7-18-2016

Measuring CRM Aptitude: Is NOTECHS a Suitable Tool for Pilot Selection?

Hans-Joachim K. Ruff-Stahl
Embry-Riddle Aeronautical University, ruffha@erau.edu
Daniel Vogel
Embry-Riddle Aeronautical University and City University London, daniel.vogel.1@city.ac.uk
Nicolaus Dmoch
Embry-Riddle Aeronautical University, dmochn@my.erau.edu
Alexander Krause
Embry-Riddle Aeronautical University, krausea@my.erau.edu
Andrea Strobl
Embry-Riddle Aeronautical University, andrea.strobl@de.bertrandt.com
Dirk Farsch
Embry-Riddle Aeronautical University, farschd@my.erau.edu
Robert Stehr
Embry-Riddle Aeronautical University, stehrr@my.erau.edu

Follow this and additional works at: https://commons.erau.edu/ijaaa

Part of the Other Social and Behavioral Sciences Commons

The human factor is widely considered as the most important, but likewise as the most complex and often vulnerable component in any aviation system. Harris (2011), for example, argued that the analysis and optimization of human performance constitute *the* major challenge for future aviation research, while Langer and Braithwaite (2012) consider aviation safety to be heavily dependent on "the management of human error in all parts of the complex system" (p.1). In light of these considerations, it is no surprise that aviation human factors are indeed cited as the major factor in context with aviation accident causation statistics (Martinussen & Hunter, 2010). However, whereas focus on the existing pilot corps constitutes one part of consideration, selection and training of future pilots should also be reviewed. In fact, it is argued that with the recent introduction of the Multi-Crew Pilot License (MPL) by the ICAO, the primary piloting skills for airline pilots have officially shifted from traditional stick-and-rudder skills to more non-technical, so-called ´soft` skills such as Crew Resource Management (CRM) and conflict resolution strategies (Skybrary, 2016). Already in 2004, the Federal Aviation Administration (FAA) discussed the interdependency of CRM skills and high technical proficiency in its Advisory Circular (AC) 120-51E and concluded "high technical proficiency cannot guarantee safe operations in the absence of *effective crew coordination*" (emphasis added) (FAA, 2004). For this reason, professional pilots are required to be proficient in CRM and other non-technical or soft skills, which must be continuously assessed during an airline pilot's professional life.

In the European Union, airline pilots' CRM skills are being mandatorily evaluated once per year by use of the NOTECHS (Non-Technical Skills) method after 2002 (JAR TEL). Yet, whereas recurrent training and annual check flights of active airline pilots have been focusing on CRM skills since roughly 10 years, the evaluation of non-technical skills during the selection of ready entry airline pilot still offers room for additional data. More precisely, and this is the research question for this paper, could the NOTECHS method be employed as a stand-alone instrument for the selection of already trained, "ready entry" airline pilots?

Review of Literature

Airline pilot training normally starts with the selection of suitable candidates. Effective pre-selection psychological testing is the better and certainly more cost-effective alternative to failing a student pilot in advanced training or even later as an airline pilot (Goeters & Maschke, 2004). The design of selection processes should, therefore, be based on a job analysis consisting of an input (personnel requirements/ specifications) to output (job requirements/ description) comparison (Martinussen & Hunter, 2010). Goeters, Maschke, and Eissfeldt

(2004) described the following top ten job requirements for airline pilots: The effective use of time, spatial orientation, problem-solving, motor control, and stress resistance. The remaining five skills are interaction and social skills, such as cooperation, communication, decision-making, leadership and management, and situational awareness. Further, 77% out of 141 interviewed airline pilots rated interactive and social skills as very relevant pilot job requirements, while no other skill area was assessed with such high ratings (Goeters et al., 2004). Notably, stick-and-rudder skills are absent.

Consequently, if CRM skills have become the most important skill set to fly a multi-crew airliner safely and effectively, then an airline pilot selection method should, firstly, focus on these skills. Secondly, an early selection of airline pilot candidates based on the right aptitude to fly multi-crew airliners is highly cost-effective for individuals, airlines, and society.

Yet, airline accidents have frequently been attributed to shortcomings in the crews' non-technical skills rather than their technical abilities. To improve the non-technical skills of flight crews, Crew Resource Management (CRM) training was introduced and has since found widespread acceptance. As the next logical step, ways were sought to objectively assess non-technical skills in operational settings, leading to European legislation that required formal assessment of pilots' non-technical skills in addition to their technical proficiency (Flin & Maran, 2004).

Some airlines, air forces, and flight schools make extensive use of aptitude tests and psychometric assessments to evaluate the suitability of prospective pilots (Hoermann & Goerke, 2014). However, only a few efforts were made to directly assess non-technical skills of prospective pilots, and even those were restricted to paper-and-pencil tests rather than based on observed behavior in operational settings (Mavin et al., 2013). Given the importance of non-technical skills in multi-crew environments, trying to assess the non-technical skills of applicants for ab-initio flight training programs would only be logical. If the NOTECHS system could be applied in that way, then it would present a pragmatic, straightforward tool that addresses the actual interpersonal skill requirements of prospective airline pilots.

