THE IMPLEMENTATION OF A CONFIDENCE-BASED ASSESSMENT TOOL
WITHIN AN AVIATION TRAINING PROGRAM

Paul F. Novacek, M.A.S.

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS
August 2015

APPROVED:
Lin Lin, Committee Chair
Greg Jones, Committee Member
Scott Warren, Committee Member
Cathie Norris Interim Chair of the Department of Learning Technologies
Costas Tsatsoulis, Interim Dean of the Toulouse School
Novacek, Paul F. *The Implementation of a Confidence-Based Assessment Tool Within an Aviation Training Program.* Doctor of Philosophy (Learning Technologies), August 2015, 131 pp., 6 tables, 4 figures, references, 25 titles.

Traditional use of the multiple-choice question rewards a student for guessing. This technique encourages rote memorization of questions to pass a lengthy exam, and does not promote comprehensive understanding or subject correlation. This begs the question; do we really want question memorizers to operate the machinery of our industrialized society? In an effort to identify guessing on answers during an exam within a safety-critical aviation pilot training course, a qualitative research study was undertaken that introduced a confidence-based element to the end-of-ground-school exam followed by flight simulator sessions. The research goals were twofold, to clearly identify correct guesses and also provide an evidence-based snapshot of aircraft systems knowledge to be used as a formative study aid for the remainder of the course. Pilot and instructor interviews were conducted to gather perceptions and opinions about the effectiveness of the confidence-based assessment tool. The finding of overall positive interview comments confirmed that the pilots and flight instructors successfully used the confidence-based assessments as intended to identify weak knowledge areas and as aids, or plans, for their remaining study time. The study found that if properly trained and administered—especially through a computer-based medium—a robust confidence-based assessment tool would be minimally-burdensome while offering worthwhile benefits.
Copyright 2015

by

Paul F. Novacek
# Table of Contents

<table>
<thead>
<tr>
<th>Chapters</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LIST OF TABLES</strong></td>
<td>vi</td>
</tr>
<tr>
<td><strong>LIST OF FIGURES</strong></td>
<td>vii</td>
</tr>
<tr>
<td><strong>CHAPTER 1 INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>Research Questions</td>
<td>4</td>
</tr>
<tr>
<td>Primary Research Question</td>
<td>4</td>
</tr>
<tr>
<td>Sub-Questions</td>
<td>4</td>
</tr>
<tr>
<td><strong>CHAPTER 2 LITERATURE REVIEW</strong></td>
<td>6</td>
</tr>
<tr>
<td>Traditional Assessment Issues</td>
<td>6</td>
</tr>
<tr>
<td>Confidence-Based Assessment Technique</td>
<td>9</td>
</tr>
<tr>
<td>Marking System</td>
<td>10</td>
</tr>
<tr>
<td>Grading Scheme</td>
<td>12</td>
</tr>
<tr>
<td>Confidence-Based Assessment Procedure</td>
<td>13</td>
</tr>
<tr>
<td>Benefits</td>
<td>14</td>
</tr>
<tr>
<td>Implementation</td>
<td>16</td>
</tr>
<tr>
<td>Considerations</td>
<td>18</td>
</tr>
<tr>
<td>Summary</td>
<td>20</td>
</tr>
<tr>
<td><strong>CHAPTER 3 RESEARCH METHODS</strong></td>
<td>22</td>
</tr>
<tr>
<td>Strategy for Inquiry</td>
<td>22</td>
</tr>
<tr>
<td>Research Study Facility</td>
<td>23</td>
</tr>
<tr>
<td>Study Elements</td>
<td>25</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>Table 1</td>
<td>Gardner-Medwin and Curtin (2003a) and Bruno (1995) Confidence Scale Models</td>
</tr>
<tr>
<td>Table 2</td>
<td>Gardner-Medwin and Gahan (2003b) Grading Scale</td>
</tr>
<tr>
<td>Table 3</td>
<td>Composite Score Grading Scheme</td>
</tr>
<tr>
<td>Table 4</td>
<td>Subject Confidence-Weighted Score Reporting</td>
</tr>
<tr>
<td>Table 5</td>
<td>Pilot Demographic Data</td>
</tr>
<tr>
<td>Table 6</td>
<td>Instructor Demographic Data</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Treatment timeline</td>
<td>27</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Sample exam question</td>
<td>30</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Sample graded exam report</td>
<td>34</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Data element consolidation</td>
<td>39</td>
</tr>
</tbody>
</table>
CHAPTER 1
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. Their odds are even better when guessing on a true/false choice; 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. But 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.

Research has been performed with explorations into the implementation of confidence-based assessments across various disciplines. Most notably, the work of Hunt (2003), Bruno (1995) and Gardner-Medwin (2006) has provided some research in the field, with many of their assessment techniques already moving into commercial operations. Hunt began his pioneering research in the early 1980s, discovering highly correlative measures between human self-assessment and learning. His work since then has continued to provide 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.

Hunt (2003) found evidence for a common-sense observation, that when a student is given a selection of answers for a simple arithmetic problem, he or she still has 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.

Guessing on a few questions in Math-101 may be perceived as somewhat benign, but taken to its eventual limit, the most problematic aspect of this guesswork manifests itself within a safety-critical, or high-consequence environment. Many industrial accidents, injuries and even deaths have been caused by misinformed operators who held a steadfast belief that a factoid of knowledge was correct when it was indeed wrong.

In a highly-critical safety environment, it is essential that correct decisions are made and the knowledge supporting these decisions is complete and highly correlative. Incorrect or mistaken information built upon a foundation of guessing undermines 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.

This dissertation reviews the current body of research regarding the use of traditional multiple-choice assessments, and the use of confidence-based assessment techniques to mitigate these problems. Additionally, an experiment was conducted which 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 this 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 approaches.

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 each pilots’ weak areas, so as to modify the remaining lectures and/or simulator briefings to address those weaknesses.
- Determine if confidence-based assessments are a useful tool to enhance learning efficiency in an aviation training environment.
Research Questions

The objectives of the study were straightforward, but to achieve these goals, specific questions had to be explored. The extent to which answers were found depended on the expressive self-awareness and articulation of the research participants. Responses to interview questions were used to explore the following questions.

**Primary Research Question**

How does the addition of a confidence-based assessment element to a mature exam change the habits of both students and instructors within an aviation training environment?

**Sub-Questions**

- To what extent does the addition of confidence-based assessments change student habits?
- To what extent does the addition of confidence-based assessments change instructor habits?
- What is the perception by both students and instructors that confidence-based assessments may be a useful tool to enhance learning efficiency in an aviation training environment?
- What are any limitations of the implementation of confidence-based assessments into an aviation training environment?

The answers to these questions were investigated through qualitative inquiry methods using semi-structured interviews of pilots and instructors at an academy-style flight training facility. An analysis was conducted to provide a rationale either supporting or opposing the
implementation of a confidence-based assessment scheme within this particular aviation training situation.

This section offered a brief introduction into the primary limitations of traditional multiple-choice assessments, that guessing needs to be tracked and misinformation needs to be corrected. The use of confidence-based assessments was identified as a possible strategy to mitigate those limitations. The pioneering researchers that explored confidence-based assessments were also introduced along with a brief overview of the study objectives and research questions. The following literature review section explores the previous research efforts in an attempt to learn the state of the confidence-based assessment knowledge base.
CHAPTER 2
LITERATURE REVIEW

There have only been a handful of research studies and review papers that have attempted to address the issue of confidence-based assessments (Adams & Ewen, 2009; Florian, 2010). But those few studies were thorough and significant enough that a solid base of knowledge exists. This literature review explores the body of research, over the previous 40 years, of the problems associated with traditional assessments, and the use of confidence-based assessment techniques to mitigate those problems. An explanation of confidence-based assessment techniques and their implementation considerations is also included. The review concludes with the expected benefits of implementing confidence-based assessment techniques and a suggested position for further research.

Traditional Assessment Issues

Traditional knowledge assessment methods attempt to focus on the recall of previously presented information. Often these assessments rely on multiple-choice exams, tests or quizzes that only measure knowledge by a right or wrong answer. 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 (2009) laments, “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 long 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, p. 1).

Exploring the multiple-choice test in greater detail, much has been written about the benefits and the limitations of a system that has been in widespread use for a long time. 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).

Another limitation that was identified with traditional assessments is the lack of knowledge retention even after a student successfully passed the test. In a study that explored the importance of confidence in improving educational outcomes, Adams (2009) found that, “…even when students pass these [traditional] assessments, they lack the necessary skill set to perform well in the work place” (p. 1). This infers that even a high test score is not a sufficient indicator of subsequent job performance, although as a society, we generally hold this view.

Many variables are involved when attempting to predict job performance, with knowledge
retention being a major contributor. The challenge is how to accurately measure a level of understanding in a quantifiable way. Adams (2009) continued,

For years, we have been teaching and assessing using traditional models that encourage guesswork in the testing process. There are even strategies on test taking provided to students to maximize their score, rather than gain an understanding of their knowledge acquisition, skills and competency to perform. (p. 1)

An entire cottage industry has evolved around the “just get me through the test” cram courses that just promote rote memorization and not a sound education (Joy, Miyagawa, Suami & Weisselberg, 2006). A common practice among entrance exam or certification test takers is to purchase test-prep books or software that are just reprints of test question banks. Many U.S. Government certification exams are considered public domain, therefore the Freedom of Information Act allows the public to obtain, word-for-word reprints of all the questions (FAA, 2015b). This rather defeats the purpose of an exam in the first place, since answers can be easily memorized.

To mitigate this possibility, the Federal Aviation Administration (FAA), for instance, employs a few tactics to achieve the intended knowledge assessment. As observed in an FAA test prep manual for aircraft maintenance technicians (ASA Test Prep Board, 2015) the actual proctored exam consists of 100 questions randomized from a bank of 1,028 released questions from the FAA. Additionally, the FAA does not release the answers. This scheme returns the exam experience to a true knowledge assessment to some extent because of the difficulty in memorizing over 1,000 question and answer combinations.

An educator’s intentions are usually virtuous, but good intentions without efficient tools often result in a learner’s perspective that lasts only until the end of the course. This observation is not a new one. Ahlgren (1969) wrote more than 40 years ago,

Most instruction is intended to have long-term effects—over months at least and perhaps over an entire lifetime. Yet most achievement testing is done immediately after an
instructional sequence. I suspect (in fact, I know very well) that a substantial part of knowledge measured on an achievement test is ephemeral knowledge, stored fleetingly for the purpose of taking the test. (pp. 2-3)

This level of transient knowledge does little for a student within a university curriculum that employs a series of courses taken over a few years. The use of scaffolding techniques within the university system requires the use of knowledge learned in the previous courses (Gardner-Medwin, 1995). Therefore, ephemeral knowledge is somewhat akin to taking two steps forward and one step back. It is a widely held belief that a comprehensive self-assessment scheme greatly improves knowledge retention and complete understanding of course material. Gvozdenko and Chambers (2007) confirmed this belief,

There is a difference in the knowledge of a test taker who answers correctly but is very unsure of it or has made a lucky guess, and a test taker who is correct and is extremely sure of their answer. On a regular test this difference is not evident to the learner, who will not realize the real status of his/her knowledge and whose further learning may be jeopardized by flawed foundations. A correct/incorrect scoring system with no indication of certainty fails to collect this information and thus robs lecturers and tutors of a chance to identify and address weaknesses in students’ knowledge. (p. 2)

One self-assessment scheme has been studied that addresses those shortcomings. It is referred to by a few names, notably, certainty-based assessment, confidence-based testing, certainty-based marking (CBM) or confidence-based learning. Each of these terms are defined slightly differently by the researcher who first coined the name, but hereafter throughout this report, the overlying term confidence-based assessment will be used.

Confidence-Based Assessment Technique

It is a desire to want 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, 2003a). Self-assessment material offers a middle ground between the traditional multiple-choice test and lengthy essays. “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” (Gardner-Medwin & Gahan (2003b), 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. A confidence-based assessment element added to an existing multiple-choice test has many advantages. Hunt (2003) concluded,

It provides a more comprehensive measure of a person’s knowledge; detects and identifies topics in which people are misinformed; measures the retainability of learned material; may reduce gender and perhaps ethnic bias in the assessment of knowledge; encourages study and enhances learning, identifies and provides practice to individuals who over or under estimate their own knowledge; and helps identify test items which may be misleadingly constructed. (p. 112)

In a research study about formative and summative confidence-based assessments for adult medical students at University College London, Gardner-Medwin and Gahan (2003b) reported that for a testing scheme of either right or wrong objective answers, the confidence-based assessment model provides easily-quantifiable results. Additionally, they found the marking scheme is 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.

Marking System

The instrument of a confidence-based assessment tool uses a double-tiered marking system to gain two distinct but interrelated scores in response to a specific question. Test takers
select the answer that they believe to be correct, then mark how confident they feel in their answer before moving on to the next question (Gardner-Medwin, 1995; Gardner-Medwin & Curtin, 2003a).

Once test takers mark the answer that they believe is correct, they are asked to mark their level of confidence, usually consisting of a multi-level ranking choice. Previous studies have used either a three-level choice (Bruno & Dirkzwager, 1995; Gardner-Medwin & Curtin, 2003a; Hevner, 1932) or a five-level choice (Hassmen, Hunt & Dybeck, 2002; Hunt, 2003). The three-level confidence scale consists of an easy to understand and remember coding format, assigning either a number or a color, as illustrated in Table 1.

Table 1

<table>
<thead>
<tr>
<th>Confidence Level (degree of certainty)</th>
<th>Numeric Marking</th>
<th>Color-Coded Marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>3</td>
<td>Green</td>
</tr>
<tr>
<td>Medium</td>
<td>2</td>
<td>Yellow</td>
</tr>
<tr>
<td>Low</td>
<td>1</td>
<td>Red</td>
</tr>
</tbody>
</table>

The five-level model used by Hassmen et al. (2002) and Hunt (2003) includes the following numbered list. Although, Gardner-Medwin and Gahan (2003b) was critical of using, “…language terms such as ‘very sure’, ‘uncertain’, etc., which may mean different things to different people” (p. 3).

1. Extremely sure
2. Very sure
3. Somewhat sure
4. Very unsure
5. Not sure at all

These scales have been used in experimental trials, but in a study that specifically monitored the range of certainty markings, Gvozdenko and Chambers (2007) reported that 85% of their student subjects preferred a three or four-level scale over a five-level certainty scale. It is clear the participants in this study preferred a rating scale with less gradation, inferring that the less complex rating scales were more easily understood.

Grading Scheme

Table 2 shows the grading scale used by Gardner-Medwin and Gahan (2003b) in both their formal research studies and within an active biomedical student curriculum at University College London. In essence, the cumulative grade using the certainty-based grading scale assigns neither penalties nor bonus points for low certainty, but in the medium and high certainty levels, bonus points are awarded for correct answers, while a greater proportion of points are deducted for wrong answers.

Table 2

*Gardner-Medwin and Gahan (2003b) Grading Scale*

<table>
<thead>
<tr>
<th>Degree of Certainty</th>
<th>C=1 Low certainty</th>
<th>C=2 Medium certainty</th>
<th>C=3 High certainty</th>
<th>No reply</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark if Correct</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Penalty if wrong</td>
<td>0</td>
<td>-2</td>
<td>-6</td>
<td>0</td>
</tr>
</tbody>
</table>

In a research study with adult medical students, Gardner-Medwin (2003b) successfully employed this three-level confidence scale, where the upper two certainty levels apply negative weights for wrong answers, effectively assigning a greater penalty for each higher level of
misinformation. This gradation is critical, and weighs heavily in the test taker’s decision for each answer, offering motivation for a more reliable self-assessment. It is also important that students realize that confidence-based assessments are designed to help them identify their weak areas within a non-threatening environment. During their analysis of test takers’ thought processes, Gardner-Medwin and Gahan (2003b) found that, “It is important to recognize that the objective of confidence-based marking is not to reward or discourage self-confidence. The aim is to encourage reflection, self-awareness, and the expression of appropriate levels of confidence” (p. 3). In a related study that specifically addressed reflective learning, Gardner-Medwin and Curtain (2003a) reported that “Certainty-based marking rewards reflection that leads to the conclusion that an answer is less certain than initially thought” (p. 3). They continue that the approach is based in probabilistic decision theory, aiding the student in judging the value and level of sureness.

In his ground breaking work on confidence-based assessments, Hunt (2003) declared, “A computer analysis of the person’s answers and self-assessment certainty responses provides multidimensional scores about a person’s knowledge that remedy some deficiencies of knowledge assessment and achievement tests now employed” (p. 100). He also found that, “…the more certain the knowledge or belief then the more likely, more rapid and more reliable is the response” (p. 107).

Confidence-Based Assessment Procedure

The combination of asking test takers to mark what they believe is the correct answer as well as their confidence in their selection captures a retention profile of an individual’s
knowledge, separating what they “think” they know versus what they “actually” know. The following is a systematic procedure of the confidence-weighted scoring scheme.

1. A student is given an explanation (verbal or handout) of how confidence-based assessment is used, how to mark their answers and a suggested strategy to achieve the most effective score. The point assignment system as shown in the preceding Table 2 is also explained to the student. It is imperative at this point that the student fully understands the consequences of marking an answer incorrect with a high confidence level.

2. The student is given the exam in either paper-based or computer-based format.

3. With each question, the student marks the answer they believe is the most correct.

4. The student then indicates their level of certainty in their answer before moving on to the next question.

5. During the grading phase, each answer is assigned the weighted score according to the scheme as illustrated in Table 2.

6. The final composite grade for the full exam can then be used as a summative grade for the entire course, or individual questions can be analyzed as part of a formative scheme to direct the topics for the remaining part of the course.

An important element within this confidence-based assessment scheme is that test takers know and understand the marking methodology, especially that points will be deducted from their final score for highly-confident wrong answers. If the confidence-based scheme is going to be used as a formative assessment, then the test taker must be informed of its use as a self-assessment tool and how it could be used effectively in their own motivation and appraisal of future learning parameters, as well as informing the instructor of what subject matter to focus their efforts.

Benefits

As with any new process within an organization, the benefits must be weighed against the cost before any serious effort can be applied to the initiative. Implementing a confidence-based
assessment scheme is no different, as the effort will require time, manpower, equipment, logistics and, most importantly, a shift in the basic teaching philosophy. Before all this can begin though, a thorough understanding of the benefits is paramount in the effort to gain acceptance and approval to explore how a confidence-based assessment scheme will improve the overall product—a better-educated student released into society.

