This paper provides background information on the development and testing on the NOTECHS system (1), designed by pilots and psychologists for assessing European pilots’ non-technical skills. The international aviation regulators have, in recent years, mandated non-technical skills training for pilots in the form of CRM courses (2,3,4); consequently, the aviation industry is particularly advanced with regard to NTS training and assessment methods. (5) In the 1990s, the Federal Aviation Administration in the USA introduced the Advanced Qualification Program (AQP). (6) This enabled airlines to develop their own CRM training programmes but they also had to demonstrate to the regulator that these were evaluated. In Europe, the regulator had the following requirement: “the flight crew must be assessed on their CRM skills in accordance with a methodology acceptable to the Authority and published in the Operations Manual. The purpose of such an assessment is to: provide feedback to the crew collectively and individually and serve to identify retraining; and be used to improve the CRM training system.” (7)

The European regulator (then Joint Aviation Authorities (JAA), now European Aviation Safety Agency, EASA) sought a generic method for evaluating pilots’ CRM/ non-technical skills that could be applied on a pan-European basis. The technique would have to possess minimal sensitivity to cultural and corporate differences, and be practical and effective for airline instructors and examiners. In response, the JAA Human Factors group commissioned, in 1996, a research project that was sponsored by four European Civil Aviation Authorities. A research consortium consisting of airline pilots and psychologists from Germany, France, the Netherlands and the UK was established to work on the NOTECHS study. This consortium had to identify or develop a feasible, efficient method for assessing an individual pilot’s non-technical (CRM) skills. The system was to be used to assess the skills of an individual pilot, rather than a crew, and it was to be suitable for use across Europe, by both large and small operators, that is, it was to be culturally robust.

A review of behaviour-rating systems for pilots, that were already being used by the larger European and American airlines, (8) showed that none of them could be adopted in its original form. In addition, no single system provided a suitable basis for simple amendment that could be taken as an acceptable method to comply with the regulations because the scrutinised systems were either too complex to be used across Europe, too specific to a particular airline or had been designed to assess whole crews rather than individual pilots. Therefore, a new taxonomy and behaviour-rating system for pilots’ NTS was required.

The development method undertaken by the NOTECHS project group employed various forms of task analysis, including literature reviews and examination of existing behavioural marker systems for assessing CRM skills of pilots. Special attention was paid to Helmreich et al.’s system (9) due to its systematic development process. Airline captains, who had considerable experience of using behaviour-rating methods to assess pilots, acted as subject matter experts and advised on practical requirements. The resulting NOTECHS system (1,10) had four categories: Situation Awareness; Decision Making; Cooperation; Leadership and Managerial skills, each with component elements. Examples of observable behaviours (behavioural markers) were provided to illustrate good and poor performance on each element.

To ensure that each pilot would receive as fair and objective an assessment as possible with the
NOTECHS system, a set of operational principles was established (1,10) and these are listed below.

1. Only observable behaviour is to be assessed – The evaluation must exclude reference to a crew member’s personality or emotional attitude and should be based only on observable behaviour. Behavioural markers were designed to support an objective judgement.

2. Need for technical consequence – For a pilot’s NTS to be rated as unacceptable, flight safety must be actually (or potentially) compromised. This requires a related objective technical consequence.

3. Acceptable or unacceptable rating required – The JAR-OPS (European aviation operational requirements) requires the airlines to indicate whether the observed NTS are acceptable or unacceptable.

4. Repetition required – Repetition of unacceptable behaviour during the cheque must be observed to conclude that there is a significant problem. If, according to the JAR-paragraph concerned, the nature of a technical failure allows for a second attempt, this should be granted, regardless of the non-technical rating.

5. Explanation required – For each category rated as unacceptable, the examiner must: a) indicate the element(s) in that category where the unacceptable behaviour was observed; b) explain where the observed NTS (potentially) led to safety consequences; and c) give a free-text explanation on each of the categories rated unacceptable, using standard phraseology.

It was acknowledged that this type of NTS evaluation based on observing a pilot’s in-flight task performance, which requires judging behaviours, would, in general, be more subjective than judging technical performance data (e.g., speed or flap settings). The NOTECHS designers endeavoured to produce a rating system that would minimise ambiguities in the evaluation of NTS and emphasised that several considerations should be noted by system users. (1,10)

1. The first related to the unit of observation and measurement, that is, the NOTECHS system was designed to be used to assess individual pilots rather than a crew composed of a captain and a co-pilot. When rating an entire crew as a unit, it can be difficult to disentangle and rate the individual contributions to overall crew performance. It should be noted that the same issue already exists during operational and licensing checks when considering technical performance of pilots. The NOTECHS system did not solve this problem in some magical fashion; rather, the designers proposed that the system should assist the examiners to objectively point to behaviours that are related more to one crewmember than the other, thus allowing them to differentiate their judgements of the two crew members.

2. A second factor related to the possible concern that raters might not be judging the pilots’ NTS on an appropriate basis. However, the NOTECHS system requires the instructor/examiner to justify the ratings and any associated criticisms at a professional level, and with a standardised vocabulary. Furthermore, the assessor’s judgement of NTS should not be based on a vague global impression or on an isolated behaviour or action. Therefore, it was advised that repetition of the behaviour during the flight is usually required to explicitly identify the nature of the problem reflected in the rating scores.

