FY2005 FITS Instructor Education Research:
Examining the Learner Centered Grading Approach

Principle Investigator:
Elizabeth Blickensderfer, Ph.D

Graduate Research Assistants:
Mark Bartosiewicz
Nidia Doherty
Kern Jennison
Justin Jennision

January 2007

The series of reports contained in this document were the deliverables under the FITS Program FY2005: FITS Instructor Education Research, "Examining the Debriefing Process Including the Learner Centered Grading Approach."

Project Overview

The purpose of this project was to investigate the "Learner Centered Grading" debriefing approach. The approach was multi-pronged and included both field and laboratory investigations. The results appeared in a series of project reports. The first report described current practices regarding use of the learner centered grading and debriefing in industry (Blickensderfer and Doherty, 2006--FY2005 FITS Instructor Education Research Report 1: Investigation of Current Practices Relating to Learner Centered Grading and Debriefing Procedures). The next report reviewed related research done in non-aviation domains (Blickensderfer, 2006--FY2005 FITS Instructor Education Research Report 2: Literature Review and Annotated Bibliography of Research Relating to the Learner Centered Grading Debriefing Approach). The third report provided recommendations for training flight instructors to use the learner centered grading debrief procedure (Blickensderfer, 2006-- FY2005 FITS Instructor Education Research Report 3: Recommendations for Instructor Training Protocol for the Learner Centered Grading Debriefing approach). The final report described an empirical evaluation of the Learner Centered Grading Debriefing approach (Blickensderfer and Jennison, 2006—FY2005 FITS Instructor Education Research Report 4: Empirical Investigation of the Learner Centered Grading Debriefing Approach). This document contains the four reports from this body of work.

Table of Contents

FY2005 FITS Instructor Education Research
Report 1: Investigation of Current Practices Relating to Learner Centered Grading and Debriefing Procedures

June 30, 2006
(Revised according to Government comments)

Submitted by:

Beth Blickensderfer, Ph.D.
Nidia Doherty, M.H.R.I.R.
Embry-Riddle Aeronautical University

This effort was the first deliverable under the FITS Program FY2005: FITS Instructor Education Research, "Examining the Debriefing Process Including the Learner Centered Grading Approach"

Task/Problem Statement

One key component of the FITS approach to aviation instruction is the notion of Learner Centered Grading (LCG) and the accompanying debriefing technique. In the LCG approach, following each instructional flight, the learner completes self-assessments of his or her performance. Next, the instructor incorporates these self-assessments into a debrief. Overall, the debrief is designed to differ from a traditional debriefing (usually a one-sided matter of instructors running through a checklist of significant errors the student made) by encouraging the learner to think actively about his/her performance and how to correct performance problems. The purpose of this report is to examine how the learner centered grading approach and debriefing strategy is being conducted in the field/industry and to identify any difficulties instructors are having in implementing this approach.

Method

Participants

Ten certified flight instructors for instrument aircraft were interviewed. The instructors represented three private companies (two aircraft manufacturers and one aviation training company). The individuals' total number of flight hours ranged from 1200 to 3800, and the average number of flight hours was 2200 (SD = 946). The individuals' experience as flight instructors ranged from three years to thirty years with an average of 8.61 years (SD = 8.5). Dual given hours ranged from 1002 to 3200 with an average of 1713.75 hours (SD = 786). All participants were certified flight instructors for instrument aircraft. In addition, eight of the individuals had multi-engine ratings, and two had airline transport pilot certificates. The experience in using the FITS approach ranged from having trained two to 300 students using the FITS approach with an average of having trained 127 students (SD = 112) using the FITS approach. Two of the instructors had also developed curriculum based on the FITS strategy.

Procedure

Interviewees were solicited from industry partners in the FITS program. Prospective interviewees were informed of the purpose of the interviews and given sample interview questions (see Appendix A). To promote the elicitation of honest, candid responses, all interviewees were ensured that their responses would remain anonymous. Interviews were conducted via the telephone. Each interview lasted approximately 1 hour. All except three of the interviews were conducted by a team of two interviewers. The remainder were conducted by an individual interviewer. Interviews were not audio-taped, and exact transcription was not possible. The interview responses have been consolidated and are presented in the next section.

Results

Please describe your debriefing process prior to FITS.

In general, interviewees reported that their prior debriefing processes were one sided and very much instructor led. They were short (no more than 15 minutes), and not standardized. For example:

- Instructor Pilot (IP): "This is what we need to work on and this is what is going well". Student would sit there and nod his head.

- The IP gave a “lecture” of what was right and wrong.
- The students tended to be exhausted (due to learning new things, stress, noise in aircraft) and not listening.
- The student and instructor had the mentality that the “Instructor knows and student doesn’t.”
- The IP informed students of good or not so good things they did with the goal of setting up for next lesson.

Some interviewees did report using a less one sided, more questioning approach. For example:
- The instructor began by asking students how they thought maneuvers went, however, most students already knew how the flight went. (This interviewee noted that he thought this questioning procedure was common sense.)
- Upon returning from the lesson, the instructor would ask the student how s/he thought things went, and would then give his own view.

Other comments included:
- The debrief occurred while taxiing in.
- The IP used a guide, syllabus, or took notes to help keep on track during the debrief.
- While instructing at universities, IPs were frequently on a tight time schedule with very limited time for the debrief.
- The IP brought the student into the office and sat down together for the debrief.
- On the occasions when their students had the opportunity to self-assess, the younger ones thought they were better than their actual skill level.
- The debriefing process prior to FITS was usually done by just randomly discussing different parts of the lesson and trying to remember certain issues with a flight. It wasn’t uncommon to forget something we (the instructors) were supposed to cover and think of it after the lesson was over and the student had left.

With the FITS training now implemented, is there a debriefing component used currently? Please describe.

The first characteristic that some interviewees noted was a longer amount of time allotted to the LCG debrief.
- A little longer, 20-30 minutes.
- The main change is time: 10-35 minutes.
- At least 5 minutes spent explaining the LCG scale.
- The time runs 20 to 30 minutes depending how the lesson went overall, but at a minimum of about 15 minutes.

Another characteristic noted was the level of detail:
- Similar format (to pre-FITS debrief), but more in-depth.
- More specific--it goes segment by segment. The debriefing is broken down, step by step.
- Discussed everything in the lesson in order; and whether or not performance was up to standards.

- Before LCG items that needed to be discussed during the debriefing were often overlooked or forgotten. Now there is a method of debriefing that actively engages the pilot in training and the instructor pilot to evaluate the flight or lesson.

Many interviewees emphasized the less one-sided nature of the conversation, with the learners assessing their own performance and answering probing questions asked by the IP. For example:

- The student gives input before the IP does, and the student is often stricter on themselves than is the IP.
- "I noticed that you thought you were on the ___ level." Student then explains their self-assessment.
- The student is usually good at self-grading, but may be harder on themselves and/or may leave some blank.
- The instructor asks the student how s/he felt things went.

Differences did appear in the exact protocol for the student self-assessment. That is, sometimes students filled out the form by themselves, and in other cases the IP and the student did this together. For example:

- The students evaluate themselves (alone) using the LCG form, then the Instructor makes his/her evaluation, then the two compare the evaluations and discuss the differences.
- The customer grades initially, IP grades separately (not in the room with customer)
- The instructor sits there while student rates him/herself. When s/he gets to a question point, the Instructor and the student discuss it.
- Sitting down with the customer, the instructor says, "Look at the first one, what does it say?" Then they read it together, and the student makes his/her rating. Then the instructor gives a rating, explains why, and gives a specific example.
- The instructor gives the grading sheet to the student and tries to leave them alone.

Issues with the LCG scales:

- The students don't always understand the scale, and the instructor must give examples.
- The debriefing begins by explaining the syllabus and scale to students in lay-man's terms.
- At the start of each class, the Instructor explains the LCG scale; for example, Manage/Decide is an "A", then Describe is an "F."
- The LCG process is a "rigid grading process."

What is your understanding of Learner-Centered Grading? What do you think each level means?

The purpose of this question was to determine how much consensus exists regarding the current LCG scale. Results of this question are presented in the form of the original interview notes in Appendix B.

Do you think that the implementation of Leaner-Centered Grading is difficult? Why or why not?

One theme that appeared in response to this question was the newness of the debriefing style for the trainees:

- Most trainees don't know enough to answer questions, because they are learning new concepts.
- The trainees need some guidance at first, and they gradually become more willing to ask questions as they go and become more receptive to the process.
- LCG is not difficult to implement with a bright, educated, motivated audience.
- The customers do not always see the value.

Other responses had to do with the newness of the method for the instructors:

- It's difficult to get instructors to do it, because it is not way the way they were taught; instructors must overcome old habits.
- It can be difficult for instructors to "convert" to "hands-off", but students will follow the instructor's lead.
- It is difficult at first, because anything is difficult to adopt when you are used to a different practice. But once people have used it, most feel that it is very effective.

Other comments included:

- In many training environments severe time constraints exist and lengthy debriefs are not currently feasible.
- The LCG form is more complicated than it needs to be. It takes several minutes to explain and the IP needs to give examples. Even other FITS practitioners disagree on what the levels mean (this observation was made at a FITS instructor workshop).
- Overall, the LCG approach is not too difficult, although the "describe" and "explain" categories get confused, and students get confused between "practice" and "performance".
- Asking someone to critique themselves requires a diplomatic approach.

How do students/customers react to the LCG-oriented debriefing process?

- The customers get better and more willing to ask questions as training progresses. With customers who are reluctant to self-assess, the IP has to continuously encourage, e.g. "How do you think it went?" "What areas are still confusing?"
- One instructor divided students up: 25% - Why is it this complicated? Why is it worded this way?; 80% - Most like grading themselves, take it seriously for the most part. 20% - Some just take it because it is paid for in cost of the plane.
- The students are usually on target if the form is explained correctly to them; do seem to know their limitations.
- The students like it and understand it, but IP needs to set them up to do it correctly
- 50% really like it, feel more involved; 50% want traditional format probably to get done faster.
- It is hard to get students engaged; need sufficient time to explain LCG to get students engaged; requires more time to do the debrief; generally pretty favorable; some students are shocked that they get to grade themselves.

Do all learners (customers) perform equally well on the LCG-oriented debrief?

- Everyone understands it with help; a pre-brief is needed to introduce self assessment; some don’t care about the grading process.
- There are differences in attitudes about the grading process and motivation; some don’t respond well with filling out the form on their own.
- If the learner is not honest with him/herself, that is a problem; but 60%-70% understand it and like to use it.
- 50% have a good understanding of ability; 25% grade too high; 25% grade too low.
- Yes, for the most part; most grade lower than where IP thinks they are. Insecurity, self confidence can be an issue, although most take it seriously.
- LCG works as long as you explain the meaning of each level.
- 60%-80% are pretty accurate with self- assessment. Some have difficulty understanding what levels are. A small percentage really have trouble and need to walk through with multiple examples.

How much time do you think the debrief takes, on average?

Each interviewee’s response is listed below.

- 20-25 minutes
- 15-20 minutes
- 15-35 minutes
- Flight 1: 20-30 minutes, Flight 2: 15-20 minutes, Flight 3: 15-20 minutes
- 45 minutes
- 30-45 minutes
- 30 minutes
- 30 minutes
- 25-30 minutes
- 15-30 minutes

What changes to the debrief have you noticed as the FITS training evolves in your organization?

- We revised the grading form to remove Describe/Explain.
- Revised to be activity-oriented (not FITS approved, but each activity has own specific form).
- A gradual acceptance of the philosophy occurred; intentionally working on letting the learner go first.
- When first using FITS, I still used the traditional debrief. Now I put much more effort into using the new style debrief.
- Haven’t changed anything on debrief since instructor training.
- After “checked out” to use FITS, the IP was never contacted; they never had an instructor sit in on any of their debriefs.
- Some instructors in my organization were giving the debriefing form to the customer to take home, but this was corrected.
- Form is being modified.
- Instructors are identifying key words to help convey to the customers what the LCG form means and why we are using the self-assessment process.
- Early on felt this approach was “cheesy.”

- Instructors need more information on the reasoning for LCG; Instructors also need examples and scenarios of how to explain it.
- This method shows the need for continual IP self assessment regarding their own quality.
- I have attended multiple FITS conferences, and I deliberately make changes based on those meetings. I have a better understanding overall and have developed additional questions to get students thinking.
- Staff has come up the learning curve.
- We reduced the number of assessment items to focus on the most important items.
- At first, I did not use the LCG sheet. Now I have student self-assess using the sheet.

How would you rate the effectiveness of the LCG debrief?
- About the same, perhaps a bit better; the additional communication helps the instructor assess what the trainee does and does not know.
- One IP rated it a “9 out of 10” with the comment that the form is too vague; Always room for improvement.
- Another IP rated it a “6-7 out of 10”. It would be a 10 if the students already knew how to do the debrief (i.e., if primary instruction was also done using the FITS strategy). In other words, the debrief would be more effective if the students had more practice in the technique itself. Also, students need an attitude change to be honest.
- “7/8 out of 10”; The discussion format is good in this setting, not sure if it works with novices; Likes customer input; believes their inputs give the instructors real insight into the student skill level; giving feedback can be difficult when self assessments are off.
- 8-9/10.
- 9/10, there is always room for improvement.
- Received more positive feedback from customers.
- Too many assessment items - get rid of repetitive questions and fluff (25%).
- Only as effective as application, time, and degree that student is engaged. If you take the time and the student engaged this approach can be a 10. If the instructor was exhausted and the student was exhausted, it will have low effectiveness. Traditional debrief can also make 10 if instructor is good.
- Very effective; students get much more out of it than they used to.
- 7; Overall, pretty effective. In a traditional debrief, it is easier to pass over on routine items that you still spend a lot of time on in a FITS debrief. (Not as efficient if you have to spend time on everything).

If you were to design a new debriefing procedure, what would you be sure to include?
- Levels would be more user-friendly for learners, and possibly include examples--especially towards a rating; Reuse the form for advance skill training; Make the scale easier for learners to understand. Describe/Explain are not applicable to all training.
- Training syllabus is vague; give examples for different levels of understanding.
- Keep LCG form – it’s a good tool; Change outcomes to something more common sense.
- Keep candidness, customer/student input.

- Help IP get insight into “Perform” level.
- It can be a tough situation if the student rates themself higher—need a delicate approach to criticism.
- Need to clarify categories (e.g., Describe, Perform) with examples
- Fundamentally no change. Cosmetically, individual assessment items need to be consolidated and prioritized (Ex: There are 6 assessment items currently for a climb out - change perceptions of number of boxes).
- Keep the LCG sheet (keep on track); do not eliminate anything; IP hopes that all flight instruction moves to FITS (even the way PTS are written).
- Descriptors of LCG outcomes could use better explanation; they are not intuitive. Concrete examples would help.
- Redesign task list to be specific to lesson.
- Have the pilot in training evaluate themselves every time.

Conclusion

Overall, the instructor pilot response to the LCG debrief is quite favorable. Three challenges emerged, however: 1) the need to instruct the learner to become an active participant in the debrief, 2) the difficulty in conveying the meaning of the current LCG scale to the learner, and 3) the need to give the instructors additional guidance on transitioning from the old style debrief to the new style.

Report 1: Appendix A

Dear FITS colleague,

We are researchers in the Human Factors Department at Embry-Riddle Aeronautical University and we are investigating the debriefing procedure used as part of the FITS training protocol. Through research and experimentation, we aim to design a debriefing training that will engage the pilot in learner-centered techniques that allow them to optimize their learning environment and continue self-improvement beyond the instructor-led training session.

We need to hear your real-world perspective on debriefing procedures. Your comments and insight will help us to design a debriefing procedure that is meaningful and applicable to the flight training environment. We are currently conducting phone interviews on this topic and are looking for participants. Specific results from the interviews will be kept anonymous. Thank you for your assistance with this research.

Example interview questions:
Please describe your debriefing process prior to FITS.
With the FITS training now implemented, is there a debriefing component used currently? Please describe.
What changes to the debrief have you noticed between the former training and the new?
What changes to the debrief have you noticed as the FITS training evolves in your organization?
How much time do you think the debrief takes, on average?
Do you think that the debriefing procedures associated with this training are standardized among all instructors? Among every instance of the training, even with the same instructor?
How would you rate the effectiveness of the debrief?
If you were to design a new debriefing procedure, what would you be sure to include?
What is your understanding of Learner-Centered Grading?
Do you think the implementation of Learner-Centered Grading is difficult? Why or why not?
What questions do you have regarding the FITS debriefing process?

Beth Blickensderfer
Assistant Professor
Human Factors & Systems
Embry-Riddle Aeronautical University
357 Lehman Building
600 S. Clyde Morris Blvd
Daytona Beach, FL 32114
Phone: (386) 323-8065
Fax: (386) 226-7050
E-mail: blick488@erau.edu

Nidia Doherty
Research Assistant
Human Factors & Systems
Embry-Riddle Aeronautical University
343 Lehman Building
600 S. Clyde Morris Blvd
Daytona Beach, FL 32114
Phone: (386) 226-7744
Fax: (386) 226-7050
E-mail: nidia.doherty@erau.edu

Report 1: Appendix B

Each interviewee's responses to the question: "What is your understanding of Learner-Centered Grading? What do you think each level means?"

Instructor Pilot 1 (IP1)

- IP assumes that ground school gets through Describe/Explain
- Practice – still need IP to help complete task
- Performance – completion of maneuver was never in doubt, student can complete the activity
- Manage/Decide – complete understanding of activity leg
- Landing aircraft, different environments, understand screens in GC used
- Uses example, but not consistently; specific to activity
- Feels grey between Describe/Explain
- May mess it up but will self correct and explain

IP2

No response recorded

IP3

- Never had Describe/Explain form
- Autopilot climb
    - Describe (thinks vague) – low level of understanding; IP provides massive instruction; customer couldn't complete
    - Explain (thinks vague) – could give information about it, but couldn't perform
    - Practice – closer, but IP provides assistance for completion
    - Perform – on own, making mistakes; IP points out and customers makes corrections
    - M/D – customer does on own; IP doesn't provide any input

IP4

- Explain – you understand and can explain underlying concepts and procedures
- New form eliminated Describe
- Explain: understand, explain underlying concepts
- Practice: helping them, can't fix themselves
- Performance: practical test standard – check ride
    - Can mess up but have to recognize and fix themselves
    - Outcome never in doubt
- Manage/Decide – (SPRM) make decision for <u>safe</u> outcome of flight
- good choices
    - ex: rate of climb or ground radar

IP5

- Describe/Explain- covered in ground school
- Practice- with some input they are able to get tasks done. Grading at first three proper process, then make corrections to process
- Perform- without input, do what needs to be done. Can be different from IP, but outcome is successful

- Manage/Decide- on tasks where there are lots of options. Weed out unacceptable, pick acceptable answer.

IP6

- Uses example for every student- e.g. takeoff
- Describe- physical numbers regurgitate from memory
- Explain- what VX means, affect take off like flap settings
- Practice- practice with input from instructor
- Performance- normal takeoff without instructor
- Manage/Decide- having a concept of the big picture, e.g. how are you going to perform a takeoff in face of an obstacle?

IP7

- Explain
  - you can regurgitate and can explain under-flying principles
  - can verbally identify what is going on
  - can determine he has a problem
  - can explain what he is going to do
  - can explain inherent risks with alternate failure
- Practice
  - you practice it… instructor gives guidance on how to correct as necessary (how to correct altitude deviation on a steep turn)
  - can practice that activity
  - example: alternator 1 failure (shut it off- simulator condition)
  - has to respond in real-time
  - instructor corrects and guides as needed
- Perform- steep turn- individual will be able to perform without instructor; can identify and correct any deviation at skill level appropriate to whatever rating he holds
- Manage/Decide- pilot gathers all info available to him
- Identify possible courses of action (go around, altitude)
- Make decision and respond in real-time

IP8

- Ground school
  - Perceive- explain and describe
  - Understand- practice and perform
  - Correlate- manage and decide
- Draw figure
- Explain- Function on an intellectual level. Can talk about situation where it may occur and where it interfaces with aircraft. Not ready to go and do it
- Practice- going and doing it (may be able to do it at 3000 ft, 50% of time)
- Perform- choir recital/ play this is the performance to PTS standards, can pass check ride (able to replicate closer to 100% pass check ride)
- Manage/Decide- I can give you the keys and be fully confident that their knowledge and understanding will not get them into trouble

IP9

- Describe- describe what it is only (e.g. what primary flight display is); use a chart- visible

- Explain- explain how it works, better understanding but can’t apply
- Practice- use it, but not 100% confident. Need to continue to practice (could move on, but need to revisit it)
- Perform- Is on their own without assistance- meet or exceed PTS
- Manage/Decide- master of concept- all perfect, no question they understand, even if something goes wrong. Instructor does not expect this for most tasks

IP10

- Describe- student should recognize task being performed, what information they are looking for
- Explain- also where one might find information, he might do it
- Practice- implement explanation, try to work through process
- Perform- successfully accomplish the task,
- Manage/Decide- perform, understanding elements involved, how they relate to other tasks, building a framework of order of events; correlate level of learning. See tasks as a whole. See bigger picture of flight. Planning ahead.