After the European Commission (EC) implemented CRM assessment as a mandatory part of recurrent training and annual performance evaluation, a consortium of airlines, universities, research centers, and the European Cockpit Association (ECA) developed the NOTECHS evaluation tool as an experimental validation process (JAR TEL, 2002). In their Final Report to the EC, the

consortium emphasized that the tool is designed for non-psychologists and shall improve effective behavior through debriefing and communication without evaluating the flight crew personality or being "a tool for introducing 'psychobabble' into the evaluation process" (JAR TEL, 2002, p. 26).

NOTECHS is divided into two social and two cognitive assessment categories: Cooperation, Leadership and Managerial Skills, Situational Awareness, and Decision Making, which are subdivided into several elements and from there into 15 behavioral markers, rated on a 5-point scale (JAR TEL, 2002). The simplicity, its proved practicability, and a high level of objectivity to test for inappropriate aptitudes support the proposal to use NOTECHS – not only as an evaluation tool for curing line checks but also as pilot selection tool (JAR TEL, 2002).

Development of NOTECHS

The development and introduction of the NOTECHS behavioral marking system to train and assess the Crew Resource Management (CRM) skills of airline pilots was triggered by the implementation of respective legislation by the European Joint Aviation Authorities in the mid-1990s (Flin et al., 2003). In 1996, the JAA Research Committee on Human Factors tasked a research consortium to identify existing behavioral marking schemes that could be adopted for that purpose, or, if no such system was found to be adequate, to develop a new methodology for assessing pilots' non-technical skills (NTS) (Flin et al., 2003). It is noteworthy that from the outset, the Committee's requirements for the rating scheme included its suitability for assessing the skills of individual pilots, rather than crews, and that it had to be usable by airlines of any size across Europe. The reason for the second requirement was that smaller airlines, unlike many of the larger European airlines, who had at that time already developed their own behavioral marker systems, lacked the resources to develop their own systems; therefore, a need for a basic, generic system was identified. The research group came to the conclusion that existing behavioral markings systems, such as the system developed by the University of Texas, were unsuitable, as they assessed the performance of flight crews and not the non-technical skills of individual pilots. (O'Connor et al., 2002).

Therefore, the research group decided to design a new behavioral marker system. This new system, named NOTECHS, offered a systematic approach for assessing pilots' non-technical skills in simulators and in real flight. In its original form, the non-technical skills were grouped into four categories: Cooperation, Leadership and Managerial Skills, Situation Awareness, and Decision-Making.

These categories were further subdivided into three to four elements each. For example, Cooperation consists of the elements Team Building and Maintaining, Considering Others, Supporting Others, and Conflict Solving. For each element, behaviors were defined that could serve as examples for desired and undesired behaviors. For example, for the element Team Building and Maintaining, a positive behavior is labeled 'helps other crew members in demanding situations,' while a negative behavior is classified as 'competes with others' (Flin, 2010).

To test the NOTECHS system's usability, reliability, and validity as an assessment tool, the cooperation of five aviation research centers and four commercial aviation enterprises was solicited. The Joint Aviation Requirements - Translation and Elaboration of Legislation (JARTEL) project began in January 1998 and lasted for about three years. It included an experimental study in which 105 examiners, which had been recruited from fourteen European airlines, were asked to rate the non-technical skills of pilots in a series of scenarios. These scenarios had been scripted by experienced subject matter experts and were recorded during simulator sessions in a full flight simulator. The examiners had previously received some background information and were given a standard briefing on the NOTECHS method and a practice session, lasting half a day in total. The results indicated that 80% of the instructors were consistent in their ratings and that 88% of them were satisfied with the consistency of the method. On average, their ratings differed by less than one point on the five-point scale from a reference rating that had been established by the same subject matter experts that had created the scenarios. This was taken as a confirmation of an acceptable level of accuracy. Furthermore, 98% of the instructors expressed their satisfaction with the NOTECHS rating system (Flin et al., 2003).

System Requirements and Limitations

Flin (2010) described five fundamental principles that should be adhered to ensure the accurate and objective assessments of pilots using the NOTECHS system: First of all, only observable behavior shall be assessed; references to a pilot's personality or emotional attitude are not permissible. Secondly, for a pilot's non-technical skills to be rated as unacceptable, flight safety must be actually or potentially compromised. This requires a related objective technical consequence. Thirdly, to meet JAR-OPS requirements, the results of an assessment of a pilot's non-technical skill must include an indication whether the observed non-technical skills were acceptable or unacceptable. Fourthly, any unacceptable behavior must be observed repeatedly during a single evaluation before any conclusions can be drawn. Finally, for each category that is rated as unacceptable, the evaluator must explain how the observed behavior led to actual

or potential safety consequences, and to which NOTECHS category and element it was related.

Furthermore, the new system was designed to be used by flight instructors and examiners rather than psychologists, and the only pre-requisites for using the NOTECH framework are thorough training in the NOTECHS methodology and knowledge of basic CRM concepts (Flin, 2010).

