from simulator to flight.

Both pilots felt that the simulator was most deficient for the Accel-Decel (in the ACAH configuration) and the Lateral Reposition. For both manoeuvres, this was reported to be due to restricted visual cueing of longitudinal translational rate (VCRs [9] of 1.5 in flight and 3 in the simulator from Pilot B, for example). This is judged to be a result of a difference in the vertical field of view (FoV) available directly ahead of the pilot. In the ASRA, the FoV is good, whereas in HELIFLIGHT-R the instrument panel is somewhat higher, and this restricts the downwards FoV by approximately 5 degrees (Figure 4). 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  $+30^\circ$ . In the Lateral Reposition MTE, the pilots' view of the final hover point cues was restricted in the simulator, but not in flight.

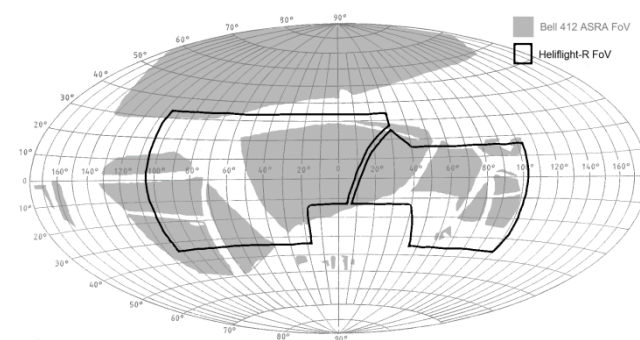


Figure 4: Fields of View - ASRA and HELIFLIGHT-R

The SFR results are in line with the previously reported numerical predicted and perceptual fidelity analyses of the four manoeuvres (Figure 1, [3]), which has shown a greater degree of control activity in flight (e.g. Figure 5, showing the cut-off frequency [4] in flight and simulator lateral reposition MTEs with the RCAH configuration), and therefore confirm a restricted utility of the simulator for training to complete the manoeuvres in flight.

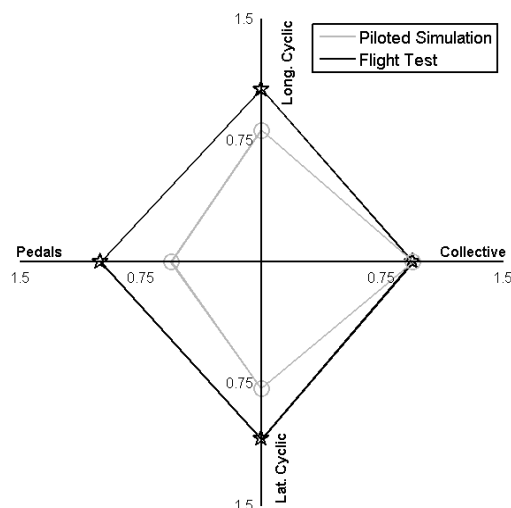


Figure 5: *Quantitative Analysis of Lateral Reposition Task: Cut-Off Frequency*

## CONCLUDING REMARKS

This paper has presented a new rating scale for the subjective assessment of flight simulator fidelity – the SFR Scale, its development, and initial use in an evaluation of the fidelity of the Liverpool HELIFLIGHT-R research flight simulator. Some of the preliminary conclusions that can be drawn from the work to date are as follows:

1. Existing simulator certification processes are lacking in the rigor of their subjective assessments of the fidelity of the overall, integrated simulation experience.
2. Previous attempts at development of a fidelity rating system have not led to widely accepted practice.
3. The handling qualities rating scale structure has proved a firm foundation on which to build a new SFR scale.
4. 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.
5. 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. Even with an evaluating pilot who is highly experienced with the test aircraft, prolonged exposure to the simulator restricts the pilot’s ability to identify the impact of fidelity deficiencies on his strategy. Simulator evaluation should ideally be limited to short periods, and re-familiarisation with the test aircraft should occur between the simulator sessions.
4. Use of the SFR scale should be limited to trained practitioners to ensure accurate ratings.

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 full-flight rotorcraft simulators. It is envisaged that the SFR scale will have a much wider utility, encompassing fixed wing simulation, part task and procedural trainers (potentially all the way down to the level of evaluating the fidelity of pressing a button). 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 is 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 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., "JAR-STD 1H, Helicopter Flight Simulators", Joint Aviation Authorities, April 2001.
- [3] 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", 66<sup>th</sup> Annual Forum of the American Helicopter Society, Phoenix, AZ, USA, May 2010.
- [4] Perfect, P., White, M.D., Padfield, G.D., Gubbels, A.W. and Berryman, A.C., "Integrating Predicted and Perceived Fidelity for Flight Simulators", 36<sup>th</sup> European Rotorcraft Forum, Paris, France, September 2010.
- [5] Szalai, K.J., "Validation of a General Purpose Airborne Simulator for Simulation of Large Transport Aircraft Handling Qualities", NASA TN-D-6431, October 1971.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] anon., "ADS-33E-PRF, Handling Qualities Requirements for Military Rotorcraft", US Army, March 2000
- [10] 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.
- [11] 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: the importance of unified metrics", 35<sup>th</sup> European Rotorcraft Forum, Hamburg, Germany, September 2009.
- [12] 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.
- [13] Timson, E., Perfect, P., White, M.D., Padfield, G.D. and Erdos, R., "Pilot Sensitivity to Flight Model Dynamics in Flight Simulation", 37<sup>th</sup> European Rotorcraft Forum, Gallarate, Italy, September 2011.
- [14] anon., "Proceedings of Workshop on Simulation Fidelity", Hosted at the 67<sup>th</sup> Annual Forum of the American Helicopter Society, 2<sup>nd</sup>-3<sup>rd</sup> May 2011. May be accessed through <http://flightlab.liv.ac.uk/fidelity.htm>
- [15] anon., "Manual of Criteria for the Qualification of Flight Simulation Training Devices, Volume 2 – Helicopters (Draft)", ICAO 9625, Volume 2, 3<sup>rd</sup> 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: Unlicking 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] DuVal, R.W., "A Real-Time Multi-Body Dynamics Architecture for Rotorcraft Simulation", The Challenge of Realistic Rotorcraft Simulation, RAeS Conference, London, November 2001.