FY2005 FITS Instructor Education Research
Report 2: Literature Review and Annotated Bibliography of Research Relating to the Learner Centered Grading Debriefing Approach

August 22, 2006
(Revised according to government comments)

Submitted by:

Beth Blickensderfer, Ph.D.
Embry-Riddle Aeronautical University

This effort was the second deliverable under the FITS Program FY2005: FITS Instructor Education Research, "Examining the Debriefing Process Including the Learner Centered Grading Approach"

Review of Research Relating to the Learner Centered Grading Debriefing Approach

Background

Although aviation has moved into a technologically advanced era, flight training has largely stayed the same. Emphasis on stick-and-rudder skills and repetition of standard flight maneuvers are most likely not the most effective strategy way to train pilots to fly the new technically advanced aircraft (TAA) (French, Blickensderfer, Summers, Ayers, & Connolly, 2005).

In response to this need, the FAA/ Industry Training Standards program (FITS) was born. FITS is not a regulatory entity, but rather it is a voluntary collaboration of industry leaders representing aircraft manufacturing, training, universities, insurance companies and trade associations, as well as the General Aviation Center of Excellence and the United States Federal Aviation Administration (FAA, n.d., b). The FITS team has been working together to develop training guidelines that fall within the boundaries of current regulations and yet incorporate the latest findings in training research.

Briefly, the FITS approach utilizes a scenario-based training strategy, wherein flight instruction is given in the form of realistic scenarios. One component of the FITS method, following each flight scenario, is to use the concept of "learner-centered grading." Learner centered grading includes two parts: learner self assessment and a detailed debrief by the instructor. The purpose of the self assessment is to stimulate growth in the learner's thought processes and, in turn, behaviors. The self-assessment is followed by an in-depth discussion between the instructor and the student which compares the instructor ratings to the student's self-assessment. Previous laboratory research (French et al., 2005) indicated that the FITS approach, including the notion of learner centered grading, is effective for training certain skills.

Implementing a training strategy in the field, however, is different from laboratory tests. Thus, in an effort to see how the learner centered grading debrief was being used in the field, Blickensderfer and Doherty (2006) interviewed certified flight instructors (CFIs) on this matter. Overall, the CFIs reported that the in-depth discussions were an improvement over traditional instructor-led debriefings (Blickensderfer & Doherty, 2006). Difficulties, however, were also evident, and a variety of questions were raised including how to use the provided scale, accuracy of the self-assessments, and the type of student who would most benefit from this approach.

In an effort to offer recommendations for instructor training regarding use of the learner centered-grading debriefing strategy, this report presents related research from non-aviation domains. This review begins with findings in the literature regarding metacognition and self-assessment, and concludes with a review of a structured debriefing strategy used recently in team research.

Metacognition

Metacognition is defined as the ability to reflect upon, understand, and control one's learning (Schraw & Dennison, 1994). Metacognitive activities that help control one's thinking or learning is what is referred to as regulation of cognition, and usually includes three essential skills: planning (strategy and resource allocation), monitoring (awareness), and evaluation (appraising the products of learning) (Schraw & Moshman, 1995). The literature on metacognition is vast and a full review is beyond the scope of this paper. Interested researchers are invited to refer to the Annotated Bibliography in this paper to review a portion of the available research. Some areas of the metacognition research, however, are of particular importance to the topic of LCG and will be reviewed. These include the following findings: a) self-assessment tends to be inaccurate and, thus, using a self-assessment scale may not be enough of an intervention to affect performance, b) generally, people do not naturally develop good metacognitive strategies spontaneously, and c) assessing metacognitive strategies for the purpose of learning new information and skills has been shown to positively affect performance on either tests of that knowledge or an increase in the effectiveness or efficiency of that skill.

Self-assessments tend to be inaccurate.

First, the social psychology literature provides evidence of inaccuracies in self-assessments: Novices in a domain tend to overestimate their performance largely and experts slightly underestimate their knowledge. This finding is consistent over tasks and over time (Maki, Shields, Wheeler, & Zacchilli, 2005). In a study done by Maki et al (2005), volunteer general psychology students were divided into three ability groups based on their SAT/ACT verbal scores. After reading six texts in random order, each student was asked to indicate the percentage of the text (0%, 20%, 40%, 60%, 80%, 100%) that s/he believed s/he understood. Next, each student indicated how many of six test questions that s/he believed s/he would answer correctly, and then s/he completed those test questions. Subsequently, each participant indicated how many test questions (0-6) that s/he thought s/he answered correctly. Results of this study showed that low-ability students were overconfident, especially in their predictions of how well they would perform on the tests. High ability students were under-confident, especially in their post-test confidence judgments. In other words, accuracy did not predict performance.

Next, consider a field study done by Nietfeld, Cao, & Osborne (2005). In this work, a psychology class gave themselves self-directed feedback. This occurred via a 100-pt self scoring scale. Each student marked the scale after each test item as well as overall for each test during a semester course. No formal training in metacognitive strategies was given. The results showed that self-monitoring accuracy remained relatively stable throughout the course, indicating that merely prompting students to think about their performance is too passive of a manipulation, even when applied over the course of several exams during a semester. The research team felt that more likely, an explicit training module with practice, feedback, and strategies would be needed to effect change in the accuracy of students' self-monitoring.

In addition to a lack of self-monitoring skills, judgments may also be affected by test difficulty. In a study by Nietfield, Cao & Osborne (2005), it was found that judgment bias increased as the test difficulty increased. Students tended to be less confident on easy items and overconfident on difficult items. Furthermore, poor students generally did not predict their performance as well as better students predicted their respective performances. It may be that poor students understand that they don't know very much, but don't know how to improve, whereas better students are aware of where they need to apply themselves to improve.

In these two studies as well as others (see annotated bibliography), self-assessments are notoriously inaccurate. Monitoring accuracy may be affected by various factors, including task difficulty, age, comprehension instruction, background knowledge, performance level, level of detail of learned information, amount of expertise in the task domain, incentive to self-monitor, and inclusion of prompting questions and feedback as part of the task. Thus, FITS instructors should be aware that self-assessments tend to be highly inaccurate, and the act of self-assessing alone (on any given scale) will not necessarily foster learning.

People are not born with a self-monitoring skill, but this skill can develop over time.

Research indicates that self-monitoring ability develops slowly. Unsurprisingly, it is quite poor in children. By college, most students have metacognitive knowledge about their learning, but often do not choose to use this knowledge to increase their performance. Even skilled adult learners can display poor monitoring under certain conditions (Schraw, Dunkle, & Bendixen, 1995).

Metacognitive strategies do not appear to be spontaneously generated, but rather evolve through developing domain-specific monitoring experience or through training (Schraw, 1997). Monitoring accuracy may be increased by introducing strategy instruction prior to testing, and it also appears to improve with training and practice (Nietfeld, Cao, & Osborne, 2005; Schraw & Moshman, 1995). For instance, Delclos and Harrington (as referenced in Nietfeld & Schraw, 2002) found that a group trained in both monitoring and problem solving performed better than groups trained in only problem-solving or not given any training at all.

Thus, a student pilot may seem to use poor self-monitoring methods. Through practice in self-monitoring and feedback from/discussions with the instructor, however, it is likely that these skills will improve.

Metacognitive awareness seems to help performance.

It is generally thought that those who self-regulate their own learning perform better and are better able to take advantage of what they have learned (De Carvalho, Moises, & Yuzawa, 2001). In fact, regulation of metacognition may be more important to performance outcomes than domain knowledge. For example, Schraw & Dennison (1994) found that differences in metacognitive awareness were more related to differences in performance than intellectual aptitude. It is argued that metacognitively aware learners are more strategic and perform better than unaware learners because they are able to plan, sequence, and monitor their learning in a way that directly improves performance.

One strategy used in the research to access metacognition is summarization. Summarizing has been found to help build relationships among concepts within a text as well as linking concepts to prior knowledge. This may occur via focusing attention, but it may also be because summarization promotes self-testing. As the person summarizes, they judge their success in retrieving information and incorporating it into the summary (Theide & Anderson, 2003). It is likely that the FITS LCG debriefing strategy elicits similar propensities for self-testing due to its synthesizing nature.

Additionally, Volet (1991) used a strategy that combined teaching task skills and teaching metacognitive learning strategies. This is somewhat similar to the FITS LCG debriefing approach, but the research occurred in the domain of computer programming rather than flight training. Volet (1991) showed that coaching relevant metacognitive strategies in conjunction with the learning of introductory computer science material resulted in improved short and long term learning. The intervention took place over 13 weeks, and consisted of a one hour per week tutoring session. Instructional manipulation included a 5-step metacognitive strategy relevant for computer programming, modeling and coaching instructional techniques, and a social support network (Volet, 1991).

The results of Volet's study (1991) showed that more students who received the 5 step strategy passed the computing course and also obtained significantly higher grades than did students who did not receive the metacognitive coaching. In terms of measures of learning, while control and the experimental groups did not differ in their course knowledge, experimental students' demonstrated greater ability to apply their knowledge to solve a novel problem. This indicated that those students' computer science knowledge was more accessible and generalizable to novel situations. In addition, long term effects also appeared, as more students who had been in the experimental condition passed a subsequent advanced course, and they performed better overall than students who had been in the control group in the subsequent course.

In summary, although the research is limited, students in non-aviation domains have demonstrated considerable performance gains when taught metacognitive strategies. It is likely that if this type of a skill synthesis and metacognitive strategy is incorporated into the FITS LCG debriefings, pilot performance will improve in the flight domain as well.

Part II: The Team-Training Paradigm

One environment where a facilitative debriefing process similar to that used in the FITS approach has been investigated is the team training domain. Briefly, in efforts to improve performance in team tasks occurring in dynamic and potentially high-stress environments where stress can have a debilitating effect on decision making, researchers working with the U.S. Navy developed a facilitative debriefing approach (Tannenbaum, Smith-Jentsch, & Behson, 1998). The facilitative debriefing approach (i.e., "team self-correction" and "team dimensional training") was used for teaching teamwork skills (e.g., communication, supporting behavior, information exchange, and leadership). While these teamwork skills are not all directly applicable to general aviation, it is likely that the facilitative debrief strategy would easily generalize to the training of many other types of skills, such as those needed in general aviation.

Team Self-Correction and the TDT Model

Team Dimensional Training (TDT) is a strategy for enhancing teams' ability to self-correct by incorporating guided team self-correction to develop teamwork-related knowledge and skills (Smith-Jentsch, Zeisig, Acton, & McPherson, 1998). Use of guided team self-correction process refers to using a facilitator who "(a) keeps the team's discussion focused, (b) establishes a positive climate, (c) encourages and reinforces active participation, (d) models effective feedback skills, and (e) coaches team members in stating their feedback in a constructive manner." Helping the team determine what and how they should discuss certain topics is the main focus of the facilitator (Smith-Jentsch, Zeisig, Acton, & McPherson, 1998).

Team self-correction has a number of key elements, including: what content the team should discuss, and how the discussion should proceed. The content should contain "explanations of and expectations concerning: mission objectives and goals, roles and responsibilities of team members, knowledge of behavioral sequences, response patterns, interaction patterns and procedures, and scripts and frameworks for performing the task" (Blickensderfer, Cannon-Bowers, & Salas, 1997).

The Impact of Self-Correction on Performance.

Multiple studies have indicated positive results from using the team self-correction method. First, in a laboratory study using a low-fidelity simulated radar task, Blickensderfer et al. (1997) indicated that teams who received the team self-correction training exhibited self-correction behavior significantly more than the control teams. Furthermore, the teams who participated in the self-correction debriefs had more shared knowledge and more efficient communications than did the control teams.

In addition, studies with Naval submarine teams have also shown positive results (Smith-Jentsch et al., 1998). In terms of performance enhancement, one case study showed that a TDT trained team accurately diagnosed teamwork-related problems, demonstrated immediate improvement on targeted goals, and generalized lessons learned to new exercises. Other experimental teams showed that, while performance was equal to control groups for an initial, easy exercise, only teams exposed to TDT training were able to generalize the lessons learned to novel situations and continued to exhibit superior performance while the control groups' performance declined with increasingly difficult tasks.

In terms of how the debrief style can change from traditional to "self-correction" based on training, Smith-Jentsch et al. (2001) reported that after receiving training in the TDT method of team self-correction, a variety of differences appeared. First, the pre- and debriefs were somewhat longer (increasing from 5 to 7.5 minutes and from 16 to about 32 minutes for the pre-brief and the debrief, respectively). Also, there was a reduction in team leader evaluation (from 83% in the control condition to 7% in the training condition), and a reduction in the leader sharing expertise (from 73% to 41%). The training intervention also led to a greater proportion of team members contributing to the self-correction process. In terms of goal setting, the control group was less likely to write team goal statements that were mastery-oriented than were those in the experimental condition. Based on all the results, teams within the training intervention condition could better critique their own processes (e.g. diagnose and solve teamwork problems) and generalize the lessons learned to future performances (Smith-Jentsch et al, 2001).

Additional Information on Pre-briefing and Debriefing

As noted previously, the TDT debrief is designed to enhance teamwork knowledge and performance by alerting teams to teamwork goals, soliciting feedback from team members regarding teamwork goals and performance, managing the flow of the discussion, and helping the team diagnose strengths and weaknesses and goals for improvement. This is done through a pre-brief and a de-brief.

The pre-brief. The first opportunity for feedback comes in the pre-brief (Smith-Jentsch, Milanovich, & Merket, 2001). As described in Smith-Jentsch et al. (2001), by using pre-briefings, information such as role expectations and task strategies can be emphasized as a means to improve future performance. The pre-brief may also be used to align leader and teammates' mental models, which should also lead to superior performance outcomes. For maximal learning, it is recommended that, during a prebrief, instructors: 1) remind members of the performance objectives, 2) explain how the objectives and structure will pertain to the upcoming tasks, 3) set goals, as well as 4) establish a positive learning climate (Smith-Jentsch et al, 1998). This preparation should increase learning and enhance future performance.

The debrief. The primary source of intervention is through the debriefing process. The debriefing can be used to engage in self-correction that involves role clarification, self-evaluation of task strategies, planning, and goal setting (Smith-Jentsch et al, 2001). To this end, how feedback is approached, and when it is given is of tantamount importance to the learning process. In order for feedback to be useful for both informational and motivational purposes, it should be designed systematically. For example, the facilitator should avoid lecturing the learner, and should withhold their observations and opinions of the exercise until the learner has given their opinion. The use of closed-ended questions may stymie the usefulness of the feedback process as well, as they encourage one-word/yes/no types of answers that do not elicit opinions of performance or suggestions for improvement. It is more effective to use open-ended questions that probe the learner to assess their own performance. Allotting enough time for the feedback is also important. Debriefs that are rushed often turn into one-way "lectures" due to time constraints (Tannenbaum et al, 1998).

Thus, to improve the effectiveness of the debrief, the facilitator should show that they are genuinely interested in the observations of the team by pausing and making eye contact with all team members after asking a question. They should ask specific, open-ended, and behavioral questions before offering their own critique. Asking for both positive and negative examples of behaviors is important because generalization of training can come from both effective and ineffective examples of behavior. Also, if no one in the team offers a good solution to a team problem, or if the team cannot agree to a reasonable solution, the facilitator should offer a reasonable explanation for why the suggested solution is not viable, and recommend a solution before moving on to the next topic (Smith-Jentsch et al, 1998).

Summary of Team Debriefing Research

A number of important points appear in the team debriefing research. First, the facilitative debrief / team self-correction is effective—teams who used this method performed better than those who did not use it. It is vital to note, however, that teams did not self-correct "naturally" nor did team leaders "naturally" lead teams in facilitative debriefs. In other words, all of the reviewed research included *training* the team leaders or instructors (the Smith-Jentsch body of work) or the team members themselves (Blickensderfer et al., 1997) before the positive results were demonstrated across many teams. Second, much of this research advocated the use of a guide for instructors to follow as they led the learners through a debrief (e.g., Tannenbaum et al., 1998; Smith-Jentsch et al. 2001). The guides included what to talk about as well as how to ask probing questions. In the case of the FITS LCG debrief, it is highly likely that additional instructor training as well as the development of such a guide would assist flight instructors transition to using the new style debrief most effectively.

Conclusions

Considerable information is available in current research to provide flight instructors using the FITS LCG approach guidance about the method. This includes research on self-assessment as well as research demonstrating the efficacy of probing debriefing strategies and, finally, guidance on how to conduct this style of debrief. It is recommended that this information be utilized in designing instructor training regarding use of the FITS LCG strategy.

Report 2: References

Blickensderfer, E L., Cannon-Bowers, & J.A., Salas, E. (1997). Training Teams to Self-Correct: An Empirical Investigation.

Blickensderfer, E L., & Doherty, N. (2006). *FY 2005 FITS Instructor Education Research Report 1: Investigation of Current Practices Relating to Learner Centered Grading and Debriefing Procedures*. Florida, USA: Embry-Riddle Aeronautical University, Human Factors and Systems Department.

Ciccarelli, S.K. & Meyer, G.E. (2006). *Psychology*. Upper Saddle River, NJ: Pearson Prentice Hall.

De Carvalho, F., Moises, K., & Yuzawa, M. (2001). The Effects of Social Cues on Confidence Judgements Mediated by Knowledge and Regulation of Cognition. *Journal of Experimental Education*. 69(4)

Ford, J.K., Smith, E.M., Weissbein, D.A., Gully, S.M., & Salas, E. (1998). Relationships of Goal Orientation, Metacognitive Activity, and Practice Strategies with Learning Outcomes and Transfer. *Journal of Applied Psychology*, 83 (2), 218-233.

French, J., Blickensderfer, E L., Summers, M., Ayers, F., & Connolly, T. (2005). *FITS Combined Task 1 & 2 Final Report*. Florida, USA: Embry-Riddle Aeronautical University, College of Arts and Sciences & College of Aviation.

Glaser, R. (1990). The Reemergence of Learning Theory Within Instructional Research. *American Psychologist*. 45 (1), 29-39.

Guglielmino, P.J. (2006, February). Opening Session. Symposium conducted at the meeting of the International Self-Directed Learning Symposium, Cocoa Beach, Florida.

Greveson, G.C. & Spencer, J.A. (2005). Self-directed learning – the importance of concepts and contexts. *Medical Education*, 39, 348-349.

Maki, R.H., Shields, M., Wheeler, A.E., & Zacchilli, T.L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology*. 97 (4), 723-731.

Nietfeld, J.L., Cao, L., & Osborne, J.W. (2005). Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom. *The Journal of Experimental Education*, 74(1), 7-28.

Nietfeld, J.L., and Schraw, G. (2002). The Effect of Knowledge and Strategy Training on Monitoring Accuracy. *The Journal of Educational Research*. 95(3), 131-142.

Oser, R.L., Cannon-Bowers, J.A., Salas, E., & Dwyer, D.J. (1999). Enhancing Human Performance in Technology-Rich Environments: Guidelines for Scenario-Based Training. *Human/Technology Interaction in Complex Systems*. 9, 175-202.

Parson, S. (2006, February). Making the Grade: The four R's of collaborative critique. *NAFI Mentor*, 12-13.

Peltier, J.W., Hay, A., and Drago, W. (2005). The Reflective Learning Continuum: Reflecting on Reflection. *Journal of Marketing Education*. 27 (3), 250-263.

Roberson, D. N., Jr. & Merriam, S.B. (2005). The Self-Directed Learning Process of Older, Rural Adults. *Adult Education Quarterly*, 55(4), 269-287.

Schraw, G. (1997). The effect of generalized metacognitive knowledge on test performance and confidence judgments. *Journal of Experimental Education*. 65(2), 135-147.

Schraw, G., & Dennison, R.S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, 19, 460-475.

Schraw, G., Dunkle, M.E., and Bendixen, L.D. (1995). Does a General Monitoring Skill Exist? *Journal of Educational Psychology*. 87 (3), 433-444.

Schraw, G. & Moshman, D. (1995). Metacognitive Theories. *Educational Psychology Review*. 7 (4), 351-371.

Smith-Jentsch, K.A., Milanovich, D.M. & Merket, D.C. (2001, April). Guided team self-correction: A field validation study. In S. Kozlowski & R. DeShon (Chairs), *Enhancing team performance: Emerging theory, instructional strategies, and evidence*. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Smith – Jentsch, K. A., Payne, S.C., & Johnston, J.H. (1996). Guided team self-correction: A methodology for enhancing experiential team training. In K.A. Smith-Jentsch (Chair), W*hen, how, and why does practice make perfect?* Paper presented at the eleventh annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., & McPherson, J.A. (1998). Team dimensional Training: A Strategy for Guided Team Self-Correction. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 271 – 297). Washington, D.C.: American Psychological Association

Stevens, C.K. & Gist, M.E. (1997). Effects on Self-Efficacy and Goal-Orientation Training on Negotiation Skill Maintenance: What are the Mechanisms? *Personnel Psychology*. 30, 955-978.