Two limitations were identified during the development and introduction of the NOTECHS ratings system. First of all, some behaviors will not always be observable. For example, conflict resolution skills are very important but are only required to be used rarely. Secondly, given that evaluator usually have to perform other tasks in addition to assessing the pilots' non-technical skills, such as programming the simulator, simulating air traffic control, and assessing technical performance, they may not always have the capacity to accurately observe and evaluate the non-technical skills of individual pilots (Klampfer, et al, 2001). Besides, the NOTECHS system was not intended to be used as a tool for judging a pilot's personality on the basis of raters' personal opinions (Flin et al., 2003).

Validation of the Pass/Fail Rating

For the purposes of the using NOTECHS as a pilot selection tool, the pass/fail rating becomes important. Normally, the system's categories and elements are intended to provide pilots with feedback about their non-technical skills and potential areas of improvement, while the pass/fail rating is intended to fulfill regulatory requirements. For the selection of ab-initio pilots, the feedback becomes inconsequential, as their evaluation will be a unique event. In contrast, pilots, who are evaluated on a regular basis, may use the feedback to modify their behavior and improve their non-technical skills. In light of this significant difference in emphasis, it is of interest whether these pass/fail ratings have been proven to be accurate and reliable.

As O'Connor et al. (2002) highlighted, the accuracy and inter-rater reliability has been tested during the initial experimental study by the JARTEL group. It was shown that the pass/fail ratings of the 105 participating examiners matched the reference ratings, which had been established beforehand by a group of very experienced examiners. Across all eight evaluated scenarios, the mean accuracy, meaning the level of agreement with the reference ratings, was 83% for the captain and 84% for the first officers. However, close examination of the inter-rater reliability (IRR), which is the level of agreement between raters, revealed drastic differences in some of the scenarios. Regarding the observed

behavior in five of the eight scenarios, the raters agreed to a high extent, while their level of agreement was very low in the three remaining cases. O'Connor et al. offered three possible explanations for this phenomenon. They speculated that the training in applying the NOTECHS system was too short to enable rater to judge these scenarios accurately, that the general rationale for assigning pass and fail ratings required further clarification, and that the overall length of the videos may have been too short to enable raters to accurately assign pass/fail ratings. O'Connor et al. added that under normal circumstances, raters would observe pilots over a longer period, which they expected to increase the accuracy of the ratings.

Interrelation of Technical and Non-Technical Skills

Another important question is whether technical skills, which candidates for ab-initio pilot programs lack by definition, are actually important enablers for non-technical skills. Mavin, Roth, and Dekker (2013) argued that the advent of the modern, highly automated flight deck makes it impossible to separate technical skills from non-technical skills. The authors claim that automation handling skills are as much technical as they are non-technical skills, and require "the careful coordination of flight parameters, navigation, associated call-outs, tasks, and double-checks" (pp. 53). Furthermore, they explain that the ability of pilots to handle unforeseen situations requires crews both to recognize existing deficiencies in the system and to develop feasible strategies to deal with it. This makes non-technical skills inseparable from technical skills, and any attempt to assess these skills independently of each other may lead to an incomplete performance assessment. According to Mavin et al. (2013), attempts are being made to develop non-technical skill evaluation systems that take the role of technical skills into account.

Consequently, the evaluation of potential ab-initio pilots may be hampered by the fact that they had, no or very little, opportunity to acquire relevant technical skills. However, although their reasoning makes intuitive sense, it must be noted that Mavin, Roth, and Dekker (2013) did not offer any quantitative evidence to substantiate their claims.

Differences in Interpretation

Furthermore, Mavin, Roth, and Dekker (2013) argue that raters may differ in their interpretation of observed facts. In their study, the authors invited three pairs of pilots, consisting of two first officers, two captains, and two examiners, to rate recorded simulator flights. None of them had received inter-rater reliability

training, but all of them were trained in using their company's customized non-technical skills rating form. In addition to the finding that differences in ratings were correlated to differences in the experience and roles of the raters, the researcher discovered that raters were generally in alignment when the recording showed good crew performance; however, when there were problems in the course of a flight scenario, raters differed from each other in respect to how the observed actions were rated in terms of skill level, in the facts they picked to justify their ratings, and in the interpretation of the facts as an explanation of overall performance. For example, in one of the recorded simulator flights, the flying pilot made an utterance regarding the direction in which he had to turn while performing a missed approach. This utterance was interpreted in three different ways. In one interpretation, it was seen as a verbalization affirming the direction of the turn and considered desirable communication. Other raters interpreted the utterance as a sign of lack of situational awareness, while a third interpretation was that the utterance was a question that showed a lack of proper planning. In each case, the raters provided valid reasons for the particular assessment they made, but they clearly intertwined.

While the researcher's findings seem to indicate a lower inter-rater reliability than previously thought, the study is hampered by a low sample size and the fact that raters had no prior inter-rater reliability training.

Evaluation of Teams or Individuals?

Gontar and Hoermann (2014) conducted a study in which sixty pairs of airline pilots had to fly a challenging scenario in a full flight simulator. Afterward, both pilots and the examiner rated the non-technical skills of both pilots using the airline's customized version of the NOTECHS rating system. The researchers then examined the correlation coefficients of all ratings, for example, the first officer's and the examiner's ratings of the captain's non-technical skills. To facilitate the analysis, the researchers grouped the four basic non-technical skills into two general groups: Social aspects, consisting of the categories Communication and Leadership & Teamwork, and cognitive aspects, consisting of the categories Work Organization and Situation Awareness & Decision Making.