In their studies with first- and second-year medical and biomedical students, Gardner-Medwin and Curtin (2003a) found that, “To measure knowledge, we must measure a person’s degree of belief… -and encourage an awareness that uncertain but correct answers, or lucky guesses, are not the same as knowledge, and that confident wrong answers deserve special attention” (p. 1). They also found that confidence-based assessments improved the statistical reliability of the exam data as a measure of student performance, compared with conventional marking.

In a study about the retrieval of learner knowledge after a 1, 6 or 12-week period, Hassmen et al. (2002) found that a higher certainty was associated positively with retention. In a related study, Hassmen, Sams and Hunt (1996) found,

Participants took either a performance or written test, with or without making self-assessments about how sure they were that their selected answer was correct. Participants who learned with self-assessment selected a greater percentage of correct responses, both during the learning trials and on the tests that followed, than did those who learned without self-assessment. (pg. 1091)

In a different study about the behaviors of cadets who were given an online formative and summative certainty-based assessment Gvozdenko and Chambers (2007) found that, “Statistics on the test questions were found to be consistent across the samples for two consecutive years, which suggests that it is a sensitive measure for evaluating the current state of the knowledge and knowledge acquisition” (p. 215). Additionally, Gvozdenko also found that certainty-based assessment “…can provide a long term monitoring of skills retention within a cohort” (p. 206).
Summarizing his efforts, Hunt (2003, p. 112) offers a list of notable findings when a confidence-based assessment scheme was applied to the learning process:

- Provides a more comprehensive measure of a person’s knowledge.
- Detects and identifies topics in which people are misinformed.
- Measures the retention of learned material.
- May reduce gender and perhaps ethnic bias in the assessment of knowledge.
- Encourages study and enhances learning.
- Identifies and provides practice to individuals who over or under estimate their own knowledge.
- Helps identify test items that may be misleadingly constructed.

Reducing the findings of a series of research studies into a few overarching concepts is worthwhile to a school administration or private training provider when contemplating the use of a new assessment scheme. As with most enterprise initiatives, an implementation team seeking approvals need those research-backed features condensed into easily understandable and manageable points. The following section addresses those implementation issues and provides the supporting research.

Implementation

A confidence-based assessment scheme can generally be thought of as a tool for both the student and teacher, with each using it for slightly different purposes (Gardner-Medwin et al. 2003a, 2003b & 2006). From the student’s perspective, when offered the use of an interactive computer-based assessment that integrates confidence awareness, the instant feedback allows them to quickly confirm their belief in selecting the right answer or to correct their perception of wrong answers. Likewise, the teacher can obtain a report of the results from a learning
management system of a confidence-based assessment given to the whole class. This can then be used to determine the teaching effectiveness of the courseware materials or teaching techniques across groups. Although that effectiveness is dependent on close control of group variables, it can easily provide an all-important check of a student’s, or group’s, knowledge level at the beginning of a course or at pivotal points within the course.

With its use as a self-awareness tool for both student and teacher, a confidence-based assessment is best implemented within a framework of a formative assessment process, alongside coursework (Gardner-Medwin & Curtin, 2003a). The eventual goal is that the individual student completes the course with a higher retention rate and/or is fully prepared for the next course in the series. Adams and Ewen (2009) concluded in their research about improving educational outcomes that, “…confidence and knowledge are correlated and are both critical determinants in evaluating future performance” (p. 2). Hunt (2003) also supports that self-assessment, computer-analyzed testing provides measurements more closely related to a person’s later performance than the common multiple-choice test.

A confidence-based assessment scheme could also be used in a summative assessment role, but does not have as much impact, because the curriculum cannot be modified after the course is already complete. Knowing students’ level of understanding at the end of a course does supply a more realistic assessment of their understanding—confidently justifying a final grade—but unless acted upon at that point by the student, the added information of a confidence-based assessment scheme would be wasted.

A confidence-based assessment scheme is ultimately implemented as a tool to stimulate learning and place emphasis on weak spots, replacing the dread-inducing test of a learner’s mental recall ability and aptitude to test taking. Toward that end, the confidence-based
assessment tool needs to be woven intricately within the framework of a course, whether it be an academic course teaching basic knowledge or within a task-specific training program.

The confidence-based assessment scheme can be applied either as a stand-alone enhancement to an existing course, or as a tool within an all-encompassing effort using more contemporary (web/computer-based) educational tools. Some commercial training consultants have already used confidence-based assessment schemes within an overlying program called Competency-Based Training (Bruno, 1995).

Considerations

Basic human traits must be considered during the implementation of a new confidence-based assessment scheme, as it would most likely cause a bit of confusion and highlight personality traits in those who would try to beat the system. In his study exploring the personality bias of a confidence-weighted scoring scheme, Ahlgren (1969) found that, “Personality has an effect on confidence-marking. Most researchers have adopted weighting systems that give students the greatest promise of reward for honest marking. But betting against the house is a persistent human trait, and there is ample evidence that characteristic levels of confidence and risk-taking play a large part in confidence-marking” (p. 5).

A critical element necessary for success is a clear communication strategy that encompasses a clear understanding of the expected outcome and a detailed plan of each individual process involved. Both learners and administrators need succinct, unambiguous, concise and complete information about what to expect and how to use the eventual outcome in the pursuit of that higher level of topic understanding.
Special attention needs to be given to the reduction, or elimination, of any incorrect beliefs. As Gardner-Medwin and Curtin (2003a) found that the bias towards a wrong answer is a huge obstacle when trying to build higher levels of knowledge, and the more confidence surrounding a learner’s belief, the worse the problem becomes. In a task-specific type training program, a deep-rooted misconception can contribute greatly to an error, slip or mistake, with possible disastrous results.

When creating an implementation plan, presenting the composite score to the learner should occur as soon as possible after completion of the confidence-based assessment, as the realizations of a learner’s misconceptions are more effective when presented while the effect of the evaluation (exam, test or quiz) is still fresh of mind. The greater the time between taking a test and learning the results, the greater the disconnect and lost opportunity to correct any misconceptions. As Gardner-Medwin and Curtin (2003a) expressed, “This [timely assessment feedback] is especially important when high confidence has been expressed for a wrong answer. A mark of -6 stings, even though it should be expected on up to 20% of the occasions when a student takes the risk of entering C=3 [high confidence]” (p. 4).

The use of a confidence-based assessment scheme appears to be very popular with students, especially when offered as a formative tool during a course, as they see it as a method to improve their overall grade. Gvozdenko and Chambers (2007) quantified this when monitoring the use of the certainty scale in formative and summative testing. The study found that, “…the majority of students (67%) on a summative test and nearly all students (96%) on a formative test chose to use the certainty scale as an additional tool in the testing procedure” (p. 215).
Summary

From the existing literature reported in this chapter, 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. Additionally, the confidence-level marking scale is easily understood and accepted by students when it was used in secondary and post-secondary schools.

The goal of an effective self-assessment scheme is to produce students who retain the requisite knowledge long after finishing a course, and 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)

Much of the functionality of a confidence-based assessment platform is already integrated into popular learning management systems. A few well-established academic platforms (Blackboard, Moodle, Schoology, etc.) contain assessment modules. But Moodle (www.Moodle.org) has already implemented a robust confidence-based assessment module, offering a wide variety of development options, and has been used by Gardner-Medwin in both his formal research studies and within an active biomedical student curriculum at University College London.
The researchers throughout this review 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 will attempt to explore those scantily-researched 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 aviation training center by providing a more efficient and comprehensive training experience for a set of pilot and instructor participants.

This Literature Review chapter explored the previous research conducted on the limitations of traditional multiple-choice assessments and a confidence marking scheme to mitigate those limitations. The next chapter, Research Methods, will explain the strategy for inquiry, research facility, participants and specific procedures. The chapter will also describe the specific data gathering techniques, interview coding methods and analysis used in the research study.
CHAPTER 3
RESEARCH METHODS

In light of the limited research conducted on confidence-based assessments within a pure training-type situation, this dissertation research conducted a qualitative inquiry into the feasibility of a confidence-based assessment tool applied to an aircraft training curriculum. The study involved adding a confidence-based assessment element to existing ground school exams and delivering reports to both the student (pilot) and instructors. These were used as formative self-assessment tools for the remainder of the respective courses. Upon course conclusion, the participating pilots were interviewed, using semi-structured questions, to collect qualitative data about their perceptions and opinions of the new learning tool. Additionally, the instructors for those courses were interviewed using the same semi-structured method. The analysis centered on the experiences and perceptions of the participants and it was concluded upon whether further investigations into full implementation would be warranted. The specifics of this exploration are described in greater detail within this chapter.

Strategy for Inquiry

In an effort to discover how a confidence-based assessment tool affects learning in a safety-critical training environment, the design of this study implemented a phenomenological qualitative approach. This particular research design was chosen because of the purely experimental nature of introducing a confidence-based assessment tool into an aviation training program, which based on the literature research, has not been attempted previously.
The use of a phenomenological research approach (Creswell, 2014; Lester, 1999) was chosen because of its use as a discovery method into how a person makes sense of a phenomenon or experience. And since this study is an inquiry into the overlying question about how students and instructors experience a new learning tool, the answers may be best discovered using a method designed to tease out those answers from individual perceptions. Lester (1999) summarized the research strategy by defining its purpose within the viewpoint and perceptions of its participants.

The purpose of the phenomenological approach is to illuminate the specific, to identify phenomena through how they are perceived by the actors in a situation. In the human sphere this normally translates into gathering ‘deep’ information and perceptions through inductive, qualitative methods such as interviews, discussions and participant observation, and representing it from the perspective of the research participant(s). (p. 1)

The specific data collection method used for this research study involved conducting semi-structured interviews of all the participants to gain a deep understanding of their experiences and perceptions. The analysis sought to find commonalities among utterances coded from the interview transcripts. Significant statements were expected to arise that sufficiently explained the participants experiences, which would then offer, or not offer, support to continued implementation of the new assessment technique. Creswell (2014) supports this approach, explaining how phenomenological research describes the personal experience from the words of the participant, and also across the group perspective of participants who have all experienced the same phenomenon.

Research Study Facility

Data collection was performed at a training facility that specializes in corporate, military and commercial pilot training. For the remainder of this report, this training facility will be referred to as the Flight Training Academy. Employing comprehensive curricula, the facility
trains to proficiency using multi-media courseware materials and full-motion simulators. Pilots take courses at the facility to become proficient at operating a particular model airplane, generally an international business executive type (Learjet, Gulfstream, Falcon, etc.) or utility airplane designed for small backcountry airports carrying only a few people or cargo.

Two specialized courses are offered; the “initial” aircraft course (FAA-regulated) introduces the pilot to the aircraft having never flown it before, and the “recurrent” course is a refresher required by the insurance underwriters to be completed each year. For a pilot hired to operate one of these specific aircraft, he or she must first take the “initial” course, then the “recurrent” course each year thereafter while still operating that aircraft. In a typical professional career, a pilot would often be hired to operate four to eight different aircraft types during the course of a lifetime. Therefore, this training facility would be visited many times during the course of a pilot’s career.

For the data collection of this study, two “initial” courses were selected that typified the cross-section of courses offered throughout the entire training facility. The long (3½-week) course consisted of two full weeks of lectures (called ground school) in a multi-media classroom, followed by another week and a half of full-motion simulator sessions. The short (two-week) course consisted of a single week of ground school lectures, followed by a single week of simulator sessions. At the completion of the ground school session, an exam is administered to ensure the pilot had absorbed the required aircraft systems and operational knowledge before proceeding with the simulator sessions. Upon successful completion of the simulator training sessions, a final “checkride” is given to the pilot, which is administered by either an FAA inspector or FAA-designated company evaluator. These checkrides are often very high pressure
situations, because a pilot who fails a checkride is usually terminated by the aircraft operator who just spent upwards of $60,000 to send the pilot through training.

Study Elements

The Flight Training Academy conducts pilot training programs that are approved and monitored by the FAA and foreign aviation authorities. Each curriculum is individually approved and includes a tightly-controlled final exam given at the end of the ground school portion of the course. The FAA requires corporate and commercial pilots to pass these multiple-choice exams with a grade of 80% or greater; grades below 80% require retaking a different version of the exam. After the exam is passed with a grade of 80% or higher, the instructor reviews each missed question with the pilot to eliminate any gaps in knowledge. 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. The inclusion of confidence-based assessment tools into the curricula did not result in changes to the existing FAA-approved exams. None of the original exam questions were modified to prevent endangering any regulatory certification.

Participants

Eleven pilots and five instructors participated as test subjects. Each participant had been previously certificated by the FAA or foreign aviation regulatory authority as a professional in their discipline. The group of participants was entirely male, as expected, since the professional pilot cadre is overwhelmingly male and females in these specific positions are rare.
Each pilot was sent to the Flight Training Academy by his company to gain the knowledge, skills and initial experience to operate the specific aircraft he had been hired to operate. The majority of the pilots had attended a previous course at the Flight Training Academy specific to other aircraft models. This lent credibility to the experimental construct, determining if the confidence-based assessment treatment had changed their study habits.

The management at the Flight Training Academy had no influence over the selection or recruitment of the participants nor did they interfere with any of the data collection procedures. The participants were given the opportunity to refuse to participate, as explained verbally during the initial recruitment and pre-screening presentation and mentioned twice in the Client Information Form and Survey (Appendix A).

Treatment Procedure and Data Collection Instruments

The data collection period was conducted over a two-month period wherein data collection from both the 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 good cross section of the population was achieved.

The aircraft initial ground school portion for the two courses was either one week or two weeks long and generally consisted of in-classroom, lecture-type environments. An exam was typically given at the end of ground school followed by approximately seven simulator sessions (varied depending on specific course) that continued through the remainder of the course (Figure 1). For each of the two-to-four hour simulator sessions, a lengthy briefing was conducted both
before and after the session to discuss the particulars of the flight and any knowledge or operational areas in which the instructor considered the pilot to be weak or deficient. For the study, a confidence-based assessment component was added to the normal exam given at the end of ground school.

Figure 1. Treatment timeline

The research study procedure progressed from pre-treatment recruitment, through administration and grading, to data collection interviews, and finally data reduction and analysis. The following sections describe the details of each step in the process.

Pre-Treatment Recruitment and Screening

Within the Flight Training Academy, 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 all the benefits and risks were explained to each possible instructor, they were asked for their permission to participate. 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.

At the conclusion of the data collection period, each of the instructors was interviewed to gather opinions about the use of the prototype confidence-based assessment tool. Additionally, blank journals were given to each instructor with instructions to record any thoughts or notable experiences throughout the data collection period. Specifically, the instructors were asked to note any changes in their teaching methods due to the implementation of the confidence-based assessment tool into the classroom.

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.
The pilots were left with the Client Information Form and Survey handout (Appendix A), which contained a recruitment letter, exam marking directions, demographic survey and University of North Texas Institutional review Board (IRB) informed consent form. The handout contained a survey that collected information about gender, age, aviation experience, training background and cultural nationality. The pilots were left with the instructions to review the packet, and if they agreed to participate in the study, to complete the demographic survey questions and sign the informed consent form before the exam was administered. During this two-day period, the principal investigator and ground school instructor were available to answer any questions or address any concerns.

_Administration of Confidence-Based Exam_

The pilots normally received knowledge exams at a set point within a specific curriculum schedule, which was at the end of the ground school session. To adhere to the premise that the confidence-based assessment tool was to be examined from a formative standpoint, the pilots took the exam at a point that allowed sufficient time to examine their confidence-weighted scores and modify their study habits before the final simulator checkride, which is the performance assessment needed to complete the course.

The exam administered at the end of ground school consisted of 50 multiple-choice questions, each with four answer choices. This knowledge-based exam had to be passed with a score of 80% or greater before the pilot was allowed to proceed with the simulator portion of the course. Because aviation is a safety-critical industry, it is imperative that any false or misleading knowledge be addressed. Therefore, after a pilot passes the exam with a score of 80% or greater, it is required by regulation for the instructor to review the missed exam questions with the pilot
to ensure any knowledge gaps are closed. This method is called “graded to 100%” and is somewhat effective, but limited in coverage because the instructor is only covering the missed question(s), not the knowledge behind the question(s). Unfortunately, this method still leaves gaps in knowledge, as those questions that were guessed on and scored to be correct are not addressed.

For the purpose of this study, an additional element was added to the ground school exam. Next to each of the multiple-choice questions, the pilot was asked to mark his perceived level of confidence using a three-level scale (Figure 2). This arrangement allowed an immediate selection of confidence level while the pilot was still engaged in the knowledge recall activity for the specific question.

![Confidence Level](image)

**Figure 2.** Sample exam question

For this study, a three-level confidence scale was chosen over a longer scale due to its ease of use and more clearly defined levels. Researchers support this method, as Gvozdenko and Chambers (2007) compared different scales, and concluded, “Comparison of the use of a five grade scale and a three grade scale suggested that the three grade scale is sufficient for most test
takers” (p. 215). Additionally, Gardner-Medwin et al. (2003a, 2003b, 2006) and Gardner-Medwin, Jenner & Roberts (2010) also used the three-level scale effectively in all their studies.

An additional inserted page (Appendix B) was included within the normal directions located in the front matter of each exam and offered a reminder of the confidence-based assessment marking procedure. It was not a requirement to select a confidence level before being able to move on to the next question. If the pilot chose not to indicate the confidence level for a sizable amount of questions, or exhibited any idiosyncratic techniques in completing the assessment, their reasoning was addressed during the post-treatment interview.

Grading of Exam and Report Generation

Upon completion of the exam, the pilot handed the answer sheet to the instructor for secure grading. After the raw scores were graded by the instructor, the principal investigator entered both the answers from the multiple choice questions and the certainty level for each answer into a spreadsheet for calculation of the confidence-based composite scores. The computations produced both a composite-weighted score for the entire exam, and a composite score for each of the individual ground school subjects within the exam using the grading scheme identified by Gardner-Medwin (2003b) as noted in Table 3.

Table 3

<table>
<thead>
<tr>
<th>Degree of Certainty</th>
<th>Low certainty</th>
<th>Medium certainty</th>
<th>High certainty</th>
<th>No reply</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark if Correct</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Penalty if wrong</td>
<td>0</td>
<td>-2</td>
<td>-6</td>
<td>0</td>
</tr>
</tbody>
</table>
The total confidence-weighted score was the summative result of each question response graded either with a confidence-weighted correct response, or a confidence-weighted penalty if the knowledge response was wrong. For the individual subject scores, the confidence-weighted score was reported using the summative result for questions within each subject area as noted in Table 4.

