2017

Exploration of a Confidence-Based Assessment Tool within an Aviation Training Program

Paul F. Novacek Ph.D.
avianpaul@gmail.com

Introduction

Traditional use of the multiple-choice question rewards a student for guessing. Students are often told when preparing for an exam that even if they are unsure of the correct answer, they should answer it anyway because with a multiple-choice selection there is a 20% (5-choice) to 25% (4-choice) chance of guessing the correct answer. Hence the reason students poke fun at the process, calling it a "multiple-guess" exam. There is an effort to maximize the score instead of gaining an understanding of the course material. Yet in this world of number-crunching rationalizations, it requires much less effort to assign a number (numeric test score) to represent a level of understanding that can be quantified, studied, and managed. It is regrettable that the reliance within the educational system on this technique infers that a student who provides a correct answer purely through guesswork possesses knowledge equivalent to a student who actually knows the correct answer. Is this really an effective way of measuring a student's comprehension?

Guessing on a few questions in a beginning math course may be perceived as somewhat benign, but the most problematic aspect of this guesswork manifests itself within a safety-critical, or high-consequence environment. It is generally understood that misinformed operators who held a steadfast belief that a factoid of knowledge was correct when it was indeed wrong have contributed to many accidents, injuries, and even deaths. In a critical safety environment, such as aviation, it is essential that the correct decisions are made and the knowledge supporting these decisions are complete and highly correlative, as incorrect or mistaken information built upon a foundation of guessing can undermine these goals. Therefore, a mechanism that eliminates, or at least reduces, guessing on an exam will support the effort to ultimately reduce

accidents and injuries. Confidence-based assessments may offer the needed solution to combat those limitations.

Confidence-based assessments include students' self-reported level of certainty in the marking of the answer. While taking multiple-choice tests, students indicate which answers they believe are correct while also indicating how confident they feel with their selections. A combined composite score is the result, with a rating scale technique used to reduce the variables.

Research has been performed with explorations into the implementation of confidence-based assessments across various disciplines. The work of Hunt (2003), Bruno (1995), and Gardner-Medwin (2006) have provided some research in the field, with many of their assessment techniques already moving into commercial operations. Hunt discovered highly correlative measures between human self-assessment and learning. His work also provided evidence that a confidence-based approach to knowledge assessment provides a more comprehensive measure of a student's knowledge, including the retainability of learned material. But further research is needed.

This study reviewed the current body of research regarding the limitations of traditional multiple-choice assessments, and the use of confidence-based assessment techniques to mitigate these problems. Additionally, an experiment employed the use of a confidence-based assessment tool integrated into an existing exam within a safety-critical aviation training course.

The target population of this study was professional pilots enrolled in a comprehensive training course that was designed to enhance the knowledge, skill and decision-making abilities required to command a specific aircraft model. The facility specializes in training professional pilots using first a lecture-based classroom environment for systems-specific knowledge, and

then progressing into a high-fidelity, full-motion simulator to teach the hands-on operational skills required to safely operate the aircraft in the National Airspace System.

This study embarked on an effort to explain the implementation of a confidence-based assessment strategy using a purely qualitative approach. Since the use of a confidence-based assessment would be new to the participants, and in addition to the small amount of research performed on the particular subject base, a qualitative approach would be a prudent choice to examine the situation from the perspective of the participants (Creswell, 2014). Moreover, the pilot participants arrived at the training center from all over the world with very disparate levels of age, experience, qualifications, and skills. Because of the great number of uncontrollable variables such as these, a homogeneous population sample could not be achieved, and it would have been difficult to rationalize the use of an inferential study. The analysis of a non-homogeneous sample group in a quantitative context would lead to conclusions that would be speculative at best. Therefore, because of these factors, the qualitative strategy would be the more prudent choice of inquiry.

The objectives of the study were as follows:

- Determine qualitatively, through interviews, if confidence-based assessments affected pilot study habits, increasing aircraft systems knowledge of those subjects that were identified as weak.

- Determine if confidence-based assessments enhanced the instructor's understanding of pilots' weak areas, so as to modify the remaining lectures and/or pre-simulator briefings to address those weaknesses.