Gontar and Hoermann's (2014) research led to two findings that are pertinent to the research question of this literature review. One finding was that the correlation coefficients of cognitive skill ratings were higher than those of social skills for any rater combination. The researchers pointed out that this result contradicted previous observations by O'Connor et al. (2002) that cognitive

aspects are more difficult to observe than social skills. They speculated that raters' preconceptions of good social non-technical skills may vary to a high extent. Furthermore, the evaluation of social non-technical skills may require more interpretation of observed behaviors, which in turn may make the resulting ratings more vulnerable to variability. There is a noteworthy shortcoming in their approach, however, as they compared ratings provided by examiners with ratings provided by the pilots themselves, who had not previously received respective training. Clearly, this approach may make the study vulnerable to the influences biases and interpersonal factors. Gontar and Hoermann (2014) continued their research regarding the inter-rater reliability in another study that is discussed in the following section.

Another important result of the simulator study was the finding that examiners' ratings for captains correlated to the examiners' ratings of first officers. This indicates that they were not so much rated as individuals, but rather as crews. Another possible explanation is that pilots with good non-technical skills enabled their colleagues to perform well. Either way, the result showed that individual contributions are difficult to be attributed to single pilots and that the differentiation between the two crew members in respect to their non-technical skills is challenging (Gontar & Hoermann, 2014).

Flin (2010) explained that this issue was identified early in the development of the NOTECHS system. According to her, assessing individual contributions to overall team performance is also a problem in evaluating technical skills, and is not eliminated by the NOTECHS system. However, she claims that NOTECHS can provide assistance in objectively pointing to "behaviors that are related more to one crewmember than the other, therefore allowing them to differentiate their judgment of the two crewmembers" (p. 184). Used this way, the system is designed to remove ambiguities and to require sound justification using a standard vocabulary rather than obscure references to vague impressions.

Nevertheless, the inability of raters to differentiate between individual members of a crew may present a significant obstacle when using NOTECHS for the purposes of ab-initio pilot selection.

Rater reliability

In another study, Gontar and Hoermann (2015) investigated the inter-rater reliability (IRR) of experienced instructor pilots when rating the non-technical skills of the pilots of the same German airline that participated in the previous

study. This time, two well-trained examiners simultaneously rated the non-technical skills of six crews (twelve pilots) in flight simulator sessions. Each crew was presented with the same scenario. The consistency of the ratings of the two examiners was taken as an indication of inter-rater reliability. Four weeks after all simulator flights had been completed, one of the raters assessed the performance of eight pilots a second time, this time based on video and audio recordings. The comparison between the two ratings of the same rater was taken as a measure of test-retest reliability. The rater had not retained any notes, documents, or rating sheets, and since he had rated another 72 pilots in the meantime, Gontar and Hoermann (2015) assumed that he did not remember any particularities of the individual pilots or his previous ratings.

Gontar and Hoermann (2015) chose to judge the reliability of raters by calculating intra-class correlation coefficients (ICC). Calculating ICCs will result in values between 0 and 1. While 1 represents fully explained variance and is probably only attainable in an ideal setting, it is generally agreed that values greater than 0.75 may be interpreted as excellent agreement. The researchers followed current practice by setting the threshold for acceptable reliability at a value of 0.7.

The results indicate that the inter-rater reliability is dependent on the rating dimension. Concerning the two social skills Communication and Leadership & Teamwork, inter-rater reliability reached values of 0.95 and 0.71, respectively. The cognitive skills Work Organization and Situation Awareness & Decision Making only reached values of 0.49 and 0.32, respectively. The researchers speculated that the results could be explained that the two social aspects are directly observable without further interpretation while rating the two cognitive skills may require raters to interpret observed behaviors, which will lead to higher variances in ratings (Gontar & Hoermann, 2015).

Furthermore, the results of the re-test reliability study showed that the following correlations: Communication 0.47, Leadership & Teamwork 0.52, Work Organization 0.45, and Situation Awareness & Decision Making 0.74. The re-test reliability was therefore 0.6 on average (Gontar & Hoermann, 2015). These results are indicating a rather low re-test reliability and raises the accuracy of ratings. However, since the second rating was assigned based on video and audio recordings, the differences in the ratings could also be explained by the different setting, in particular by the restrictions of audio and video recordings compared to being present in the actual simulator. Furthermore, the sample size is again quite low. In particular, only one rater was asked to rate a limited amount of pilots a

second time. This may have resulted in rating variations that a larger sample may have smoothed out.

Summary of Existing Research

The literature review revealed a few particular aspects that indicate that the NOTECHS framework could be a usable tool for the selection of ab-initio student pilots. In particular, the system was designed to capture the non-technical skills of the individual and to be suitable for pan-European settings, which, given the spread of multi-national airlines in Europe, may become useful. Furthermore, the NOTECH system is intended to be a practical tool for the use of non-psychologist, in particular of examiners and instructor, with relatively little training, but operational experience.

However, several concerns remain. First of all, it has been pointed out that not all behaviors will always be observable. This may not be of significant consequence for the evaluation of pilots' non-technical skills, but it has major implications for the selection of student pilots given the limited observation time available during a selection process.