Table 4

Subject Confidence-Weighted Score Reporting

<table>
<thead>
<tr>
<th>Student (Pilot) Mark</th>
<th>Calculated</th>
<th>Report Legend (average weighted score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer</td>
<td>Confidence Mark</td>
<td>Weighted Score</td>
</tr>
<tr>
<td>Correct</td>
<td>3 (High)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2 (Medium)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1 (Low)</td>
<td>1</td>
</tr>
<tr>
<td>Wrong</td>
<td>1 (Low)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2 (Medium)</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>3 (High)</td>
<td>-6</td>
</tr>
</tbody>
</table>

* The 2.4 threshold was used to accommodate multiple questions covering a single aircraft system topic. 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.
From this process, a well-formatted, color-coded report of the graded exam was generated as indicated in Figure 3. Multiple scores were presented on the report. The overall raw score, in percentage, was displayed at the top and used for course certification purposes. A score of 80% or above had the word “PASS” immediately after the raw score. Next, the total confidence-weighted score, in percentage, was reported, as a summation of all the exam answers. A textual representation of that score was also given and color coded to indicate severity. Below the total scores section, the individual subject area scores were reported as a color-coded composite score.
Figure 3. Sample graded exam report

**Distribution of Exam Report**

Within an hour of taking the exam and after the scores were processed, each pilot was given the exam report. The intent of the report 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 so as to assure passing of the simulator checkride at the end of the course. As the research into confidence-based assessments suggests, the opportunity for self-assessment at the formative stage of a course provides an added tool to promote further, and focused learning.

In addition, the course simulator instructors also received the exam report for each pilot who participated in the study. The instructor used the report as he wished, but the intent was to identify the individual pilot knowledge discrepancies and use it as a tool to modify, if necessary, the training in the simulator briefing or flight 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 actually weak, instead those areas that the pilot “appeared” to be weak.

Pilot Interview

During the simulator session period of each course, each pilot was interviewed by the principal investigator using a semi-structured questioning technique. The interview took place in a small room normally used to brief and de-brief the pilots during the simulator sessions. The principal investigator used an interview script and question guide (Appendix C) throughout the data collection process to achieve consistency across the transcribed data. As the interviews progressed, however, additional issues arose that needed further investigation. Therefore, sub-questions were added for latter interviews.
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
- Opinion about benefits to training

These questions form the basis of the post-treatment interview of the pilot. 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 construct categories. If the initial answer in any of the construct categories was not sufficiently answered, 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 following written script, in Appendix C, was read at the beginning of each interview session to ensure consistent instructions were given to each participant.

Thank you for participating in this short interview. We are looking into your experience and opinions about confidence-based assessments and how they could be used during a course. Please be assured that we are not asking for any personal information, and once your responses are transcribed, your answers cannot be traced back to you. We are just interested in your anonymous responses. Please answer the following questions as honestly as you can and elaborate as much as you would like.

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. Once the alias-identified textual files were obtained, the audio recordings and completed Client Information Form and Surveys (Appendix A) were transferred to the University for secure storage.

_Instructor Interview_

Before the data collection period began, the participating instructors were given an Instructor Information Form and Survey packet (Appendix D), which contained a recruitment letter, demographic survey and IRB informed consent form. At the end of the data collection period, each participating instructor was interviewed to gain their opinions and experiences. The interviews were an opportunity to gather any procedural or pedagogical issues that either contributed or detracted from the data collection or program success. The interview took place in a small room normally used to brief and de-brief the pilots during the simulator sessions. An interview script and question guide (Appendix E) was used for each instructor interview. The post-treatment instructor interview instrument was designed to gather data on the following constructs.

- Familiarity with confidence-based assessments
- Ease of use
- Pilot acceptance
- Instructor insight into pilots
- Improved learning
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 instructor was used to identify both the audio recordings and transcribed text files. Once the alias-identified textual files were obtained, the audio recordings and completed Instructor Information Form and Surveys (Appendix D) were transferred to the University for secure storage.

Analysis

The primary objective of this 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.

To insure that the findings of the research were relevant, the analysis phase incorporated credibility strategies, such as triangulation, to validate the qualitative interview data. As shown in Figure 4, all the data collection elements were combined in the final analysis to cross-support views from all the participants.
The data collected during the interview phase of the study consisted of interview transcriptions of the semi-structured interviews. The transcript utterances were organized within the online research application Dedoose.com 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. If there were multiple utterances in a single participant’s transcript text that answered a single specific question, only a single code was applied.

The coding process was a melding of two methods described in Saldaña (2013), evaluation and magnitude coding. 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 far-reaching 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. 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 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).

This chapter identified the specific methods to be used during the data gathering phase of this study. The strategy for inquiry was explained along with the research facility and participants. In addition, the step-by-step data collection phase was explained in detail, including the analysis method. The data collection phase was conducted over a two month period, with the results of the demographic data and interview transcripts presented in the next chapter.
CHAPTER 4
RESEARCH FINDINGS

The previous chapter outlined the methods and instrument this study used to collect qualitative data. The data collection phase consisted of two distinct components: demographic data collected from paper-based surveys, and transcribed interviews. Each of the pilot and instructor participants was given both components and underwent similar experiences. The data collection component that sought to answer the primary research questions was the interview, as the participants’ own words are the instrument of qualitative research. This chapter will first cover the demographic answers to the survey questions, then present the pilot and instructor responses to each interview question, along with additional comments not associated with a particular question.

Demographic Data

The pilots received the Client Information Form and Survey handout (Appendix A), which was required to be completed before the confidence-based assessment exam was administered. The handout contained a survey that collected information about gender, age, aviation experience, training background and cultural nationality. The instructors were also given a similar demographic survey (Appendix D) before the data collection phase began. The demographic data from both surveys was quantified and is presented as follows.
Pilot Participants

The pilot group consisted of eleven participants. Their specific responses are presented in tabular format in Table 5. They were all male, as was expected before the study began, since females in these specific positions are rare. The average age of the pilots was 37, with the youngest having been 30 and the oldest having been 52 years old.

Table 5

Pilot Demographic Data

<table>
<thead>
<tr>
<th>Alias</th>
<th>Gender</th>
<th>Age</th>
<th>Aviation Years</th>
<th>Employed Years</th>
<th>Courses Taken</th>
<th>Flight Hours</th>
<th>Primary Training</th>
<th>Country Raised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roger</td>
<td>Male</td>
<td>33</td>
<td>11</td>
<td>9</td>
<td>0</td>
<td>6200</td>
<td>Academic</td>
<td>Croatia</td>
</tr>
<tr>
<td>Aviator</td>
<td>Male</td>
<td>32</td>
<td>13</td>
<td>7</td>
<td>1</td>
<td>3700</td>
<td>Civilian</td>
<td>India</td>
</tr>
<tr>
<td>Barry</td>
<td>Male</td>
<td>34</td>
<td>16</td>
<td>13</td>
<td>1</td>
<td>4500</td>
<td>Civilian</td>
<td>U.S.</td>
</tr>
<tr>
<td>Dick</td>
<td>Male</td>
<td>47</td>
<td>30</td>
<td>26</td>
<td>5</td>
<td>7000</td>
<td>Military</td>
<td>U.S.</td>
</tr>
<tr>
<td>Chris</td>
<td>Male</td>
<td>40</td>
<td>33</td>
<td>15</td>
<td>1</td>
<td>7500</td>
<td>Academic</td>
<td>U.S.</td>
</tr>
<tr>
<td>Randy</td>
<td>Male</td>
<td>52</td>
<td>30</td>
<td>27</td>
<td>7</td>
<td>10000</td>
<td>Civilian</td>
<td>U.S.</td>
</tr>
<tr>
<td>Jimmy</td>
<td>Male</td>
<td>32</td>
<td>14</td>
<td>9</td>
<td>3</td>
<td>3500</td>
<td>Academic</td>
<td>U.S.</td>
</tr>
<tr>
<td>Cirrus</td>
<td>Male</td>
<td>30</td>
<td>25</td>
<td>10</td>
<td>0</td>
<td>1800</td>
<td>Civilian</td>
<td>U.S.</td>
</tr>
<tr>
<td>Pete</td>
<td>Male</td>
<td>33</td>
<td>18</td>
<td>15</td>
<td>2</td>
<td>5500</td>
<td>Academic</td>
<td>U.S.</td>
</tr>
<tr>
<td>John</td>
<td>Male</td>
<td>32</td>
<td>11</td>
<td>8</td>
<td>0</td>
<td>3200</td>
<td>Academic</td>
<td>U.S.</td>
</tr>
<tr>
<td>Homer</td>
<td>Male</td>
<td>42</td>
<td>20</td>
<td>15</td>
<td>2</td>
<td>6800</td>
<td>Civilian</td>
<td>U.S.</td>
</tr>
<tr>
<td>Average:</td>
<td></td>
<td>37</td>
<td>20</td>
<td>14</td>
<td>2</td>
<td>5427</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The pilots were asked in the survey, “How many years have you been employed in the aviation industry?” (indicated in table as Employed Years). The average response was 14, with the shortest duration having been seven (7) years and the longest having been 26 years. The intent of gathering this information was to establish the overall level of experience in the participants’ professional aviation career and to possibly discover an association with the coded interview data. A similar question was asked about their involvement in aviation throughout their
lifetime; “How many total years have you been an active participant in aviation (including building models as a child)? In other words, how long have you had the ‘aviation bug’?” (indicated in table as Aviation Years). The average was slightly over 20 years, with the least having been 11 years and the longest having been 33 years.

The intent of gathering this experience duration was to establish the amount of years dedicated to the craft and to identify any “lifers” whose interview responses may have more credibility. There were two pilots who were considered “lifers,” reporting that they were five (Cirrus) and seven (Chris) years old when they first became involved in aviation. The rest of the pilot participants began their aviation involvement during their late teen years. The amount of years that the pilot participants had dedicated themselves to aviation was on average six (6) years longer than the length of their professional careers. This number concurs with the typical amount of years necessary to complete flight training before gaining employment as a professional pilot.

The next two survey questions lend further credibility to the interview responses by gaining an insight on the participants’ amount of professional type-rating courses and total hours operating an aircraft. The pilots were asked, “How many full courses have you taken at [this Flight Training Academy]?” The average response was two (2), with the highest having been seven (7) while three participants reported that this was their first course at the academy.

Naturally, the pilot who reported that this was his seventh course had the highest amount of time employed in the aviation industry (27 years), and was also the oldest.

The use of “flight hours” in the aviation industry is a method used to quickly establish a benchmark number representing pilot experience. Generally, each hour a pilot is at a “piloting station,” conducting flight operations, is considered an hour worthy of recording in the pilot’s personal logbook. Total flight hours are generally stated as minimum requirements of job
postings, and are used extensively by both the regulatory authorities (FAA, etc.) and insurance underwriters to establish minimum levels of experience. Therefore, gathering total flight time in this study was an important metric to establish credibility of the participants’ interview responses. When asked in the survey, “Approximately how many total flight hours do you have?” The average was slightly over 5,400 hours, with the lowest time pilot reporting 1,800, and the highest time having reported 10,000+ hours, once again from the oldest pilot in the group. For comparison sake, career airline pilots can easily exceed 25,000 total flight hours during a lifetime.

The last two survey questions documented each respondent’s nationality and method of primary training. They were asked, “In what country, or countries, did you grow up in?” All of the pilots reported having grown up in the U.S., except two, who grew up in India and Croatia/England respectively. In a related question, the survey asked, “Where did you receive your primary aviation training?” The radio buttons allowed one of four exclusive responses: Military, Civilian (club, flight school), Academic (university) or Other. The overall result was nearly split between two responses. Five (5) pilots had trained through a university-type academy, five (5) had trained through civilian means (small flight schools generally), and the remaining pilot had trained in the military. The method by which each participant received his primary flight training is important to this study, as the approach and level of thoroughness experienced during his initial flight training typically dictates his learning method preference throughout his career. Teaching philosophies differ greatly between, for example, a highly-structured military academy versus a more informal civilian school.
Instructor Participants

The instructor group consisted of five participants, each a highly-qualified pilot himself. Their specific responses are presented in tabular format in Table 6. They were all male, as was expected, as there are very few female commercial aviation instructors. The average age of the instructors was 42, with the youngest having been 28 and the two oldest having been 47 years old.

Table 6

Instructor Demographic Data

<table>
<thead>
<tr>
<th>Alias</th>
<th>Primary Training</th>
<th>Age</th>
<th>Instructor Years</th>
<th>Flight Hours</th>
<th>Aviation Years</th>
<th>Improvement Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sawyer</td>
<td>Civilian</td>
<td>47</td>
<td>5</td>
<td>5800</td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>Edrod</td>
<td>Academic</td>
<td>42</td>
<td>4</td>
<td>3100</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Bob</td>
<td>Civilian</td>
<td>28</td>
<td>3</td>
<td>4000</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Nick</td>
<td>Civilian</td>
<td>46</td>
<td>4</td>
<td>3700</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td>Tommy</td>
<td>Academic</td>
<td>47</td>
<td>6</td>
<td>10000</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Average:</td>
<td></td>
<td>42</td>
<td>4.4</td>
<td>5320</td>
<td>22.4</td>
<td>12.6</td>
</tr>
</tbody>
</table>

The first survey question documented each instructor’s method of primary training. They were asked the identical question that had been asked of the pilots, “Where did you receive your primary aviation training?” The radio buttons allowed one of four exclusive responses: Military, Civilian (club, flight school), Academic (university) or Other. Two (2) of the instructors disclosed they had trained through a university-type academy, while the remaining three (3) instructors had trained through civilian means. As with the pilots in the study, the method by which each instructor received their primary flight training is important to this study.

The next question asked, “Approximately how many total flight hours do you have?” The responses averaged slightly over 5,300 hours (only 100 hours less than the pilots, on average),
with the least experienced instructor reporting 3,100, and the most experienced instructor having reported 10,000+ hours. The flight hour average and high-end amounts for the instructors were nearly identical to the pilots. The pilots were then asked, “How many years have you been employed as a flight instructor?” (indicated in table as Instructor Years). The average among the responses was 4.4 years, with the shortest duration having been three (3) years and the longest having been six (6) years.

It is noteworthy that the average amount of years the instructors were active participants in aviation exceeded those of the participant pilots. When the instructors were asked, “How many total years have you been an active participant in aviation? (including building models as a child) In other words, how long have you had the ‘aviation bug’?” (indicated in table as Aviation Years). The average of the closely-spaced responses was slightly over 22 years, with the least having been 11 years and the longest having been 26 years. There were not any “lifers” (i.e., interest began in early childhood) among the instructors.

The last survey question was designed to gather the amount of training each instructor had received specific to improving their skill as a flight instructor. The instructors were asked, “How many courses have you taken to improve your craft as an instructor?” (indicated in table as Improvement Courses). The average of the responses was slightly greater than 12, with the highest two responses having been 20, and the least having been six (6). Knowing the amount of improvement courses taken is telling of an instructor’s dedication to his profession and lent additional credibility to his interview responses.
Interview Responses

Eleven (11) pilots and five (5) instructors were interviewed with the sessions recorded for transcription. Approximately 97 minutes of interviews were recorded. The 16 interviews took an average of six minutes to conduct, with the shortest at 2 minutes 35 seconds and the longest at 8 minutes 34 seconds. The pilot interview question script is reproduced in Appendix C, and the instructor interview question script is reproduced in Appendix E.

Familiarity

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. Most of the responses were simple and to the point, as this first question evoked a somewhat binary response. The following is a typical response to the question from those who did not have any prior experience with confidence-based assessments.

Chris (pilot): I never had experience with a confidence-based assessment.

Homer, Roger, Aviator and John responded similarly, with short comments such as “none” or “zero,” whereas the rest of the participants who were coded as being somewhat familiar with the tool had the following typical response.

Randy (pilot): Not a whole lot…. I mean, I heard about them, but not really on any other tests that I’ve been through here at [Flight Training Academy] for the past 20 years or anywhere else.
Barry (pilot): College level classes and mostly surveys. Just surveys in college-level classes, but nothing in regards to actual tests.

These responses, along with those from Pete, Cirrus, Dick and Jimmy indicated a theme that they had at least heard of confidence-based assessments but just had not used them in an official self-awareness capacity. Four of the instructors (Tommy, Nick, Bob and Edrod) also responded with short answers, such as “none” or “zero,” whereas Sawyer answered that he was very familiar with confidence-based assessments and responded with the following exchange.

Researcher: What prior experience do you have in confidence-based assessments?

Sawyer: I was a police officer in Little Rock and I had opportunities to use similar confidence-based testing during our rookie school, and it was utilized on me, not as me being the instructor, but I was a student… This was in the early 90s… At the time, they didn’t ask us opinions like you are asking the students now. They just put the results down and we never heard anything about it, but it was similar to that, but it was a 1, 2, 3, and 4 score instead of a high, medium, or low.

Researcher: Did they give you the report afterwards?

Sawyer: No. This was in the late 80s, early 90s.

Researcher: So it was really just for their use.

Sawyer: Correct, for their use, not for ours.

This exchange from Sawyer’s interview illustrates the use of a confidence-based assessment tool to help the instructors or administration. However, in this case it was clearly not used to further the student’s self-awareness of their weak areas or used as a tool for further study. It is also unknown whether the assessments in this situation were used in a formative or summative context. This response from Sawyer offered a noteworthy insight into the use of confidence-based assessments, at least from a comparison to previous experience standpoint.
Exam Procedure Acceptance

Another important factor during the study was to confirm if the exam instructions were adequate enough so as to not hinder the actual assessment process. To confirm that the procedures were clearly understood and that any misunderstandings would not have contributed to any faulty data, the pilots were asked the question, “How easy or hard was it to understand the directions about how to mark your exam confidence levels?” A strong theme emerged, as all the pilots responded with simple answers, such as “yeah” or “it was easy,” reflecting that they clearly understood the directions and that there was not any confusion about how to mark their confidence levels. This confirmed that the procedure of explaining the exam directions three times (two written and once verbally) successfully prepared the pilots prior to the exam.

The instructors were asked a similar question, although focused on their understanding of the specific classroom, or briefing room, procedures as told to them by the principal investigator. During the interviews, they were asked, “How easy or difficult did you find the confidence-based assessment classroom [or briefing room] procedures?” A theme similar to the pilots emerged, as all the instructors responded with comments such as “it was simple” or “it was pretty easy,” confirming that they clearly understood the procedures and there was not any confusion.