Tannenbaum, S.I., Smith-Jentsch, K.A.& Behson, S.J. (1998). Training Team Leaders to Facilitate Team Learning and Performance. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 247 – 270). Washington, D.C.: American Psychological Association

Theide, K.W. & Anderson, M.C.M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*. 28, 129-160.

United States Federal Aviation Administration. (n.d.,a) *Learner Centered Grading*. Retrieved March 9, 2006, from http://www.faa.gov/education_research/training/fits/media/ program%20plan.doc

United States Federal Aviation Administration. (n.d., b) *FITS Program Plan*. Retrieved March 9, 2006, from http://www.faa.gov/education_research/training/fits/media/program%20plan.doc

United States Federal Aviation Administration. (n.d., c) *FITS Questions and Answers*. Retrieved March 9, 2006, from http://www.faa.gov/education_research/training/fits/media/fits_qa.pdf

Volet, S.E. (1991). Modelling and Coaching of Relevant Metcognitive Strategies for Enhancing University Students' Learning. *Learning and Instruction*, 1, 319-336.

Annotated Bibliography

Part I. Learner Self-assessment and Metacognition Related Articles

A Less Structured, More Learning-Centered Environment. (2005, June/July). *Teaching Professor,* 19(6), 6-7.

This article comments on the publication, "Being in the Classroom." 'Being' classes focus on teachers and students being equal sources of learning, the premise being that teachers cannot learn content for the students, nor can they force students to learn. From one learning experience, different students will come away with different interpretations depending on their prior experience. An issue with 'being' classes is that teachers have to be willing to give up control and accept ambiguity and uncertainty in the classroom. Teachers have to be confident and versed in the topic well enough to answer any questions that may arise out of the students' control of the discussion. For the students, "compliance is not enough" and neither is cooperation. Students need to commit to their own learning and the learning of others.

De Carvalho, F., Moises, K., & Yuzawa, M. (2001). The Effects of Social Cues on Confidence Judgements Mediated by Knowledge and Regulation of Cognition. *Journal of Experimental Education*. 69(4)

It is generally thought that those who self-regulate their learning perform better and are better able to take advantage of what they have learned. Self-regulation encompasses the active participation of metacognition, motivation, and behavior in the learning process (De Carvalho, Moises, & Yuzawa, 2001).

In this study, 77 Hiroshima University undergraduate students; half of which participated for extra credit consideration, the other half were volunteers. Participants completed a 10-question 'knowledge of cognition' checklist and six Japanese literature texts, each followed by four, 5-option multiple-choice questions and an overall performance accuracy rating scale. The cognition checklist asked the participants to indicate on a 5-point Likert scale (ranging from always true to always false) whether they applied each factor to their daily academic routine. The confidence scales were also 5-point Likert-type scales labeled 0%, 25%, 50%, 75%, and 100%, for participants to indicate how many test questions they thought they had answered correctly on the entire test. A 21-point scale (0%, 5%, 10%, 15%, etc) was used for participants to indicate overall confidence in their test performance (De Carvalho et al, 2001).

Participants were divided up into groups based on high and low knowledge of cognition, based on the checklist, and also divided into high and low regulators of cognition, according to the mean accuracy score of their judgments (De Carvalho, et al, 2001).

Results of the study indicated that participants with higher level of knowledge and regulation of cognition are more confident of their performance, which may be explained by prior successful experiences. High knowledge and high regulators both performed better than their counterparts on the performance tests, indicating that both knowledge of cognition and regulation of cognition are essential for improving performance.

Regulation of cognition may compensate for knowledge, in that knowledge of cognition affected the performance of low regulators, but not of the high regulators (De Carvalho, et al, 2001).

Glaser, R. (1990) The Reemergence of Learning Theory Within Instructional Research. *American Psychologist*. 45 (1), 29-39.

Note: This summary combines Glaser (1990) with a summary of

Ciccarelli, S.K. & Meyer, G.E. (2006). *Psychology*. Upper Saddle River, NJ: Pearson Prentice Hall.

The focus of this paper is the discussion of several theoretical perspectives in learning and competence for complex knowledge and skills garnered from various domains. All the instructional programs are based on one of three aspects of competence: a) proceduralized an automatized knowledge for skills; b) internalized self-regulation for comprehension; and c) structuring of knowledge for the purpose of problem solving (Glaser, 1990).

The first application of learning theory is aimed at the acquisition of proceduralized skills through Anderson's ACT model (as discussed in Glaser, 1990) where learning transforms from declarative knowledge to procedural knowledge, or from knowing 'what' to knowing 'how'. Problem solving with only declarative knowledge and general methods demands intense cognitive resources. As knowledge increases and automation progresses, cognitive workload lessens, increasing the amount of working memory available for new information. Declarative knowledge is then incorporated into a production-rule architecture that is used to compare performance to both an "ideal" model and a "buggy" model, depending on the student's level of skill (Glaser, 1990).

A second application of learning theory is based on the development of expertise and self-regulatory skills. Experts are known for their ability to monitor their performance and predict the outcomes of their actions, which contributes to their value and utility. In the Reciprocal Teaching program, students learn metacognitive strategies concurrently with specific knowledge. Students are exposed to an expert model of performance and metacognition via an instructor, then practice executive strategies that help develop their ability to self-monitor (e.g. questioning, summarizing, clarifying, and predicting) in an environment designed for social interaction and participation. The theory behind this strategy is borne of the developmental psychology work of Vygotsky, who believed that cognitive development, specifically in children, was aided by scaffolding – a process where leading questions and examples are provided to novices, and slowly scaled back as the learner gains expertise (Ciccarelli & Meyer, 2006). Other techniques like Reciprocal Teaching include metacognitive strategies such as giving explicit prompts to facilitate the process of sophisticated writing strategies, and heuristic methods for managing math problems by generating alternative courses of action, evaluating those alternative, and self-assessing progress (Glaser, 1990).

A third application of learning theory is centered on mental models for problem solving. An example of this method is the GUIDON intelligent tutoring system for medical diagnosis led by William Clancy (as referenced in Glaser, 1990). The learning process for GUIDON is the detection and explanation of problem-solving failures. Learners attempt to solve a specific medical problem by applying their incomplete mental model to the GUIDON model, which then detects inconsistencies between the two models and suggests repairs to the learner model (Glaser, 1990).

Greveson, G.C. & Spencer, J.A. (2005). Self-directed learning – the importance of concepts and contexts. *Medical Education*, 39, 348-349.

Focusing on the medical field, this editorial outlines some of the articles that criticize self-directed learning (SDL). Much of the medical education field is based on the premise that doctors are, or should be, self-directed learners. One author draws attention to the idea that self-directed learners need to be effective at self-assessment, however most medical practitioners have difficulty with this. (note: is there any research that generalizes the ability of people to self-assess?) Another author claims that external guidance would always be required for unfamiliar practice, which conflicts with the principles of SDL. More interestingly, the editorial notes that the context of learning is an important factor. SDL is viewed most often as a relatively stable trait or measurable personal attribute (e.g. the Self-Directed Learning Readiness Scale.) Some authors argue that SDL varies based on motivation cued by context such as 'subject matter, the social, cultural and educational setting, past experiences, self-concept, and relevant study skills.

Maki, R.H., Shields, M., Wheeler, A.E., & Zacchilli, T.L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology*. 97 (4), 723-731.

In educational psychology literature, metacognition is often measured through *bias*, usually calculated as the difference between confidence judgments and actual performance on some task. A positive value indicates overconfidence, as the judgments are higher than performance; a negative value shows underconfidence (Maki, Shields, Wheeler, & Zacchilli, 2005).

There is support in the social psychology literature for novices in a domain to overestimate their performance largely and for experts to slightly underestimate their knowledge, and that this is consistent over tasks and over time (Maki et al, 2005).

For this experiment, 159 general psychology undergraduate students volunteered to participate. SAT/ACT verbal scores were used to divide the students into three ability groups. The came to a laboratory session for one session lasting approximately one hour, and read six texts presented over the computer in a random order. Participants read one sentence at a time and then pressed a key for the next sentence to appear. After reading each text, students were asked to indicate the percentage of the text (0%, 20%, 40%, 60%, 80%, 100%) they thought they were able to comprehend. They then indicated how many from six test questions they thought they would be able to answer correctly, and then they actually completed the test questions. Following the questions, the participants indicated how many (0-6) they thought they answered correctly (Maki et al, 2005).

Results of this study showed that low-ability students were overconfident, especially in their predictions of how well they would perform on the tests. High ability students were underconfident, especially in their post-test confidence judgments (Maki et al, 2005).

Nietfeld, J.L., Cao, L., & Osborne, J.W. (2005). Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom. *The Journal of Experimental Education,* 74(1), 7-28.

Metacognitive monitoring is the key to self-regulated learning; however it is unlikely that students are always accurate with their monitoring. Monitoring accuracy may be affected by test difficulty, age, comprehension instruction, background knowledge, and performance level. Most students have metacognitive knowledge about their learning, but don't choose to use this knowledge to increase their performance. Skilled adult learners can display poor monitoring under certain conditions, as well. Monitoring accuracy may be increased by introducing strategy instruction prior to testing.

Metacognition assists in learning by increasing the efficiency of attentional resources, allowing for deeper processing of information, and increasing accuracy of performance monitoring. Metacognition has two components: knowledge of cognition (accumulated knowledge about general cognitive processes, tasks, and strategies), and regulation of cognition (ongoing, active tracking of mental processes and regulatory strategies, e.g. monitoring)

Judgments of performance can be made either at the "local" level (following a single item) or at a "global" level (a single judgment following an entire test).

In this field study of a psychology class of 27 upper class students, the students' self-monitoring accuracy remained relatively stable throughout the course. Self-directed feedback (in the form of a 100-pt self scoring item after each item and each test) was not sufficient of an intervention to improve accuracy. There was no explicit training in monitoring. This indicates that merely prompting students to think about their performance is too passive of a manipulation, even when applied over the course of several exams during a semester. More likely, an explicit training module with practice, feedback, and strategies will be needed to effect change.

Student performance was strongly linked to accuracy of predicting performance. The more difficult the test items, however, the more judgment bias there was. Students tended to be less confident on easy items and overconfident on difficult items. Globally, poor students do not predict performance as well as better students. It may be that poor students understand that they don't know very much, but don't know how to improve, whereas better students are aware of where they need to apply themselves to improve.

Nietfeld, J.L., and Schraw, G. (2002). The Effect of Knowledge and Strategy Training on Monitoring Accuracy. *The Journal of Educational Research.* 95(3), 131-142.

The purpose of this study was to determine whether strategy training and prior knowledge improved monitoring accuracy and confidence judgments. Prior experimentation revealed that monitoring accuracy is unlikely related to ability or domain specific expertise, and that it is malleable and can improve when a learner gains metacognitive knowledge. Delclos and Harrington (1991) found that a group trained in both monitoring and problem solving performed better than groups trained in only problem-solving or not given any training at all. (Nietfeld & Schraw, 2002).

The debilitative hypothesis (Glenberg & Epstein, 1987; Koria, 1994; Schwartz 1994; Wiley, 1998) states that individuals make monitoring judgments not on a specific problem-by-problem basis, but rather on their general perception of confidence within a domain. (Nietfeld & Schraw, 2002).

The results of this experimentation show that brief strategy training can improve performance, confidence, and monitoring accuracy, but only for the short-term, as a 1-week delayed posttest did not replicate the same results. (Nietfeld & Schraw, 2002).

Oser, R.L., Cannon-Bowers, J.A., Salas, E., & Dwyer, D.J. (1999). Enhancing Human Performance in Technology-Rich Environments: Guidelines for Scenario-Based Training. *Human/Technology Interaction in Complex Systems*. 9, 175-202.

In order for training to be effective, the training environment must support the trainee's ability to develop and maintain the knowledge, skills, abilities, and attitudes (KSAAs) necessary to perform the trained tasks. Particular attention should be paid to the deliberate and systematic approaches applied to the learning environment. These learning approaches should be applicable to all phases of training development, as well as performance measurement and feedback. (Oser et al, 1999).

In the past, training design has largely been driven by technology. Only in its more recent history has training been centered on the learner. Learner-oriented design focuses on systematically applying sound learning principles (both cognitive and behavioral) to training design as well as systematically measuring outcome and process learning objectives. (Oser et al, 1999).

Scenario-based training is an answer to complex training systems. It begins with the collection of historical performance data and a skill inventory to determine the necessary KSAAs. This information is combined with task lists to create specific task objectives or general competencies. Learning objectives and competencies form the basis for the construction of scenario events, the core of scenario-based training. Each event is a condition within a scenario that evokes the need to use an objective or competency.

The scenario should have a script, or Master Scenario Event List (MSEL) that lists the events in order of occurrence, the learning objective associated with each event, the performance measurement for the event, and any resources needed to complete the event. Following the scenario, performance is diagnosed, and feedback is given on both process and outcome through a debriefing. (Oser et al, 1999).

Peltier, J.W., Hay, A., and Drago, W. (2005). The Reflective Learning Continuum: Reflecting on Reflection. *Journal of Marketing Education*. 27 (3), 250-263.

In the marketing education literature, a recent trend suggests that learning environments are perceived to be most effective when students are proactive and engaged and when students and instructors produce outcomes together (Peltier, Hay, & Drago, 2005).

Reflection, as defined in the literature, stems from Dewey's pioneering article in 1933 (as cited in Peltier, Hay, & Drago, 2005) where thinking came from a state of doubt or mental difficulty and information was sought to relieve that state. That theme is consistent in the literature, where reflection is thought to be a mechanism through which people make sense of their experiences, as a basis for critical thought, as a portal for change, and as an internalization of experience that triggers exploration. These are all very personal revelations, and so it should be true that learning comes from personal motivation.

Because we can not analyze merely from a position of understanding, reflection is an integral part of problem solving. Critical reflection involves content (what we perceive, think, or feel) reflection and process (examination of how we reflect on the content) (Peltier et al, 2005).

In a learning environment, one of the keys to reflection is the instructor-to-student interaction. In order for reflection to occur, the environment must be facilitative for the student. They must be able to express doubt and uncertainty, raise points of disagreement, and explore the differentiation between the instructor mental model and the student model. Instructors, in addition to creating the open atmosphere required for this, must also help the student build metacognitive strategies to become aware of their own cognition and to improve regulation of it (Peltier et al, 2005).

In this study, 323 recent graduates of an MBA program (within 3 years) were given a $2 incentive to complete a questionnaire consisting of 56 questions across six dimensions (intensive reflection, reflection, understanding, habitual action, instructor-to-student interactions, and student-to-student interactions) and six global dependent measures designed to describe the student's overall value of the MBA program (Peltier et al, 2005).

The results of the study showed that reflecting and intensive reflection are perceived as higher levels of learning and correlate positively with program outcomes. Instructor-to-student interaction as well as student-to-student interactions were identifies as conditions related to reflection (Peltier et al, 2005).

Pressley, M., & Ghatala, E.S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, 25, 19-34.

Roberson, D. N., Jr. & Merriam, S.B. (2005). The Self-Directed Learning Process of Older, Rural Adults. *Adult Education Quarterly*, 55(4), 269-287.

While the population studied here is not exactly the population of our interest, this article does identify an important aspect to self-directed learning: motivation of the learner. SDL "begins with an incentive to learn plus an interest, leading to accessing resources." In their review of the literature, SDL is defined as "intentional, self-planned learning, where the individual is responsible and in control of the learning." It is also recognized that SDL is a product of the person's context – a person's situation in life is linked to their motivation. The context gives some internal or external motivation that provides the *incentive* to engage in SDL. In the model put forth by the authors, the second stage of SDL is personal *interest*. Without it, the activity may be put aside and not completed. After there is motivation and interest, the SDL student seeks out *resources* needed to move forward with the learning. *Systematic attention*, or more than random time spent with the resources, gives the object of SDL priority. At some point during the learning, *adjustments* are needed to correct errors or overcome obstacles. Participants referred to "trial and error" to move past mistakes in their learning. The last part of the model is *resolution*, where the SDL reaches its natural conclusion.

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology*, 19, 143-154

Schraw, G. (1997). The effect of generalized metacognitive knowledge on test performance and confidence judgments. *Journal of Experimental Education*. 65(2), 135-147.

This study looked at supporting either a domain-specific or domain-general hypothesis in terms of monitoring function. The domain-specific hypothesis states that the ability to accurately monitor one's performance is a natural by-product of increased learning in a particular domain. The domain-general view is that self-evaluation takes place through a more metacognitive approach, involving more general skills such as checking for comprehension and comparison of test information to information stored in memory. (Schraw, 1997).

Participants took several short tests over different domains, as well as a General Monitoring Strategies Checklist. Participants recorded their confidence judgment for each individual item by drawing a slash through a 100mm line at the location that best corresponded to their perceived confidence related to that test item. (Schraw, 1997).

It was found that confidence judgments and performance were not correlated for a particular test or between tests. This low correlation supports the domain-general hypothesis, or more general metacognitive approach. (Schraw, 1997).

Based on their results, the researchers infer that individuals having access to general metacognitive knowledge use it to their advantage to make more accurate judgments of performance and monitor their performance in difficult and unfamiliar domains. It may be that general metacognitive skills evolve through developing domain-specific monitoring experience. (Schraw, 1997).

Schraw, G., & Dennison, R.S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, 19, 460-475.

Metacognition: the ability to reflect upon, understand, and control one's learning. It is made up of two components: knowledge about cognition (three subparts: declarative knowledge- knowledge about self and strategies; procedural knowledge – how to use strategies; conditional knowledge – when and why to use the strategies) and regulation of cognition (includes planning, information management strategies, comprehension monitoring, debugging strategies, and evaluation).

Metacognitively aware learners are more strategic and perform better than unaware learners because they are able to plan, sequence, and monitor their learning in a way that directly improves performance. Differences in metacognitive awareness are more related to differences in performance than intellectual aptitude.

This study tested a 52-item metacognitive inventory called the MAI. Results of the two experiments showed that it positively correlated with theoretical predictions of the two factors of metacognition (knowledge and regulation), indicating that the MAI measured the two kinds of metacognitive knowledge reliably. Also, they both showed a statistically significant relationship between knowledge and regulation of cognition, which also is congruent with previous theoretical findings; although they appear to work independently of each other, as is shown in experiment 2, where knowledge of cognition was related to higher test performance, but regulation of cognition was not.

The MAI was shown to have good predictive ability for future performance. Knowledge of cognition was related to pre-test judgments, and those judgments were related to test performance. Those who expressed a greater degree of confidence in their cognitive monitoring ability showed better test performance overall.

Surprisingly, there was no statistical relationship between monitoring accuracy and MAI or pre-test judgments. It was expected that regulatory aspects of cognition would be correlated with performance, but that was not found to be the case in this study.

Schraw, G., Dunkle, M.E., and Bendixen, L.D. (1995). Does a General Monitoring Skill Exist? *Journal of Educational Psychology*. 87 (3), 433-444.

Prior research has shown that self-monitoring can be affected by various factors, including task difficulty, level of detail of learned information, amount of expertise in the task domain, incentive to self-monitor, and inclusion of prompting questions and feedback as part of the task (Schraw, Dunkle, & Bendixen, 1995).

Two competing hypothesis about the nature of metacognition emerged from this prior research: 1) Metacognition is domain-general; there are general metacognitive strategies that exist regardless of the domain the knowledge pertains to, and 2) Metacognition is domain-specific; executive strategies applied to enhance learning are dependent on the domain knowledge(Schraw et al, 1995).

These hypotheses were tested in two experiments. Experiment 1 had one hundred and thirty-four Intro to Psychology students participate as a part of their grade. They took multiple choice question tests in seven different domains ranging from geography, nutrition, special tests, to general knowledge questions. Participants indicated how much domain knowledge they felt they had for each of the tests by recording it on a 100-mm line that ranged from none to extensive background knowledge. Monitoring proficiency was measured by a discrimination index calculated by taking the difference between the average confidence scores for correct and incorrect items for each of the eight tests, and also by taking a bias score consisting of the signed difference between the average confidence and average performance for each of the eight tests. Confidence ratings for each item were recorded on a 100-mm line that followed each question (Schraw et al, 1995).

The results of the first experiment supported both the domain-general and the domain-specific hypotheses. Successful performance and discrimination were in large part domain-specific. The confidence and bias scores were highly correlated across domains, meaning that participants were generally over- or underconfident, no matter what the domain, which supports the domain-general view. It appears that when multiple, diverse domains are tested, both types of metacognitive processes are active. This lends support to the view that metacognition is tied to domain-specific strategies, and not a general skill (Schraw et al, 1995).