Furthermore, care must be taken not to use the NOTECHS system as a tool to assess a prospective pilot's personality. This is of particular concern considering that ab-initio pilot candidates will most probably not have received prior CRM training or technical training. The latter is of significance, as some researchers argue that technical skills may be important enablers for non-technical skills. In the absence of any related training that could have led to the development of these skills, there may be a risk that raters will end up measuring a person's personality.

Finally, recent research has cast doubt on the reliability and accuracy of ratings, in particular of pass/fail ratings, which are of special importance for the selection. The most important obstacle may that some research showed that raters might be prone to evaluate team performance rather than the non-technical skills of individual pilots. However, most of the research that seems to question the accuracy and reliability of the NOTECHS system was based on studies that employed small samples.
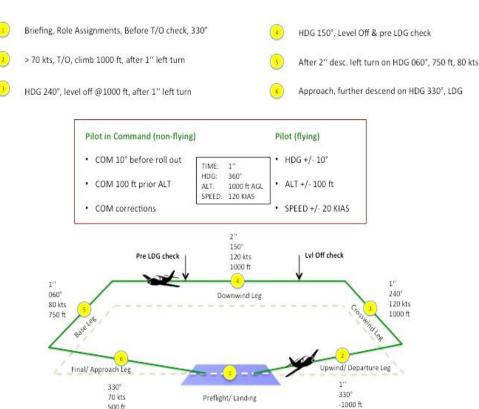
Overall, while some attributes make the NOTECHS system viable as a pilot selection tool, some research exist that casts doubt about the accuracy and reliability of NOTECHS for any use. This obstacle is, however, not insurmountable. Rather, it invites further research.
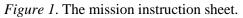
Hypothesis

Given the research reviewed above, the following hypothesis is stated: The NOTECHS framework will predict the success of ab-initio student pilots at least as accurately as common psychometric tests, if the following requirements are met: Firstly, raters have been thoroughly trained and given inter-rater reliability training, with special emphasis on the pass/fail ratings, and the need to take particular care not to rate the team, but rather the individual. Secondly, raters need to evaluate the candidates in an operational setting, such as a flight simulator, and not by means of video and audio recordings. Thirdly, the observations must provide sufficient opportunity to observe all behaviors. This may require several simulator sessions and more than one examiner. Finally, if a simulator is used, then it should require low technical skills, and some technical training should be provided beforehand.

Method

Schmidt and Hunter (1998) tested several selection procedures on their predictive validity concerning later job and training performance, whereas the predictive validity of *work sample tests* demonstrated the highest figure with 0.54. In combination with a General Mental Ability (GMA) test, the validity rises even to 0.63. Hence it was decided for this study to evaluate individual CRM skills according to the NOTECHS method in combination with an experimental work sample test. A local airline pilot training facility, the *Verkehrspilotenschule Berlin* offered a Flight Navigation Procedure Trainer 2 (FNPT2, simulating a DA 42 – TDI, Twin Star) to conduct this experiment as work sample body. Eight test subjects completed the experiment, functioning as a pilot study. The test subjects, all of whom were inexperienced student pilots (SP), were tasked to fly two traffic patterns (Figure 1) while being assisted by an instructor pilot (IP) in the right seat. The IP first explained the basics of flying this type of airplane (five minutes) and then continued to offer verbal instruction during the evaluation phase. The first traffic pattern flown by the SP was closely supported by the IP, while the second pattern was only influenced verbally by the IP when certain flight parameters were not met according to the mission instruction sheet (Figure 1). Two independent evaluators, sitting behind SP and IP assessed and rated the test person according to JAR TEL standard evaluation markers (see Appendix), then discussed and averaged their individual results after the flight. It should be emphasized that this assessment did not rate flying performance but instead focused on the SP's CRM skills.

Figure 1. The mission instruction sheet.
Note: This template explains all the parameters, which the test person has to meet and maintain.

A five-point Likert-type scale rating was used to assess the SPs CRM performance (see Appendix). Unobserved markers were rated as "NA". Within the completed final evaluation sheet an overall mean was calculated to describe the overall level of non-technical (CRM) skills with a single numeric grade. The goal of this experiment was to demonstrate the practical usability of the NOTECHS evaluation framework during a work sample test for the purpose of airline pilot selection.

Results

The results of the experiment are listed in Table 1-1. A number of behavioral markers, generally 1 or 2 per cognitive assessment category, were not applicable for this work sample test scenario. The overall SP grades ranked between poor (2) and good (4) in those categories that were tested in the

experiment. No SP showed overall very poor (1) performance and only one SP showed very good (5) performance overall.