- Determine if confidence-based assessments are a useful tool to enhance learning efficiency in an aviation-training environment.

Existing Research

Research has been performed with explorations into the implementation of confidence-based assessments across various disciplines. Most notable, the work of Hunt (2003), Bruno (1995), and Gardner-Medwin (2006) have provided research in the field, yet research using confidence-based assessments applied to an aviation curriculum do not exist.

Hunt began his pioneering research in the early 1980s, discovering correlative measures between human self-assessment and learning. Finding evidence of a common-sense observation, that when students are given a selection of answers for a simple arithmetic problem, they still have a chance to select the correct answer even if they do not know how to add two numbers. It is regrettable that the reliance within the educational system on this technique infers that a student who provides a correct answer purely through guesswork possesses knowledge equivalent to a student who actually knows the correct answer. This situation is even more damaging (Adams & Ewen, 2009), as it presents numerous obstacles for academic institutions in their attempt to offer a fair and representative evaluation of a student's knowledge that can be compared against a standard.

In a paper that explored a definition of personal knowledge, Hunt (2003) explained that to be useful, knowledge must be learned and retained before having an effect on behavior, and although it cannot be seen, knowledge must be inferred from observing performance. Whether that performance is a grade from a written test or through physical observations, advancement through an educational system requires successful passing of performance milestones.

But many of the traditional techniques are often ineffective or burdensome. Adams and Ewen (2009) lament, "Many institutions recognize the ineffectiveness of standard assessment processes for measuring individual knowledge, they have had a difficult time identifying better

solutions" (p. 1). Creating large multiple-choice tests is somewhat effective in reducing the negative aspect of missing a few questions, but the lengthy test takes a lot of time to complete and can be overwhelming to apprehensive students. In addition, multiple-choice tests "…fail to measure the degree of confidence that students have in their knowledge or the amount of information they retain" (Adams & Ewen, 2009, p. 1).

The multiple-choice test has been in widespread use for a long time, and much has been written about its benefits and limitations. Hunt (2003) extols some of the benefits of the multiple-choice test, which include, "…objectivity, ease and economy of administering and scoring, reliability, and the ability to measure simple and complex knowledge" (p. 108). Whereas, he also recognizes the limitations: "The knowledge of a person has more characteristics than is represented by the percentage correct score on a multiple-choice test. …a correct answer on a test is not sufficient to conclude that the knowledge has been learned" (p. 109).

A consistent objective within education circles is for learning to be more effective and efficient. As stated previously, multiple-choice assessments have limitations, while lengthy essays or face-to-face assessments are a burden on staff resources (Gardner-Medwin & Curtin, 2003). Self-assessment material offers a middle ground between the traditional multiple-choice test and lengthy essays. In a related study, Gardner-Medwin and Gahan (2003) suggest, "One of the major limitations of computer-aided assessment is that it generally implements little of the subtlety of face-to-face assessment. Confidence-based assessment is one way in which it can catch up" (p. 3).

Confidence-based assessments include students' self-reported level of certainty in the marking of the answer. While taking multiple-choice tests, students indicate which answers they

believe are correct while also indicating how confident they feel with their selections. A combined composite score is the result, with a rating scale technique used to reduce the variables.

In a research study about formative and summative confidence-based assessments for adult medical students at University College London, Gardner-Medwin and Gahan (2003) reported that for a testing scheme of either right or wrong objective answers, the confidence-based assessment model provided easily-quantifiable results. Additionally, they found the marking scheme was appropriate in formative exercises that are true/false, multiple-choice, extended matching sets, text, numbers or quantities. Although at the time of the published research report, they had only used true/false type answers in a summative setting.

From a review of the existing literature, the implementation of a confidence-based assessment scheme should be beneficial for achieving the objectives of a course when used as a formative evaluation tool. Confidence-based assessments offer a middle ground between the traditional multiple-choice answer and a lengthy essay response, resulting in a quality measure of a student's knowledge retention while still being able to quantify the results against a standard.