Seeking further insight into the pilots’ use of the procedures, especially their attitudes toward those procedures, the pilots were asked, “Were you reluctant about revealing your confidence level for each of the exam questions?” All but one (Randy) of the pilots responded with short comments such as “no” or “not at all,” with some expanding only briefly on their thoughts. These common responses indicated that they did not experience any reluctance when marking their confidence to each exam question. The lone individual diverged only slightly from the rest, reporting that he was somewhat reluctant to reveal his confidence.
Randy (pilot): When I was taking the test I probably put some things down on the right that probably should have been more confident, you know, where I was kind of going through the question and I just put in a line question putting the answer where, you know, 1, 2, 3. So I probably should have paid more attention to that, in my opinion.

Randy appeared to understand the concept of marking his confidence for each question, but just mistakenly marked a higher confidence on a few questions that upon reflection should have been marked lower. This issue was most likely due to inexperience with this type of exam, even though he reported on a previous question that he was somewhat familiar with the confidence-based assessment concept. Randy’s situation may have also been compounded by a little anxiety, as he was taking an exam that was crucial to his piloting career.

The instructors were also asked whether they detected any reluctance by the pilots during the course of pre-screening or the administration of the exam. The instructors were asked to reflect on the question, “Overall, how did you find the clients accepted the addition of confidence-based assessments to their normal routine?” A common theme did not emerge as the observations were split, with two instructors (Nick and Bob) reporting their pilots’ exhibited very little or no reluctance, while the remaining three instructors (Tommy, Edrod and Sawyer) reported that their pilots were somewhat reluctant.

Nick (instructor): I just had one client in class, so it may be a little skewed there, but he was more than willing to do it. [researcher: “Did he have a lot of questions about it or not, that it seemed easy to carry it out?”] He seemed to understand it very well, yes.

Bob (instructor): I thought most of them were very receptive to it and to use it to kind of fine-tune their study habits to their different weak areas and whatnot because otherwise they don’t know if they are targeting their time to the right material.

Nick and Bob reported very little reluctance from their pilots, with Bob even reporting that he observed his pilots using the exam report to its full intended purpose. The remaining three instructors though observed some reluctance from the pilots.
Tommy (instructor): …they didn’t seem to care for it much, just because the one guy, he had yellow in here [referring to a line in the graded exam report] and he was very competitive and he was like, ‘Why did I have yellow if I got it right? I only missed one.’ So I told him, ‘Well, my understanding is that you didn’t have a high confidence on it.’ And he goes, ‘But I got it right.’ And I said, ‘Well, that’s how the test is. It’s to build confidence, to show the areas you might need some more instruction.’ So I don’t know. They were hit and miss on it.

Edrod (instructor): I kind of felt like most of them didn’t completely get the ideas behind it and there was a lot of mumbling about different perceptions of it or what have you. I didn’t feel like I got a very positive impression from most of them, until you explained to them what you were trying to accomplish and then talked about it a little bit more, and then it seemed like the light started to come on and they were kind of okay with it and they got the principle behind it, but really you could tell their main concern was how good or bad was this going to look in my file?

Sawyer (instructor): Right. Um, for our particular purposes, or our particular class, and the level at which it is at, I’m not going to say it’s overwhelming, but it was an added pressure that I think that they felt on average.

These three instructors reported some reluctance on the part of the pilots, with Tommy and Edrod observing some confusion about how to use the graded exam report while Sawyer revealing he thought the confidence-based portion of the exam may have been just a little too much for his pilots to shoulder in addition to the comprehensive exam. All three of these observations relate to a common issue, that the pilots did not fully understand the concepts of confidence-based assessments and how it could help them improve their course experience.

During the interview process an additional question was asked of the pilots to learn if they thought the actual act of marking their confidence for each question was a procedural burden, and how much time was spent marking each question. This was not a pre-planned question, but emerged as a natural progression during the discussions. When asked, “How long did you hesitate when marking your confidence level?” or “Do you feel the added procedure of marking your confidence level for each answer extended the exam to the point that you became fatigued at all?” The majority of the pilots (8) responded with a common theme, that the time and
effort to mark a confidence level was inconsequential. Responses were heard such as “not much
time at all,” “a couple seconds,” or “I breezed through it.” The remaining three pilots responded
that it varied and was dependent on the particular question. Chris responded the most articulately
about his experience. An abridged excerpt of his long response follows.

Chris (pilot): A question that was fact based, that gave me 3 or 4 short answers to choose
from, was very short [time to answer]. I know the answer and I am confident, and
obviously the ones that I had to delve into a little bit more, questions and answers
that you had to read into and make sure there weren’t “gotchas” or things like that
in there, obviously that was where I would lower my grade [confidence level] a
little bit, only because there were so many complications in the question…. I was
definitely more confident with something that was not a quick response than I was
with something that was a little bit more wordy. So I was probably more likely to
put a 2 [confidence level] down on something, a nice wordy answer, than I was on
something that was, you know, a choice between 100, 90, 80, 70.

This response from Chris indicated that he felt his confidence marking strategy depended
on the specific situation and there was not a simple response to the question. The other two
pilots, Dick and Pete, provided shorter answers, focused on their experience that some questions
took longer to answer than others. These three pilots may have represented the thoughts of the
rest of the pilots, but the responses from the remaining eight pilots does not offer enough insight
into any possible associations. Although, Chris may have just been more articulate with reporting
his cognitive thought processes. As is often the case with overarching interview questions, the
pilots may have been thinking, “well, it depends,” which reduces the amount of detail that could
be collected about their thoughts.

Score Expectations

The interviews addressed the use of the graded exam report to learn how they were used
by each pilot. The pilots were asked, “After receiving the report of your confidence level and
grade for each of the knowledge sections, did the scores surprise you or were they about what
you expected?” A common theme did not emerge, as the comments were split, with five (5) responding that the report scores were what they expected, with comments such as “it didn’t surprise me,” or “it was what I expected,” while the remaining six (6) pilots suggested that the report surprised them at least a little. Of the pilots who responded that the report was not what they expected, Randy and John typify the responses.

Randy (pilot): No. I think some of it needs improvement. I probably put maybe some of the wrong responses down.

John (pilot): Well, I really thought it was a sea of green, but yeah, I felt okay about it. I didn’t expect to come out of here knowing everything. If anybody does, I think that’s kind of pretty cocky. [interviewer: “…these areas here, the Ice & Rain, did you feel weak in these areas before you went into the test at all?”] Um, maybe the Ice Protection and Rain Protection a little, but the Air Conditioning kind of surprised me. It seems like a fairly easy system to understand. Why I missed that question I don’t recall.

The responses were similar among the pilots who were surprised with the results; mostly focused on the awareness of their own weaknesses, instead of the confidence-based assessment procedures. This was the expected result of the new tool, to provide a self-awareness of a pilot’s weak areas. Although, Roger mentioned that he marked a few of the questions incorrectly by mistake, and this may be in response to hindsight, rather than actually remembering the question. In his defense, he blamed the exam procedure for his mistake, saying “it could be a little unreliable,” while at the same time praising the confidence-based assessment tool as “…great and I enjoyed getting the information.”

Study Habit Change

Questions to both the pilots and instructors sought to gather experiences or opinions about the use of the graded exam report (see Figure 3). The report presented a complete breakdown of the exam results, including: the raw exam score, the total confidence-weighted
score and the individual color-coded subject area scores. The pilots were asked, “Did knowing your combined confidence level from the report for each of the knowledge sections provide useful information for your remaining study time before the simulator briefing sessions?” A clear theme emerged, as most of the pilots responded that the graded exam report gave them a better insight into their knowledge level for the various aircraft system subjects. The responses from Cirrus, John and Roger typify the responses seen across all the pilots.

Cirrus (pilot): Yes, and these were the areas that I had thought previously that I may be weak on, but yeah, it gave me a better idea of areas that I can focus on before the simulator and going forward.

John (pilot): I grabbed the yellow and the red areas and then dug into those a little bit better. Just so when it came time to do the simulator sessions I’d have a more clear understanding to make sure that I was more successful.

Roger (pilot): Yeah. It definitely helped, I mean, seeing [individual subject] things that I needed to improve on definitely helped.

The results were overwhelmingly positive, with nine (9) pilots reporting that their study habits did change, and for the better. 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. In contrast, two pilots reported during the interview that the new tool did not dramatically change their habits. Jimmy reported that his study habits changed only somewhat for the better, although getting only a single question wrong on the exam, his study habits were most likely already very effective. The remaining pilot (Homer) 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.
The instructors were also asked a similar question to discover if they felt a greater awareness of the pilots’ strong and weak areas after receiving the graded exam report for each individual. There were only two instructors in the sample who conducted simulator session training with the pilots. Of the five participating instructors, one taught both ground school and in the simulator, one taught only in the simulator, and the remaining three taught only ground school. The simulator instructors were asked, “Overall, did you gain a greater insight into the level of aircraft system knowledge of your clients after receiving the grade report?” Both instructors disclosed that the graded exam report was very helpful, giving them an insight into their pilot’s weak areas, so as to focus their simulator scenarios to address those weaknesses. The following two excerpts summarize their responses.

Tommy: In the briefing, I might ask or go into, for instance, Ice and Rain Protection, because of the red…. I went through it and covered some more things in it… without their knowledge.

Bob (instructor): I like it from an instructor’s standpoint in the simulator and being able to see the report, because it gives me a lot more ideas in how to manipulate the malfunctions in the simulator to try and focus in on the weak areas, because you tend to learn more from an airplane, especially when it comes to things like electrical and powerplant when the airplane is broken than when it is operating properly. So we can kind of target that knowledge a little bit better by introducing different malfunctions.

Training Benefits

At the end of each interview, a discussion was conducted to gather the overall attitudes toward the implementation of confidence-based assessments. The pilots and instructors were asked slightly different questions, as their specific use of the exam tool resulted in unique experiences and would offer differing perspectives from each group’s point of view. The responses, however, were viewed as a whole, since the questions were seeking a somewhat binary response—that the confidence-based assessment tool was either beneficial or not in the
particular situation. The pilots were asked, “How beneficial do you think confidence-based assessments are during a [Flight Training Academy] course?” The instructors were asked a similar question, “After having now worked with confidence-based assessments, how comfortable do you feel with their use as a tool for learning improvement?” Each of the responses from both the pilots and instructors indicated that the confidence-based assessment tool was a positive addition to their normal course exam, as they provided an abundant amount of comments to explain their experiences. The following responses were typical of both the pilots and instructors.

John (pilot): I think they are very beneficial. I felt like it was pretty accurate in determining my knowledge level and allowed me to correct the areas that were deficient or, like it says in here, needs improvement.

Bob (instructor): I think it was a positive experience all the way around, both from the classroom and the simulator instructor’s perspective in trying to gauge, not just how well the clients are, not necessarily how well they are studying, but how well they are receiving the knowledge, so it can lead you to modify your techniques as well and the way you deliver instruction. So it’s not just a positive for them, but it is a positive for us as well to go throughout and kind of have some own self-assessment as to how we can better deliver instruction to the clients.

A clear trend emerged, as the results were overwhelmingly positive by a large margin. During the pilot interviews, there were 21 individual comments extolling the confidence-based assessment tool as very beneficial, with just a single comment stating the tool was only somewhat beneficial. The instructors were equally encouraging; with all five stating they felt very comfortable with the use of confidence-based assessments and that the tool offered many benefits. There was not a single instance of an outwardly-negative response. Additionally, both the pilots and instructors used the process as it was intended, as a technology-based enhancement tool to improve the learning efficiency of the pilots and evaluation ability of the instructor.
Marking Strategies

Due to the nature of the semi-structured interviews, many comments were recorded from the pilots that were not in response to a specific interview question. A notable sub-category of these revealed the pilots’ efforts to interpret the exam directions and generate their own confidence marking methods. The comments ranged from short passages that revealed technique, to long explanations about their test-taking philosophies. There were multiple occurrences of these types of excerpts throughout the transcripts, which were grouped separately and referred to as marking strategies for the purpose of discussion. Two categories of strategies were observed, the Boolean decision and the holistic approach. John offered a good explanation of the use of a Boolean method.

John (pilot): After I read the questions, if I had the answer in my head, that’s where I felt a 3 was necessary. If I had to dive into the question, because some questions you have to reread the choices in order to get the answer…. But if I had to read the answers of the choices that were given to me to derive an answer, that then was the one that I put a 2 down.

John and Chris used a marking strategy that was logical and clear in its execution. This technique of marking: answer-confidence, answer-confidence is a common technique that is easy to understand and was the intended technique expected from the participants when the research study was designed. In contrast, Roger, Randy and Jimmy used a holistic technique where they only marked the “non-high (3)” confidence scale concurrently with selecting the knowledge component answer, then returning after completing the exam to mark all the “3s”. Jimmy typifies the response from these three pilots who used this method.

Jimmy (pilot): As I was going through it, I knew the answers as I went through every single one. There were literally 3 questions, I think, that I took a matter of 5 seconds to be like “Let me read that again,” and those are the ones that I marked 2s. Everything else I left completely blank, so once I’d answered every question, I just went back and put a 3 on every other one. Like I answered every question,
and honestly, I was very confident about every single one of those answers, except for the ones that I marked 2s.

This technique was not expected, but since three pilots offered their unsolicited responses about using this holistic technique, it was worth investigating. A more detailed analysis of this finding is explored in the next discussion chapter.

After the first few pilot reports were generated, it was noticed that a couple of the pilots responded with mostly “3”s as their confidence mark, and it was suspected that those individuals may have been trying to “beat” the test. This phenomenon was cited as a possible limitation in a previous study (Ahlgren, 1969) that explored personality bias of a confidence-weighted scoring scheme, “…personality has an effect on confidence-marking…. betting against the house is a persistent human trait, and there is ample evidence that characteristic levels of confidence and risk-taking play a large part in confidence-marking.” The same complication was also suspected in the current study. Therefore, an additional question was added to the interview script.

To determine if this “beat the test” phenomenon was happening, an additional metric was calculated from the confidence scores for all the pilots, called Confidence Sum. This single metric was the sum of all the confidence marks for the entire exam. With the 50-question exam, if the pilot marked a “3” for the confidence level on each question, the maximum Confidence Sum score could be 150. If the pilot scored over 120 for the Confidence Sum, it was suspected that he may have been trying to “beat the test,” thinking a high confidence mark would result in a higher score. All but one of the pilots in the study scored above 120, but only four pilots were left to be interviewed when an additional question about high confidence was asked. Those last four pilots were asked, “I noticed that your confidence marks showed a tendency to be mostly 3s, or high confidence. Thinking back to when you were taking the test, please reflect on your thoughts as to your use of the three confidence levels.” All four of the pilots responded that they
did not dwell too much on their marking philosophy, but Homer seemed to encapsulate the thoughts of the four pilots.

Homer (pilot): I think I had a pretty good knowledge of the question that was asked. So to me that was a 3. There wasn’t much guessing or decision.

Homer’s response and a more detailed analysis of the “beat the test” phenomenon from the other three pilots are analyzed in greater detail in the discussion chapter.

Exam Schedule Placement

As stated earlier in the literature review, the ground school portion for the two courses was either one week or two weeks long, culminating in a multiple-choice exam that the pilots had to pass before moving on to the simulator sessions. The confidence-based assessment tool was added to this exam. During the data collection phase of the study, some comments from the ground school instructors suggested that the stakes for passing the exam were very high, and the pilots were singularly-focused on just passing the exam, rather than interested in learning of their weak areas through a confidence-based assessment report. In light of this, an additional question was added to the remaining interviews. Both the pilots and instructors were asked, “During this experiment, the confidence-based assessment exam was given at the end of ground school. Is this placement in the schedule effective, or at what other point would you like to see an exam or quiz administered and the report given to the client?” An overarching trend was not discovered, as the comments were split, with five (5) responding that the current placement (end of ground school) was effective, while the remaining six (6) participants suggested that a self-assessment type of report would be more effective if placed earlier in the curriculum schedule. Bob embodied the typical response from those who thought the confidence-based assessment was placed correctly at the end of the ground school.
Bob (instructor): I think during the [final ground school] exam is good for them to gauge or respond to how comfortable they are with the material. I mean, it’s a good time for the report as well because they can focus on those items before they head into the simulator.

Of those who believed likewise, their responses were short, such as “I think it’s in a good spot.” It appears these five participants feel that having a confidence assessment report was correctly used as a formative self-awareness tool, but for the remaining simulator sessions, not as a tool to improve the ground school sessions. The opposing six participants supported the use of a confidence-based assessment tool earlier in the course schedule, during the ground school sessions. Dick expressed the typical view of those participants.

Dick (pilot): Or maybe even like smaller tests throughout the week, after some systems are covered. So maybe if you had flaps and hydraulics on day one, well maybe on day two or three you have a flaps and hydraulics test, and maybe that’s a little more focused or concentric on that area, and then you could keep doing that throughout the week and maybe you have the bigger database and a more centralized focus on what you’re testing.

Dick focused on the use of smaller tests (quizzes) throughout the week(s) of ground school. Along with similar viewpoints from the other participants in that camp, Dick felt the confidence-based assessment tool would be better used in smaller chunks so as to solidify knowledge through assessments while it was still fresh of mind. The group felt that smaller, reflective-type exercises throughout the week would aid in the reduction of the “drinking from a fire hose” effect.

The interviews progressed smoothly and all the participants were forthcoming with their experiences or opinions. The participants’ own words embodied their thoughts about how the confidence-based assessment tool was accepted and applied to their training situation. This chapter first covered the demographic answers to the survey questions, then presented the pilot and instructor responses to each interview question, along with additional comments not
associated with a particular question. A cursory review of the interview statements discovered there were many instances of positive responses throughout, but a more detailed analysis is necessary to find any overlying trends. The following chapter will offer a discussion of the interview responses, coded and analyzed for any trends or recommendations that may need to be considered during the implementation of a confidence-based assessment program.
CHAPTER 5
DISCUSSION

The proceeding chapter reviewed the transcribed interviews from eleven pilots and five instructors during the data collection phase. Each offered succinct insight about his experience with a confidence-based exam given at the end of a ground school course. The transcriptions provide clear evidence that the following primary research objectives were achieved, while positively supporting the implementation of confidence-based assessments within an aviation training environment.

Overarching objectives —

- 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 simulator briefings to address those weaknesses.

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

Specific questions were asked of the participants to achieve these objectives. The extent to which answers were identified depended on the expressive self-awareness and articulation of the research participants. Participant responses to interview questions were used to answer the following questions posed by the research study.