Table 1-1
The evaluation results of the work sample test using the NOTECHS score form

		Cand 1	Cand 2	Cand 3	Cand 4	Cand 5	Cand 6	Cand 7	Cand 8	Mean	Median	Mode
Coop	a	3	3	4	4	4	4	3	4	3.62	4	4
	b	3	3	4	4	4	4	2	3	3.37	3.5	4
	c	3	3	4	3	3	3	3	4	3.25	3	3
	d	NA	NA	NA	NA	NA	NA	NA	NA			
Sub-Mean		3	3	4	3.67	3.67	3.67	2.67	3.67			
Leadership and Managerial Skills	a	3	2	3	4	4	3	2	4	3.12	3	3
	b	NA	NA	NA	NA	NA	NA	NA	NA			
	c	3	2	3	3	3	4	2	3	2.87	3	3
	d	3	NA	NA	NA	NA	NA	NA	3			
Sub-Mean		3	2	3	3.5	3.5	3.5	2	3.67			
Situation Awareness	a	3	2	4	3	4	4	3	4	3.37	3.5	4
	b	3	2	3	3	4	4	2	4	3.12	3	3
	c	NA	NA	NA	NA	NA	NA	NA	NA			
Sub-Mean		3	2	3.5	3	4	4	2.5	4			
Decision Making		3	2	4	4	4	4	3	4	3.5	4	4
		4	2	4	3	3	5	3	4	3.5	3.5	4
		NA	NA	NA	NA	NA	NA	NA	NA			
		NA	NA	NA	NA	NA	NA	NA	NA			
		3.5	2	4	3.5	3.5	4.5	3	4			
Sub-Mean Mean		3.125	2.25	3.265	3.42	3.67	3.92	2.54	3.83			
Median		3	2	4	3	4	4	2.5	4			
Mode		3	2	4	3	4	4	3	4			

Note: The candidates were evaluated by using a Likert-type scale divided into five performance levels: (1) very poor, (2) poor, (3) acceptable, (4) good, (5) very good. Skills not observed were rated (NA). Cand = Candidate; Coop = Cooperation.

The lowest scoring individual category was *Leadership and Managerial Skills*, while the highest score was achieved in *Decision Making*, closely followed by *Cooperation*. *Decision Making* showed the greatest variance, with the lowest scoring participant achieving a 2 and the highest scoring participant a 4.5. *Cooperation* was the least variant category. In general, category scores showed great variety from one SP to another and some categories were scored only by a few SPs or by none at all. The reason for these deviations resulted from

differences in the actual work sample simulator flight, in spite of the rigid mission design.

Discussion and Conclusion

Among the non-observable NOTECHS markers was *Workload Management*, which was only observed and scored in two cases. Both SPs were the only ones who showed signs of stress but proceeded to ignore these signs. While technically the additional score in this category affected those SPs' overall score, the resulting influence on their overall grade was minimal. Hence the grade in this sub-category was included in the final grade calculation.

Other sub-categories, such as *Conflict Solving* were not observed and hence not scored, simply because no conflict occurred during the experiment. Even here, the impact of some non-scored categories did not result in skewed overall grades.

The results of this pilot study demonstrate that it *is* possible to score beginner student pilots on the NOTECHS framework, which originally has been designed for trained airline pilots as a re-currency evaluation tool. Further, the NOTECHS method offers clear and individually discernible results for each candidate, which could serve as an individual's evaluation of his or her CRM aptitude. The clarity and unambiguousness of the results in terms of CRM aptitude makes NOTECHS an easily comprehensible tool for non-psychologists. Hence, this framework offers great usability for Human Resource personnel and airline pilots for the selection of their future colleagues.

If NOTECHS is used for the screening of airlines pilot candidates, then a number of psychological markers will not be testable and should be taken off the list. These include under "Cooperation": *Conflict Solving,* under "Leadership & Managerial skills": *Providing & Maintaining Standards,* under "Situation Awareness": *Awareness of Time & Anticipation,* and under "Decision-Making": *Risk Assessment & Option Selection* and *Outcome Review*. Although only one test subject has been graded on *Workload-Management* under the category "Leadership & Managerial Skills", it is believed that workload-management strategies can be learned in many other life contexts other than an airline cockpit and hence might be testable in some occasions and with some test subjects during a simulator CRM screening test. Even if this item cannot be graded in some or even most cases, it will not significantly skew a candidate's overall CRM aptitude grade in either case. Therefore, this marker should remain in a "NOTECHS for pilot selection framework" as an optionally graded item.

Another peculiarity in this study is slight deviations in the conduct of one simulator flight to the next – in spite of the rigid design of the experiment and its flight profile. Notably, this resulted in varied social interactions between SP and IP from flight to flight. Yet, social interaction is a highly complex issue and varies fundamentally from one individual to another, regardless of the situation. Moreover, this evaluation method is based on social interaction and hence must allow for free play in CRM. In summary, these deviations are not considered as a weakness in this study's experimental design. On the contrary, these variations show that NOTECHS is a usable tool in *realistic* CRM scenarios. Yet it is recommended to validate the results of this study in a larger scale experiment.

In conclusion, a *modified* NOTECHS score form (see Appendix) provides an excellent addendum to traditional methods of airline pilot selection procedures. It is an efficient and highly usable tool for CRM aptitude evaluation. In sum, this pilot study indicates that NOTECHS holds potential for the selection of future airline pilots.