The goal of an effective self-assessment scheme is to produce students who retain the requisite knowledge long after finishing a course and to hold that knowledge in high confidence. Even though this goal may be lofty in its expectations, the implementation of a confidence-based assessment program elevates the teaching profession smartly toward that goal. Adams and Ewen (2009) may say this best:

> The confidence-based assessment and learning methodology provides numerous benefits
>
> to educational institutions by accelerating the student's *time to competency and*
>
> *knowledge mastery* and allows knowledge transfer to take place through technology,

which allows educators more time to work on application of knowledge and critical thinking in the classroom. …The connection of confidence and knowledge provides an acceleration of learning and improves student performance by creating a more confident and productive student. (p. 4)

The researchers in these studies primarily explored the effects of a confidence-based assessment scheme within secondary and post-secondary school environments. Further studies outside of academia may uncover additional benefits or limitations associated with these non-traditional situations. This study explored one of those areas by monitoring the implementation of a confidence-based assessment tool within a performance-based training curriculum at an aviation training facility. It was predicted that the outcome of the study would determine if the use of a confidence-based assessment methodology was beneficial to an aviation training center by providing a more efficient and comprehensive training experience for a set of pilot and instructor participants.

Research Method and Analysis

Since the use of a confidence-based assessment would be new to the participants, and in addition to the small amount of research performed on the particular subject base, a phenomenological qualitative approach was selected to examine the situation from the perspective of the participants (Creswell, 2014). Moreover, the pilot participants arrived at the training center from all over the world with very disparate levels of age, experience, qualifications and skills. Due to the number of uncontrollable variables, a homogeneous population sample could not be achieved, and it would have been difficult to rationalize the use of an inferential study. The analysis of a non-homogeneous sample group in a quantitative

context would lead to conclusions that would be speculative at best. Therefore, because of these factors, the qualitative strategy would be the more prudent choice of inquiry approaches.

This research study added a confidence-based assessment element to existing ground school exams. The results were graded and a paper report was given to both the pilots and instructors to be used as formative self-assessment tools for the remainder of their course. Researchers conducted interviews with pilots and instructors to collect their thoughts about how they accepted and used the reports.

A large aviation academy was selected for the study that trains both pilots and maintenance personnel for corporate, owner-flown and utility business jets as well as helicopters. Each curriculum is FAA (Federal Aviation Administration)-approved individually and includes a tightly controlled final exam given at the end of the ground school portion of the course. The FAA requires pilots to pass these multiple-choice exams with a grade of 80% or greater with each missed question reviewed with the instructor. This technique is called, "graded to 100%." Unfortunately, this technique does not catch those questions that were marked correct by purely guessing. Confidence-based assessments are designed to identify knowledge gaps so they can be sufficiently addressed during exam reviews.

Eleven pilots and five instructors participated as test subjects; all were male. Over the data collection period of two months, two specific initial corporate aircraft courses were selected, a two-week and three-and-a-half-week course respectively. A confidence-based assessment component was added to the normal exam given at the end of ground school. Figure 1 shows the typical schedule for the courses.
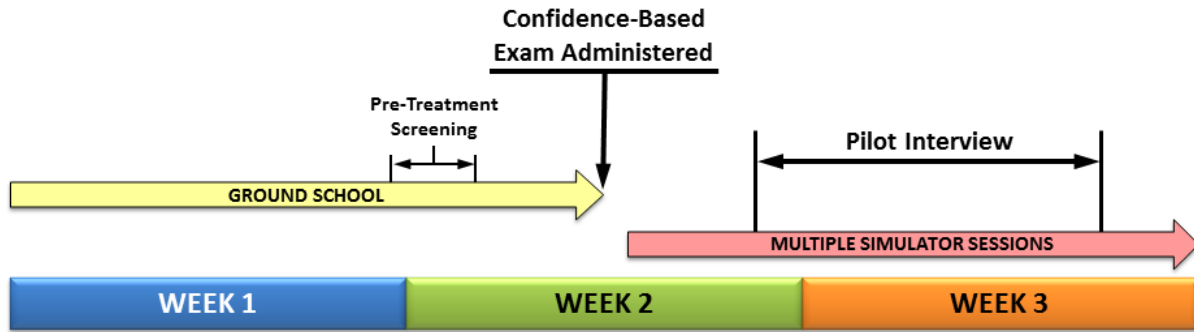
Figure 1. Research treatment timeline.