Primary Research Question
How does the addition of a confidence-based assessment element to a mature exam change the habits of both students and instructors within an aviation training environment?
Sub-Questions

- To what extent does the addition of confidence-based assessments change student habits?
- To what extent does the addition of confidence-based assessments change instructor habits?
- What is the perception by both students and instructors that confidence-based assessments may be a useful tool to enhance learning efficiency in an aviation training environment?
- What are any limitations of the implementation of confidence-based assessments into an aviation training environment?

The study procedures were specifically designed to explore the particular methods as an analogue to an electronically-delivered exam. This discussion section is organized in the following order to reflect the sequential procedure of the interview data collection phase. This section finishes with the conclusions, implications, recommendations, limitations and ideas for future research.

1. Credibility
2. Acceptance
3. Accessibility
4. Marking Strategies
5. Score Expectations
6. Study Habit Change
7. Training Benefits
Credibility

The primary reason to gather demographic data was to lend credibility to the interview responses, which in this case were collected from highly-professional, experienced participants. The pilots averaged 20 years of aviation experience (5400 average flight hours), and the instructors averaged 22 years of experience (5300 average flight hours). In the aviation industry, 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 (FAA, 2015a). 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.

Resulting demographic data was compared in terms of age, aviation experience, training background and cultural nationality against the various interview responses. Each one of these indicators, however, turned out to be relatively flat across the ranges. Except for nationality (all but two from the U.S.), the responses were diverse across the scales. This result was expected, as the pool of customers that the Flight Training Academy generally attracts are from a large and diverse spectrum of the aviation population. In addition, the sample size was too small and a cursory comparison of the demographic data against the interview responses did not show any trends. This is acceptable, as this study was not intended to be a mixed methods effort, where interview responses are verified, or refuted, against a statistical analysis of survey data.
Prolonged engagement is another measure of credibility. This was achieved through a protracted relationship and understanding of the cultural experiences of the participants. In the case of this study, the principal investigator has over 35 years of experience within several areas of aviation (pilot, technician, engineer, researcher) and more relevant to this study, over 20 years specifically within the aviation training arena as an instructional designer and flight/maintenance instructor. Therefore, prolonged engagement was achieved by an extensive association with the study participants. This was highly beneficial during the interviews, as the principal investigator could gain the trust and catch the nuances or specific lingo of the participants, enabling deeper probing of issues that required additional questions or discussion.

A point that should be addressed is if the investigation was thorough enough during the interviews with the use of pre-written questions. A critic might argue that since the interviews were short (six minutes average), they may not have uncovered all the pertinent issues. But at the end of each interview, the principal investigator specifically asked if the interviewee had any additional thoughts or observations. Each time, except one, the interviewee admitted—after a thoughtful pause, not just a reaction—that he did not have any additional comments. One pilot (John) even said, “No. I think you’ve covered most of my thoughts. You asked some pretty pointed questions.” So it can be clearly reasoned that saturation was achieved from the participants and there was not anything additional that could have been discovered.

Acceptance

The topic of acceptance involves the way the participants agree to, and embrace, the use of confidence-based assessments as an interruption to their normal processes while offering
worthwhile benefits. Determining the extent of participant acceptance was a primary objective of this study.

The use of a confidence-based assessment tool is definitely an interruption to a normal routine. As the literature review highlighted, there has not been much incorporation of confidence-based assessments into the mainstream culture of education. The interviews also bear this out, with seven of the participants having no previous knowledge of these assessments and eight admitting just a passing familiarity with them. The single participant, Sawyer, who had actual experience in completing a confidence-based assessment did not benefit because of the assessment design, which did not offer any self-awareness of his weak areas.

Gathering awareness about the baseline experience of the study participants is important, since prior experience influences perceptions and attitudes. For instance, the U.S. military has been using advanced computer-based training and electronic delivery of assessments since the 1970s, which may influence a participant’s attitude, either good or bad, depending on the experience. Whereas another participant who learned to fly at a very rural flight school would most likely not have experienced those advanced learning tools. In this study, the only military-trained pilot, Dick, responded, “I think maybe there was something that we might have used there, maybe some multiple-choice things that might have identified some weak areas, but nothing that they used in particular to structure your training or anything like that.” Therefore, it appears that in this instance, the pilot was not negatively, or positively, swayed by any previous military training experiences.

The pilots and instructors who were not familiar with confidence-based assessments responded with simple comments such as “none” or “zero,” whereas the rest of the participants who had heard of the tool were at least somewhat familiar to varying degrees. It seems that
integrating confidence levels into learning is danced around a little to varying degrees, if even sometimes just contemplating their confidence in their head about certain topics. This makes sense, as the learning process naturally involves reflection and the conquering of self-doubt.

From a protagonist point of view, general awareness is always beneficial, for the introduction of something new always entails some reluctance, and the more pilots, maintainers and instructors are aware of the available learning tools, the easier the path would be to implementation. Although from the antagonist point of view, if that awareness is from a bad experience, the reluctance may be insurmountable if trying to implement a new tool into a high-stakes training situation. Gardner-Medwin et al. (2010) also discovered the conflicting viewpoints, “In evaluation surveys, a majority of students have always said they like CBM [Confidence-Based Marking], finding it useful and fair. They asked to include it in exams, and after five years of exam use at UCL they voted 52%:30% to retain it (in 2005/6), though this was rejected by the conservative medical establishment” (slide 14).

The instructors were also asked about their experiences after the confidence-based assessment exam was administered. Bob’s class appeared to accept the confidence-based assessment tool outright and saw the benefits it could provide, while the three other instructors noticed a little bit of hesitation on the part of the pilots. Tommy reflected on his experience with one pilot who seemed a little confused about how the new exam tool was supposed to work. Granted, confidence-based assessments may be something new, but after it had been explained three times, it would appear this pilot may have not been paying enough attention.

Edrod’s comments reflect a somewhat similar experience, where it took a little time and multiple explanations before his pilots fully understood the concepts. But once they did, they seemed to embrace the new tool, even though their primary concern was how their performance
on the exam looked to the next (simulator) instructor. No doubt, Edrod’s pilots were concerned about any extra work they might need to do if the report looked bad.

Sawyer’s experience was somewhat different. Unlike all the other instructors who taught one-week ground school classes, he taught the long ground school class, which consisted of a very intensive two weeks of non-stop aircraft systems classroom lectures. The pilots’ fatigue level approaching the exam was very high, and Sawyer reflected that they already felt overwhelmed, even before hearing about the experiment that modified their exam.

In addition to Sawyer’s general comments, an interesting incident during one of his classes is worthy of examination. It involved a class that contained nine (9) professional pilots; quite a coup for a qualitative data collection effort. As already mentioned in the research findings section, two days before administering the confidence-based assessment-enhanced exam, the principal investigator addressed the class and explained how confidence-based assessments work and the nature of the study. The nine pilots were left with the Client Information Form and Survey handout (Appendix A), which contained a recruitment letter, exam marking directions, demographic survey and an IRB informed consent form. The pilots were left with the instructions to complete the form and sign the informed consent statement. During this two-day period, the principal investigator and ground school instructor, Sawyer, were available to answer any questions or address any concerns.

A couple days later, the instructor informed the principal investigator that all of the pilots had refused to take the confidence-based portion of the ground school exam. Upon further investigation, the instructor revealed that during a couple class discussions, the pilots were willing to help the Flight Training Academy test a new assessment technique and saw its advantages, but as a group, balked abruptly when reading the IRB informed consent statement,
realizing the research was part of a thesis or dissertation. It seems unlikely that all nine pilots felt the same way about the situation, so it is assumed there were likely only a couple who were fervent about not helping in the research, but being the emergent leaders of the class, they became the spokesmen for the entire group.

The instructor explained to the principal investigator that a couple of the pilots were heard talking about the exam and saying, “…we already have to worry about the test, we don’t need this thing added on.” The instructor continued to explain that it was typical for the pilots during this two-week intensive class to be so overwhelmed and nervous about passing their exams that any added burden may just tip them over the edge. The instructor empathized with the pilots, but also saw the long-term benefits of confidence-based assessments. This particular situation also indicates a bigger problem, that the course may just be so massively-intense that it takes students to the ultimate limit of cognitive fatigue.

But the tipping point was the IRB form, with the pilots fearing it was some “lame” Master’s thesis research and they did not want to be a part of it. A second pre-treatment talk to the class was conducted in an attempt to reduce their fears and explain that the research was sponsored by the Flight Training Academy to help them. The principal investigator explained that high confidence is an attribute of knowledge, and if a person actually knows the information but has low confidence about that information, there may be a reluctance to use it in the real situation. It was further explained that the goal was to elevate all the subject areas so the pilot has enough confidence in each to actually use them instead of second guessing their decisions in the cockpit. This concept relates to a statement from Hassmen et al. (1996) where, “Unsure knowledge can in this respect be regarded as knowledge that is less likely to be employed in solving real problems” (p. 1093).
But the effort was to no avail, even after additional assurances by the academic Program Manager. In the end, it was assumed that even though the nine pilots who declined to participate may have understood the reason for the experiment, they may have felt the added burden on the exam was not worth any short-term benefit, which is a research finding in itself. As is typical nearing the end of any intensive, safety-critical course, the primary goal of the student is to just pass the test and move on. Unfortunately, the nine pilots were also unwilling to be interviewed about why they did not want to take the confidence-based part of the exam, even after repeated requests by the principal investigator to interview them.

Accessibility

To achieve success in the experiment, it was vitally important that the pilots fully understood the process and how to mark the exams with their confidence levels. As stated in the research methods section, the experiment process was verbally explained to them twice, during the pre-treatment screening session and just prior to administering the exam. Additionally, written directions were given to the pilots with the Client Information Form and Survey (Appendix A) and Exam Directions Insert (Appendix B) documents.

The comments from the interviews showed high levels of comprehension. When asked in the interview if the exam procedures were easy to understand, all of the pilots responded that they clearly understood the directions and there was not any confusion about how to mark their confidence level. Therefore, a robust exam procedure was achieved and the use of a three-level marking scheme was appropriate. The three-level confidence marking scale was chosen because of its simplicity and its success in previous research (Gardner-Medwin & Curtin, 2003a and
Bruno, 1995), in addition to its preference by the students (85%) in the study by Gvozdenko and Chambers (2007).

The instructors were also asked about their experiences with the procedures. Similar to the pilots, all the instructors responded that they clearly understood the procedures and there was not any confusion. The positive responses from both the pilots and instructors clearly establish that the directions about how to administer and take the exam were comprehensive and straightforward. Enough so that further analysis of the data would not be constrained or complicated by any problems related to a misunderstanding of the process.

A related interview question sought to gauge the pilots’ attitudes toward any resistance to marking their confidence level for each question. This is an important concept, as any reluctance on the part of the learner to honestly admit their shortcomings is in contradiction of the principles of a confidence-based assessment. Without honest admissions, especially of pure guesses, the resulting report would not accurately reveal their weak subject areas and offer a guide to explore reflective thinking. Therefore, it was important to discover if the pilots had any resistance, either fundamentally or culturally, to marking their confidence for each question. Only one of the pilots reported that he was somewhat reluctant to reveal his confidence, whereas the rest seemed comfortable, seeing how the benefits would outweigh the additional effort. In a related issue, it was important to explore whether the time taken to mark their confidence level was excessive. The two-dimensional process of measuring the knowledge (correctness – logical choice) and confidence metric (emotional state) of their answer is no doubt a disruption over the normal test taking process. Ever since early childhood we have been accustomed to a single-dimensional answer scheme; provide a single mark for your answer and move on to the next question.
Introducing a second step to each question is not only a disruption, but also adds time to the exam duration.

This concern has been addressed by previous researchers as a possible deterrent to any implementation into a curriculum. In a case study about the administration of a confidence-based exam to officer candidates, Shuford and Messengill (1969) concluded,

Even though it must be true that it does take longer to write down a degree of confidence for each of the answers than to choose among the answers and to indicate this choice, and even though confidence testing tends to make students think more carefully about test questions, the data shown indicates it is quite feasible to give a valid confidence test within the time limits usually allowed for choice tests. This is so because most of the students finished early and had time left over before the time limit was reached. (p. 1)

This concern though has been largely quashed by researchers who specifically studied the issue, and found the burden to be minimal. Hunt (2003) found, “The additional time usually required to mark ‘How sure are you?’ is minimal (1-5 sec.) since the test takers’ assessment of their certainty about the correctness of an answer is often largely completed during the process of selecting or producing the answer” (p. 109). In addition, Yen, Ho, Chen, Chou, and Chen (2010) found that, “…only an additional six (6) seconds were required to mark each confidence option, and that time was not a drawback during their experiment comparing a standard Computerized Adaptive Test (CAT) against one that implemented the addition of confidence-weighting” (p. 174). This current study found roughly the same result as the previous research, that the time for the pilots to mark their confidence level was minimal, even somewhat inconsequential. Although a couple of the pilots responded that it depended on the specific situation, which required a brief moment of reflection before marking their confidence level. But overall, the additional time to mark their confidence was minimally burdensome.
Marking Strategies

Sometimes the simple act of deciding on a “yes/no”, or A, B, C, D answer becomes complicated. A person’s career may be at stake or critical knowledge development can hinge on the final score of an exam. Otherwise, the decision can be much simpler, for instance an inconsequential quiz or survey answer. In either case, the agony over the decision is complicated by all the different strategies undertaken. The pilots were asked, but most of the time just offered, their strategy about marking the confidence level for each answer. Many of the pilots reported using simple techniques, while others employed a more intricate strategy. Some of the more interesting techniques were as follows.

John: After I read the questions, if I had the answer in my head, that’s where I felt a 3 was necessary. If I had to dive into the question, because some questions you have to reread the choices in order to get the answer…. But if I had to read the answers of the choices that were given to me to derive an answer that then was the one that I put a 2 down.

Chris: A question that was fact based, that gave me 3 or 4 short answers to choose from, was very short, I know the answer and I am confident. And obviously the ones that I had to delve into a little bit more—questions and answers that you had to read into and make sure there weren’t “gotchas” or things like that in there—obviously that was where I would lower my grade [confidence mark] a little bit. Only because there were so many complications in the question. Not necessarily in the answer, but there were so many “gotchas” in there it was like, I am pretty sure this is it, so put it in there.

Both John and Chris stated they had a very logical way of generating their confidence marks. It mimics the technique used in computer languages, or logic statements, called IF-THEN-ELSE. This conditional statement postulates what to do if a Boolean (binary) condition is either true or false. In this instance, John and Chris were first deciding whether they knew the answer to the question or not, a Boolean yes or no. If they had decided that their sole primary answer was correct, then they would have marked a confidence level of “3.” Whereas, if they had to think about it further, or needed to look more carefully at all the available answers, then
they would have marked a confidence level of “2.” They did not mention their logical process when deciding what constituted a confidence level of “1.” Incidentally, neither John nor Chris used a “1” in any of their responses. It seems they were both so certain of their ability to ace the exam, a low-confidence mark was not even an option. But their assuredness was not misplaced, as John scored a 98% and Chris a perfect 100%.

Looking over the numeric occurrences of each confidence marking level across all of the exams, the “1” or “guessing” mark was only used 20 times, or 0.04% of the time. This low number is indicative of test takers who may be a tad over-confident in their responses, and would rather refer to themselves as making an “educated guess” than a flat out guess. But this phenomenon is not without rationalization, as pilots are specifically taught to make decisions very quickly and confidently, since delaying an action may cause a very serious accident. Correspondingly, the piloting vocation attracts those personality types who are more prone to have high regard for themselves and act more on instinct than reflective thought. Chris may have said this best.

I think I might be the wrong person to test, because I have pretty good test-taking skills, so I go through and you could probably give me a test on something that was not even a subject matter that I knew, and if it was a multiple choice test, I could probably get myself through it pretty well.

Both John and Chris marked their confidence level for each of the questions concurrently while answering the question, but three other pilots used a different method that involved looking at the exam from a holistic viewpoint.

Roger: I think maybe I was a little confused with my “low confidence,” “medium confidence,” and “high confidence” answers, because I initially go through every question, and if I am 100% sure I am going to get this question right, I mark it and say, “Okay, I am definitely going to do this,” and so I go through the whole exam. If I do not know the question 100%, I am not even going to answer it, and then when I finished the exam, I count how many questions I have, and this tells me what I am pretty sure I am going to get as my final score. So pretty much, I had I think 80… it was like pretty close to 86. It was just above 80 that I was 100% sure
of, and then I went back and then I do all the ones that I didn’t do and I just chose the best answer [knowledge] I am sure with…. So when I put in “high confidence” it was just all the ones I was confident of. I only put a few that were low confidence because medium confidence I just thought “Hey, I am just doing an educated guess here, so…”

Randy: Overall, I thought I was pretty confident. Like I said, I probably should have gone over, after the test, and reviewed the responses, 1, 2, 3…. To make sure they were kind of accurate on what I was looking at…. So maybe after the fact I could have said what would be better versus going question-response, question-response, question-response, because you know you are taking the test and you want to sit there and try to get through…. Maybe when you’re concentrating and taking the test and getting the answers and trying to make sure you got it, maybe it would be better for somebody to after the test then go ahead and grade the response [confidence mark]. I did my question then response, question-response, question-response, question-response. Then when you get done with the test, when you know you’ve got everything done, then go back and just kind of maybe review that.

Jimmy: As I was going through it, I knew the answers as I went through every single one. There were literally 3 questions, I think, that I took a matter of 5 seconds to be like “Let me read that again,” and those are the ones that I marked 2s. Everything else I left completely blank, so once I’d answered every question, I just went back and put a 3 on every other one. Honestly, I was very confident about every single one of those answers, except for the ones that I marked 2s.

It appears Roger, Randy and Jimmy used a strategy where they first went through all the questions as they would for a typical exam, marking the confidence level on only those where there was some doubt. By default, all the remaining questions would become “high-confidence” marks, or “3”s. Then after reaching the end of the exam, they would return to their questions and mark a “3” for the confidence level they had not marked previously. This technique is in contrast to the recommendations from Gardner-Medwin and Gahan (2003b), where they state,

Each time an answer is entered, this is followed up with a request for the confidence level. It is important for formative use (study and revision) that all questions be marked individually one at a time (i.e. not in batches), with immediate presentation of feedback and explanations: this ensures that the feedback arrives while the student still has in mind the reasons for selecting an answer. This is especially important when high confidence has been expressed for a wrong answer. (p. 3)
This holistic strategy by a few of the pilots is commendable, and somewhat surprising, as it was not predicted during the design phase of the study. It surely would need to be addressed in the future design of a computer-based assessment delivery system, as the pilots had used a different strategy than what the experienced researcher, Gardner-Medwin had recommended.