In the second experiment, the tests were made more homogenous in terms of length, difficulty, sensitivity to guessing, and allowances for inferences made based on information included in the test; however the topics were still diverse and ranged from U.S. presidents, vocabulary, celebrity athletes and musicians, to U.S. geography. Ratings were collected in the same manner as in the first experiment (Schraw et al, 1995).

When the tests were made more homogenous, all performance scores and confidence scores were intercorrelated, and with two exceptions, all discrimination scores and half the bias scores were intercorrelated. This supports the domain-general view of metacognition. However, the pattern of correlations between performance and confidence scores supports the domain-specific hypothesis, because the correlation in each domain was significant. This is different from the pattern in the first experiment, where only half the domains showed correlations between performance and confidence scores. Collectively, performance and confidence are correlated both within and across domains, supporting both domain-general and domain-specific views of metacognition (Schraw et al, 1995).

Schraw, G. & Moshman, D. (1995). Metacognitive Theories. *Educational Psychology Review*. 7 (4), 351-371.

What individuals know about their own cognition, or about cognition in general is called knowledge of cognition, and usually includes three different kinds of metacognitive awareness: declarative (knowing what/about), procedural (knowing how), and conditional (knowing why and when) knowledge (Schraw & Moshman, 1995).

Metacognitive activities that help control one's thinking or learning is what is referred to as regulation of cognition, and usually includes three essential skills: planning (strategy and resource allocation), monitoring (awareness), and evaluation (appraising the products of learning). Research indicates that monitoring ability develops slowly and is quite poor in children and even adults; however, it appears to improve with training and practice (Schraw & Moshman, 1995).

Stevens, C.K. & Gist, M.E. (1997). Effects on Self-Efficacy and Goal-Orientation Training on Negotiation Skill Maintenance: What are the Mechanisms? *Personnel Psychology*. 30, 955-978.

When looking to improve performance outcomes, one could look to the mastery orientation of the people completing the task. Those individuals who are performance-oriented tend to view outcomes as resulting from their ability levels – if performance is poor, then there is little use in changing effort or tact. In contrast, those who are master-oriented test to view outcomes as feedback. When performance is poor, they will change strategies or level of effort to try and change the outcomes. Whether a person adopts a mastery- or performance-oriented view is in part related to their disposition, but it can also be influenced by the situation. Research has shown support for the ability to elicit mastery orientation through explicit instruction in metacognitive activities associated with self-regulation during post-training interventions (Stevens & Gist, 1997).

In this study, voluntary participants were randomly assigned either a performance or a mastery-oriented intervention. The study consisted of four phases: 1) 4 hours of content training, including lecture, discussion, and modeling demonstration; 2) first practice negotiation, 3) post-training intervention, and 4) a second negotiation paralleling the first practice negotiations, which took place 7 weeks after the content training. The mastery-oriented intervention introduced self-management as a skill-maintenance strategy by 1) identifying performance obstacles, 2) planning how to overcome obstacles, 3) setting goals, 4) monitoring progress, and 5) rewarding themselves for meeting their goals. Results supported the hypothesis that master-oriented participants would engage in more skill-maintenance activities, plan on using more effort, and show more positive affective responses than those who were performance-oriented (Stevens & Gist, 1997).

Theide, K.W. & Anderson, M.C.M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*. 28, 129-160.

According to many models of self-regulated study, one first starts the learning process by setting a desired goal for the level of expertise in the material. Monitoring must take place as studying progresses to compare current levels of knowledge against the desired state, and the outcome of the monitoring determine whether studying continues or is terminated. Self-monitoring is critical for the regulation of study, and has been shown that accurate monitoring is associated with the improved ability to differentiate items for study, more effective regulation of study, and higher performance on tests based on the learned material (Theide & Anderson, 2003).

Summarizing has been found to help build relationships among concepts within a text as well as linking concepts to prior knowledge. It may do this through focusing attention, but it may also be because summarization promotes self-testing. As the person summarizes, they judge their success in retrieving information and incorporating it into the summary (Theide & Anderson, 2003).

In this experiment, participants were randomly assigned to one of three groups – a no summary group, a group that summarized each text immediately after it was read, and a group that read all seven texts first (taken from various domains in an encyclopedia), then completed summaries for each text. All groups were asked to rate their ease of learning of each text by viewing the title of the text and judging how easily they thought they could learn information from a text on that topic from 1 (very difficult) to 7 (very easy), then rated their comprehension of the test and took a six-question test to measure their comprehension on each text. Metacomprehension accuracy was measured by computing the median proportion of correct text response and comprehension rating across the six texts (Goodman-Kruskal gamma correlation) (Theide & Anderson, 2003).

Results showed that when accuracy is operationally defined as the correlation between comprehension ratings and total test performance, the delayed-summary group performed better than the immediate summary or no summary groups. Additionally, the groups were allowed to select some texts for re-reading, and the delayed-summary group chose to reread less-learned texts to a greater degree than the other groups, showing that more accurate monitoring leads to more effective regulation of study (Theide & Anderson, 2003).

Winne, P.H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551-572.

Winne, P.H., & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, 28, 259-276.

Part II: Team Self-Correction Related Articles (U. S. Navy research)

Blickensderfer, E L., Cannon-Bowers, & J.A., Salas, E. (1997). Training Teams to Self-Correct: An Empirical Investigation.

Mental models are explored as a construct to measure team coordination and effectiveness. Teams build mental models as they observe correlations between behaviors, events, and actions, and as they figure out how and why the team operates as it does. These shared mental models enable team members to be on the same page and have similar predictions about future events.

Team self-correction, as conceptualized in the paper, has a number of key elements, including: what content the team should discuss, and how the discussion should proceed. The content should contain "explanations of and expectations concerning: mission objectives and goals, roles and responsibilities of team members, knowledge of behavioral sequences, response patterns, interaction patterns and procedures, and scripts and frameworks for performing the task."

The training was designed in a behavioral role-modeling format, and was unrelated to the actual tested task; thus it could be generalized to train on any type of task. The training time was approximately one hour, and included an audio-taped lecture focused on the definition and process of team self-correction, with an emphasis on why the team self-correction process is useful and how to have an effective team self-correction discussion, including a step-by-step review. The lecture was followed by behavioral vignettes that displayed effective and ineffective team self-correction behaviors.

The results of the study showed that the teams who received the team self-correction training exhibited self-correction behavior significantly more than the control teams, with the second data collection showing the highest frequency of the behaviors and the third collection being approximately equal to the first. The control team actually displayed fewer self-correction behaviors in the third data collection than either the first or the second tests.

Smith-Jentsch, K.A., Milanovich, D.M. & Merket, D.C. (2001, April). Guided team self-correction: A field validation study. In S. Kozlowski & R. DeShon (Chairs), *Enhancing team performance: Emeergind theory, instructional strategies, and evidence*. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Although this study focuses on team self-correction, there are many parallels between team and individual self-correction. The strategy that this study was looking to validate focused on improving team coordination and performance through team briefings where members "share ideas and feedback, clarify roles and expectations, critique their own performance, and solve their own problems."

By using pre-briefings, information such as role expectations and task strategies can be emphasized as a means to improve future performance. The pre-brief may also be used to align leader and team's mental models, which should also lead to superior performance outcomes.

The debriefing can be used optimally to engage in self-correction that involves role clarification, self-evaluation of task strategies, planning, and goal setting. Goal setting can be further manipulated by setting master-oriented goals as opposed to performance or outcome-oriented goals, which increases the effectiveness of the debrief.

The teaching strategy was enhanced by adopting a computer-based training module and a structured pre-and debriefing guide. This maintains a consistent teaching environment, as well as provides an outline for the facilitator to follow during briefings.

Of the many measures taken, one measure taken was that of the debriefing climate. Participants completed an 8-item measure by circling a number that indicated the degree to which they felt sets of bipolar adjectives described their debriefing climate. Another measure of interest was that of the self-report of team self-correction. Participants reported on their use of 11 team self-correcting behaviors (e.g. admitting a mistake, pointing out a team weakness) during the debriefs.

Results of the briefing training intervention included an increase in pre-brief time from 5 to 7.5 minutes, and an increase of 15 minutes over the control debrief – from 16 to about 32 minutes in length. Also, there was a reduction in team leader evaluation from 83% in the control condition to 7% in the training intervention condition, and a reduction from 73% to 41% in the leader sharing expertise. There was a modeling effect from instructors to team leaders as well - positive correlations were found between the number of time the instructors asked for critiques and team leaders and members asking for critiques of their performance. The training intervention also led to a greater proportion of team members contributing to the self-correction process and committed significantly fewer tactical errors as compared to the control condition. In terms of goal setting, the effect of the training intervention on individual self-correction was greatest for those with a high mastery goal orientation as opposed to a performance goal orientation. The control group was also less likely to write team goal statements that were mastery-oriented.

Based on all the results, teams within the training intervention condition can better critique their own processes (e.g. diagnose and solve teamwork problems) and generalize the lessons learned to future performances.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., & McPherson, J.A. (1998). Team dimensional Training: A Strategy for Guided Team Self-Correction. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 271 – 297). Washington, D.C.: American Psychological Association

Team Dimensional Training (TDT) is a strategy for enhancing teams' ability to self-correct by incorporating guided team self-correction to develop teamwork-related knowledge and skills. Use of guided team self-correction refers to using a facilitator in place of a traditional leader who "(a) keeps the team's discussion focused, (b) establishes a positive climate, (c) encourages and reinforces active participation, (d) models effective feedback skills, and (e) coaches team members in stating their feedback in a constructive manner." Helping the team determine what and how they should discuss certain topics is the main focus of the facilitator.

In order for self-correction to occur, members must accurately determine which processes they use well and which they do not. Organizing these processes as a team requires a shared mental model (knowledge structures or cognitive representations used to organize new information; to describe, explain, and predict events, and to guide their interactions with others.) Sharing accurate mental models with other team members should allow them to "(a) uncover performance trends and diagnose deficiencies, (b) focus their practice appropriately on specific goals, and (c) generalize the lessons they learn to new situations."

One aspect of TDT is the prebrief. During this prebrief, instructors remind members of the TDT structure, remind them of performance objectives, explain how the objectives and structure will pertain to the upcoming tasks, and establish a positive learning climate.

During the actual task, which should be relatively brief (30-60 minutes), instructors should record representative samples of strengths and weaknesses, rather than record every positive behavior and every error. The representative samples should relate to the task/teamwork objectives. Discussing every part of the task would not be conducive to time constraints on the debrief.

The TDT debrief is designed to enhance teamwork mental models and performance by alerting teams to teamwork goals, soliciting feedback from team members regarding teamwork goals and performance, managing the flow of the discussion, and helping the team diagnose strengths and weaknesses and goals for improvement. To improve the effectiveness of the debrief, the facilitator should show that they are genuinely interested in the observations of the team by pausing and making eye contact with all team members after asking a question. They should ask specific, open-ended, and behavioral questions before offering their own critique. Asking for both positive and negative examples of behaviors is important because generalization of training can come from both effective and ineffective examples of behavior. Also, if no one in the team offers a good solution to a team problem, or if the team cannot agree to a reasonable solution, the facilitator should offer a reasonable explanation for why the suggested solution is not viable, and recommend a solution before moving on to the next topic.

TDT instructor-training workshop involved approximately 4 hours of classroom training, followed by hands-on practice lasting between 4 and 8 hours. The instructors' ability to correctly use TDT improved dramatically after 3 or 4 experiences with the process.

In terms of performance enhancement, one case study showed that a TDT trained team accurately diagnosed teamwork-related problems, demonstrated immediate improvement on targeted goals, and generalized lessons learned to new exercises. Other experimental teams showed that, while performance was equal to control groups for an initial, easy exercise, only teams exposed to TDT training continued to exhibit superior performance while the control groups' performance declined with increasingly difficult tasks. Only the TDT trained teams were able to generalize the TDT lessons learned to novel situations

Smith – Jentsch, K. A., Payne, S.C., & Johnston, J.H. (1996). Guided team self-correction: A methodology for enhancing experiential team training. In K.A. Smith-Jentsch (Chair), W*hen, how, and why does practice make perfect?* Paper presented at the eleventh annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Self-correction doesn't happen just because teams are given the opportunity to do so. There is no guarantee that teams will use that opportunity to have discussions that are constructive or relevant to performance. When presented with self-correction opportunities that are unguided, teams will often focus on outcome goals, which are not as conducive to generalizing training concepts as process goals. They also tend to focus on problems that are easier to identify and solve, rather than difficult and more central problems.

An improvement over unguided self-correction is guided self-correction. Here, a facilitator questions a team following their performance using regarding specific processes. The team identifies their own problems, which then become the focus for future team training exercises.

From this research, an instructor debriefing guide was created. The instructor primes the group for the debriefing by restating the learning objectives, outlining the organization of the debrief, and recapping the major scenario events. Then, the instructor poses open-ended questions that ask for specific information regarding the targeted behaviors under significant scenario events. The debriefing guide contains questions that target both positive and negative behaviors. Finally, instructors would summarize the strengths from the debriefing, areas identified for improvement, and a specific goal for the next practice session.

For the instructor training, instructors received a one-day workshop on how to use guided team self-correction, which included lecture, demonstration of both effective and ineffective debriefings, practice evaluating a sample team, and a role-play of a debriefing using the guide.

In this experiment, the guided self-correction team showed an increase in performance when moving from a low-stress exercise to a higher-stress exercise, while 6 control groups showed a decline in performance in the same situations. Improvement was linked to the debriefing sessions in which specific behaviors were addressed by the team, indicating that the improvement was due to the team's self correction, and not simply a Hawthorne effect.

Some difficulties encountered in this type of training include a culture change for instructors and team leaders. Some leaders tend to want to provide their own observations of performance before allowing the team to self-assess. Unfortunately, this tends to stifle active participation in the team self-correction process. Also, some leaders want to ask general questions (e.g. how did this activity go?) to which the team responds with general answers (e.g. fine). Specific open-ended questions forced team members to come up with focused answers and specific examples that enhanced learning. Many team leaders also tended to focus their questioning on higher ranking team members, which underutilizes the amount of input from the rest of the team.

Recommendations for creating a guide for team self correction include first identifying critical skills needed by the team, then identify critical events that would or should naturally occur in their environment for training scenario purposes. Next, identify specific, observable positive and negative behaviors associated with the critical skills. The guide should summarize each critical event, provide specific open-ended questions that elicit examples of positive and negative behaviors, and require goal setting for future training.

Tannenbaum, S.I., Smith-Jentsch, K.A.& Behson, S.J. (1998). Training Team Leaders to Facilitate Team Learning and Performance. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 247 – 270). Washington, D.C.: American Psychological Association

Because of the dynamic nature of naval environments, it is important to move away from a traditional *control and command* style of leadership and be more like coaches or facilitators. This is especially true of environments where high levels of stress can have a debilitating effect on decision making. In the traditional environment, the supervisor provides explicit directions and makes most of the decisions. In a facilitative environment, the coach helps the learner to set goals, roles and procedures, as well as to synthesize lessons from past events to apply learning to a new, dynamic situation.

It should be noted that some instructors may resist the change from traditional supervisory roles to that of a facilitator, as it may be perceived as a sign of weakness or lack of authority.

To improve learning, it is recommended that learners prepare to learn from their experiences both before and after key events. This preparation should increase learning and enhance future performance. Pre-briefs are essential for setting goals. [note: if we want to add another element to LCG, we could add a pre-brief component on the goal of LCG.] During key events, especially those that require high levels of attention, there may be little time for learning; most individuals allocate the bulk of their cognitive resources to performing the actual task; however, they may also dedicate some cognitive resources to self-monitoring, learning, and correction.

How facilitation and feedback occur is important to the learning process. In order for feedback to be useful for both informational and motivational purposes, it should be designed systematically. For example, the facilitator should avoid lecturing the learner, and should withhold their observations and opinions of the exercise until the learner has given their opinion. The use of closed-ended questions may stymie the usefulness of the feedback process as well, as they encourage one-word/yes/no types of answers that do not elicit opinions of performance or suggestions for improvement. It is more effective to use open-ended questions that probe the learner to assess their own performance. Allotting enough time for the feedback is also important. Debriefs that are rushed often turn into one-way “lectures” due to time constraints.

Referring to prior pre-briefs when conducting subsequent debriefs provides a sense of continuity, reliability, and consistency, all of which are desirable attributes of a feedback source. Reminding learners of goals and lessons learned from prior exercises helps them plan for future events. Learners may also be more receptive to feedback during a debrief if they were appraised of the goal criteria in a pre-brief.

Debriefs were more effective after leaders in the experimental condition received a mere 2 hours of training on effective pre-briefing and debriefing skills. As part of their training, leaders were shown videotapes of examples of effective and ineffective debriefs. Following each videotape debrief, team leaders were asked to critique the model's debriefing skills, using the effective briefing behaviors as criteria. Leaders then conducted a practice debrief with two scripted, role-playing team members. Finally, the leaders critiqued themselves on their use of the critical briefing behaviors and received feedback from the instructor.

FY2005 FITS Instructor Education Research
Report 3: Recommendations for Instructor Training Protocol for the Learner Centered Grading Debriefing Approach

November 13, 2006
(Revised according to government comments)

Submitted by:

Beth Blickensderfer, Ph.D.
Embry-Riddle Aeronautical University

This effort was the third deliverable under the FITS Program FY2005: FITS Instructor Education Research, "Examining the Debriefing Process Including the Learner Centered Grading Approach."

Recommendations for Instructor Training Protocol for the Learner Centered Grading Debriefing Approach

Introduction

Recently, a report on current practices regarding use of the learner centered grading and debriefing portion of the FITS program was completed (Blickensderfer and Doherty, 2006--FY2005 FITS Instructor Education Research Report 1: Investigation of Current Practices Relating to Learner Centered Grading and Debriefing Procedures). In addition a review of related research in other domains was also completed (Blickensderfer, 2006--FY2005 FITS Instructor Education Research Report 2: Literature Review and Annotated Bibliography of Research Relating to the Learner Centered Grading Debriefing Approach). The purpose of the current report is to provide recommendations to help enable flight instructors to use the learner centered grading debrief procedure.

Overall, the Blickensderfer & Doherty (2006) report indicated that the instructor pilot response to the LCG debrief is quite favorable. Three challenges emerged, however: 1) the difficulty in conveying the meaning of the current LCG scale to the learner, 2) the need to instruct the learner to become an active participant in the debrief, and 3) the need to give the instructors additional guidance on transitioning from the old style debrief to the new style.

Issue 1: The difficulty in conveying the meaning of the current LCG scale to the learner

Both the results of the interviews with CFIs (Blickensderfer & Doherty, 2006) as well as the results from surveys from the FITS instructor seminars, indicate that the current LCG scale is confusing and difficult to explain (i.e., describe, explain, practice, perform, manage/decide). A number of possibilities exist regarding why this is the case. Before discussing this, however, it should be noted that while the notion of self-assessment followed up by a probing debrief is a fundamental component to the FITS approach, this particular LCG scale was intended to be a model rather than an absolute guide (personal communication Tom Connolly and Charlie Robertson, September 2005). Thus, instructors should have the freedom to modify the self-assessment scale.

It is likely that one reason the scale is complicated for learners to use is inherent to its Likert-type rating scale format. In the case of the LCG scale, users are required to give a rating (describe, explain, practice, perform, manage/decide) essentially on a scale of 1 -5 (although the numbers are not presented) for each of the listed tasks. This is a much more difficult task cognitively than, for instance, checking off whether a behavior occurred during flight (i.e., using a behavioral checklist). In short, a behavioral checklist simply asks for a count of how many times a certain behavior occurred. Both positive behaviors and negative behaviors can be included. In other words, the instructor/observer simply checks yes or no regarding whether or not the Pilot in Training (PT) demonstrated each behavior on the list.

Given the many other demands on the instructors observing a student pilot perform a complex a task, using a behavioral checklist would likely to be more accurate as well as require less diversion of attention from the learner's performance. From the learner perspective, given the many demands on the student him/herself in performing the tasks, using a behavioral checklist would also likely to be more accurate and feel less overwhelming to rate oneself.

Importantly, using a simpler format such as a behavioral checklist could still be used to generate long term, mastery-oriented goals rather than performance goals. Mastery (i.e., learning goals) are goals aimed at increasing one's ability or competence in task. This type of goal helps learners to view errors and mistakes as part of the learning process. On the other hand, performance goals focus on task performance and how an individual compares to others (e.g., winning the game; or make the highest grade in the class). Students with mastery/learning goals are motivated to <u>learn</u>, and these students put more effort into learning than do students with performance goals.

Thus, one option regarding the LCG scale is to generate a simpler format such as the behavioral checklist described above. It is likely that a simpler format may be just as effective, if not more, as it would eliminate considerable confusion.