References

Eurocontrol (n. d.). *Multi-Crew Pilot Licence (MPL).* Retrieved from http://www.skybrary.aero/index.php/Multi-crew_Pilot_Licence_(MPL)

FAA (2004). Regulations & Policies, Advisory Circulars (ACs): AC 120-51E - Crew *Resource Management Training*. Retrieved from http://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/22879

Flin, R., Lynne, M., Goeters, K.-M., Hörmann, H.-J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the notechs (non - technical skills) system for assessing pilots' crm skills. *Human Factors and Aerospace Safety*, *3*(2), 95-117. Retrieved from http://www.abdn.ac.uk/iprc/uploads/files/NOTECHS HFAS proof copy.pdf

Flin, R., & Maran, N. (2004). Identifying and training non-technical skills for teams in acute medicine. *Quality & Safety in Health Care*, *13*(1), doi: 10.1136/qshc.2004.009993

Flin, R. (2010). Crm (non-technical) skills - applications for and beyond the flight deck. In B. Kanki, R. Helmreich & J. Anca (Eds.), *Crew Resource Management* (2nd ed.). San Diego, CA: Academic Press.

Garland, D., Wise, J., & Hopkin, V. (2010). *Handbook of aviation human factors* (2nd ed.). Mahwah, NJ: Lawrence Earlbaum Associates.

Goeters, K. M., & Maschke, P. (2004). Cost-benefit analysis of pilot selection: The economic value of psychological testing. In Aviation psychology: *Practice and research*, ed. K. M. Goeters, 203-208. Aldershot: Ashgate Publishing Limited.

Goeters, K. M., Maschke, P., & Eißfeldt, H. (2004). Ability requirements in core aviation professions: Job analyses of airline pilots and air traffic controllers. In *Aviation psychology: Practice and research*, ed. K. M. Goeters, 203-208. Aldershot: Ashgate Publishing Limited.

Gontar, P., & Hoermann, H.-J. (2014, September). *Flight crew performance and crm ratings based on three different perceptions*. Paper presented at the 31st European association of aviation psychology conference, Valetta, Malta. Retrieved from https://www.researchgate.net/profile/Patrick_Gontar/publication/272027651_Flight_Crew_Performance_and_CRM_Ratings_Based_on_Three_Different_Perceptions/links/5509abce0cf26198a639195e.pdf

Gontar, P., & Hoermann, H.-J. (2015). Reliability of instructor pilots' non-technical skills ratings. In *Proceedings of the 18th International Symposium on Aviation Psychology*. doi: 10.13140/RG.2.1.2152.5921

Harris, D. (2011) *Human performance on the flight deck*. Farnham, UK: Ashgate Publishing Ltd.

Hedge, J. W., Bruskiewicz, K. T., Borman, W. C., Hanson, M. A., Logan, K. K., & Siem, F. M. (2000). Selecting pilots with crew resource management skills. *The International Journal of Aviation Psychology, 10*(4), 377-392. doi: 10.1207/S15327108IJAP1004_5

Hoermann, H.-J., & Goerke, P. (2014). Assessment of social competence for pilot selection. *The International Journal of Aviation Psychology, 24*(1), 6-28. doi: 10.1080/10508414.2014.860843

Klampfer, B., Flin, R., Helmreich, R. L., Haeusler, R., Sexton, B., Fletcher, G., Field, P., Staender, S., Lauche, K., Dieckmann, P., & Amacher, A. (2001, July). *Enhancing performance in high risk environments: Recommendations for the use of behavioural markers*. Presentation at the Behavioural Markers Workshop of the Kolleg Group Interaction in High Risk Environments (GIHRE), Zurich, Switzerland.

JAR TEL (2002). *Final Report Consolidation of Results: WP7 Draft Report.* Retrieved from http://www.transport-research.info/Upload/Documents/200310/jartelrep.pdf

Langer, M., & Braithwaite, G. (2012). *The Development of the Maintenance Operations Safety Survey: Challenges in transferring a predictive safety tool from flight operations to aircraft maintenance*. Retrieved from http://www.isasi.org/Documents/library/technical-papers/2012/8-Maintenance-Safety-Survey-Transferring-Predictive-Safe.pdf Malik, F. (2004). *Führen, leisten, leben. Wirksames management für eine neue zeit*. Frankfurt, Germany: Campus Verlag.

Martinussen, M., & Hunter, D. R. (2010). *Aviation Psychology and Human Factors*. Boca Raton, FL: CRC Press, Taylor and Francis Group.

Mavin, T. J., Roth, W.-M., & Dekker, S. (2013). Understanding variance in pilot performance ratings. *Aviation Psychology and Applied Human Factors, 3*(2), 53-62. doi: 10.1027/2192-0923/a000041

O'Connor, P., Hoermann, H.-J., Flin, R., Lodge, M., Goeters, K. M., & the JARTEL Group (2002). Developing a method for evaluating Crew Resource Management skills: a European perspective. *The International Journal of Aviation Psychology, 12*(3), 263-285.

Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and Theoretical Implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.

Shaban, K. B., & Abdulwahed, M. (2012, August). *Research-based learning in computing courses for senior engineering students*. Paper presented at the 2012 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Hong Kong. doi: 10.1109/TALE.2012.6360355

Appendix

The NOTECHS score form page one (Adapted from JAR TEL, 2002)

NOTECHS Scoreform

Name of trainee/ applicant: _____   Date:

Gender: M/F   Age: Assessor (initials): _____

Please follow the key below and circle the number corresponding to the applicant's performance.