The rest of the pilots did not expand on any interesting marking strategies, often redirecting the discussion toward a self-disclosed mistake. But a few responded with either the, “I was honest with it.” type of response or explaining how they, “should have gone back and reviewed some of them.” This discussion about strategies illustrates that some pilots are less expressive and just may revert to ingrained processes without giving it much thought, while other pilots are more introspective about their thoughts and can offer detailed explanations about their confidence marking processes. Overall, a consistent marking strategy was not discovered, but the exercise did lead to a greater insight about the techniques that must be accommodated during the design of future confidence-based assessment platforms.

After the first few pilot reports were generated, it was noticed that a couple of the pilots responded with mostly “3”s as their confidence mark, and it was suspected that those individuals may have been trying to “beat the test.” This phenomenon was cited as a possible limitation in a previous study (Ahlgren, 1969) that explored personality bias of a confidence-weighted scoring scheme, “…personality has an effect on confidence-marking…. betting against the house is a persistent human trait, and there is ample evidence that characteristic levels of confidence and risk-taking play a large part in confidence-marking” (p. 5). The same complication was also suspected in the current study. Therefore, an additional question was added to the interview script.
To determine if this “beat the test” phenomenon was happening, an additional metric was calculated from the confidence scores for all the pilots, called Confidence Sum. This single metric was the sum of all the confidence marks for the entire exam. With the 50-question exam, if the pilot marked a “3” for the confidence level on each question, the maximum Confidence Sum score could be 150. If the pilot scored over 120 for the Confidence Sum, it was suspected that he may have been trying to “beat the test,” thinking a high confidence mark would result in a higher score. All but one of the pilots in the study scored above 120, but only four pilots were left to be interviewed when an additional question about high confidence was asked. Those last four pilots were asked, “I noticed that your confidence marks showed a tendency to be mostly 3s, or high confidence. Thinking back to when you were taking the test, please reflect on your thoughts as to your use of the three confidence levels.” The pilots who were asked this question responded with the following.

Pete: I was honest with it. I mean, that’s the whole point of it.

Homer: I think I had a pretty good knowledge of the question that was asked. So to me that was a 3. There wasn’t much guessing or decision.

John: After I read the questions if I had the answer in my head, that’s where I felt a 3 was necessary. If I had to dive into it, well, it depended on the question, because some questions you have to reread the choices in order to get the answer, but if I had to read the answers of the four choices that were given to me to derive an answer, that then was the one that I put a 2 down.

Jimmy: As I was going through it, I knew the answers as I went through every single one. There were literally three questions, that I took a matter of five seconds to be like “Let me read that again,” and those are the ones that I marked 2s. Everything else, I left completely blank, so once I’d answered every question, I just went back and put a 3 on every other one. Like I answered every question, and honestly, I was very confident about every single one of those answers, except for the ones that I marked 2s.

The comments from these four pilots were compared against their exam scores to explore any correlations. The first two pilots, Pete and Homer earned an 82% and a 90% respectively,
which may have been an indication of their rudimentary marking strategy as reflected by their short, quick and to-the-point responses. Whereas, John and Jimmy revealed how they used very complex strategies to achieve their 98% (both) scores, but they may just have been better at articulating their thoughts.

As stated previously, all but one of the pilots had a high Confidence Sum score (over 120). It would appear from a purely numeric standpoint that they were either trying to beat the test or were just overly confident on their choices. An excessive amount of confidence, maybe, but it is not the first time a pilot has been labeled “overconfident” by any means.

In a related study about knowledge measurement, Hunt (2003) related that, “Men were particularly overconfident when incorrect…. And in many other life situations in which people are dependent upon the judgments and decisions made by overconfident people, the consequences can be quite negative” (p. 111).

A few possible theories may be in play here. First, it may seem that they were not being humble in their own self-assessment, a reflection on their over confidence. Secondly, it just may be that they are not self-aware enough to differentiate their own confidence between the three levels. Or possibly, a supposition is not even warranted in these cases, because their high raw scores may have just indicated they were simply “appropriately” confident, not over confident. And still another explanation may be that they just did not understand the process. It is interesting how these theories may play out when a confidence-based assessment is fully implemented in an electronically-delivered exam, as most likely all these theories will be true to some extent.
Score Expectations

The intent of confidence-weighted scoring is to enhance understanding and learning through self-awareness and reflective thinking, by employing a mechanism—via the graded exam report—that identifies those subject areas that score either low confidence, false confidence or a lucky guess. False or misleading knowledge can have serious ramifications within a safety-critical environment, such as operating an aircraft. The graded exam report mechanism has been designed to offer the pilot a clear and concise understanding of areas that need improvement.

The interviews addressed the use of the exam report and attempted to gain insight into how they were used by each pilot. The first interview question within that thread sought to determine how the report was received and more specifically, if it was welcomed. The polarity of the comments was split down the middle and a succinct theme did not emerge. Five interview responses were categorized as “the report was as suspected,” appearing that the graded exam report was about what they expected. This attitude is reflective of the extensive pre-exam explanations about how confidence-based assessments were supposed to work and the overall acceptance of their weak areas. These five pilots were generally open to seeing their weaknesses in full color and accepted their weaknesses as a useful tool to help improve their knowledge. It is also an indicator that they may have already known what subject areas they were weak in and the graded exam report just offered confirmation. On the other hand, the remaining pilots seemed to be somewhat surprised at what the report was telling them, or there was a bit of confusion that the report uncovered. There were also a few missteps along the way, where the pilot realized an error. The following excerpts illustrate their frustration with the process or defensiveness over
realizing they did not do well on the exam. And it appears each of the pilots had a slightly different take on their experience.

Jimmy: I mean the red one [color coded on graded exam report] is what I’d expected because that is the question I missed. The rest of them that are yellow, I don’t quite understand why they would be yellow because I didn’t miss a question on them and they all had to be…[researcher then explained the color scheme] I’d have to go with like I specifically remember I thought I only marked three of them as 2s, but I’d have to go back and look, and I know two of them had to deal with ADS-B [a new Air Traffic Control system], which I feel like would be in Avionics, so I mean… I don’t know.

Jimmy was surprised to see so many yellow marks, as he thought that because he marked the knowledge selection right, it should have been a green. But then after hearing the explanation that the yellow mark was because of a low confidence mark, he realized the coded color was a combination of both the knowledge and confidence marks. This realization triggered a frustration/defensive response, in which he concluded that there must have been something wrong with the exam. Considering that Jimmy was annoyed at scoring only a 98% (one wrong) on the exam, blaming the grading process was definitely true to his character.

Being defensive about an undeniable realization is counterproductive to the principles of a confidence-based assessment tool, as without a little humility leading to reflective thinking the exercise is wasted. Gvozdenko and Chambers (2007) commented on this realization about how confidence plays into the foundation of knowledge, “A correct/incorrect scoring system with no indication of certainty fails to collect this information and thus robs lecturers and tutors of a chance to identify and address weaknesses in students’ knowledge” (p. 2).

Similar to Jimmy, John scored a 98% on the exam, but was a little more humble in his attitude toward the graded exam report.

John: Well, I really thought it was a sea of green, but yeah, I felt okay about it. I didn’t expect to come out of here knowing everything. If anybody does, I think that’s kind of pretty cocky. [interviewer: “…these areas here, the Ice & Rain, did you
feel weak in these areas before you went into the test at all?"] Um, maybe the Ice Protection and Rain Protection a little, but the Air Conditioning kind of surprised me. It seems like a fairly easy system to understand. Why I missed that question I don’t recall.

John appeared to understand that the report was generated in response to his own admissions. He accepted the fact that he went into the exam thinking he was weak on the Ice & Rain Protection subject, and the report confirmed that assumption by showing red. But he was surprised that he did not do as well as he thought he would on the Air Conditioning subject, which showed yellow on the report. The realization that his knowledge of the Air Conditioning subject was weak may have encouraged further study about the subject before the end of the course. At any rate, the discovery of weak areas through the use of confidence weighting is the point of the assessments, so it was an appropriate finding in John’s case.

Roger, on the other hand, was less concerned about his own performance, having scored an 86% on the exam, but questioned the grading process as a reason for what he perceived as a low score.

Roger: Well, some of the questions there was only, some of the topics, I had just a few questions, and some of the topics that I felt I was less confident about before the exam, and I happened to get all the questions right. It acknowledged I was proficient. There was one question here that actually I got wrong by accident. I totally just misread the question. It said it was on Limitations and it said “deficient knowledge.” You know, I wrote on a piece of paper, on my question paper, what question [answer] I was going to choose, and then when I went to actually put it on the answer paper, I put C instead of B or something, and then there was only one Limitation question and it said “deficient knowledge.” So that was something where it could be a little unreliable, but still, I thought it was great and I enjoyed getting the information.

Roger felt lucky that while for some of the subjects he went into the exam not knowing very well, he, “…happened to get all the questions right,” which meant he guessed, but guessed correctly, or at least made plenty of educated guesses. His confidence-weighted score of 66% confirmed his low confident, but successful, educated guessing technique. This result then
reinforced that even though he had low confidence going in, his hunches were somewhat correct, and in the end this actually gave him a boost in confidence over those subjects, as indicated by his admission, “It acknowledged I was proficient.”

Roger also had some confusion over a couple “Deficient Knowledge” results on the report. In the first instance, he meant to mark one answer, but mistakenly marked the wrong answer. He felt that the first red mark was justified, but the second triggered a defense response. Roger felt that receiving the second (Limitations question) red “deficient knowledge” was unjustified by stating, “… [the test] could be a little unreliable.” So therefore, his marking of the wrong answer must have been, in his mind, the result of a mistake and not the undeniable evidence that he was misinformed about a topic, and the grading process should have somehow accounted for his error. Being called out on this question with a strong belief of a wrong answer is also reflected in the findings from Hunt (2003).

The person may be extremely sure that the incorrect answer which he/she selected is correct and, thus, may be misinformed— which is much worse than being uninformed. A sure-but-wrong belief, used confidently as a basis for making decisions and taking actions, may lead to surprising errors in performance—sometimes with tragic results. (p. 105)

However, Roger finished his evaluation on a high note, saying, “…but still, I thought it was great and I enjoyed getting the information.” So, it appears Roger became a little frustrated with the process, but overall appreciated the greater information the confidence-based assessment provided.

Though Chris actually responded that his confidence-based score did not surprise him—and therefore belonged in the first group of five pilots—he contradicted his statement by saying the results were not necessarily what he expected. Furthermore he, like Roger and Jimmy, was frustrated with the assessment.
Chris: It didn’t surprise me, but I wouldn’t say it was necessarily what I expected. I knew that you would obviously get some results out of it so it was something that was going to show up. I couldn’t have guessed in which area was going to show up, and like I said, it was almost question based. And the harder the questions were constructed, the more likely I was to grade it down than the easier the questions were constructed. If I had probably seen on that same test all questions that weren’t wordy but were short questions, I probably would have graded them all 3’s.

Chris expounded very little on his performance, most likely because it was not in question, for he scored a perfect 100% on the exam. Chris is one of those aviation lifers, having been involved in aviation since he was seven years old. Since the exam score was not a point to discuss, he began revealing his frustration in marking the confidence level for the questions, and how that frustration was based on the “wordy” questions. The critique of question misunderstanding should be more directed toward the writer of the exam itself, but the issue of a badly-worded question is of great concern and most definitely does have an effect on the ability to correctly mark a confidence level. If test takers do not fully understand the question, how can they ever have high confidence in their answer?

Study Habit Change

One of the primary objectives of the study was to determine qualitatively if confidence-based assessments affected pilot study habits, by encouraging greater emphasis to be placed on those subject areas that were identified as weak. The following research sub-question sought to gain the pilots perspective about this issue.

To what extent does the addition of confidence-based assessments change pilot habits?

During the initial design phase of this study, it was not a forgone conclusion that confidence-based assessments were beneficial in the context of an aviation training environment. Since there have not been any research studies into the subject that even approached safety-
critical training situations, the research questions declared a neutral assumption as to whether the effects would be positive or negative.

The topic of study habits involves the processes used to gain knowledge and skills through a solo effort in preparation for a knowledge or performance assessment. These assessments can be in the form of a written exam, an oral discussion or an observed performance demonstration. All of these techniques are typically used throughout the aviation training environment, and the courses examined for this study include all three of these techniques. The ground school exam, oral discussion and simulator performance checkride must all be passed before the overall course is considered passed. Various study habit techniques are used by students, such as reading books or manuals, online search investigations, video tutorials, computer/web-based training, flash cards, peer discussions, virtual visualizations, audio recordings, muscle memory drills, practice exams, etc. The extent to which the research pilots changed, modified or employed these study techniques is the subject in this section of the discussion.

Since the majority of the pilots had attended a previous course at the Flight Training Academy, they were able to use those experiences as a comparison with the current course and its confidence-based assessment component. The pilots were asked the question, “Did knowing your combined confidence level from the report for each of the knowledge sections provide useful information for your remaining study time before the simulator briefing sessions?” Two pilots responded with a simple “Yes, it did” comment, while Homer hadn’t started studying yet by the time of the interview, and he stated, “I can’t say that I’ve studied a whole lot, since I’ve been concentrating on the simulator session, but I will keep it in mind when I hit the books again.
One pilot, Jimmy, admitted that the graded exam report did not change his study habits, as he stated the following.

For me personally, I don’t think so. I’m not saying that it can’t work for people. I just feel like I know the areas I’m deficient in and I’m the type of person that if I know that I’m lacking in something, I’m studying that myself. Like I just have that kind of willpower about me that I don’t let anything go undone. Like, if I don’t know it, I don’t just say, “Oh, let’s move on to the next thing. I got a 98.” Like, I want 100. So I personally already know to study that just myself.

Jimmy appears to be the independent sort, and feels quite impressed with his own performance, even though his confidence-weighted score was an 81% with four yellow individual subject area scores and a single red one that indicated a highly-confident wrong answer. Jimmy seemed very competitive, at least with himself, as it was unknown whether he was competitive with his colleagues. So for Jimmy, the confidence-based assessment exercise didn’t seem to affect his study habits, because it is his habit to increase study in weak areas. The rest of the pilots had much more favorable responses, saying that the graded exam report did change their study habits for the better by offering a tool for self-reflection. This is a point addressed within the literature, as Gardner-Medwin and Gahan (2003b) stated that “Reflection strengthens the links between different strands of knowledge… and it strengthens the ability to justify an answer, one of the essential elements in an Aristotelian definition of knowledge that is often missing in students who prefer rote-learning to understanding” (p. 2).

A common theme was observed among the majority of the pilots, as they had very positive things to say about their experience and the addition of the graded exam report. Many of the comments 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 employed a new study technique, but just continued with more of their usual techniques. Although it might be said 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.

Training Benefits

The primary objective of this study was to determine if the implementation of a confidence-based assessment tool into an existing aviation training program provided benefits beyond any risks or burdens. The last series of interview questions sought to determine those overall benefits and if any participants were fervently against the program. Additionally, the interviews were a platform for the participants to give their opinions about how their colleagues would accept confidence-based assessments.

The pilots and instructors were asked similar questions about training benefits, first to learn their open-ended opinions and second, to understand the degree to which their opinions differed as to which particulars were beneficial to each group. The pilots were directly asked if they thought confidence-based assessments were beneficial during the course. A couple of the responses were short, such as, “I think it’s a good idea.” and “I really like this.” while nearly all the others had much more to say, which was an indicator itself that they perceived the benefits. The more elaborate responses are as follows.

John: I think they are very beneficial, at least in my personal opinion. In this particular case, I felt like it was pretty accurate in determining my knowledge level and allowed me to correct the areas that were deficient or, like it says in here, needs improvement.
Pete: I think that would be brilliant. I think that would be great….Well, a few of these deficient knowledges [sic] I knew the correct answer. I just mismarked my answer sheet. So I believe that’s why I have the deficient knowledge on some of these, but with my low confidence on some of the other ones that I got right, I want it to help me review those and to study up on them…. There were a few that the test surprised me with some of the questions and the difficulty, and it made me realize how much I didn’t know.

Cirrus: I think it’s good for the client, to have a better understanding, just guessing and answering and getting it right doesn’t do the client any good if they don’t go back and review that. They say, ‘Well yeah, I answered that and got it right, but I had no idea what the answer was or I was, you know, kind of fifty-fifty.’ It gives them a better idea of the areas that they can improve on.

Roger: I mean, it does not take that much time, and if it’s going to get you 5% extra in your studying what you should study, that is something that is worth it to me.

Chris: It can help me understand what my areas I am deficient in, even though I may have gotten the right answer.

Barry: I thought they were excellent. I thought that was a good way to get me more focused on the subject areas that I needed to be on. So yes, I agree that I benefited from that, and the ability to know that nobody else was going to see that but me, made me confident that I just need to focus on study, not that I need to be graded or… that it was a help for me.

Dick: I think it’s a good idea. I think especially younger guys will get a lot out of it. You know, I’m not saying that I wouldn’t either, but you know, guys that are newer to the airplanes or younger that don’t have as much flying experience are going to get a lot out of it too.

Within the excerpts from the pilots, there were plenty of “very beneficial,” “helpful,” “brilliant” “excellent,” and “great” comments. Many of the comments personalized the experience, offering details about for whom the addition of confidence-based assessments would be the most beneficial. John, Pete, Roger, Chris and Barry directly experienced the benefits for themselves, while Cirrus and Dick felt the concept was good, and beneficial, but that it would be more effective on other pilots. Homer and Aviator commented positively, but in more third party terms. In addition, the comments confirmed that the pilots used the graded exam reports in the
manner in which they were intended, to highlight areas in which they were weak and as an aid, or plan, for self-reflection during their remaining study time.

This positive encouragement has been found in other studies where students were polled about their preference, as Gvozdenko and Chambers (2007) found, “…the majority of students (67%) on a summative test and nearly all students (96%) on a formative test chose to use the certainty scale as an additional tool in the testing procedure” (p. 215).

The pilots were given plenty of opportunities to be totally honest with their opinions, as they were assured only a pseudonym would be used to identify them in any reports and they were alone with the interviewer. During the course of 11 pilot interviews, not a single disparaging comment was heard. There were a couple comments saying, “I don’t think this is for me…,” but in each case the confidence-based assessment tool was not directly criticized, saying that it would benefit “others” a lot more.

Although they did not openly say so, Pete, Roger and Randy alluded to the additional time and effort involved, but in a roundabout way saying, “…it would be worth it.” This circles back to the previous 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.