If a simpler form is not developed, then more work is needed regarding clarifying the terms (describe, explain, practice, perform, manage/decide). This includes giving examples of the many different tasks and skills used in flying the TAA and how they would be classified in terms of the current LCG categories.

Issues 2 and 3: The need to instruct the learner to become an active participant in the debrief, and the need to give the instructors additional guidance on transitioning from the old style debrief to the new style.

The existence of these two issues, along with the description of the instructor training required for the "team self-correction" method employed by the U. S. Navy (e.g., Tannenbaum, Smith-Jentsch, & Behson, 1998) indicates that flight instructors most likely need more in-depth training on how to use the LCG debrief strategy most effectively. Recommendations for a training protocol will be described next.

Recommended Training Protocol for Training CFIs to Use the LCG Debrief

This protocol is based heavily on related research and, in particular, the team debriefing methodology (Smith-Jentsch, Zeisig, Acton, & McPherson, 1998) in addition to Smith-Jentsch (2002). Smith-Jentsch et al. (1998) advocated a training process to transition instructors from traditional debriefs to the new debriefing strategy. In some of the Naval research, debriefs were shown to be more effective after leaders in the experimental condition received a mere 2 hours of training on effective pre-briefing and debriefing skills. Other research, however, indicated that more time was necessary.

The following instructor training outline could probably be accomplished in a 2 hour training session. Additional time may be needed for adequate practice and feedback, however.

Section 1 (Lecture): The "what" and "why" of the LCG debrief and the rationale behind the approach.

- What is it? Brief Description of LCG debrief
    - A method of debriefing
    - Designed to foster flight skills as well as skills needed for self-assessment and continuous learning once the student-pilot is no longer receiving flight instructor
    - Guided by a trained facilitator (CFI)
    - Follows a structure (e.g., skills to be trained)
    - LCG debrief lasts on average 30 minutes. This time would depend on the complexity of the lesson and the PT performance.
- Why do it? The research supporting the LCG approach.
    - Discuss traditional grading systems and the focus on
        - outcome versus process feedback
        - performance versus mastery goals
    - Define Self-assessment and Learner centered grading
    - Discuss the difference between effective vs. ineffective learners
    - Discuss the difference between Experts vs. novices
    - Present current research data demonstrating this approach works
        - FITS research
        - Other domains (e.g., team training research, computer science)
    - Discuss the parallel of this approach with the old axiom, "Give a man a fish and he will fish for a day. Teach a man to fish, and they will eat for a lifetime."
    - Self Monitoring and Self-Assessment
        - Demonstrate examples of self-assessment scales to use (e.g., the LCG scale, and others if possible)
        - Discuss how to tailor the LCG form
        - Discuss problems with self-assessment and how to teach a learner to self-monitor
        - Could include Susan Parson's 4 Rs method here
- How to do it: The LCG Debrief Step by Step
    - Identify learning objectives for scenario

- Identify the LCG form you will use (i.e., use the example provided by FITS or develop your own)
- Observe student pilot performance in scenario
  - Focus on the primary learning objectives
  - Record concrete examples
  - Focus on most critical examples
  - Aim for a representative snapshot
  - Note both effective and ineffective behaviors
- Have student complete self-assessment form (without instructor input!!)
- Facilitate Debrief
  - Began by asking learner for a brief recap of events
  - Ask learner for concrete behavioral examples of key behaviors
  - Discuss discrepancies between your ratings and the student pilot's self-assessment
  - Use questioning techniques described in Section II of this training.
- Help student pilot set concrete goals for improvement

Section 2 (Lecture): Facilitation Skills - Setting the stage and asking effective questions.

For an effective LCG debrief, it is crucial that the instructor establishes a positive learning environment where the learner feels comfortable offering his/her opinion, critiquing his/herself, and asking questions. This is done, in part, by the instructor asking effective questions. Suggestions on how to accomplish this are as follows (from Smith-Jentsch, 2002, Smith-Jentsch et al., 1998):

1) Asking for learner's opinions/view.

Instructors should ask the learner for his/her opinion before offering the instructor's views.

For example:
"Can you give me an example your demonstrating an effective or ineffective response regarding the _____ system"

Rather than:
"This is what I saw"

Important points regarding asking the learner for his/her opinion first:

- The learner will begin to realize that s/he needs to be aware of his/her own performance during flight and be ready to give the instructor examples during the debrief. Hopefully, this type of self-awareness during training will carry on in post-training flight.
- As the learner articulates his/her own examples, some of these will match the instructors observations and others will not. This enables the instructor to get a feel for whether or not the learner is able to identify his/her own errors.
- If the learner does not identify exactly the same examples as the CFI, the CFI can always state his/her own observations later in the debrief.

- Instructor should give the student pilot time to respond, by pausing for 5-10 seconds after asking questions. The learner may be slow at first to respond (particularly since this approach is new to general aviation). If the instructor rushes on with little pause, the leaner may get the message that the CFI does not really care about his/her response.
- In general, the instructor should follow the outline of the instructor/self-assessment tool. Don't skip around, but keep the discussion moving. Focus on the learning objectives.

2) Asking General Questions Versus Concrete Questions

Questions that are too general will not elicit beneficial responses from the learner.

For example:

Instructor:	How did you do?
Student pilot:	"Pretty good" "Ok" "Sloppy" "Bad"
Instructor:	Do you have any comments about this scenario?
Student pilot:	(may think to him/herself "What type of comments is s/he looking for?")
Instructor:	Did that instance of incorrectly programming the GPS/MFD cause problems?
Student Pilot:	"No"

A more effective questioning process includes questions that

- Focus on concrete examples
- Encourage the learner to think beyond this scenario to think of the potential impact on future flights
- Include an emphasis on both positive and negative examples

Examples:

Instructor:	Give me an example of <u>improper</u> GPS/MFD programming. (Pause/wait for response) How could that lead to problems during future flights?
Instructor:	Give me an example of an <u>effective</u> use of cockpit resources. (Pause/wait for response) How could that same behavior help/ be used in future flights?
Instructor:	Give me an example of when you <u>missed</u> a pre-flight checklist item. (Pause/wait for response) How could that lead to problems during future flights?

3) Providing behavior based feedback.

As any instructor knows, providing feedback is crucial for learning to occur. All feedback is not created equal, however. Considerable research has investigated feedback and related issues (e.g., Dweck, 1986; Earley, Northcraft, Lee, & Lituchy 1990; Nadler, 1979; Ilgen, Fisher, & Taylor, 1979; Jacoby, Mazursky, Troutman, & Kuss, 1984; Wexley & Latham, 2002). Based on the feedback literature, here are some tips to help instructors to avoid pit-falls while giving feedback during the LCG debrief.

- *Provide behavior based feedback rather than personality based feedback*

When personality based feedback is given, the recipient tends to get defensive. So, strive not to make generalizations regarding the personality of the individual. Stay focused on the behaviors at hand.

Personality based example: "You're sloppy"
Behavior based example: "You missed the "exterior lights" checklist item. This is a problem because _____."

When feedback is <u>specific</u> to a behavior and an explanation is given as to <u>why</u> the behavior caused a problem, the individual can understand exactly what the problem was and why it was a problem. He/she knows what to practice to improve.

- *Offer both positive and negative points.*

It is important for learners to recognize when they exhibit a positive behavior, and positive, specific feedback will reinforce the leaner to use that behavior again. In addition, providing positive feedback first before providing negative feedback helps encourage the recipient to listen to the negative feedback.

Receiving only negative feedback can be highly de-motivating, and the learner may actually stop listening to the message. Additionally, when errors are corrected in a rude manner and only negative feedback is given, the individual may not believe the feedback was correct.

Also, instructors should refer back to previous comments from previous scenarios. This can help cement in improved performance or the continued need for improvements.

- *Encourage, guide, and reinforce the learner as s/he participates in the debrief.*

Most instructors have encountered instances where a learner gives a wrong answer. This can also occur in the case of the LCG debrief. When this occurs, it is vital for the instructor to respond in a manner that corrects the mistake while at the same time does not discourage the leaner from future input in the debrief.

For example, if a student pilot admits a mistake:

- Don't make light of the mistake (e.g., that wasn't a big problem, don't worry about it)
- Don't just move on after an awkward silence

- DO reinforce the individual for identifying the mistake and his/her willingness to tell the CFI
- DO make sure the student pilot knows the correct answer or response

For example, if a student pilot gives a bad solution or suggestion:
- Don’t ridicule or make fun of it
- Don’t shoot down the idea without explaining why
- Don’t let the learner think it was a good idea
- DO reinforce the individual for offering the idea, but explain why it is incorrect.

Section 3: Demonstration and Hands on Practice.

At this point in the training, the instructors need to see an example of the LCG debrief process. Both ineffective debriefing behaviors as well as effective debriefing behaviors should be demonstrated. The instructor should then be asked to critique the debrief. This could be accomplished with a live model or via example debriefs shown via video recording.

Next, the instructor needs to practice leading the LCG debrief and be given feedback. This could be a role-play situation or with an actual flight student. It is crucial, however, for the instructor-trainee him or herself to be observed by another instructor familiar with the LCG debrief who can then give the instructor-trainee effective feedback on his/her performance.

Report 3: References

Blickensderfer, E L., & Doherty, N. (2006). *FY 2005 FITS Instructor Education Research Report 1: Investigation of Current Practices Relating to Learner Centered Grading and Debriefing Procedures*. Florida, USA: Embry-Riddle Aeronautical University, Human Factors and Systems Department.

Dweck, C. S. (1986). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.

Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal, 33,* 87-105.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences on individual feedback on behavior in organizations. *Journal of Applied Psychology, 64,* 349-371.

Jacoby, J., Mazursky, D., Troutman, T., & Kuss, A. (1984). When feedback is ignored: Disutility of outcome feedback. *Journal of Applied Psychology, 69,* 531-545.

Nadler, A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance, 23,* 309-338.

Tannenbaum, S.I., Smith-Jentsch, K.A.& Behson, S.J. (1998). Training Team Leaders to Facilitate Team Learning and Performance. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 247 – 270). Washington, D.C.: American Psychological Association.

Smith- Jentsch, K. A. (2002). *Team dimensional training facilitator workshop. Videotape.* Sponsored by NAVAIR Orlando. Produced by Sonalysts, INC. Orlando, FL.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., & McPherson, J.A. (1998). Team dimensional Training: A Strategy for Guided Team Self-Correction. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 271 – 297). Washington, D.C.: American Psychological Association

Wexley & Latham (2002). *Developing and training human resources in organizations* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

FY2005 FITS Instructor Education Research
Report 4: Empirical Investigation of the Learner Centered Grading Debriefing Approach

December 15, 2006
(Revised according to government comments)

Submitted by:

Beth Blickensderfer, Ph.D.
Justin Jennison
Embry-Riddle Aeronautical University

This effort was the final deliverable under the FITS Program FY2005: FITS Instructor Education Research, "Examining the Debriefing Process Including the Learner Centered Grading Approach."

Empirical Investigation of the Learner Centered Grading Debriefing Approach

Introduction

Although aviation has moved into a technologically advanced era, flight training has largely stayed the same. Emphasis on stick-and-rudder skills and repetition of standard flight maneuvers may not be the most effective strategy to train pilots to fly the new technically advanced aircraft (TAA) (French, Blickensderfer, Summers, Ayers, & Connolly, 2005). In response to this need, the FAA/Industry Training Standards program (FITS) was born. FITS is not a regulatory entity, but rather a voluntary collaboration of industry leaders comprised of aircraft manufacturing, training, universities, insurance companies and trade associations, as well as the General Aviation Center of Excellence and the United States Federal Aviation Administration. The FITS team has been working together to develop training guidelines that fall within the boundaries of current regulations and yet incorporate the latest findings in training research.

FITS and the Learner Centered Grading Debrief

Briefly, the FITS approach utilizes a scenario-based training strategy, wherein flight instruction is given in the form of realistic scenarios. One component of the FITS method is to use the concept of "learner-centered grading" (LCG) following each flight scenario. LCG includes two parts: learner self assessment and a detailed debrief by the instructor. The purpose of the self assessment is to stimulate growth in the learner's thought processes and, in turn, skill acquisition. The self-assessment is followed by an in-depth discussion between the instructor and the student which compares the instructor ratings to the student's self-assessment. Thus, the debriefing process changes from an instructor-led critique of performance to a student-led and instructor-facilitated analysis of the student's performance. Previous laboratory research (French et al., 2005) indicated that the FITS approach, including the notion of LCG is effective for training certain skills. However, it is not clear how much of the effect of the FITS approach is associated with the LCG debrief alone. Furthermore, research was needed to assess how the LCG process was being used in the field.

LCG in the Field

In an effort to see how the learner centered grading debrief was being used in the field, Blickensderfer and Doherty (2006) conducted in-depth interviews with several certified flight instructors (CFIs). Overall, the CFIs reported that the in-depth discussions were an improvement over traditional instructor-led debriefings. Difficulties, however, were also evident, and a variety of questions were noted.

Of particular interest to this study were questions regarding the type of scale to use during the student- pilot self-assessment and the instructor assessment portion of the approach (i.e., when both the instructor and the student rate the student pilot's skills and abilities). The scale commonly used consists of 5 categories: describe, explain, practice, perform, and manage/decide. The student pilot's performance on each of a list of tasks is assessed using those categories (both by the student him/herself and by the instructor). This is equivalent to making ratings on a scale of 1 -5 (although the numbers are not presented) for each of the listed tasks. Ratings of this nature are much more cognitively complex than, for instance, checking off whether or not (yes/no) a behavior occurred during flight (i.e., using a behavioral checklist). Given the many other demands on the instructor when observing a student pilot perform a complex a task, using a behavioral checklist would likely to be more accurate yet require less diversion of attention from the learner's performance. From the learner perspective, given the many demands on the student while performing the tasks, using a behavioral checklist would also likely to be more accurate and feel less overwhelming to rate oneself. Research is needed, however, on the effectiveness of the LCG debrief using a simplified assessment tool.

The purpose of this study was two fold: 1) to isolate the LCG debriefing approach and demonstrate, empirically, the value added of the LCG grading debrief (to the scenario based training), and 2) to examine the efficacy of the LCG approach using a behavioral checklist as the tool for the instructor and the learner to assess the learner's performance. The study focused on improving the task management and single pilot resource management related skills of low hour pilots and student pilots who were close to taking their certification exam.

Hypothesis

When using the FITS approach (scenario based training) to train task management and single pilot resource management related skills, the individuals who are debriefed using the LCG debrief strategy will demonstrate better performance on task management and single pilot resource management related skills than will the individuals who are debriefed using the traditional debrief strategy.

Method

Participants

The 31 participants (27 male, 4 female) were students at a southeastern university. The ages ranged from 17 to 26 years with an average age of 19.6 (SD = 1.9) years. All participants had between 31 and 200 logged flight hours. The average logged visual flight hours (VFR) was 86.2 (SD = 33.4). Logged instrument flight hours (IFR) ranged from zero to 50 hours, with an average of 10.2 (SD 13.3) hours.

Level of comfort with flight simulators. Participants indicated their comfort level with flying flight simulators on a Likert scale of 1 – 7 (1 = very uncomfortable, 7 = very comfortable). The mean responses was 4.8 (SD = 1.7), and the range was 1.50 to 7.0.

Level of comfort with the Piper Arrow. Participants also indicated their comfort level with flying the Piper Arrow on a Likert scale of 1 – 7 (1 = very uncomfortable, 7 = very comfortable). The mean responses was 4.5 (SD = 1.4), and the range was 1.50 to 7.0.

Apparatus

The Elite G501 iGATE flight simulation device was used in the study. The device offers advanced PC-based Instrument Flight training on the traditional display systems. The device was configured to emulate the Piper Arrow aircraft and included displays, yokes, instrument banks and rudders. An overhead projector casts the outside view on a large projection screen directly in front of the pilot. The system is driven by Microsoft Flight Simulator (2002). A nearby computer running the MSFlightSim instructor station allowed the experimenter and instructor to control the scenarios.

Scenarios

Two scenarios were used in the study. The scenarios were designed to be parallel in terms of difficulty and skills required but not to be identical. To accomplish this, the two scenarios took place in different geographic locations and included variations of, yet similar, events. The instructor scenario guides are shown in Appendix A. The scenarios were designed to emphasize pilot skills related to task management and use of resources in the cockpit. In both scenarios, two research assistants alternated playing the role of the passenger (described to be a non-pilot, friend).

Scenario 1 (Pre-test). In Scenario 1, the participant and a friend flew from Daytona Beach, FL to Tampa, FL to attend a family reunion. Events in the scenario included a problem with retracting the landing gear after take-off, a vacuum failure, an ATC re-route, and a temporary closing of the destination airport due to weather.

Scenario 2 (Post-test). In Scenario 2, the participant and a friend flew from Daytona Beach, FL to Jacksonville, FL to attend the superbowl. Events in the scenario included a pitot tube blockage, closure of the destination airport, an engine failure, and a problem with the landing gear.

Flight instructor

A CFI performed the duties of the flight instructor during the experiment.

Debriefing Strategy (Independent Variable).

The debriefing strategy was the independent variable (traditional debrief vs. LCG debrief).

Traditional Debrief (Control condition).
The traditional debrief was designed to be highly instructor led, with the instructor providing feedback in a straightforward style regarding what was right and wrong about the participant's performance. If the participants asked questions, the instructor answered the questions. The traditional debrief style was based on the comments of CFIs in Blickensderfer & Doherty (2006) as well as the opinions of other CFIs.

LCG Debrief (Experimental condition). The LCG debrief began with participants completing a self-assessment form by him/herself (see Appendix B). When the participant had completed the self-assessment, the debrief began. While the debriefs varied somewhat from participant to participant, in general the instructor emphasized the following**:** asking for the learner's opinions/views of his/her respective performance, discussing discrepancies between the instructor assessment and the student pilot's self-assessment, focusing on concrete examples, encouraging the learner to think beyond this scenario and consider the potential impact on future flights, discussing both positive and negative examples, providing behavior based feedback, and reinforcing the participant for participating in the debrief itself. In general, the instructor used the structure of the scenario to guide the debrief (i.e., the debrief was organized event by event).

Questionnaires

The questionnaires used in the study were the self-assessment questionnaire, a self-efficacy questionnaire, and a reactions survey regarding the debrief.

Self-Assessment Questionnaire. The self-assessment questionnaire consisted of a one page list of scenario events and yes/no questions regarding the student pilots' performance on the events. The assessment was designed to parallel the instructor's assessment measure (see appendix B). The assessment questionnaire was tailored to each scenario. This questionnaire was used only for the LCG debrief itself; no data from the self-assessment was analyzed in the results section of this paper.

Self-efficacy Questionnaire. Self-efficacy is an individual's belief in their ability to succeed at a specific task (Bandura, 2000) and is an important consideration in training evaluations (Gist, 1989). A high level of self-efficacy is linked to superior performance (Bandura, 1997; Bandura, 2000). According to Locke & Latham (1990), the mean correlation for self-efficacy and performance goal setting is also estimated to be fairly high across studies (r = .39). The participants' self-efficacy for piloting aircraft was assessed with a ten Likert question survey (see Appendix B). The scale for this study was adapted from a validated scale for self-efficacy (Riggs, 1989). A reliability analysis indicated an appropriate level of internal consistency (Cronbach's alpha = .79 pre-test, and .70 post test) for the questions to be combined together in one score, and this was done.

*Reactions Questionnair*e. This measure was designed to be a manipulation check (see Appendix B for the questionnaire). It assessed the participants' perceptions of the degree to which that during the debrief the instructor: asked for the learner's opinions regarding his/her own performance, encouraged the learner to think about his/her own performance, and encouraged the learner to think beyond this scenario and consider the potential impact on future flights. Additionally, the participants responded according to how similar/dissimilar the debrief was to other flight instruction debriefs they had experienced. Internal consistency was assessed as .79 (Cronbach's alpha; for both pre- and post- tests). The responses were reverse coded when necessary and combined into an average reaction score for the first debrief and an average reaction score for the second debrief.

Performance Metrics.

The metrics selected for the study aimed to assess *task managemen*t and *use of resources in the cockpit*. The metrics were all subjective assessment metrics (i.e., rating scales and behavioral checklists; see Appendix C). Performance was assessed by two independent raters. During the scenarios, the raters made independent assessments of each participant's performance. Following each scenario, the raters met and came to consensus on each of the ratings. The item by item ratings were later condensed into the variables described below.

Using Passenger as a Resource. The participants were rated on the degree to which they used the passenger as a resource during various events in the scenarios. Ratings were given on a 1-5 scale, with 1 being ineffective and 5 being highly effective (the exact definition of ineffective and effective varied according to the event, see Appendix C). These ratings were later combined into one average score for "using passenger as a resource" for the pre-test and one average score for "using passenger as a resource" during the post-test.