Approximately two days before administering the confidence-based assessment exam, the ground school class was addressed and an explanation was given about the nature of the study. Additionally, it was explained that the purpose was to identify subject areas where they may have been uninformed or misinformed with enough time to address the disparities before the end of the course. This pre-treatment screening session emphasized to the pilots that they were not under any obligation to participate, pseudonyms would be used, and that it would not affect their final outcome for regulatory purposes.

During the exam, the pilots were instructed to mark their confidence level next to each of the questions using a three-level scale. This arrangement allowed an immediate selection of the confidence level while the pilot was still engaged in the knowledge recall activity for the specific question. Previous studies (Bruno & Dirkzwager, 1995; Gardner-Medwin & Curtin, 2003; Hevner, 1932) have found success in using this three-level confidence marking scale, which consists of an easy to understand and remember coding format as shown in Table 1.

Table 1

Gardner-Medwin & Curtin (2003) and Bruno & Dirkzwager (1995) Exam Marking Scale

Numeric Confidence Code	Textual Equivalent	Easy-to-Remember Reference
3	High confidence	I am sure
2	Medium confidence	I am partially sure
1	Low confidence	I am not sure – I am guessing

The total confidence-weighted grade was the result of each answer given a score as a combination of correctness and self-admitted level of confidence. As used by Gardner-Medwin and Gahan (2003), the grading scheme awarded a maximum of three (3) bonus points for a highly-confident correct answer, while a confident wrong answer was penalized up to a maximum of negative six (-6) points. Based on Gardner-Medwin and Gahan's (2003) model, Table 2 shows the numeric weights given for all the combinations. For the individual aircraft subject scores, the confidence-weighted score was reported similarly, but only using one of three colors.

Table 2

Subject Confidence-Weighted Score Reporting

Student (Pilot) Mark		Calculated	
Answer	Confidence Mark	Weighted Score (reward/penalty)	Report Legend (average within subject)
Correct	3 (High)	3	**>2.4* to 3.0 "Knowledge Proficient"**
	2 (Medium)	2	
	1 (Low)	1	**>0 to 2.4* "Needs Improvement"**
Wrong	1 (Low)	0	**0 to -6.0 "Deficient Knowledge"**
	2 (Medium)	-2	
	3 (High)	-6	
* The 2.4 threshold was used to accommodate multiple questions covering a single aircraft system subject. This allowed for a confidence mark of 1 or 2 for a single correct answer to still achieve a "Knowledge Proficient" score on the report.			

A well-formatted, color-coded report (Figure 2) of the graded exam was generated from this process and given to the pilot within an hour of taking the exam. Multiple scores were presented. The overall raw score was the graded knowledge responses without any confidence weight added, which was used for course certification purposes. The confidence weighted score, in percentage, represented the combination of knowledge and confidence across all the subjects. The individual subject area scores were reported as a color-coded composite score using the grading scheme presented in Table 2.
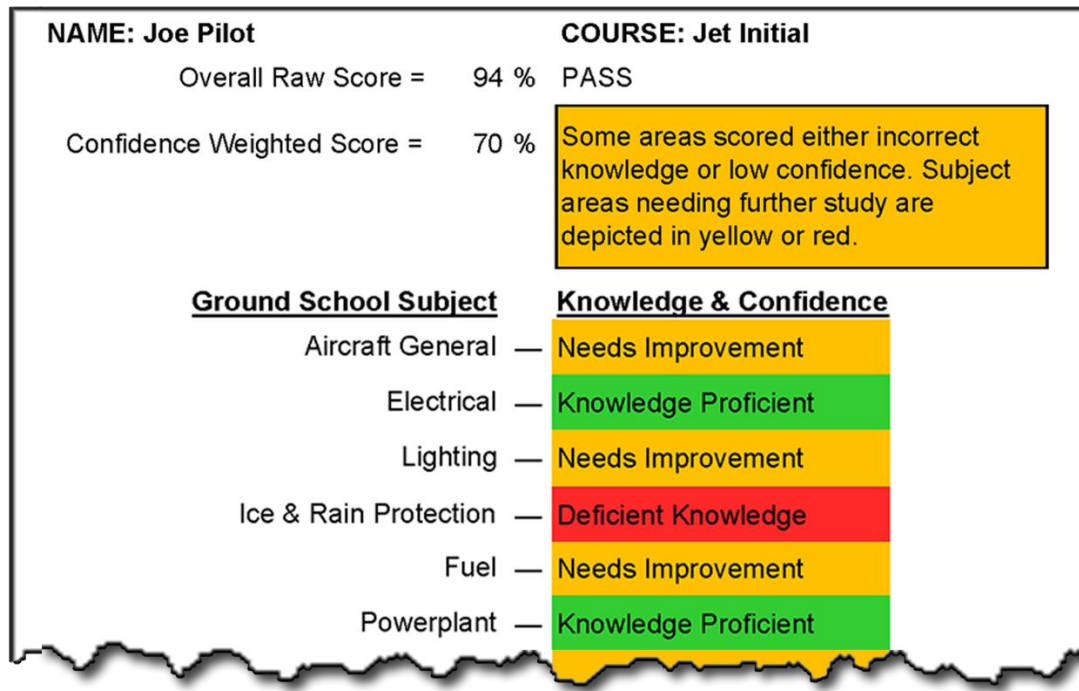
Figure 2. Sample graded exam report.