Another previous topic, accessibility, was specifically studied to learn if the confidence-based assessment process was easy to understand and administer. The responses to the specific questions (covered previously) about usability from both the pilots and instructors showed high levels of comprehension. These last comments about the pilots’ overall experiences show a definite lack of comments about any accessibility issues. This supports the notion that if properly
administered and trained—especially through computer-based means—a robust confidence-based assessment tool would be minimally-burdensome while offering worthwhile benefits.

The instructors looked at their overall experience in a slightly different way, as the confidence-based assessment effort was designed to give them a more objective viewpoint of their students’ progress. They were asked two interrelated questions about their insight into their students and how comfortable they felt using the new tool. The first question asked if they gained a greater insight into the pilots’ level of aircraft system knowledge, while the second related question sought their comfort level with the process. Each instructor expressed at length their opinions about the confidence-based assessment tool, with a common theme emerging.

The instructors were notably 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. They each had a slightly different take on the experience, describing at length their individual usage of the reports. A couple saw the tool as a knowledge-based snapshot of the pilot’s weak areas and how it could offer ideas about specific failures to concentrate on in the simulator sessions. This particular finding was one of the primary intended benefits of the graded exam report, to supply the instructors with an evidence-based plan to cover those areas in which pilots were actually weak, instead those areas in which the pilots “appeared” to be weak.

A couple of the instructors, Bob and Sawyer, also saw how a long-term analysis of the reports could be used to modify classroom delivery techniques. Ground instructor Bob stated that these graded exam reports are self-assessments for the instructors, to better deliver instruction. The results of these assessments could be one criterion in measuring instructor teaching quality, since it would improve on current subjective techniques.
The all-encompassing theme is that the confidence-based assessment exam is a worthwhile tool that the instructor can use to improve the training experience and possibly increase the efficiency of the curriculum schedule.

Conclusions

In a highly-critical safety environment, it is essential that correct decisions are made and the knowledge supporting these decisions is complete and highly correlative. Incorrect or mistaken information built upon a foundation of guessing undermines those goals. A mechanism that eliminates, or at least reduces, guessing on exams, called confidence-based assessments, has been shown in previous research to be effective.

These confidence-based assessments have been implemented and studied in only a very few independent situations. The principles of the teaching technique have been researched in secondary, post-secondary and on a limited scale, in tightly-controlled academy-type settings. These controlled studies all contained relatively homogeneous subject populations, such as a class of second-year medical students (Gardner-Medwin & Curtin, 2003a), or a public service personnel academy (Bruno, 1995). But as the literature search uncovered, a rigorous research study has not been performed within a purely adult-oriented, safety-critical technical training environment.

This study performed a qualitative investigation of a formative confidence-based assessment within a specialized environment, a flight training academy with a non-homogeneous (in terms of age, experience and societal factors) sample of pilot and instructor participants. Multiple areas were researched in an effort to view the use of the confidence-based assessment tool from perspectives pertinent to eventual implementation into an existing aviation training curriculum.
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 simulator briefings to address those weaknesses.

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

The objectives of the study were straightforward, but to achieve these goals, specific questions had to be answered. The extent to which these answers were found depended on the expressive self-awareness and articulation of the research participants. The responses given to the specific questions asked of the study participants were used to answer the following high level questions.

How does the addition of a confidence-based assessment element to a mature exam change the habits of both students and instructors within an aviation training environment?

Within that overlying primary research question, sub-questions were explored that targeted specific usability aspects and attitudes about the process.

Sub-Questions

- To what extent does the addition of confidence-based assessments change student habits?

- To what extent does the addition of confidence-based assessments change instructor habits?

- What is the perception by both students and instructors that confidence-based assessments may be a useful tool to enhance learning efficiency in an aviation training environment?
What are any limitations of the implementation of confidence-based assessments into an aviation training environment?

From those sub-questions, focused interview questions were created to garner rich and descriptive responses from both the pilot and instructor participants. The answers to these questions were investigated through phenomenological qualitative inquiry methods. The all-encompassing theme that was found throughout the study was that the confidence-based assessment exam is a worthwhile tool within the advanced aviation training environment.

Due to regulatory restrictions for the end-of-ground school exam, the experiment had to use a paper-based process to explore the new confidence-based assessment tool. But the experiment procedures were specifically designed to explore the particular methods as an analogue to an electronically-delivered exam.

The conclusions are organized in the following order to reflect the sequential procedure of the interview data collection effort.

1. Acceptance
2. Accessibility
3. Marking Strategies
4. Study Habit Change
5. Training Benefits

Acceptance

The use of the term “acceptance” in this study is described as the way the participants agreed to, and embraced, the use of confidence-based assessments despite the interruption to their normal processes while offering worthwhile benefits. Or from another perspective, the agreement to perform extra tasks to achieve an acceptably higher return on investment. For it
was only if the pilots and instructors fully accepted the interruption, that they could begin to acknowledge how it would benefit their experience at the Flight Training Academy.

It appeared there was a somewhat warm response from the participants. Some took to the new assessment right away, understanding that the added benefits were worth a little uncomfortable effort. Others took a little time and multiple explanations before they fully understood the concepts. But once they did, they seemed to embrace the new tool.

Of special note was one instance in which a class of nine pilots refused to participate in the study. The pilots’ fatigue level approaching the exam was very high, and they already felt overwhelmed after having experienced two full weeks of intense lectures. This may explain the reluctance on the part of the pilots to accept the confidence-based assessment tool, as the course required them to learn too much within a very compressed schedule.

Accessibility

This research study was commissioned by the Flight Training Academy to uncover any issues that may arise with the future implementation of a computer-administered confidence-based assessment exam. Therefore, to achieve success, it will be vitally important that the pilots (students) fully understand the process of how to mark the exams with their confidence levels and the instructors fully understand the process of administering the exams and use the assessment reports effectively. The path to successful pilot-instructor understanding rests in the careful application of proper and effective exam directions.

During this experiment, the participants received both verbal and written directions before the exam was administered. Specific interview questions sought to discover how well those directions were understood. Success was achieved, as the comments from the interviews
showed high levels of comprehension and compliance. All the pilots and instructors responded that they clearly understood the directions. Therefore, the continued use of similar directions would reduce or eliminate any confusion and lead to greater usability.

A related topic was the pilots’ attitudes toward any resistance to marking their confidence levels. This is an important issue, as any reluctance on the part of the learner to honestly admit their shortcomings contradicts the principles of a confidence-based assessment. Without honest admissions—for example, when they are purely guessing—the resulting reports would not accurately reveal their weak subject areas. The majority of the pilots responded that they did not have any reluctance admitting their confidence levels and that the time needed to mark their confidence levels was inconsequential.

The analysis of the interview responses about accessibility confirms that when thorough directions are used, the confidence-based exam is easy to use and does not add any consequential burden on either the test taker pilot or instructor administrator.

Marking Strategies

The act of marking an answer on an exam is pretty straightforward, but when a second selection is added to each question (i.e., confidence level ratings), it complicates matters and requires a strategy. It was intended that the pilots in this study proceed through the exam by selecting the knowledge part of the question then marking their confidence right next to it, before moving on to the next question. But it was discovered that the pilots used another unexpected technique that was both innovative and effective.

Beyond the aforementioned two-step process, some of the pilots implemented an interesting strategy. They would first go through all the questions as they would for a typical
exam, but only marking the confidence level on those questions where there was some doubt, usually a “2”, as a low-confidence “1” was rarely used. Then after reaching the end of the exam, they would return to each question and mark a highly-confident “3” for the confidence level. This holistic strategy, although unconventional, was successful nonetheless and would need to be accommodated for in the design of any future computer-based assessment delivery system.

Study Habit Change

The topic of study habits involves the processes used to gain knowledge and skills through a solo effort in preparation for a knowledge or performance assessment. A primary objective of the study was to determine if confidence-based assessments affected those study habits. Compared to the pilots’ previous experiences at the Flight Training Academy, this study found that the majority of the pilots had very positive things to say about their experience. Several of the pilots openly disclosed that they had indeed modified their study habits through the simple use of a confidence-based study plan. Therefore, it appears that the pilots did change their study habits noticeably from their previous course(s), and in a positive direction.

Training Benefits

The overall primary objective of the study was to determine if confidence-based assessments are a useful tool to enhance learning efficiency in an aviation training environment. In support of that objective, pilots and instructors were interviewed for their perceptions after having experienced controlled confidence-based assessment exams. The results were overwhelmingly positive by a large margin, as both the pilots and instructors praised the confidence-based assessment tool, saying it was very beneficial. There were plenty of “very
beneficial,” “helpful,” “brilliant” “excellent,” and “great” comments. Many of the comments personalized the experience, offering details about where, or for whom, the addition of confidence-based assessments would be the most beneficial, but there was not a single instance of any outwardly-negative response.

One of the research sub-questions sought to discover if the addition of confidence-based assessments had any noteworthy limitations. Even though there were a few minor concerns, which were disclosed in the discussion section, the absence of any disparaging comments indicates that those minor limitations could be easily overcome.

The comments also confirmed that the pilots used the graded exam reports in a manner for which they were intended, to highlight areas in which they were weak and as aids, or plans, for their remaining study time. The instructors also used the exam reports as they were intended, as evidence-based plans to cover those areas in which the pilots were actually weak, instead those areas in which the pilots “appeared” to be weak. In addition, the instructors even envisioned how a long-term analysis of the reports could be used to improve classroom delivery techniques, both for individual instructors and company instructors as a whole. Conceivably, long-term analysis of the assessments could also be used to critique courseware content and delivery methods.

The all-encompassing theme is that if properly trained and administered—especially through a computer-based medium—a robust confidence-based assessment tool would be minimally-burdensome while offering worthwhile benefits.

Implications

This investigation into the effects of a formative confidence-based assessment tool within a specialized, adult-oriented, safety-critical technical training environment was designed as a
phenomenological research study. The nature of introducing a new learning tool where it has not been researched, or possibly even tried, previously is a somewhat adventurous endeavor, and as such, the “let’s try this and see what happens” approach required a method that accommodated that open-ended exploration.

The use of a phenomenological research approach was chosen because of its use as a discovery method into the experiences as told through the perceptions of the actors themselves. The use of that method was successful in collecting the participant perceptions and opinions of the new tool, but a new understanding or derivation of the method was not discovered, nor was this the intention of the study.

Researchers may also take note of this study as an examination into a different aspect of confidence-based assessments as it not only applies to an adult learning situation, but to one that trains implicit knowledge for the operation of a safety-critical system. The administrators of an aviation training center would specifically find the results and conclusions helpful when designing a curriculum, as the study determined that the use of a confidence-based assessment methodology would be beneficial by providing a more efficient and comprehensive training experience under the study conditions.

This study also explored the use of a confidence-based assessment report as a tool for reflective learning, especially when a learner is deciding whether to hold a specific chunk of knowledge in either low or high confidence. Reflecting on that decision drives the learner to either seek out further information to fill the knowledge gap, or if self-judged in high confidence to move on to additional topics.

The evaluation of instructors is a feedback process used within a training organization to continually improve the customer experience through greater thoroughness and efficiency of the
courseware and delivery. In this study, the instructors themselves extolled the benefits of viewing the pilots’ confidence-based exam reports, both as aids for oral evaluations, as well as tools for long-term analysis of their own classroom delivery performance.

Another specific finding in this study relates to the design of a web or computer-based assessment delivery system that uses confidence-based assessment techniques. Many of the pilots used varying strategies in selecting their confidence marks for each question. In particular, some pilots marked only their “non-high confidence” marks while advancing through all the exam questions, then returned after they were finished to mark all the remaining with a high confidence mark. While this holistic strategy is commendable, it would need to be accounted for in the design of a computer-based assessment delivery system. In general, the electronic delivery of an exam affords the test taker only a limited ability to review questions after they have been marked. Unfortunately, many online confidence-based exams are already using the confidence marking button(s) as the link to the next question. Therefore, any future design and implementation of a computer delivery system would need to accommodate this holistic strategy.

Recommendations

Many specific recommendations that may directly benefit organizations thinking of creating new assessments or just adding a confidence-based element to existing assessments can be derived from this study. Since this research study explored the general use of confidence-based assessments, the techniques can be applied to both traditional paper-based exams and computer-based delivery mediums.

Research that explores the use of confidence-based assessments within secondary and post-secondary learning environments already exists, and the recommendations from this study
do not offer any direct insight into those environments. This study explored the experiences of professional pilots and instructors within academy-style short (2-3 week) courses who train for the initial operational aspects of specific corporate aircraft. The following recommendations were therefore developed in the context of adult training, but may have applications in other learning environments as well.

- Schedule the confidence-based exam/test/quiz early enough during the class so that the students have an ample amount of time to self-assess their progress and plan their remaining study time.

- Provide extremely clear and thorough confidence marking directions to the test takers.

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

- Clearly delineate the three confidence level labels to:
  - High Confidence (I am sure)
  - Medium Confidence (I am partially sure)
  - Low Confidence (I am not sure – I am guessing)

- Accommodate a holistic confidence marking strategy by allowing the test taker the ability to move to the next question without marking the confidence level, then having the ability to return to mark their overall confidence level.

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

- Provide a checkbox for each question allowing the test taker 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 test taking technique will most likely meet with some resistance by the administrators, instructors and students. Persistence, careful planning and patience are requirements for the successful implementation of any new system.
Limitations

*Societal/Cultural Differences*

The aerospace industry is by design an international entity, and to fully gather the impact of a confidence-based assessment tool on a worldwide market, a cross-cultural population needs to be investigated. The subject population for this study was mostly raised and lives in the U.S. Therefore, a broader study may uncover slightly different results. For instance, Asian people are known to be rather reluctant to reveal their personal limitations or feel embarrassed to reveal their true confidence of a subject. This reluctance factor could skew results if this study were carried out on a group of Asian pilots.

*Gender Differences*

The participants in this study were all male. It is generally known that less than ten percent of pilots are women, with an even lower percentage employed as professional pilots. The literature review did uncover studies that compared genders, but found either no gender differences (Gardner-Medwin & Gahan, 2003b), or males being only slightly more sure of their answers (Hassmen et al., 2002). The results of this study may have been different if female pilots or instructors were involved.

*Application Across Training Stages*

This study was applied to only a small segment of the aviation training arena, namely the corporate/utility aircraft, type-rating transition course. During a professional pilot’s career, this specific stage of training is generally the last step of a long, progressive path beginning many years, even decades, before. The progression through basic, primary, and advanced training
offers many opportunities to use confidence-based assessments. The outcome may be very
different if a study such as this is applied to pilots within those lower stages of training, since the
students would be younger, far less experienced and possibly in a wholly-different environment,
such as the military. Therefore, it would be inappropriate to assume that the results and
conclusions of this study would apply to other segments of aviation training.

Measures

This study was purely a qualitative effort. Many of the studies explored in the literature
review were quantitative in nature, statistically analyzing exam knowledge and confidence
scores. This method of analysis brought out different viewpoints than the current study in an
attempt to predict the outcome of future exam performance. If this study would have employed a
quantitative element (e.g., Likert perception/opinion measures survey) exploring the same
constructs, there would have been results that addressed an audience accustomed to quantifiable
data. The use of a quantitative survey study may also result in a greater saturation of the findings,
as it could be administered over a greater population and over many years. But it is doubtful that
the conclusions would have been markedly different, due to the specialized nature of the subject
group.

Treatment

This study was commissioned by the Flight Training Academy to explore the effects of
confidence-based assessments for their possible use in a computer-administered exam still under
development. But due to regulatory restrictions that mandate the use of paper-administered
exams in a current FAA-approved course, this study could not use computer-administered exams
and had to piggyback the confidence-based assessment element onto those paper exams. Therefore, the results and conclusions of this study may not transfer directly to the use of a confidence-based marking scheme within the intended computer-administered exam system. The pilot and instructor interview opinions may have been different if the exam had been administered over a computer, although the general concepts and use as a self-assessment tool are similar in both the computer-based and paper-based environments.

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 air crew 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.
APPENDIX A

CLIENT INFORMATION FORM AND SURVEY
Hello,

You are being asked to participate in a research study exploring the use of a new client assessment tool. Because you are a professional in the aviation community, you have a unique insight into our teaching methods, and your perceptions are greatly needed to complete this study.

We have partnered with an external researcher looking into a new tool called, Confidence-Based Assessments, which integrates the selection of multiple-choice answers with your self-perceived level of certainty, offering a middle ground between the traditional multiple-choice answer and a lengthy essay response. Research has discovered that confidence-based assessments provide a more comprehensive measure of a person’s knowledge, increases the retainability of learned material and identifies topics in which people are misinformed.

We are asking that you take your final ground school exam with the added Confidence-Based Assessment component. This new type of question will not interfere with, or influence, your score or your ability to pass the exam. Upon completion of the exam, you will receive a report identifying your specific strong and weak subject areas. Then sometime during the week of your simulator sessions, we would like you to share your thoughts and experience during a one-on-one interview that will take approximately 15 to 20 minutes.

Please be assured that your responses are completely anonymous and we will not publish your real name, nor will any outside organizations have access to your interview responses.

Your feedback will help us create a better training experience for you in the future.

Thank you for your time, and please let me know if you have any questions.

Paul Novacek
[contact info removed for privacy concerns]
Here is how the ground school exam will work.

At the completion of your ground school session, you will receive the FAA-approved multiple-choice exam much like any other course exam.

The exam packet contains:
- Instruction sheets
- Closed-book question sheets
- Open-book question sheets
- Answer sheet

Use the answer sheet to mark your individual answers to each question. This is the official answer sheet that will be placed into your record.

But to the right of each question on the question sheet, indicate your degree of confidence that your answer will be marked as correct using one of three (3) numbers.

3 = High confidence (I am sure)
2 = Medium confidence (I am partially sure)
1 = Low confidence (I am not sure – I am guessing)

For example, in the following question, after you mark the separate answer sheet with the letter you think is correct, draw the level of your confidence to the right of the question on the question sheet.

12. With the electromechanical gear system, what holds the landing gear in the retracted position?
   A. Engine driven hydraulic pump
   B. Hydraulic accumulator pressure
   C. Proximity switches
   D. Mechanical up-locks

If no confidence level score is marked, a medium confidence level is assigned.

On the top of each page a sticker reminds you of the confidence scoring scale.
It is to your best advantage that you be honest with your confidence level for each answer. There is no way to “beat” the test by answering all with high confidence. Your responses are there to help you study more efficiently and confidently by identifying your weak and strong areas.