Notifying and debriefing the passenger. The participants were rated on the degree to which they notified and briefed the passenger regarding flights events during various times in the scenarios. Ratings were given on a 1-5 scale, with 1 being ineffective and 5 being highly effective (the exact definition of ineffective and effective varied according to the event, see Appendix C). These ratings were later combined into one average score for "notifying and debriefing passenger" during the pre-test and one average score for "notifying and debriefing passenger" during the post-test.

Checklist use. The participants were rated on the degree to which they used the checklists as a resource during various times in the scenarios. Ratings were simple yes/no responses. To combine these scores, the "yes" responses were totaled and divided by the total possible yes responses to create a percentage score. This was done for both the pre-test and the post-test, yielding one percentage score for "checklist use" during the pre-test and one percentage score for "checklist use" during the post-test.

ATC use. The participants were rated on the degree to which they used the ATC as a resource during various times in the scenarios. For certain ATC events, the ratings were simple yes/no responses. To combine these scores, the "yes" responses were totaled and divided by the total possible yes responses to create percentage score. This was done only for the post-test, yielding one percentage score for "ATC use" during the post-test. In addition, for other ATC events, the ratings were given on a 1-5 scale (1 = ineffective, 5 = effective (the exact definition of ineffective and effective varied according to the event, see Appendix C). These ratings were later combined into one average score for "using ATC as a resource average" during the pre-test and one average score for "using ATC as a resource average" during the post-test.

POH and AP use. The participants were rated on the degree to which they use the Pilot Operating Handbook (POH) and Autopilot (AP) as a resource. Ratings were simple yes/no responses. To combine these scores, the "yes" responses were totaled and divided by the total possible yes responses to create percentage score. This was done for both the pre-test and the post-test, yielding one percentage score for "POH and AP use" during the pre-test and one percentage score for "POH and AP use" during the post-test.

Overall effectiveness. At the end of each event within each scenario, the participants were rated on overall effectiveness. Ratings were given on a 1-5 scale, with 1 being very ineffective and 5 being highly effective (the exact definition of ineffective and effective varied according to the event, see Appendix C). These ratings were later combined into one average score for "overall effectiveness" during the pre-test and one average score for "overall effectiveness" during the post-test.

Procedure

Upon arrival at the experimental site, the participant was greeted by the experimenter, briefed on the study, and then the participant signed the consent form. Next, s/he completed the demographics questionnaire and the self-efficacy questionnaire. Next, to familiarize the participant with the simulation device, the CFI briefed the participant on the simulation device, and the participant flew a practice flight. During the flight the CFI pointed out various aspects of the simulation device as well as characteristics of the Piper Arrow aircraft.

Once the familiarization phase was complete, the CFI pre-briefed the participant on the flight occurring in Scenario 1. Next, the participant flew scenario 1. Upon completion of the first scenario, participants in the experimental group (LCG debrief) completed the first self-assessment. The control group (traditional debrief) did not complete the self-assessment. The CFI then debriefed the participant using the LCG debriefing style for the individual's in the experimental group and the traditional debriefing style for the individuals in the control group. At the end of the debrief, all participants completed the reactions survey. The participant was then given a 10 minute break.

Next, the CFI pre-briefed the participant on the flight in Scenario 2. The participant than flew the second scenario. Again, upon completion of the second scenario, participants in the experimental group completed the first self-assessment. The control group participants did not complete the self-assessment. The CFI then debriefed the participant using the LCG debriefing style for the individuals in the experimental group and the traditional debriefing style for the individuals in the control group. At the end of the debrief, all participants completed the reactions survey once again. The experimenter then debriefed the participant on the experiment, the participant was compensated for their time ($40.00), and the participant was dismissed.

As a final note, at the approximate half-way point during each flight scenario, the instructor fast forwarded the flight to an area near the final destination. This was done due to experimental time constraints. The time required for the entire experimental session was approximately 3.5 hours. The experimental timeline is shown in Table 1.

Table1: Experimental Timeline

Time	Activity
5 minutes	Experimental pre-brief Informed consent Demographics Self-efficacy questionnaire I
10 minutes	Practice flight, familiarization with simulator
10 minutes	Scenario I Preflight
50 minutes	Scenario I Flight
5 minutes	Self assessment I (only for LCG group)
15 – 30 minutes	Scenario I Debrief
5 minutes	Reactions questionnaire I
10 minutes	Break
5 minutes	Self-efficacy questionnaire II
10 minutes	Scenario II Preflight
40 minutes	Scenario II Flight
5 minutes	Self-Assessment II (only for LCG group)
15 – 30 minutes	Scenario II Debrief
5 minutes	Reactions questionnaire II
5 minutes	Experimental debrief and Participant payment information
3.25 - 3.75 hours	Total Time

Results

The means for each of the dependent measures/performance metrics are shown in Table 2. The analyses began with the manipulation check**.**

Table 2: Means for Dependent Measures

Variable	Condition	Pre Mean	Pre S.D.	Post Mean	Post S.D.	Change Score Mean	Change Score S.D.
Reaction Survey	LCG	4.3	0.5	4.3	0.4	0.3	0.2
	Traditional	3.5	0.5	3.5	0.5	0.3	0.2
Self Efficacy Questionnaire	LCG	3.6	0.5	3.5	0.5	0.2	0.3
	Traditional	3.8	0.5	3.6	0.6	0.2	0.2
Average Used Passenger as a Resource	LCG	1.2	0.8	3.2	1.0	2.2	1.0
	Traditional	1.3	0.6	2.0	1.2	1.1	1.2
Average Notifying and Briefing	LCG	2.4	0.9	3.3	1.1	1.1	1.1
	Traditional	3.1	0.3	2.6	1.0	0.8	0.7
Percent Checklist Use	LCG	29.2	10.4	78.5	9.9	50.9	16.4
	Traditional	40.0	19.2	65.7	16.5	25.7	18.3
Percent ATC Use	LCG	n/a	n/a	35.7	24.3	n/a	n/a
	Traditional	n/a	n/a	42.2	34.4	n/a	n/a
Average ATC Use	LCG	2.8	0.9	3.6	0.6	1.0	0.6
	Traditional	3.1	0.8	3.4	0.8	0.9	0.6
Percent POH and AP Use	LCG	4.6	8.7	23.3	25.8	21.5	22.7
	Traditional	5.3	9.2	17.9	24.9	19.3	22.4
Overall Effectiveness Total	LCG	12.9	2.6	16.7	3.3	4.2	3.1
	Traditional	16.1	2.4	15.9	3.2	1.8	1.4

Manipulation Check.

The first analysis was a manipulation check to determine whether the debriefings did actually differ between the two conditions (traditional and LCG). The measure used was the reactions questionnaire. s shown in Table 2, the mean reaction score for those individuals receiving the traditional style debriefs were 3.5 (SD = .5) and 3.5 (SD =.5) for debrief 1 and debrief 2, respectively. For those individuals receiving the LCG debrief, the mean reaction scores were 4.3 (SD = .5) and 4.3 (SD = .4) for debrief 1 and 2, respectively. A two-way mixed analysis of variance (ANOVA) was used to analyze the data. The main effect for condition was significant, $F(1, 29) = 23.68$, $p = .00$. A partial eta squared of .45 indicated that 45% of the variance in the reactions survey responses was accounted for by condition. Unsurprisingly, the main effect for trial (debrief 1 vs. debrief 2) was not significant with $F(1, 29) = .40$, $p = .54$. Also unsurprisingly, the interaction effect was not significant $F(1, 29) = .79$, $p = .39$.

The means are graphed in Figure 1. It can be seen that the LCG groups rated their debriefings higher than did the traditional group, but that, within the conditions, the first and second briefs were approximately the same. Thus, the individuals who received the LCG debrief rated the debriefings significantly higher in terms of the reactions measure than did the individuals who received the traditional debriefings.

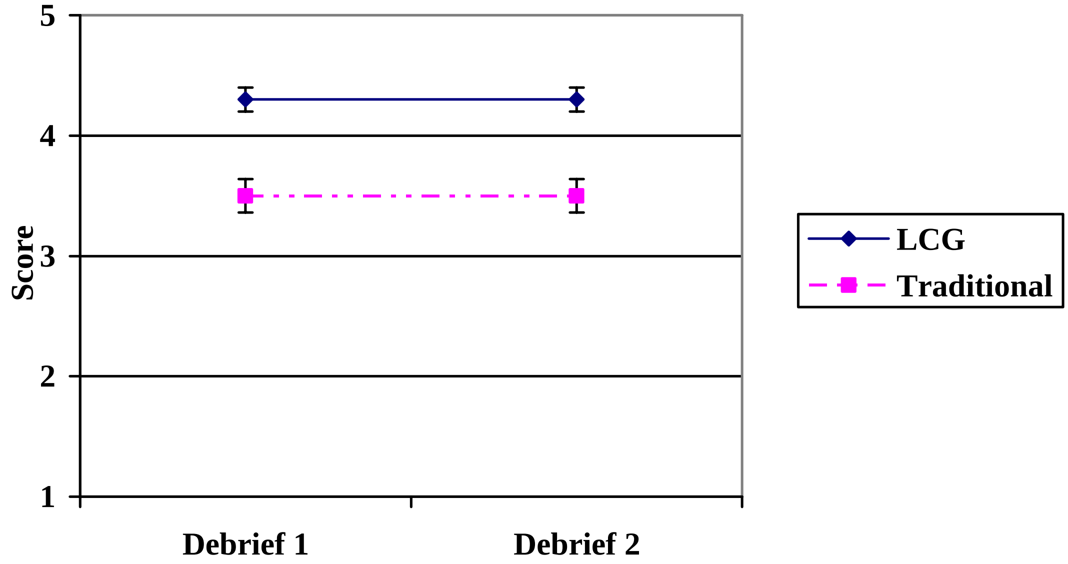

Figure 1: Manipulation Check - Reaction Survey

Analyses of the Performance Metrics

In order to assess performance differences between the two study conditions, two-way mixed (between and within) ANOVAs were used to analyze each of the dependent measures when a pre-test was available. In these analyses, the most important part of the results was the interaction effect. Thus, the interactions are emphasized in the write-up. When pre-tests were not available, independent groups t-tests were used. In addition, change scores (from the pre-test to the post test) were computed for most variables. The change scores were analyzed with independent groups t-tests.

While the authors realized the issue of inflation of the family-wise error rate associated with multiple ANOVAs and t-tests, a decision was made to forego a multivariate approach for the sake of assessing the individual contributions of each of the dependent measures.

Using Passenger as a Resource

As shown in Table 2, the mean ratings for “using passenger as a resource” for the traditional group were 1.3 (SD = .6) and 2.0 (SD = 1.2) for the pre-test and post-test, respectively. The mean ratings for “using passenger as a resource” for the LCG group were 1.2 (SD = .8) and 3.2 (SD = 1.0) for the pre-test and post-test, respectively. A two-way mixed ANOVA revealed that a significant effect existed for the interaction with $F(1, 26) = 4.90$, $p = .04$. The partial eta squared of .16, indicated that 16% of the variance was accounted for by the interaction of condition and trial (pre-test, post-test).

These means are graphed in Figure 2. Looking at Figure 2, it appears that the groups did not differ in “using passenger as a resource” during the pre-test, however, the LCG group seems to have improved during the post-test.

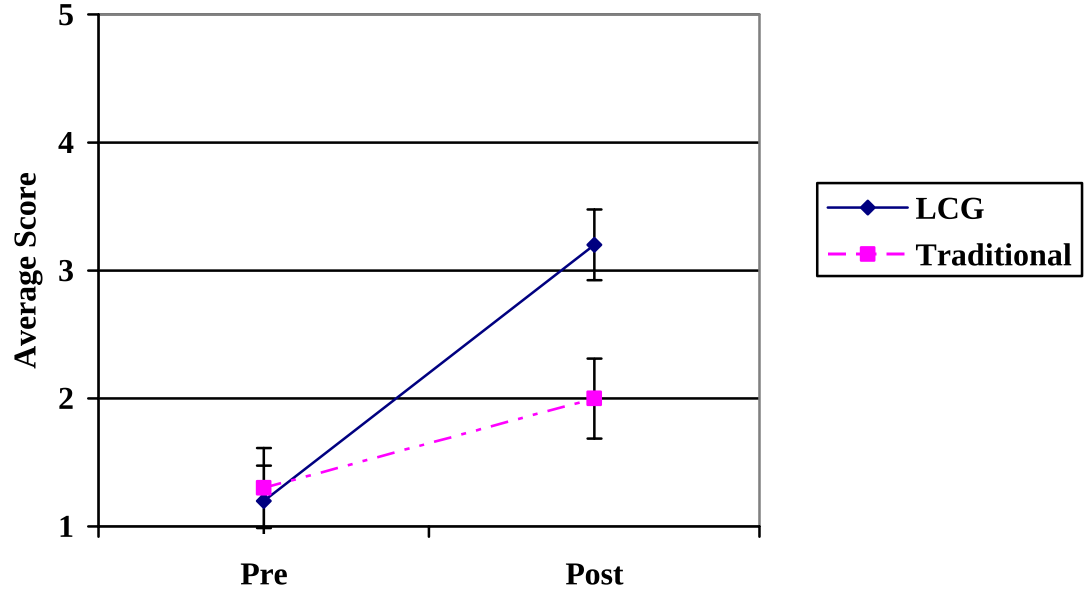

Figure 2: Using Passenger as a Resource

Subsequent post hoc comparisons using Tukey HSD (HSD = .86) revealed that the traditional debriefing group and the LCG debriefing group did not differ significantly on “using passenger as a resource” in the pre-test (scenario 1). While the mean for both groups was higher in the post-test (scenario 2), only the LCG debriefing group achieved a mean significantly higher than the pre-test mean. Furthermore, the LCG group mean for the post-test “using passenger as a resource” was also significantly higher than the traditional debriefing group mean for the post-test “using passenger as a resource.”

Using Passenger as Resource Change Score. The mean change scores for “Using passenger as a Resource” were computed for the traditional group and the LCG group. As shown in Table 2, the traditional group mean change score was +1.1 (SD = 1.2) while the LCG group mean change was +2.2 (SD = 1.0). The change scores were compared using an independent groups t-test, and a significant difference did appear $t(26) = 2.61$, $p = .015$ (two tailed). Thus, the LCG group had a significantly greater change score from the pre-test to post-test than did the traditional debriefing group.

Summary: Both the LCG and the Traditional group means increased from the pre- to the post-test. However, the LCG increased more than the traditional group.

Notifying/Briefing Passenger

As shown in Table 2, the mean ratings for “notifying/briefing passenger” for the traditional debriefing group were 3.1 (SD = 1.1) and 2.6 (SD = 1.0) for the pre-test and the post-test, respectively. The mean ratings for “notifying/briefing passenger” for the LCG debriefing group were 2.4 (SD = 0.9) and 3.3 (SD = 1.1) for the pre-test and the post-test, respectively. A two-way mixed ANOVA indicated that a significant interaction was present with $F(1, 25) = 14.24$, $p = .001$. A partial eta squared of .36, indicated that 36% of the variance in “notifying/briefing the passenger” was accounted for by the interaction of condition and trial (pre-test, post-test).

The means are graphed in Figure 3. Looking at Figure 3, it appears that the traditional group was rated higher than LCG group to begin with, but actually ended up slightly lower than the LCG group in the post-test. These apparent differences were analyzed further with post hoc comparisons using Tukey HSD (HSD = .16). The results revealed that, in the pre-test, the traditional group did perform significantly better on "notifying/briefing passenger" than did the LCG group. However, in the post-test, the traditional group scored significantly *lower* than they had in the pre-test. At the same time, the LCG group scored significantly *higher* in the post-test than they had in the pre-test. Furthermore, the LCG group mean for the post-test "notifying/briefing passenger" was significantly higher than the traditional group post-test mean.

Summary: The LCG group got better, but the traditional group actually declined.

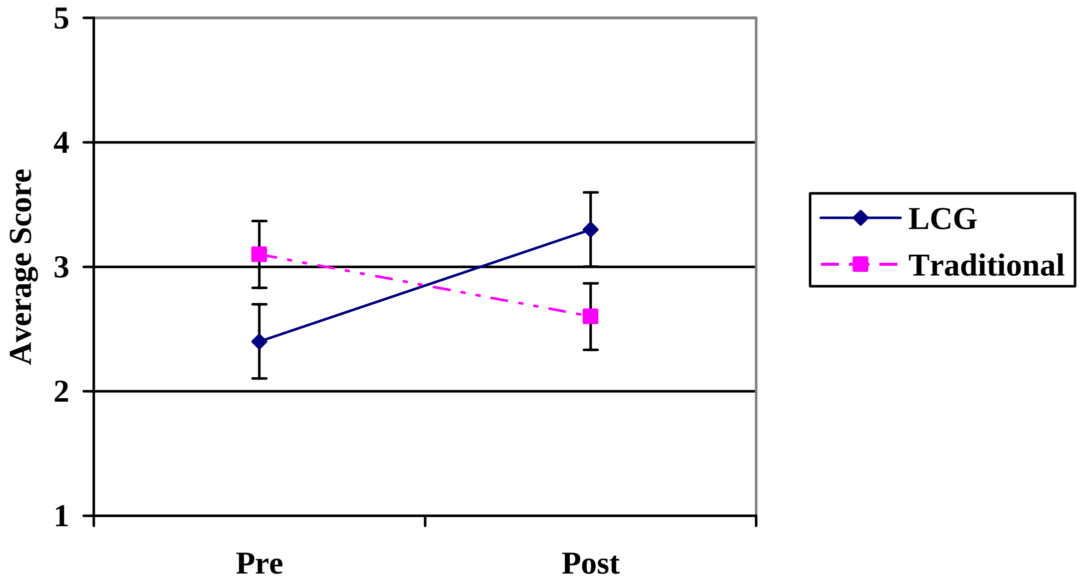

Figure 3: Notifying and Briefing the Passenger

Checklist Use

As shown in Table 2, the mean scores for "checklist use" for the traditional debriefing group were 40.0 percent (SD = 19.2) and 65.7 percent (SD = 16.5) for the pre-test and post-test, respectively. The mean scores for "checklist use" for the LCG debriefing group were 29.2 percent (SD = 10.4) and 78.5 percent (SD = 9.9).

A two-way mixed ANOVA indicated that a significant interaction effect was present with $F(1, 23) = 12.79$, $p = .002$. The partial eta squared equaled .36. This indicated that 36% of the variance in checklist use was accounted for by the interaction of condition and trial (pre-test, post-test).

These means are graphed in Figure 4. It appears that during the pre-test the two groups scored somewhat similarly, although the traditional group did score higher than LCG group. In contrast, although both groups improved from the pre-test to the post-test, the LCG group improved more than did the traditional group and ended up scoring higher than the traditional group.

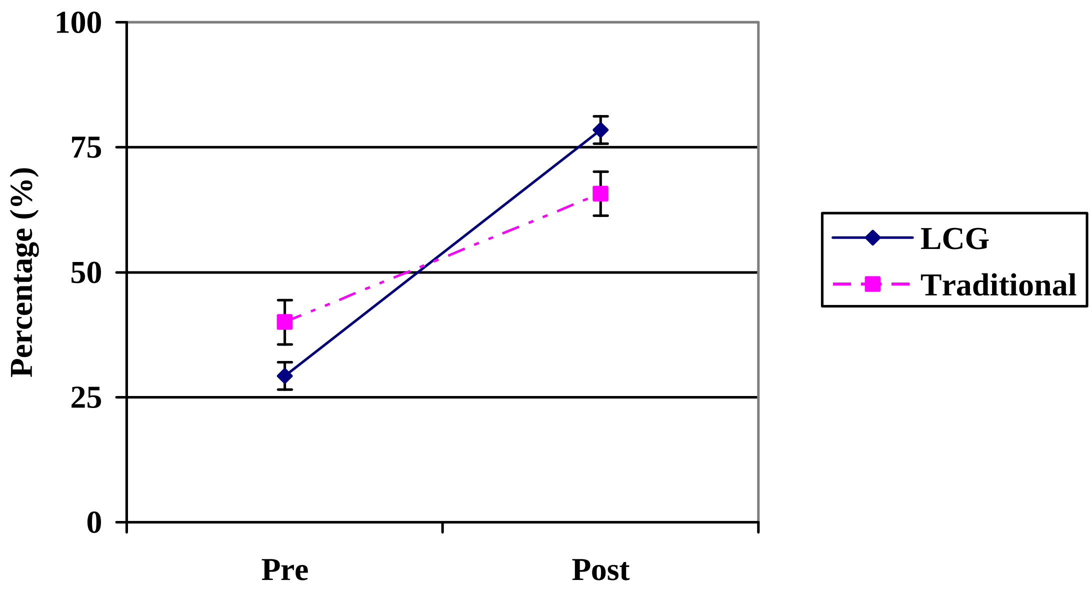

Figure 4: Checklist Use

These apparent differences were analyzed further with Tukey HSD post-hoc comparisons (HSD = 32.3). The pre-test means did not differ significantly. The post-test means did not differ significantly. However, the LCG group post-test score was significantly higher than the LCG pre-test score, whereas the traditional group's pre-test score did not differ significantly from its post-test score. *Thus, the interaction effect was driven by the greater response of the LCG group to the debriefing.*

Checklist Use Change Score. The mean change scores for "checklist use" were computed for the traditional group and the LCG group. As shown in Table 2, the traditional group's mean change score was + 25.7, while the LCG group mean change was +50.9. The change scores were compared using an independent groups t-test, and a significant difference did appear $t(23) = 3.58$, $p = .002$. Thus, the LCG group had a significantly greater change score than did the traditional debriefing group.