The intent of the report (Figure 2) was to give the pilot an objective view of his level of aircraft systems knowledge in an easy-to-use format. This hopefully encouraged the pilot to study on his own and seek greater knowledge about those weak areas. Upon analysis by the instructor, and self-realization by the student, a highly-confident wrong answer (shown as a red "Deficient Knowledge") represents a falsely-held belief (misinformation), which would deserve special attention and emphasis during follow-up learning sessions. This report also gave the instructor a focused plan for the "graded to 100%" discussion. Instead of relying on intuition to review weak spots, the instructor had an evidence-based plan to cover those areas in which the pilot was found to be weak.

As is often the case at this aviation academy, a different instructor than the ground school instructor was assigned for the simulator sessions, and the turnover between instructors rarely

covers the details of any aircraft system deficiencies. During the study, the simulator instructors also received the graded exam reports, finding them especially helpful as a tool to focus the briefing session discussions on those knowledge discrepancies.

A few days after the exam, during the simulator period, each pilot was privately interviewed by the principal investigator. An interview script was used throughout the data collection process to achieve consistency across the transcribed data. Additionally, all the participating instructors were interviewed after all the pilot interviews were completed.

The post-treatment pilot interview instrument was designed to gather data on the following constructs:

- familiarity with confidence-based assessments;

- comfort level of revealing own confidence;

- ease of use;

- perception change;

- study habit change; and

- opinion about benefits to training.

The semi-structured interview used a combination of questions to gather the pilots' opinions and perceptions about their experience using the confidence-based assessment tool. Each construct was covered by asking at least one question from each of the categories. If the initial answer in any of the construct categories were incomplete, a follow-up question from that category was asked until a rich, descriptive answer was received. The interview questions were reviewed with research colleagues to address any perceived bias and ensure the questions were not leading or influential. The interview exchange was captured using an audio recording

device, with written notes used as a back-up. Each digital audio file was transcribed into a text file by an external transcription service for the sake of expediting the research study schedule and eliminating any researcher bias during the transcription process. Only the alias/pseudonym of each pilot was used to identify both the audio recordings and transcribed text files.

The post-treatment instructor interview instrument was conducted in the same manner as the pilots and was designed to gather data on the following constructs:

- familiarity with confidence-based assessments;

- ease of use;

- pilot acceptance;

- instructor insight into pilots; and

- improved learning.

The primary objective of the analysis was to determine if the introduction of confidence-based assessment into a formative self-awareness tool had an effect on the study habits of pilots resulting in a greater understanding of aircraft systems knowledge. This analysis also sought to gain insight from the instructors about the impact of a confidence-based assessment component added to their courses.