NA-not applicable

1 Very poor	2 Poor	3 Acceptable	4 Good	5 Very Good

Category	Element		Behaviour							Comments
Cooperation	a	Teambuilding & maintaining	-Establishes atmosphere for open Communication/ participation -Encourages inputs & feedback from others -Does not compete with others	NA	1	2	3	4	5	
	b	Consideration of others	-Takes notice of the suggestions of other crewmembers even if s/he does not agree -Takes condition of other crewmembers into account -Gives appropriate feedback	NA	1	2	3	4	5	
	c	Support of others	-Helps other crewmembers in demanding situations -Offers assistance	NA	1	2	3	4	5	
	d	Conflict solving	-Keeps calm in conflicts -Suggests conflict solutions -Concentrates on what is right rather than who is right	NA	1	2	3	4	5	

Mean: _____

Category	Element		Behaviour							Comments
Leadership & Managerial Skills	a	Use of authority & assertiveness	-Advocates own position -Takes initiative to ensure involvement & task completion -Takes command if necessary	NA	1	2	3	4	5	
	b	Providing & maintaining standards	-Ensures SOP compliance -Intervenes if task completion deviates from standards -consults crew for applying non-standard procedures	NA	1	2	3	4	5	
	c	Planning & co-ordination	-Encourages crew participation in planning & task completion -Clearly states intentions & goals -Consults crew, changes plan if necessary	NA	1	2	3	4	5	
	d	Workload-management	-Allowing secondary operational tasks to interfere with primary flight duties, distributes tasks -Ignoring signs of stress and fatigue	NA	1	2	3	4	5	

Mean: _____

NOTECHS score form page two (Adapted from JAR TEL, 2002)

				NA	1	2	3	4	5	
Situation Awareness	a	Awareness of aircraft systems	-Monitors and reports changes in system states -Acknowledges entries and changes to systems	NA	1	2	3	4	5	
	b	Awareness of external environment	-Contacts outside resources when necessary -Shares information about the environment with others -comments on relevant environmental factors	NA	1	2	3	4	5	
	c	Awareness of time & anticipation	-Discusses contingency Strategies -Identifies possible/ future problems	NA	1	2	3	4	5	

Mean: _____

				NA	1	2	3	4	5	
Decision Making	a	Problem definition & diagnosis	-Gathers information and identifies problem -Reviews causal factors with other crewmembers	NA	1	2	3	4	5	
	b	Option generation	-States alternative courses of action -Asks crewmembers for options	NA	1	2	3	4	5	
	c	Risk assessment & option selection	-Considers and shares risks of alternative courses of action -Talks about possible risks for course of action in terms of crew limitations -Confirms selected course of action	NA	1	2	3	4	5	
	d	Outcome review	-Checks outcome against plan	NA	1	2	3	4	5	

Mean: _____

Overall score: _____

Modified NOTECHS score form page one (Adapted from JAR TEL, 2002)

Modified NOTECHS Scoreform

Name of trainee/ applicant: _____   Date:

Gender: M/F   Age: Assessor (initials): _____

Please follow the key below and circle the number corresponding to the applicant's performance.

NA-not applicable

1 Very poor	2 Poor	3 Acceptable	4 Good	5 Very Good

Category	Element		Behaviour							Comments
Cooperation	a	Teambuilding & maintaining	-Establishes atmosphere for open Communication/ participation -Encourages inputs & feedback from others -Does not compete with others	NA	1	2	3	4	5	
	b	Consideration of others	-Takes notice of the suggestions of other crewmembers even if s/he does not agree -Takes condition of other crewmembers into account -Gives appropriate feedback	NA	1	2	3	4	5	
	c	Support of others	-Helps other crewmembers in demanding situations -Offers assistance	NA	1	2	3	4	5	

Mean: _____

Category	Element		Behaviour							Comments
Leadership & Managerial Skills	a	Use of authority & assertiveness	-Advocates own position -Takes initiative to ensure involvement & task completion -Takes command if necessary	NA	1	2	3	4	5	
	b	Planning & co-ordination	-Encourages crew participation in planning & task completion -Clearly states intentions & goals -Consults crew, changes plan if necessary	NA	1	2	3	4	5	
	c	Workload-management	-Allowing secondary operational tasks to interfere with primary flight duties, distributes tasks -Ignores signs of stress & fatigue	NA	1	2	3	4	5	

Mean: _____

Modified NOTECHS score form page two (Adapted from JAR TEL, 2002)

Situation Awareness				NA	1	2	3	4	5	
	a	Awareness of aircraft systems	-Monitors and reports changes in system states -Acknowledges entries and changes to systems	NA	1	2	3	4	5	
	b	Awareness of external environment	-Contacts outside resources when necessary -Shares information about the environment with others -comments on relevant environmental factors	NA	1	2	3	4	5	

Mean: _____

Decision Making										
	a	Problem definition & diagnosis	-Gathers information and identifies problem -Reviews causal factors with other crewmembers	NA	1	2	3	4	5	
	b	Option generation	-States alternative courses of action -Asks crewmembers for options	NA	1	2	3	4	5	

Mean: _____

Overall score: _____