ATC Use

One of the two dependent variables assessing ATC use had a pre-test score. This was the assessment of ATC use using a 1-5 rating scale and is referred to as "average ATC use." As shown in Table 2, the means for the "average ATC use" metric for the traditional group were 3.1 (SD = .8) and 3.4 (SD = .8) for the pre-test and the post-test, respectively. The means for "average ATC use" for the LCG group were 2.8 (SD = .9) and 3.6 (SD = .6) for the pre-test and the post-test, respectively. A two-way mixed ANOVA indicated that the interaction effect was not significant, $F(1, 27) = 1.62$, $p = .21$. The graph of the means are shown in Figure 5.

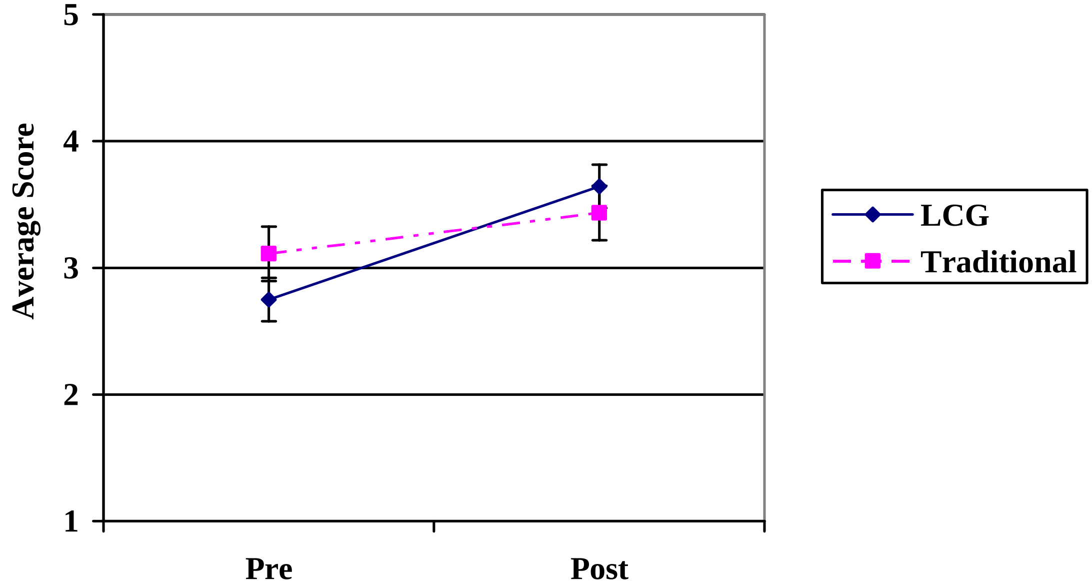

Figure 5: ATC Use

Average ATC Use Change Score. After examining the results of the ANOVA analysis of average ATC use described above, we decided to examine the mean change scores (from the pre-test to the post test). The mean change scores for "average ATC use" were computed for the traditional group and the LCG group. As shown in Table 2, the traditional group's mean change score was + .9 (SD = .6) while the LCG group mean change was +1.0 (SD = .6). The t-test for the mean change was not significant, t (27) = .310, p = .76 (two tailed). Thus, the two study conditions exhibited approximately equal change in "avg ATC use" from the pre-test to the post-test. The other metric assessing ATC use was a percentage ("Percent ATC Use"). The mean post-test score for the traditional group was 42.2 (SD = 34.4). The mean pre-test score for the LCG group was 35.7 (SD = 24.3). These means did not differ significantly, t (27) = -.59, p = .56.

POH and Autopilot Use

As shown in Table 2, the mean scores for "percent POH and Autopilot Use" for the traditional group were 5.3 (SD = 9.2) and 17.9 (SD = 24.9) for the pre-test and post-test, respectively. The mean scores for "POH and Autopilot Use" for the LCG group were 4.6 (SD = 8.7) and 23.3 (SD = 25.8) for the pre-test and post-test, respectively. A two-way mixed ANOVA indicated a main effect for trial, F(1, 25) = 10.24, p = .004, observed power = .867. Partial eta squared of .30 indicated that 30 percent of the variance in POH and Autopilot use was accounted for by trial (pre-test, post-test). Thus, both the traditional and the LCG groups improved on POH and Autopilot use from the pre-test to the post-test. An interaction effect, however, did not appear, F(1, 25) = .239, p = .629. In other words, the LCG group was no better off than the traditional group on this measure (see Figure 6).

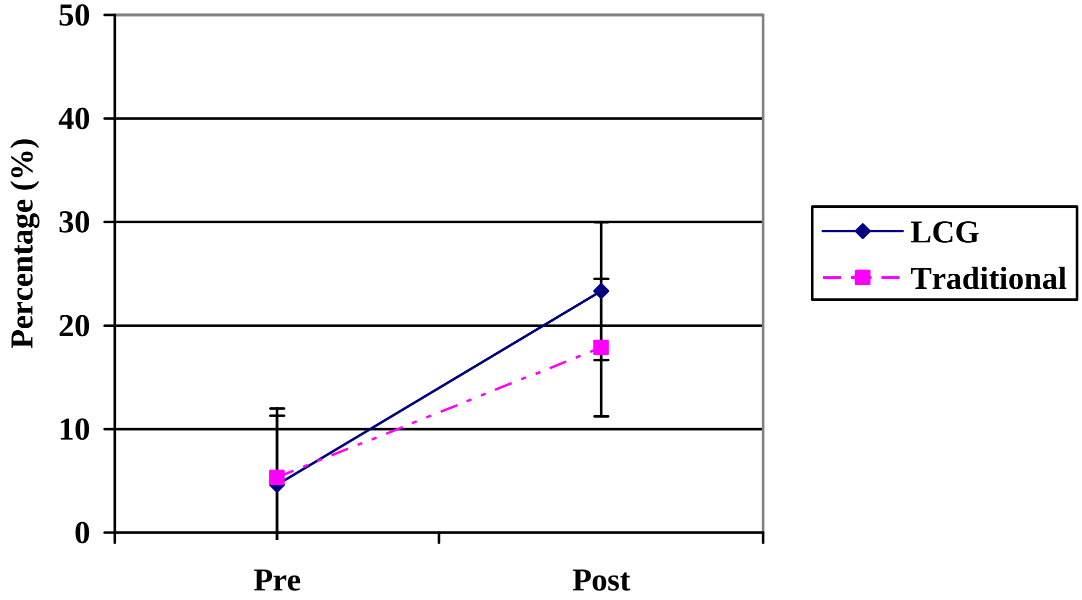

Figure 6: POH and Autopilot Use

POH and Autopilot Use Change Score. As shown in Table 2, the traditional group's mean change score was + 19.3 (SD = 22.4) while the LCG group mean change was +21.5 (SD = 22.7). These means did not differ significantly, t(25) = .260, p = .797 (two tailed). Thus, the two study conditions exhibited approximately equal change in "percent POH and AP use" from the pre-test to the post-test.

Overall Effectiveness

As shown in Table 2, the mean scores on "Overall Effectiveness" for the traditional group were 16.1 (SD = 2.4) and 15.9 (SD = 3.2) for the pre-test and the post-test, respectively. The mean scores on "Overall Effectiveness" for the LCG group were 12.9 (SD = 2.6) and 16.7 (SD = 3.3) for the pre-test and the post-test, respectively. A two-way mixed ANOVA indicated that a significant interaction effect occurred, F(1, 24) = 9.59, p = .005. Partial eta squared of .29 indicated that 29% of the variance in overall effectiveness was accounted for by the interaction of condition and trial (pre-test, post-test).

As shown in Figure 7, it appears that although the LCG group achieved a much lower score on "Overall effectiveness" than did the traditional group, their score on the post-test was actually higher than the score for the traditional group.

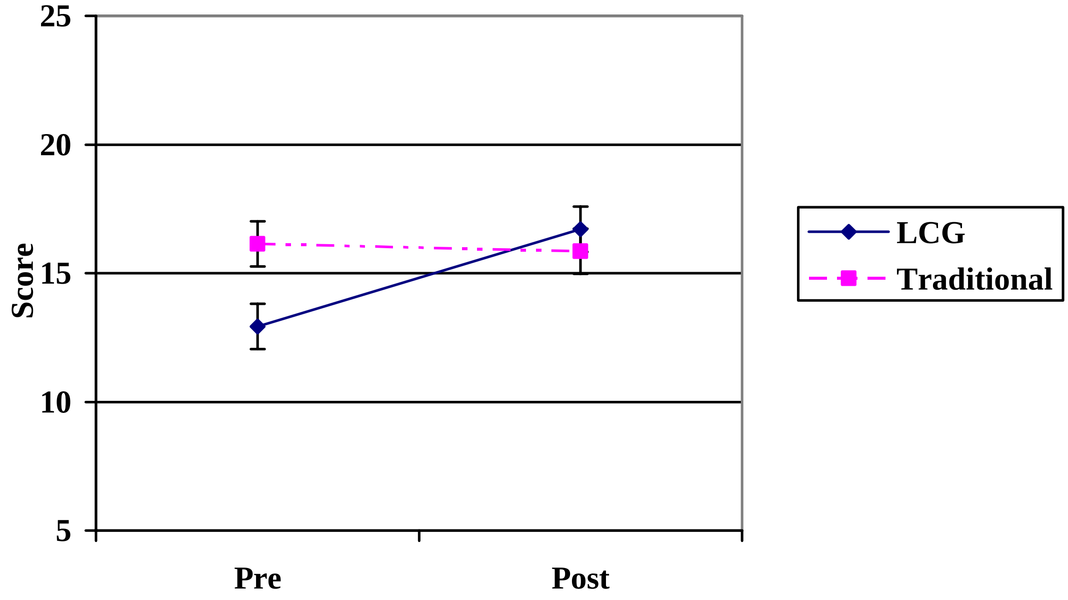

Figure 7: Overall Effectiveness

These apparent differences were analyzed further with Tukey HSD post-hoc comparisons (HSD = 1.40). The LCG group did have a significantly lower score than the traditional group on the pre-test. Additionally, the LCG group's post-test score was significantly higher than the pre-test score. However the LCG group's post-test score was not significantly higher than the traditional group's post-test score. Finally, the traditional group's pre-test and post-test scores did not significantly differ.

Summary: The interaction resulted from the response of the LCG group to the debriefing. While the traditional group was the same from pre- to post-test, the LCG group increased significantly.

Overall Effectiveness Change Score. The mean change scores for "Overall Effectiveness" were computed for the traditional group and the LCG group. As shown in Table 2, the traditional group's mean change score was -1.8 (SD = 1.4) while the LCG group mean change was +4.2 (SD = 3.1). The absolute values of the change scores were compared using an independent groups t-test, and a significant difference did appear, $t(24) = 2.54$, $p = .018$ (two tailed). Thus, the LCG group demonstrated a significantly higher change in score from the pre-test to the post-test. Furthermore, it was a positive change, while the traditional group experienced a negative change score.

Self-Efficacy

A shown in Table 2, the mean scores for self-efficacy for the traditional group were 3.8 (SD = .5) and 3.6 (SD = .6) for before and after the first scenario and debriefing, respectively. The mean scores for self-efficacy for the LCG group were 3.6 (SD = .5) and 3.5 (SD = .5). A two-way mixed ANOVA indicated that an interaction effect was not present, $F(1, 29) = .047$, $p = .83$. Nor was a main effect for condition present, $F(1, 29) = .476$, $p = .50$. However, a significant main effect for trial did appear, $F(1, 29) =$

16.89, p = .00. The partial eta squared of .37, indicated that 37 percent of the variance in self-efficacy was accounted for by trial (before scenario, after scenario).
As shown in Figure 8, the mean self-efficacy actually decreased for both groups.

Thus, participants had higher self-efficacy before the pre-test than they did at the end of their first scenario and debriefing.

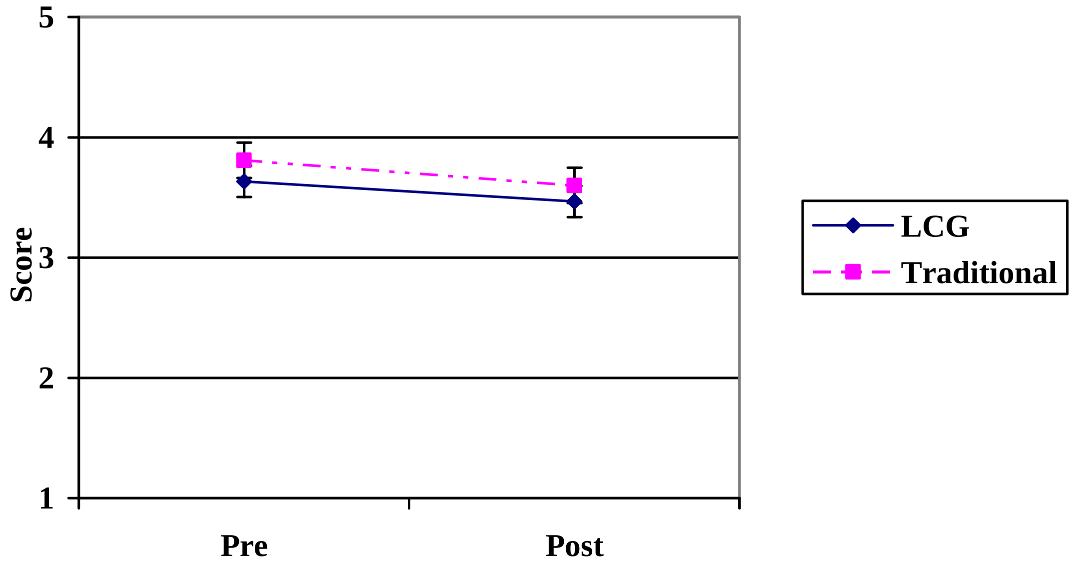

Figure 8: Self-Efficacy

Discussion

The purpose of this study was two fold: 1) to demonstrate, empirically, the value added of the LCG grading debrief to the scenario based training, and 2) to examine the efficacy of the LCG approach using a behavioral checklist as the learner self-assessment tool. It was expected that when using the FITS approach (scenario based training) to train task management and single pilot resource management related skills, the individuals who were debriefed using the LCG debrief strategy would demonstrate better performance on task management and single pilot resource management related skills than will the individuals who are debriefed using the traditional debrief strategy. The results will be discussed in two sections: support for the LCG approach and lack of support for the LCG approach. To begin, however, the manipulation check will be reviewed.

The results of the reactions survey indicated that the participants did perceive a difference between the two styles of debrief, and, hence, the manipulation was successful. Interestingly, the traditional debrief was not rated down, per se, as the average was 3.5 on the 1-5 scale (1= low, 5 = high). Thus, the traditional style of debrief, at least as given in this study was thought provoking to a degree, but the LCG style debrief was perceived as more thought provoking and as involving the participant to a greater degree. We now turn to the empirical support for the LCG debrief.

Support for LCG

Considerable support was shown for the LCG approach. Importantly, for the dependent measures, “using passenger as a resource”, “notifying/briefing passenger”, “checklist use”, and “overall effectiveness” the LCG group improved at a greater rate than did the traditional group. For those variables, the participants who received the traditional style debrief either did not improve or actually decreased in performance.

In addition, for the dependent measures “using passenger as a resource” and “notifying/briefing passenger”, the LCG group achieved higher overall post-test scores than the traditional group. Thus, for those two measures, the LCG both improved more and demonstrated a higher final score than did the traditional group.

Lack of Support for LCG

In two areas, the LCG brief did not have an effect: ATC use and POH & Autopilot use. Although a pattern of results similar to the above mentioned measures appeared, the results were not significant. The lack of effect on these variables may be due to a number of reasons. For instance, it may be that a need exists for additional instruction on these resources--more than was possible in the limited debriefing time in this study. Additionally, the lack of results may be an artifact of the experimental design—in particular that the participants did not have the opportunity to demonstrate use of these resources enough times during the flights to get an accurate measure of their performance/skill.

Study Weaknesses and Future Research

A number of questions still exist on this topic. For instance, one question is <u>why</u> the LCG approach worked. It has been argued that the LCG approach influences performance due to the learner developing metacognitive awareness and self-monitoring skills which, in turn, leads to skill acquisition. In this study, a motivational process may also have led to the performance improvements. That is, it could be that the in-depth discussions with the instructor motivated the LCG participants to try harder in scenario 2. Additionally, the LCG debriefs lasted, on average, 15 minutes longer than the traditional debriefs. Thus, it may be that the increased time of the feedback session, rather than the style of debrief, was the influence. Finally, this study was done in a flight simulation device, and as is true with other studies accomplished in similar environments, it is possible that the participants performed differently than they would in actual flight.

Implications for SBT

From the scenario based training standpoint, it is highly interesting to note that performance differences appeared based solely on the difference of debriefs. The implication is that the feedback/debriefing portion of scenario based training is crucial, as it plays a significant role in the effect of the scenario based training episode on performance (and this is true whether the influence is due to cognitive changes or motivational changes).

Summary & Conclusions

In summary, the learner centered grading style of instructional debrief appeared to be overall more effective than the traditional style instructional debrief for skills related to task management and single pilot resource management related skills. While these are not the only skills involved in flight, it is likely that the results will generalize to the acquisition of other cognitively oriented flight skills.

Report 4: References

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman

Bandura, A. (2000). Cultivate self-efficacy for personal and organizational effectiveness. In E. A. Locke (Ed.), *Handbook of principles of organization behavior* (pp. 120-136). Oxford, UK: Blackwell.

Blickensderfer, E L., & Doherty, N. (2006). *FY 2005 FITS Instructor Education Research Report 1: Investigation of Current Practices Relating to Learner Centered Grading and Debriefing Procedures*. Florida, USA: Embry-Riddle Aeronautical University, Human Factors and Systems Department.

French, J., Blickensderfer, E L., Summers, M., Ayers, F., & Connolly, T. (2005). *FITS Combined Task 1 & 2 Final Report*. Florida, USA: Embry-Riddle Aeronautical University, College of Arts and Sciences & College of Aviation.

Gist, M. E. (1989). The influence of training method on self-efficacy and idea generation among managers. *Personnel Psychology*, *42*, 787-805.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice- Hall.

Riggs, M.L. (1989). The development of self-efficacy and outcome scales for general applications. Paper presented at the 1989 meeting of the Society of Industrial and Organizational Psychology. Boston, MA.