The transcript utterances were organized for ease of coding and category reduction. A second researcher was used to verify the coding scheme and offer a peer debriefing of the transcribed interviews, which contributed to the credibility of the study. As described in Saldaña (2013), a melding of the evaluation and magnitude coding methods were used. The evaluation coding method was used because of its inclination toward seeking judgment about value, significance, and implication of specific programs to be used for policy making. In the case of

this study, the intent of the research was to determine if the implementation of confidence-based assessments within an existing curriculum would be worthwhile; therefore, the evaluation coding method satisfies that intent from a programmatic standpoint. Combined with the evaluation coding method, the magnitude coding method was used because of its focus on the amount and polarization of attitudes as it applies to specific inquiries of attitude. The application of a magnitude to specific interview answers, and stand-alone excerpts, provided the means to delicately quantify the extent of similar opinions in support of the research goals.

The first coding pass identified excerpts that were specific replies to the interview questions and also identified any comments reflecting an attitude, either positive or negative, toward the use of confidence-based assessments in the particular situation. Any emerging themes discovered were identified, coded and broken down into their basic components. The process of analyzing the qualitative text data required reducing the text data into its smallest components, then categorizing those components into any overlying themes (Creswell & Clark, 2011). Creswell and Clark (2011) also claim that "The core feature of qualitative data analysis is the coding process…. grouping evidence and labeling ideas so that they reflect increasingly broader perspectives" (p. 208).

Research Findings and Discussion

The data collection was conducted over a two-month period wherein interviews of pilots and instructors was tightly controlled. Two specific courses were selected from the course offerings of the Flight Training Academy, a two-week and three-and-a-half-week course respectively. Since the course offerings at the Flight Training Academy generally fall under these two types of schedules, a representative cross section was achieved.

Five instructors were recruited to participate in the study. Since this was a new, and possibly disruptive, method to administer end-of-course exams, the principal investigator recruited instructors who were open to try new ideas and also showed an interest in the findings of the research study. Once a cadre of instructors was in place, the treatment and data collection period began. There was not any monetary compensation to participate or any ramifications in opting out.

Approximately two days before administering the confidence-based assessment-enhanced exam, the study's principal investigator addressed the ground school class of the selected course and explained how confidence-based assessments work and the nature of the study. It was explained that this type of new exam technique was being tried by the academy for possible inclusion into future courses and their opinion on this trial experiment would be greatly appreciated. Additionally, it was explained that the purpose was to identify subject areas where they may have been uninformed or misinformed with enough time to address the disparities before the end of the course.

This recruitment and pre-treatment screening session emphasized to the pilots that they were not under any obligation to participate and that their participation would not affect their final outcome for regulatory purposes. By explaining the confidence-based assessment tool a couple days ahead of administering it, the goal was to let the pilots absorb the intent of the experiment and decide if they would like to participate. The intent of this waiting period was to "let it sink in" so as to not surprise the pilots with this new and potentially upsetting change in the normal curriculum in an effort to possibly reduce some concerns about experimental validity. Eleven pilots and five instructors agreed to participate in the study.

The pilots averaged 20 years of aviation experience (5,400 average flight hours), and the instructors averaged 22 years of experience (5,300 average flight hours). Experience in years, and specific flight hours, are base requirements that regulatory authorities (FAA, etc.) use to award operational flight privileges. For example, the requirements to take the test for an Airline Transport Pilot certification is 23 years old and 1,500 flight hours (Federal Aviation Administration [FAA], 2015). Although these are just the minimum requirements to be certificated by the FAA, the insurance underwriters often require much more experience (multiple factors) than the minimum before fully insuring a pilot to operate the aircraft used within this study. Since the participants in this study exceeded those requirements, the sample was representative of the larger aviation population.

Transcriptions of the recorded interviews were generated and used to analyze the responses from both the pilots and instructors. The overarching questions were whether the new tool was accepted and if the benefits outweighed any additional efforts.

The first interview question sought to discover what prior experience the participant had with confidence-based assessments. Both the pilots and instructors were asked the identical question, "What, if any, prior experience do you have using confidence-based assessments?" The responses from the pilots were split, with five (5) responding they were not familiar at all, while the remaining six (6) stating they were somewhat familiar. Among the five instructors, four (4) responded they were not familiar at all, with the remaining instructor stating that he was very familiar. Because of that lack of experience, it was imperative that the pilots and instructors fully understood the directions explaining the exam procedures. All the pilots and instructors responded that they clearly understood the directions and that there was not any confusion about how to mark their confidence levels or use the graded report.