The NOTECHS method was essentially designed to be a guiding tool to help the examiner/instructor captains to look beyond failure during recurrent checks or training, and to help point out possible underlying deficiencies in non-technical competence, with regard to technical failures. It was recommended that the evaluation of NTS in a pilot’s check using NOTECHS should not provoke a failed (not acceptable) rating without a related objective technical consequence, leading to compromised flight safety in the short or long term. In the event of a crew member failing a check for any technical reason, NOTECHS could provide useful insights into the individual human factors contributing to the technical failure (e.g., an ‘altitude bust’ – flying at the wrong altitude level). Used
in this way, the method could provide valuable assistance for debriefing and orienting tailored retraining. The prototype NOTECHS system offered a systematic approach for assessing pilots’ NTS in the flight simulator as well as during actual flight operations.

Testing of the basic usability and psychometric properties of the NOTECHS system was then conducted, along with an examination of the effects of national cultural differences within Europe. A consortium of psychologists and pilots from a larger group of European research centres and aviation companies (including British Airways, Alitalia and Airbus) was set up in 1998 to test the NOTECHS method, and this was known as the JARTEL project. (11)

The main JARTEL study was an experimental rating task using NOTECHS based on eight video scenarios filmed in a Boeing 757 simulator, with airline pilots as the actors. The scenarios simulated realistic flight situations with predefined behaviours (from the NOTECHS elements) exhibited by the pilots at varying standards (‘very poor’ to ‘very good’). The pilots’ behaviours were rated using the NOTECHS system by 105 instructors, recruited from 14 large and smaller airlines in 12 European countries. Many of these instructors already had experience using behavioural rating systems for pilot assessment although they had not used the NOTECHS system. Therefore, each experimental session began with a briefing on the NOTECHS method and a practice session, lasting half a day. Then, in the afternoon session, the instructors were asked to individually rate the captains’ and first officers’ behaviour displayed in each of the eight cockpit scenarios using the NOTECHS score forms. The results indicated that the majority of the instructors were consistent in their ratings, there was an acceptable level of accuracy, and they reported being very satisfied with the NOTECHS rating system. (12) Cultural differences (relating to five European regions) were found to be less significant than other biographical variables, for example, proficiency in the English language, experience with NTS evaluation and role perceptions of captain and first officer. (13) An operational trial of NOTECHS was subsequently run with several airlines confirming the applicability and feasibility of the system in real check events for pilots. (11)

In summary, these first tests of the NOTECHS system showed that it was employable by airline instructors and appeared to have acceptable psychometric properties. It should be noted that these results were achieved with a minimal training period of half a day due to difficulties in recruiting experienced instructors (airline captains) to take part in the study, in particular, from the smaller companies. This level of training would be insufficient for using the NOTECHS system for regular training or assessment purposes. The recommended basic training period is two full days or longer (depending on the level of previous experience in rating pilots’ NTS). (14)

Users of NOTECHS are expected to be certified flight instructors and authorised examiners who have been trained in the application of the method for rating performance. It is assumed that pilots will have sufficient knowledge of the psychological concepts included in their theoretical examination on human performance and limitations which is a licensing requirement. (See the Civil Aviation Authority (CAA) 2006 (4) for the current UK position on CRM Instructors and CRM Instructor Examiners). It was recommended that the majority of any training should be devoted to the understanding of the NOTECHS method, the specific use of the rating form, the calibration process of judgement and the debriefing phase. As the NOTECHS system is primarily used as a tool for debriefing and identification of training needs, it is important to ensure that, in debriefing, an emphasis is placed on skill components rather than on more ‘global’ analyses of performance.

In summary, NOTECHS was designed as: a professional pragmatic tool for instructors and authorised examiners rather than for researchers (although it has been used for this purpose). (15) It was composed in common professional aviation language, with the primary intention that the tool be used for debriefing pilots and communicating clear advice for improvements. The preliminary
evaluation of the NOTECHS system from the experimental and operational trials indicated that the basic psychometric properties were acceptable and that the method was employable and accepted by practitioners. There have been subsequent studies involving NOTECHS ratings of pilots’ behaviours in Swiss (16) and Southeast Asian (17) airlines.

In response to the regulatory requirements from JAA with regard to evaluation of CRM skills, several European airlines (e.g., KLM, Lufthansa, Alitalia, etc.) had developed their own systems. (10) Some of these made use of the basic NOTECHS framework in their design, while other airlines initially used NOTECHS, or their own versions of it, to complement their proficiency evaluation methods (e.g., Finnair, Eastern Airways, Gulf Air, Iberia, etc.). In the UK, mandatory regulations from the Civil Aviation Authority (3,4) required a formal incorporation of non-technical (CRM) skills into all levels of training and checking of flight crewmembers’ performance. Similarly, in other high-risk work settings, such as nuclear power plants, assuring competence in NTS is, at present, a key component of licensing and revalidation. (18)

References