Report 4: Appendix A

Scenario 1 Instructor Chart

SCENARIO # 1

(Flight From KDAB to KTPA For a Family Reunion)

Event	Event Description	Instructors Actions	ATC Instructions	Observed Behaviors (Pilot Will)	Pilots Action	Comments
# 1	Landing Gear Wont Retract	Fail landing gear after take off		Slow the aircraft to 129 Knots Refer to the checklist Recalculate flight plan Use passenger as a resource Contact ATC and advise Notify passengers	Y OR N Y OR N Y OR N Y OR N Y OR N Y OR N	
# 2	Vacuum Failure	Fail Pitot/Static system	If pilot does not immediately recognize failure: "N123ER, turn heading 260. Vectors for traffic at your altitude.	Recognize the failure Refer to the checklist Turn "Alternate Air" switch on Notify ATC	Y OR N Y OR N Y OR N Y OR N	
# 3	ATC Re-route due to weather	Level 3 & 4 storms along flight path	"N123ER, Be advised that there are level 4 & level 5 storms over Sanford. Stand by for a new VFR route". "N123ER, Turn heading 320 and 3500'. Proceed direct to the Ocala VOR, then on course to Tampa. Please advise with your new ETA.	Comply with ATC instructions Re-calculate flight plan Advise ATC of new ETA	Y OR N Y OR N Y OR N	
	Instructor Note: Fast Forward the event to 10 NM northeast of KTPA at 2000' heading 250					
# 4	Go Around at TPA	Have aircraft go around	"N123ER Be advised that an aircraft has crashed on Runway 18L. Runway 18L and Runway 27 are closed until further notice. Please state your intentions".	Comply with ATC instructions Analyze available options Request vectors for Runway 18R Request to fly to alternate	Y OR N Y OR N Y OR N Y OR N	
# 5	ATC Re-route due to weather	KTPA has gone IFR	"Attention all aircraft in the vicinity of the Tampa International Airport, be advised that the airfield has just gone IFR. ATIS reporting visibility of 1SM and ceiling at 500'. All aircraft please state your intentions". **If IFR:** Fly heading 360 vectors for the Runway 18R ILS Approach **If VFR:** Fly heading 180 to alternate.	Request IFR clearance into KTPA Request vectors to nearest VFR field Recalculate flight plan and ETA Use passenger as a resource	Y OR N Y OR N Y OR N Y OR N	

Scenario 2 Instructor Chart

SCENARIO # 2

(Night VFR Flight From KDAB to KCRG To Watch The Super Bowl)

Event	Event Description	Instructors Actions	ATC Instructions	Observed Behaviors (Pilot Will)	Pilots Action	Comments
# 1	Starting the aircraft from a cold/dark start. Pilot request VFR Flight Following	Open Scenario # 2 (Part 1) file. Give taxi instructions and ATC clearance to take off	"N123ER, After departure turn heading 030 climb to 3000'. Cleared for take off Runway 7L	• Give passenger brief • Use checklist to succesfully start aircraft • Request proper clearance and VFR Flight Following • Performs after take off checklists	Y OR N Y OR N Y OR N Y OR N	
# 2	Pitot Tube Blockage	Fail Pitot/Static system **IF/THEN:** If pilot does not recognize failure:	"N123ER, maintain 120Knots. Spacing for traffic at your 12:00.	• Recognize the failure • Refer to the checklist • Turn "Pitot Heat" switch on	Y OR N Y OR N Y OR N	
	Instructor Note: Fast Forward the event to 10 NM southeast of KCRG at 2000' heading 360					
# 3	ATC Notifies pilot that KCRG is closed due to an extension of the Presidential TFR	Give pilot directions after they have choosen a plan of action and a new alternate.	"N123ER, Be advised that the TFR in effect over the JAX International Airport has been extended 30 minutes. Please state your intentions. ATC instructions will be determined based upon pilots decision. ATC will vector the pilot towards the chosen airfield.	• Choose to fly to alternate or nearest suitable VFR airfield • Use ATC as a resource • Contacts ATC with intentions • Re-calculate flight plan • Notify passengers	Y OR N Y OR N Y OR N Y OR N Y OR N	
# 4	Engine Failure	Fail Primary Engine within 5 NM of the landing airfield **IF/THEN:** If pilot does not sucessfully restart the aircraft. Restart the engine for the student	Give the pilot vectors to nearest suitable landing field. Set up pilot for a pattern to the active runway.	• Establish best glide airspeed (79) • Refer to the checklist • Utilize passenger in helping locate a safe place to land. • Attempt to restart the aircraft • Notify ATC • Brief passengers on situation	Y OR N Y OR N Y OR N Y OR N Y OR N Y OR N	
# 5	Landing Gear will not extend	Fail landing gear motor **IF/THEN:** If pilot does not recognize failure. Have the aircraft Go Around	"N123ER turn final Runway 00, Cleared to land Runway 00 "N123ER, It appears as though your landing gear are not down and locked, Go around!	• Recognize the failure. • Go around • Notify ATC • Refer to the checklist • Manually extend the gear • Notify passengers of situation • Land succesfully at airport	Y OR N Y OR N Y OR N Y OR N Y OR N Y OR N Y OR N	

Report 4: Appendix B

Questionnaires

Demographics Questionnaire

_____ M   _____ F   Age: ______

Years of piloting experience: ______   Total flight hours: ______

Please list all aircraft licenses and ratings:

Please list how many instrument hours you have: __________

How comfortable do you feel flying a computer simulated Piper Arrow?

|--------|--------|--------|--------|--------|--------| N/A
Very UNcomfortable   Very Comfortable

How comfortable do you feel "flying" in Microsoft Flight Simulator or another simulation device?

|--------|--------|--------|--------|--------|--------| N/A
Very UNcomfortable   Very Comfortable

Where did you receive your training for your private pilot certificate?

Where did you receive you training for IFR?

Reactions Survey

	To no Extent				To a Great Extent
1. The debrief seemed about the same as other debriefs I've had during flight training.	1	2	3	4	5
2. The instructor asked many questions during the debrief.	1	2	3	4	5
3. During this experiment, I began thinking about my own performance more than I usually do.	1	2	3	4	5
4. The instructor mainly gave his opinion of the flight and did not ask for my opinion.	1	2	3	4	5
5. The instructor did NOT ask many questions during the debrief.	1	2	3	4	5
6. The debrief was more detailed than most other debriefs I've had during flight training.	1	2	3	4	5
7. The instructor asked me for my opinion regarding my performance.	1	2	3	4	5
8. The debrief prompted me to think about my own performance.	1	2	3	4	5
9. During this debrief I felt comfortable to ask the CFI questions.	1	2	3	4	5
10. This debrief helped me to learn.	1	2	3	4	5

Self Assessment Scenario 1

Directions: Please rate your own performance on the following items as honestly as possible.

Event 1: Landing gear won't retract

Did you slow the aircraft to the appropriate speed? What speed? ___	Yes / No / Had difficulties
Did you refer to the checklist	Yes / No / Had difficulties
Did you recalculate the flight plan correctly and in a timely manner?	Yes / No / Had difficulties
Did you use the POH?	Yes / No / Had difficulties
Did you use the passenger as a resource?	Yes / No / Had difficulties
Did you contact ATC and advise?	Yes / No / Had difficulties
Did you notify the passengers of the situation?	Yes / No / Had difficulties

Event 2: Vaccum Failure

Did you recognize the failure?	Yes / No / Had difficulties
Did you refer to the checklist?	Yes / No / Had difficulties
Did you solve the problem? How? ________	Yes / No / Had difficulties
Did you notify ATC of the problem?	Yes / No / Had difficulties
Did you use the passenger as a resource?	Yes / No / Had difficulties

Event 3: ATC Re-Route due to weather

Did you comply with ATC instructions?	Yes / No / Had difficulties
Did you recalculate the flight plan correctly and in a timely manner?	Yes / No / Had difficulties
Did you advise ATC of the new ETA?	Yes / No / Had difficulties

Event 4: Go Around at TPA

Did you comply with ATC instructions?	Yes / No / Had difficulties
Did you analyze available options?	Yes / No / Had difficulties
Did you request vectors for Runway 18R?	Yes / No / Had difficulties
Did you request to fly to alternate airport?	Yes / No / Had difficulties
Did you use ATC as a resource?	Yes / No / Had difficulties

Event 5: ATC Re-route due to weather

Did you request IFR clearance into KTPA?	Yes / No / Had difficulties
Did you request vectors to nearest VFR field?	Yes / No / Had difficulties
Did you recalculate flight plan and ETA?	Yes / No / Had difficulties
Did you use passenger as a resource?	Yes / No / Had difficulties
Did you use ATC as a resource?	Yes / No / Had difficulties

Self Assessment Scenario 2

Directions: Please rate your own performance on the following items as honestly as possible.

Event 1: Starting aircraft from cold/dark start

Did you give passenger a brief?	Yes / No / Had difficulties
Did you use the before takeoff checklist to successfully start the aircraft?	Yes / No / Had difficulties
Did you request proper clearance and VFR flight following?	Yes / No / Had difficulties
Did you perform the after takeoff checklist?	Yes / No / Had difficulties
Did you use ATC as a resource?	Yes / No / Had difficulties

Event 2: Pitot Tube Blockage

Did you recognize the failure?	Yes / No / Had difficulties
Did you refer to the checklist?	Yes / No / Had difficulties
Did you solve the problem?	
How? _______	Yes / No / Had difficulties
Did you use the passenger as a resource	Yes / No / Had difficulties

Event 3: ATC notified pilot that KCRG is closed due to extension of presidential TFR

Did you choose to fly to alternate or nearest suitable VFR airport?	Yes / No / Had difficulties
Did you use ATC as a resource?	Yes / No / Had difficulties
Did you contact ATC with your intentions?	Yes / No / Had difficulties
Did you recalculate the flight plan correctly and in a timely manner?	Yes / No / Had difficulties
Did you notify passengers?	Yes / No / Had difficulties

Event 4: Engine Failure

Did you establish the best glide airspeed? What was it? _____	Yes / No / Had difficulties
Did you refer to the checklist in a timely manner?	Yes / No / Had difficulties
Did you use the checklist correctly?	Yes / No / Had difficulties
Did you solve the problem?	
How? _______	Yes / No / Had difficulties
Did you utilize the passenger to help locate a landing place?	Yes / No / Had difficulties
Did you attempt to restart the aircraft?	Yes / No / Had difficulties
Did you notify ATC?	Yes / No / Had difficulties
Did you brief passengers on the situation?	Yes / No / Had difficulties

Event 5: Landing gear will not extend

Did you recognize the failure?	Yes / No / Had difficulties
Did you refer to the checklist?	Yes / No / Had difficulties
Did you solve the problem?	Yes / No / Had difficulties
How? _______	
Did you go around?	Yes / No / Had difficulties
Did you notify ATC?	Yes / No / Had difficulties
Did you mannally extend gear?	Yes / No / Had difficulties
Did you notify passengers of situation?	Yes / No / Had difficulties
Did you land successfully at the airport?	Yes / No / Had difficulties

Self-Efficacy

	To no Extent				To a Great Extent
1. I have confidence in my abilities to fly.	1	2	3	4	5
2. I believe I can become unusually good at flying.	1	2	3	4	5
3. I expect to be known as a high-performing pilot	1	2	3	4	5
4. I feel I can solve any problem I encounter during a flight.	1	2	3	4	5
5. I can accomplish a lot in the cockpit when I work hard.	1	2	3	4	5
6. No flight is too tough for me.	1	2	3	4	5
7. I expect to have a lot of influence on other pilots.	1	2	3	4	5
8. I feel I can be a very productive pilot.	1	2	3	4	5

Report 4: Appendix C

Performance Metrics and Rating Sheets

Scenario 1

Note to raters: Please read over the following performance measures carefully before experiment. It is important to rate pilot consistently and their actions may not perfectly match one of the available choices on the 5 pt scale. For these reasons, we ask for you to take notes if unable to reach a definite rating before the next event occurs, and then complete the rating after the test is finished.

Flight Plan: Daytona International to Tampa for a family reunion. Passenger (friend, non-pilot).

Event 1a: Starting the aircraft from a cold/dark start.
Skills tested: Use of resources (Checklist use, Pilot request VFR flight following)

Pilot completed safety brief

1 – Did not complete brief
2- –
3- -- Completed minimal brief
4- –
5- – Complete fully effective brief

Pilot followed checklist procedure (before take off) YES / NO

Pilot used ATC as a resource

Pilot requested "Flight following" YES / NO

Pilot performed after take-off checklist procedure before 1000 feet. YES / NO

Overall, the pilot responded

1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Event 1b: Landing gear does not retract. Now the flight will be delayed, and the pilot must recalculate the flight.

Skills tested: Use of resources (i.e., passenger); recalculation of flight plan

Using resources:

Pilot referred to checklist as a resource. YES / NO

Pilot referred to POH as a resource YES / NO

PILOT use autopilot as a resource YES / NO

PILOT used passenger as resource YES / NO

Pilot contacted ATC and Advised

1 - Pilot did not contact ATC
2 –
3 - Pilot contacted ATC and advised after some delay
4 –
5 – Pilot contacted ATC immediately

Pilot notified passengers

1 – Did not notify passengers of the problem or did minimal after prompt
2 –
3 – Pilot gave minimal brief to passengers before prompt
4 –
5 – Pilot immediately notified passengers

Pilot slowed/kept the aircraft at 129 knots **YES / NO**

Overall, the pilot responded

1 – Very Ineffectively
2–
3- Effectively
4–
5– Very effectively

Event 2: Vacuum failure.
Skills tested: Use of resources (checklist, passenger); recognition of problem; technical solution to the problem

Pilot recognized the failure (e.g., "heading not working") YES / NO

Pilot responded appropriately (i.e. turned on alternate air switch) without the POH YES / NO

<u>Using resources</u>

Pilot referred to checklist as a resource YES / NO

Pilot referred to POH as a resource YES / NO

PILOT use autopilot as a resource YES / NO

Pilot followed checklist or POH and resolved problem. YES / NO

Pilot contacted ATC and Advised

1 - No (e.g., Pilot did not contact ATC)
2 –
3 - Somewhat (e.g, Pilot contacted ATC and advised after some delay)
4 –
5 – Yes (e.g., Pilot contacted ATC as soon as practical and kept ATC informed)

Pilot used passenger as a resource

1- Pilot made no attempt to use passenger or only did after prompt
2 -
3 - Pilot used passenger for assistance after some delay
4 -
5 - Pilot quickly used passenger to help

Overall, the pilot responded

1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Event 3: ATC Re-Route due to Weather

Skills tested: Use of resources (interacting with ATC), recalculation of flight plan

Pilot complied with ATC instructions **YES / NO**

1 – Pilot did not comply with instructions
2 –
3 - Pilot complied after some delay
4 -
5 - Pilot complied immediately and correctly

ATC Asks: Do you have enough fuel to be diverted 60 miles to the South?

(-) Pilot responds “Stand-by”
(+) Pilot responds right away and confidently “yes”

<u>ATC asks for new ETA to Orlando</u>

Pilot knowledge of ETA:

1 - Pilot responds first with “Standby” and then gives INaccurate ETA
2 -
3 – Pilot responds first with “Standby” and then gives an accurate ETA
4 -
5 – Pilot responds immediately and with accurate ETA

Overall, the pilot responded

1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Event 4: Go-Around at Tampa. In Tampa, ATC reports that certain runways are closed. ATC says, "State your intentions"

Skills tested – Use of resources (interacting with ATC)

Pilot complied with ATC instructions (goes around and climbs to correct altitude)

1– Pilot did not comply with instructions
2 –
3 - Pilot complied after some delay
4 - -
5- Pilot complied immediately and correctly

Pilot used ATC as a resource

1 - Pilot does not talk to ATC/ does so with great delay/ or doesn't know what to do
2 -
3 - Pilot requests to fly to alternate without asking any questions
4 -
5 - Pilot requests vectors to Runway 18R OR asks "are there alternate runways"

Overall, the pilot responded

1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Event 5: ATC reroute due to weather. Tampa airport (KTPA) has gone IFR. ATC says, "State your intentions."

Skills tested: Using available resources

Pilot used ATC as a resource

(-) Pilot does not know what to do	YES / NO
(+) Pilot requests vectors to nearest VFR field	YES / NO
(+) Pilot request IFR clearance into KTPA	YES / NO

Overall, the pilot responded

1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Scenario 2

Flight plan: Daytona International to Jacksonville for Superbowl game. Passenger (friend, non-pilot).

Event 1: Starting the aircraft from a cold/dark start.

Skills tested: Use of resources (Checklist use, Pilot request VFR flight following)

Pilot completed safety brief

1– Did not complete brief
2–
3-- Completed minimal brief
4–
5– Complete fully effective brief

Pilot followed checklist procedure (before take off) YES / NO

Pilot used ATC as a resource
Pilot requested "Flight following" YES / NO

Pilot performed after take-off checklist procedure before 1000 feet. YES / NO

Overall, the pilot responded
1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Event 2: Pitot tube blockage. This happens immediately after take-off. At the same time, ATC begins vectoring the aircraft (to add workload).

Pilot recognized the failure (e.g., speed indicator goes out) BEFORE ATC says "maintain 120 knots"
YES / NO

Pilot responded appropriately (pushing Pitot heat) immediately with no need for checklist use etc.
YES / NO

Pilot responded appropriately (i.e. turned on alternate air switch) without the POH
YES / NO

Pilot refers to checklist as a resource to solve problem YES / NO
Pilot refers to POH as a resource to solve problem YES / NO

Pilot used Autopilot as a resource YES / NO

Pilot used passenger as a resource

1- Pilot made no attempt to use passenger or did only after prompt
2 -
3 - Pilot used passenger for assistance after some delay
4 -
5 - Pilot quickly used passenger to help

Pilot notified ATC

1- No
2-
3- Gave minimal info
4-
5- Kept ATC fully informed of failure and/or used ATC as a resource (e.g., asked questions)

<u>ATC asks for new ETA to Orlando</u>

Pilot knowledge of ETA:

1 - Pilot responds first with "Standby" and then gives INaccurate ETA
2 -
3 **–** Pilot responds first with "Standby" and then gives an accurate ETA
4 -
5 - Pilot responds immediately and with accurate ETA

Overall, the pilot responded

1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Event 3: ATC notifies pilot that KCRG airport is closed due to extension of the presidential TFR (flight restriction).

ATC says "State your intentions"

Use of ATC as a resource

(-)Return to DB or St. Augustine YES / NO
(+) Nearest lighted airport YES / NO
(+)Asks, "How much longer is TFR in effect?" or "Where is a close airport?" YES / NO

Pilot contacted ATC with intentions

1 - Pilot did not contact ATC
2 –
3 - Pilot contacted ATC and advised after some delay
4 –
5 – Pilot contacted ATC as soon as practical and advised of _____.

Overall use of ATC

1 - ineffective
2 -
3 - Effective
4 –
5 – Highly effective

ATC tells the pilot where to go.
Pause
Passenger says, "What's going on?"

Pilot notified passenger

1- – Pilot did not notify passengers of the problem until AFTER prompt
2- –
3- – Pilot notified passengers after some delay OR gives a vague/incomplete report
4- –
5- – Pilot immediately notified passengers after ATC re-route

Passenger says, "How long until we get there? I need to call the Limo service."

Recalculation of Flight Plan

1 - Pilot responds, "Not sure" (i.e, pilot has not recalculated flight plan)
2 –
3 - Pilot responds with a rough estimate (i.e, pilot has a vague idea of ETA)
4 –
5 - Pilot responds confidently with an actual time. (i.e., pilot knows ETA from calculations)

Overall, the pilot responded

1– Very Ineffectively
2–
3 - Effectively
4 –
5– Very effectively

Event 4: Engine Failure

Pilot recognized the failure

1 – Pilot did not recognize the failure
2 -
3 – Pilot recognized failure after some delay
4 –
5 – Pilot recognized failure immediately

Pilot established best glide airspeed YES / NO

Pilot refers to checklist as a resource YES / NO

Pilot follows checklist procedure to resolve problem.

1- Pilot did not follow checklist procedure.
2 -
3 - Pilot executed most steps correctly.
4 -
5 - Pilot executed all steps correctly.

Pilot used passenger as a resource

1- Pilot made no attempt to use passenger
2 -
3 - Pilot used passenger for assistance after some delay
4 -
5 - Pilot quickly used passenger to help find a place to land

Pilot attempted to restart aircraft YES / NO

Pilot contacted ATC with intentions

1 - Pilot did not contact ATC or only did after prompt
2 –
3 - Pilot contacted ATC and advised after some delay
4 –
5 – Pilot contacted ATC as soon as practical and advised of engine failure

Pilot briefed passengers of situation

1 – Pilot did not notify passengers of the problem or minimal AFTER prompt
2 –
3 – Pilot gave minimal notification to passengers
4 –
5 – Pilot immediately gave full notification to passengers

Overall, the pilot responded

1 – Very Ineffectively
2 –
3 - Effectively
4 –
5 – Very effectively

Event 5: Landing Gear Failure

Pilot recognized the failure

1 – Pilot did not recognize the failure
2 -
3 – Pilot recognized failure after some delay
4 –
5 – Pilot recognized failure immediately and does a go around

Pilot refers to checklist as a resource YES / NO

Pilot follows checklist procedure to resolve problem.

1. - Pilot did not follow checklist procedure.
2 -
3 - Pilot executed most steps correctly.
4 -
5 - Pilot executed all steps correctly.

Pilot used passenger as a resource

1- Pilot made no attempt to use passenger OR only after prompt
2 -
3 - Pilot used passenger for assistance after some delay
4 -
5 - Pilot quickly used passenger to help

Pilot contacts ATC.

1- Pilot does not contact ATC.
2 -
3 - Pilot contacts ATC after some delay
4 –
5 - Pilot contacts ATC immediately and informs them of what's going on

Pilot used ATC as a resource

(-) Does not use ATC or only after prompt	YES / NO
(+) Asks ATC to verify gear is down	YES / NO
(+) Tells ATC they want to hold OR go south and fix the problem	YES / NO

Pilot briefed passengers of situation

1 – Pilot did not notify passengers of the problem OR minimally after prompt
2 –
3 – Pilot gave minimal notification to passengers before prompt
4–
5 – Pilot immediately gave full notification to passengers

Pilot manually extends the gear	YES / NO
Pilot lands successfully at the airport	YES / NO

Overall, the pilot responded

1 – Very Ineffectively
2 –
3 - Effectively
4 –
5 - Very effectively