Across the group, there was a slight reluctance among the pilots in marking their confidence level for each answer, which was expected due to the new and interruptive nature of the new exam tool. They did not, however, consider it burdensome. Some of the pilots admitted to being frustrated, explaining that their frustration was not due to the addition of a confidence marking element, but that their final grade was lower than expected or having had difficulties understanding specific questions. The pilots also responded that the actual time to mark their confidence for each answer was inconsequential (2-5 seconds) depending on the complexity of the question but overall did not appreciably extend the total time to take the exam.

As for their confidence marking strategy, many reported using a simple binary technique (if-then-else), while others employed a more intricate strategy that involved only marking unconfident answers while taking the exam, then returning after completing the exam to mark all the others with high confidence. This technique, although unexpected, is commendable and will need to be addressed when developing a computer-based exam using a confidence-based element.

When asked about the graded exam report and its use, nine, out of eleven pilots, reported that their study habits changed in a positive way. A couple of the responses were short, such as, "I think it's a good idea" and "I really like this" while all the others had more to say, which was an indicator itself that they perceived the benefits. Within the excerpts from the pilots, there were specific comments of "very beneficial," "helpful," "brilliant," "excellent," and "great… I enjoyed getting the information," while not a single disparaging comment was recorded.

It appeared the pilots used the confidence-based assessment tool as it was intended first as a self-awareness tool identifying the weak subject areas and then using that awareness to focus their remaining study efforts. One pilot reported that his study habits changed only somewhat

for the better. Although, missing only a single exam question, his study habits were most likely already very effective. The remaining pilot revealed that by the time of the interview, he had not had the time yet to study since receiving the exam report. Even though these last two pilots' responses were not coded as positive responses, their circumstances most likely prevented them from gaining full advantage of the process, and were therefore unable to provide an evaluation of their experience with the confidence-based assessment tool.

A common theme was observed among the majority of the pilots: they had very positive things to say about their experience and the addition of the graded exam report. Many of the comments related to study habits consisted of: "reread chapters," "dug into more," "review chapters," "get better understanding," "look at things," and so forth. It did not appear from the comments that any of the pilots fundamentally changed their study habits or that any new techniques were employed. Although it could be surmised that merely the addition of reading and reacting to the graded exam report that identified specific weak subjects and confidence levels insinuates a new technique was indeed being used. It relates directly to the learning theory of reflection, by offering an easy-to-use tool that objectively identifies weak areas so the pilot can relive the specific learning situations to fill any information gaps or correct any misinformation.

Although they did not openly say so, three pilots alluded to the additional time and effort involved, but in a roundabout way saying, "…it would be worth it." This circles back to the topic of acceptance and the way the pilots could agree to, and embrace, the use of confidence-based assessments as an interruption to their normal processes while offering worthwhile benefits.

The instructors were also very encouraging, with all five stating they saw the new tool as worthwhile and they were very comfortable with its use. There was not a single instance of any outwardly negative responses; however, three instructors noticed some of the pilots did not fully understand the concepts of confidence-based assessments and how it could help them improve their course experience. Although, this issue could be overcome by proper and effective training.

Each instructor had a slightly different take on the experience, but all saw the tool as a comprehensive snapshot of the pilot's weak areas. This particular finding was one of the primary intended benefits, to supply the instructors with an evidence-based plan that would help focus the aircraft system knowledge discussions during pre-simulator briefing sessions. Two instructors specifically extolled the benefits of the exam reports as tools for long-term analysis of their own classroom delivery performance, which may be explored in follow-up research.

In conclusion, the findings support that the confidence-based assessment exam is a valuable tool that instructors can use to address knowledge gaps and improve the training experience. In addition, these findings support the concept that, if properly trained and administered, a robust confidence-based assessment tool would be minimally-burdensome while offering worthwhile benefits.

Recommendations

This study was conducted to explore a confidence-based assessment tool and determine whether it could effectively enhance the learning experience. Several specific recommendations could apply to the actual implementation of this new tool within existing pilot or maintenance courses and distance learning applications. The following list offers specific guidelines to be

used when designing a practical confidence-based assessment, either paper-based or electronically delivered.