After the exam is graded, you will then receive a report that breaks down each subject area by a combined knowledge and confidence score. Those areas where you either answered the questions incorrectly, or with low confidence are identified by yellow or red colors.

You can use this report as a study guide to review those weak areas and plan your remaining study time before the end of the course.

This report will also be available to your simulator instructor so those weak areas can be reviewed during the simulator briefing sessions.

During the week following your ground school exam, we would like to conduct an interview with you about your experience and thoughts about how we could better implement this new assessment technique. We will do our best to schedule the 20-minute interview right before one of your scheduled simulator sessions.

Once your training is complete, this Confidence-Based Assessment report will not be kept in your permanent records, but you are free to take home your copy of the report.

If you have any questions, please do not hesitate to ask your instructor or research coordinator.
Before the interview, we’d like to learn a little about who you are and your aviation experience. Please answer the following questions.

Are you male or female?

〇 Male
〇 Female

What is your age? __________

How many total years have you been employed in the aviation industry?

_______________________

How many total years have you been an active participant in aviation? (including building models as a child) In other words, how long have you had the “aviation bug”?

_______________________

How many full courses have you taken with this training academy?

_______________________

Approximately how many total flight hours do you have?

_______________________
Where did you receive your primary aviation training?

- Military
- Civilian (club, flight school)
- Academic (university)
- Other __________

In what country, or countries, did you grow up in?

________________________________________

In case we may need to call you for follow-up questions, please provide your name, phone number and e-mail.

Name: _______________________________________________________

Phone: _______________________________________________________

E-mail: _______________________________________________________  

In the final published report, we need to use a different name, or alias, than your real name. What name would you like us to use? (a common, single word, name is best)

Alias: ________________________________
Informed Consent Notice and Participation Form
University of North Texas Institutional Review Board

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

**Title of Study:** The Implementation of a Confidence-Based Assessment Tool within an Aviation Training Program

**Student Investigator:** Paul Novacek, University of North Texas Department of Learning Technologies.

**Supervising Investigator:** Dr. Lin Lin, University of North Texas Department of Learning Technologies.

**Purpose of the Study:** You are being asked to participate in a research study which involves gathering data about Confidence-Based Assessments and the operational feasibility of implementing a formative Confidence-Based Assessment program for the training of aviation flight crews and maintainers.

**Study Procedures:** You will be asked to answer a few questions during an interview, after you have taken an exam that includes a Confidence-Based Assessment component and received a report of the results. The interview will be recorded (audio only) and will take about 20-30 minutes of your time.

**Foreseeable Risks:** No foreseeable risks are anticipated in this study.

**Benefits:** We expect the study to benefit you by listening to your thoughts and opinions about how a Confidence-Based Assessment tool could be used in the classroom to enhance a client’s understanding of those aircraft subjects in which he/she may be deficient.

**Compensation for Participants:** None, just our heartfelt thanks…

**Procedures for Maintaining Confidentiality of Research Records:** Please be assured that we are not asking for any personal information. In addition, to secure your responses, we will ask that you create an alias for data tracking purposes. Your responses will be held in a database with high security and will be destroyed in a timely manner after we analyze the data. The confidentiality of your individual information will be maintained in any publications or presentations.
Questions about the Study: If you have any questions about the study, you may contact Paul Novacek at [contact info removed for privacy concerns], or Dr. Lin Lin, UNT Department of Learning Technologies, [contact info removed]

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants’ Rights:
Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Paul Novacek has explained the study to you and you have had an opportunity to contact him with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may obtain a copy of this form for your records.

Name: _______________________________________________________

Signature: __________________________________________________________________

Date: ______________
APPENDIX B

EXAM DIRECTIONS INSERT
Marking Confidence Levels on Exam

Use the answer sheet to mark your individual answers to each question.

To the right of each question on the question sheet, indicate your degree of confidence that your answer will be marked as correct using one of three (3) numbers.

3 = High confidence (I am sure)
2 = Medium confidence (I am partially sure)
1 = Low confidence (I am not sure – I am guessing)

For example, in the following question, after you mark the separate answer sheet with the letter you think is correct, draw the level of your confidence to the right of the question on the question sheet.

12. With the electromechanical gear system, what holds the landing gear in the retracted position?

A. Engine driven hydraulic pump
B. Hydraulic accumulator pressure
C. Proximity switches
D. Mechanical up-locks

If no confidence level score is marked, a medium confidence level is assigned.

It is to your best advantage that you be honest with your confidence level for each answer. There is no way to “beat” the test by answering all with high confidence. Your responses are there to help you study more efficiently and confidently by identifying your weak and strong areas.

Once you receive your grade report, please discuss your simulator schedule with the researcher to schedule a time for the one-on-one interview. We would like to conduct the interview 30 minutes before one of your simulator session.
APPENDIX C

PILOT INTERVIEW QUESTIONS
Instructions to the Pilot – Script

“Thank you for participating in this short interview. We are looking into your experience and opinions about confidence-based assessments and how they could be used during a course. Please be assured that we are not asking for any personal information, and once your responses are transcribed, your answers cannot be traced back to you. We are just interested in your anonymous responses. Please answer the following questions as honestly as you can and elaborate as much as you would like.”

Alias: __________________________________________

Familiarity with Confidence-Based Assessment

1. What, if any, prior experience do you have using Confidence-Based Assessments?
   a. Have you used Confidence-Based Assessments previously?
   b. Have you previously heard about similar learning tools?

Comfort Level of Revealing Own Confidence

2. Were you reluctant about revealing your confidence level for each of the exam questions?
   a. What was your comfort level in giving a confidence-level mark to each of the exam questions?
   b. How long did you hesitate after answering a question before you were able to decide on a confidence level mark to give that question? (just an average amount of time please)

Ease of Use

3. How easy or hard was it to understand the directions about how to mark your exam confidence levels?
   a. How much of a burden was the addition of these confidence-level scores to the exam?
   b. Do you feel that the added procedure of marking your confidence level for each answer extended the exam to the point that you became fatigued?

If high confidence average, then question:

4. I noticed that your confidence marks showed a tendency to be mostly 3s, or high confidence. Thinking back to when you were taking the test, please reflect on your thoughts as to your use of the three confidence levels.
Perception Change
5. After receiving the report of your confidence level and grade for each of the knowledge sections, did the scores surprise you or were they about what you expected?
   a. How confident do you now feel about your knowledge of the aircraft systems for this particular aircraft?

Study Habit Change
6. Did knowing your combined confidence level from the report for each of the knowledge sections provide useful information for your remaining study time before the simulator briefing sessions?
   a. How much did your study habits change knowing your confidence level for the various subjects?

Opinion About Benefits to Training
7. How beneficial do you think confidence-based assessments are during a [Flight Training Academy] course?
   a. If you gained a greater depth of knowledge about subjects you may have been weak on, was the extra time devoted to the exam worth it?
   b. Overall, how useful did you find the Confidence-Based Assessment experience during your course at the [Flight Training Academy]?
APPENDIX D

INSTRUCTOR INFORMATION FORM AND SURVEY
Hello,

You are being asked to participate in a research study exploring the use of a new client assessment tool. Because you are a professional instructor in the aviation community, you have a unique insight into our teaching methods, and your perceptions are greatly needed to complete this study.

We have partnered with an external researcher looking into a new tool called, Confidence-Based Assessments, which during the client ground school exam integrates the selection of multiple-choice answers with the self-perceived level of certainty. This technique offers a middle ground between the traditional multiple-choice answer and a lengthy essay response. Research has discovered that confidence-based assessments provide a more comprehensive measure of a person’s knowledge, increases the retainability of learned material and identifies topics in which people are misinformed.

We are asking that you participate in this research by administering the confidence-based assessment exam to your clients and share your opinions during a short interview at the conclusion of your participation.

Please be assured that your responses are completely anonymous and we will not publish your real name, nor will any outside organizations have access to your interview responses.

If you would like to participate, complete the rest of this form and return it to the researcher.

Remember, you are not under any obligation to participate in this study, nor will declining to participate have any effect on your job performance. But we sincerely hope that you agree to participate, as your experience will help us create a better training experience for clients in the future.

Thank you for your time, and please let me know if you have any questions.

Paul Novacek
[contact info removed for privacy concerns]
Before the interview, we’d like to learn a little about who you are and your aviation experience. Please answer the following questions.

1. Where did you receive your primary aviation training?
   - Military
   - Civilian (club, flight school)
   - Academic (university)
   - Other __________

2. What is your age? ____________

3. Approximately how many total flying hours do you have? _________________

4. How many total years have you been employed as a flight instructor? _________________

5. How many total years have you been an active participant in aviation? (including building models as a kid) In other words, how long have you had the “aviation bug”? _________________

6. How many courses have you taken to improve your craft as an instructor? _________________

In case we may need to call you for follow-up questions, please provide your name, phone number and e-mail.

   Name: _______________________________________________________

   Phone: _______________________________________________________

   E-mail: _______________________________________________________
In the final published report, we need to use a different name, or alias, than your real name. What name would you like us to use? (a common, single word, name is best)

Alias: ________________________________
Informed Consent Notice and Participation Form

University of North Texas Institutional Review Board

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: The Implementation of a Confidence-Based Assessment Tool within an Aviation Training Program

Student Investigator: Paul Novacek, University of North Texas Department of Learning Technologies.

Supervising Investigator: Dr. Lin Lin, University of North Texas Department of Learning Technologies.

Purpose of the Study: You are being asked to participate in a research study which involves gathering data about Confidence-Based Assessments and the operational feasibility of implementing a formative Confidence-Based Assessment program for the training of aviation flight crews and maintainers.

Study Procedures: You will be asked to answer a few questions during an interview, after you have experienced administering an exam that includes a Confidence-Based Assessment component and received a report of the results. The interview will be recorded (audio only) and will take about 20-30 minutes of your time.

Foreseeable Risks: No foreseeable risks are anticipated in this study.

Benefits: We expect the study to benefit you by listening to your thoughts and opinions about how a confidence-based assessment tool could be used in the classroom to enhance a client’s understanding of those aircraft subjects in which he/she may be deficient.

Compensation for Participants: None, just our heartfelt thanks…

Procedures for Maintaining Confidentiality of Research Records: Please be assured that we are not asking for any personal information for the final report. In addition, to secure your responses, we will ask that you create an alias for data tracking purposes. Your responses will be held in a database with high security and will be destroyed in a timely manner after we analyze
the data. The confidentiality of your individual information will be maintained in any publications or presentations.

Questions about the Study: If you have any questions about the study, you may contact Paul Novacek at [contact info removed for privacy concerns], or Dr. Lin Lin, UNT Department of Learning Technologies, [contact info removed]

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants’ Rights:
Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Paul Novacek has explained the study to you and you have had an opportunity to contact him with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may obtain a copy of this form for your records.

Name: _______________________________________________________

Signature: _____________________________________________________

Date: ______________
Instructions to the Instructor – Script

“Thank you for participating in this short interview. We are looking into your experience and opinions about confidence-based assessments and how they were used in the classroom and simulator sessions. Please be assured that we are not asking for any personal information, and once your responses are transcribed, your answers cannot be traced back to you. We are just interested in your anonymous responses. Please answer the following questions as honestly as you can and elaborate as much as you would like.”

Instructor Alias: ________________________________________

Familiarity with Confidence-Based Assessments
1. As an instructor, what prior experience do you have in Confidence-Based Assessments?
   a. As a student, what prior experience do you have in Confidence-Based Assessments?
   b. Have you previously heard about similar learning tools?

Ease of Use
2. How easy or difficult did you find the confidence-based assessment classroom [or briefing room] procedures?
   a. Did you find the instructions for giving the new testing procedures easy to follow?

Pilot Acceptance
3. Overall, how did you find the pilots accepted the addition of confidence-based assessments?
   a. On average per class, how many questions about the confidence-based assessment procedures did you receive from the pilots?
   b. How much reluctance from the pilots did you receive about the additional requirement (time or effort) to mark their level of confidence for each exam question?
   c. How much do you feel the addition of confidence-based assessments helped the pilots accept their own deficiencies?
Instructor Insight Into Pilots

4. Overall, did you gain a greater insight into the level of aircraft system knowledge of your pilots after receiving the grade report?
   a. To what extent did receiving the confidence-based assessment report change the class dynamic?
   b. To what extent did receiving the confidence-based assessment report change how you conducted the remainder of the class?
   c. How much feedback (good or bad) did you receive from the pilots that they changed the focus of their study habits after receiving their confidence-based assessment reports?
   d. Overall, was that feedback positive or negative?

Placement in Curriculum

5. During this experiment, the confidence-based assessment exam was given at the end of ground school. Is this placement in the schedule effective, or at what other point would you like to see an exam or quiz administered and the report given to the client?
   a. Do you believe the self-evaluation phase should be at a different point in the course?
   b. Was the current placement of the exam scheduled at a point that was too critical?
   c. Was the current placement of the exam a burden on the clients?

Improved Learning

6. After having now worked with confidence-based assessments, how comfortable do you feel with their use as a tool for learning improvement?
   a. How much do you think the addition of confidence-based assessments in the classroom improved the pilots’ knowledge level of aircraft systems?
   b. How much do you think the addition of confidence-based assessments in the classroom improved the pilots’ understanding of their own deficiencies?
   c. How much do you think the addition of confidence-based assessments in the classroom improved the pilots’ motivation to learn?

7. Do you have any additional observations or comments?
APPENDIX F

INSTITUTIONAL REVIEW BOARD APPROVAL LETTER
September 24, 2014

Supervising Investigator: Dr. Lin Lin  
Student Investigator: Paul Novacek  
Department of Learning Technologies  
University of North Texas

Re: Human Subjects Application No. 14374

Dear Dr. Lin:

As permitted by federal law and regulations governing the use of human subjects in research projects (45 CFR 46), the UNT Institutional Review Board has reviewed your proposed project titled “The Implementation of a Confidence-based Assessment Tool within an Aviation Training Program.” The risks inherent in this research are minimal, and the potential benefits to the subject outweigh those risks. The submitted protocol is hereby approved for the use of human subjects in this study. Federal Policy 45 CFR 46.109(e) stipulates that IRB approval is for one year only, September 24, 2014 to September 23, 2015.

Enclosed is the consent document with stamped IRB approval. Please copy and use this form only for your study subjects.

It is your responsibility according to U.S. Department of Health and Human Services regulations to submit annual and terminal progress reports to the IRB for this project. The IRB must also review this project prior to any modifications. If continuing review is not granted before September 23, 2015, IRB approval of this research expires on that date.

Please contact Sheila Bourns, Research Compliance Analyst at extension 2018 if you wish to make changes or need additional information.

Sincerely,

Chad R. Trulson, Ph.D.  
Professor  
Department of Criminal Justice  
Vice Chair, Institutional Review Board
Informed Consent Notice – Online Survey

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: The Implementation of a Confidence-Based Assessment Tool within an Aviation Training Program

Student Investigator: Paul Novacek, University of North Texas Department of Learning Technologies.

Supervising Investigator: Dr. Lin Lin, University of North Texas Department of Learning Technologies.

Purpose of the Study: You are being asked to participate in a research study which involves gathering data about Confidence-Based Assessments and the operational feasibility of implementing a formative Confidence-Based Assessment program for the training of aviation flight crews and maintainers.

Study Procedures: You will be asked to answer a few questions that will take about 20 to 30 minutes of your time.

Foreseeable Risks: No foreseeable risks are anticipated in this study.

Benefits: We expect the study to benefit you by listening to your thoughts and opinions about how a confidence-based assessment tool could be used in the classroom to enhance a client's understanding of those aircraft subjects in which he/she may be deficient.

Compensation for Participants: None, just our heartfelt thanks...

Procedures for Maintaining Confidentiality of Research Records: Please be assured that we are not asking for any personal information. In addition, to secure your anonymous responses, the online survey website does not record any information that can be traced back to you or your computer. Your responses will be held in a database with high security and will be destroyed in a timely manner after we analyze the data. The confidentiality of your individual information will be maintained in any publications or presentations regarding this study.

Questions about the Study: If you have any questions about the study, you may contact Paul Novacek at (866) 486-8733, e-mail: Paul.Novacek@FlightSafety.com, or Dr. Lin Lin, UNT Department of Learning Technologies, Lin.Lin@unt.edu, (940) 369-7572

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants’ Rights:
Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Paul Novacek has explained the study to you and you have had an opportunity to contact him with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may print a copy of this form for your records.
Informed Consent / Agreement Form – Interview

Before agreeing to participate in this research study, by signing your name, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: The Implementation of a Confidence-Based Assessment Tool within an Aviation Training Program

Student Investigator: Paul Novacek, University of North Texas Department of Learning Technologies.

Supervising Investigator: Dr. Lin Lin, University of North Texas Department of Learning Technologies.

Purpose of the Study: You are being asked to participate in a research study which involves gathering data about Confidence-Based Assessments and the operational feasibility of implementing a formative Confidence-Based Assessment program for the training of aviation flight crews and maintainers.

Study Procedures: You will be asked to answer a few questions during an interview, after you have experienced an exam that includes a Confidence-Based Assessment component and received a report of the results. The interview will be recorded (audio only) and will take about 20-30 minutes of your time.

Foreseeable Risks: No foreseeable risks are anticipated in this study.

Benefits: We expect the study to benefit you by listening to your opinions about the use of confidence-based assessments in the classroom to enhance a client's understanding of those aircraft subjects in which he/she may be deficient.

Compensation for Participants: None, just our heartfelt thanks...

Procedures for Maintaining Confidentiality of Research Records: Please be assured that we are not asking for any personal information. In addition, to secure your responses, we will ask that you create an alias for data tracking purposes. Your responses will be held in a database with high security and will be destroyed in a timely manner after we analyze the data. The confidentiality of your individual information will be maintained in any publications or presentations.

Questions about the Study: If you have any questions about the study, you may contact
Paul Novacek at (866) 486-8733, e-mail: Paul.Novacek@FlightSafety.com, or
Dr. Lin Lin, UNT Department of Learning Technologies, Lin.Lin@unt.edu, (940) 399-7572.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The IRB can be reached at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights:
Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Paul Novacek has explained the study to you and you have had an opportunity to contact him with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study participants may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you will receive a copy of this form for your records.

PARTICIPANT: ______________________ SIGNATURE: ______________________ DATE: ________

STUDENT INVESTIGATOR: ______________________ APPROVED BY THE UNT IRB: ______________________ DATE: 9/24/14
FROM 9/24/14 TO 9/23/15

1 of 1
REFERENCES