- Schedule the confidence-based exam/test/quiz early enough during the class so that there is ample time to self-assess progress and plan remaining study time.

- Provide extremely clear and thorough confidence marking directions.

- Restrict the available confidence level selections to a maximum of three (3).

- Delineate the three confidence level labels to:

   o High Confidence (I am sure)

   o Medium Confidence (I am partially sure)

   o Low Confidence (I am not sure – I am guessing)

- Accommodate a holistic confidence marking strategy by allowing the pilot the ability to move to the next question without marking the confidence level, then having the ability to return to mark their confidence level for each of their answers.

- Ensure the instructors are thoroughly briefed on how to effectively use the confidence-based assessment report.

- Provide a checkbox for each question allowing the pilot to select a "Don't understand question" option, so as to not confuse a low confidence mark with a misunderstanding of the question.

- Be aware that the introduction of a new testing technique will most likely meet with some resistance by all the participants. Persistence, careful planning and patience are requirements for the successful implementation of any new system.

Future Research

The results of this study provide evidence that a confidence-based assessment tool has merit for use within a corporate/utility flight-training academy, which is only a small segment within a broad breadth of aviation training programs. Many opportunities exist to adapt a confidence-based assessment tool within those other arenas. The following are some possible avenues where follow-on research may be pursued.

- A study that explores the effects of confidence-based assessments on courses taught to basic, primary or advanced pilots or maintainers.

- A study that explores the possible quantitative analysis of a group of pilots and maintainers when tested for performance improvement after the treatment of a confidence-based assessment formative exam, against a similar control group.

- A similar qualitative study that compares the use of confidence-based assessments between a short course (one week) and a relative long course of three weeks or more.

- A study that employs confidence-based assessment techniques as an element of the branching decision formulae used within a Computer Adaptive Testing scheme for aircrew certification exams.

- A study that addresses the differences between Eastern and Western learning cultures, as it relates to acceptance and accessibility of confidence-based assessments in a safety-critical training organization.

References

Adams, T. M., & Ewen, G. W. (2009). The importance of confidence in improving educational outcomes. *Proceedings of 25th annual conference on distance teaching & learning*. Aug. 4-7, Madison, Wisconsin. Retrieved from http://www.uwex.edu/disted/conference/Resource_library/proceedings/09_20559.pdf

Bruno, J. E. (1995). *Information reference testing (IRT) in corporate and technical training programs.* Los Angeles, CA: University of California Los Angeles.

Bruno, J. E., & Dirkzwager, A. (1995, December). Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic perspective. *Educational and Psychological Measurement, 55*(6), 959-966. https://doi.org/10.1177/0013164495055006004

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4a ed.). Thousand Oaks, CA: SAGE Publications.

Creswell, J. W., & Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Los Angeles, CA: SAGE Publications.

Federal Aviation Administration (FAA). (2015). *Certification: Pilots, flight instructors, and ground instructors; Subpart G—Airline transport pilots*, 14 C.F.R. § 61.153.

Gardner-Medwin, A. R. (2006). Confidence-based marking - Towards deeper learning and better exams. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education.* London: Routledge, Taylor and Francis Group Ltd. PMCid:PMC2082770.

Gardner-Medwin, T., & Curtin, N. (2003). Certainty-based marking (CBM) for reflective learning and proper knowledge assessment. *In REAP Int. online conference on assessment design for learner responsibility.* May 29-31, 2007.

Gardner-Medwin, A. R., & Gahan, M. (2003). *Formative and summative confidence-based assessment proceedings*. 7th International Computer-Aided Assessment Conference, Loughborough, UK, July 2003, 147-155.

Hevner, K. (1932). A method of correcting for guessing in true-false tests and empirical evidence in support of it. *The Journal of Social Psychology, 3*(3), 359-362. https://doi.org/10.1080/00224545.1932.9919159

Hunt, D. (2003). The concept of knowledge and how to measure it. *Journal of Intellectual Capital, 4*(1), 100-113. https://doi.org/10.1108/14691930310455414

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Los Angeles, CA: SAGE Publications.