FY2004 Task 5, Report 3:

RESULTS OF INITIAL VALIDATION OF TASKS AND COMPLETION STANDARDS IN FAA PRACTICAL TEST STANDARDS

Submitted by:

Beth Blickensderfer, Ph.D.
Michele Summers, M.S.
Embry-Riddle Aeronautical University

Peter M. Schumacher, M.S.
University of North Dakota

July 5, 2005

This effort is the third deliverable under the FITS Program FY2004 Task 5: ***Development of a methodology to justify the inclusion or removal of maneuvers from flight training curriculums.***

INTRODUCTION

The FITS Task 5 team recently completed a review of common practices used to validate the inclusion of specific tasks in training curriculum and job related evaluations (SEE FY2004 TASK 5, REPORT 1: INVESTIGATION OF CURRENT PRACTICES RELATING TO THE PTS) and proposed a methodology for justification of the tasks and completion standards in the FAA Practical Test Standards (SEE FY2004 TASK 5, REPORT2: METHODOLOGY FOR VALIDATION OF TASKS AND COMPLETION STANDARDS IN FAA PRACTICAL TEST STANDARDS). The purpose of the current report is to provide results of a demonstration (i.e., test) of this methodology. This report first describes the method used in the test. This is followed by a discussion of the results.

METHOD

Participants

One hundred and thirty-nine certified flight instructors participated in the survey. One hundred twenty-nine were from the University of North Dakota (UND), and ten were from Embry-Riddle Aeronautical University (ERAU). Due to research deadlines, the participants were limited to these two locations. Another limitation of our survey was the high number of participants from UND as compared to ERAU. The participant average total flight hours was 820 hours (SD = 618), and the average total instructor hours was 472 (SD = 545). Of the participants, 133 were certified instrument instructors, 61 were certified multi-engine instructors, and 1 was an airline transport pilot. None of the pilots surveyed were FITS trained. Incomplete surveys were not included in the data analyses. Thus, the number of participants for the different analyses ranges from 103 to 105.

Materials

A Private/Commercial Pilot Validation Survey was developed. The survey included a cross-section of tasks from the Private Pilot and Commercial Pilot Practical Test Standards (see appendix). Using Likert scales, each participant rated each task on 4 dimensions.

<u>Frequency</u>. How frequently is this task required for actual flight?
1-5 Likert Scale: 1 = Seldom, 5= Always

<u>Importance</u>. How important is this task for actual flight?
1-5 Likert Scale: 1 = Non-essential, 5 = Critical

<u>Reflective/Realism</u>: How reflective are task completion standards of performance required in actual flight?
1-5 Likert Scale: 1 = Not realistic, 5 = Realistic

Redundancy: Are the skills required for execution of this task evaluated in other tasks?
1-5 Likert Scale: 1 = Much Redundancy, 5 = Zero Redundancy

In addition, space was provided for the participants to explain their rationale for any maneuver rated a 2 or less on any of the 4 scales.

Procedure

The surveys were distributed and completed over a period of three weeks in June 2005. The surveys were collected by the participants' supervisors who then returned the surveys to the experimenters.

RESULTS

Task Overall Importance Score

A task overall importance score was computed by combining the frequency and importance score for each task. This was accomplished by multiplying each participant's frequency rating by his/her importance rating for each task. Thus, if an individual had rated a task "5" for frequency of task and "3" for importance of that task, the combined frequency/importance score would be 15. The mean overall importance scores for each task are shown in Table 1. The tasks are listed from highest to lowest mean.

The mean overall importance scores were analyzed with a one-way within subjects ANOVA. A significant difference did appear with $F(36, 3672) = 145.87$, $p = .000$. Partial eta squared of .588 indicates that 58.8% of the variance in Frequency/Importance rating depends on differences between the tasks.

Upon examining the means in Table 1, many differences appear. These apparent differences were analyzed using Tukey HSD post hoc comparisons. The Critical Difference score (.05) = 2.07. Thus, any two means whose difference (absolute value) is equal or greater than 2.07 have a significant difference at the $p = .05$ level.

To summarize some of the post-hoc comparisons, Normal takeoff/climb, normal approach/land, and traffic pattern were the highest ranking tasks. These three tasks did not differ significantly from each other but were rated as having significantly higher overall importance than most of the other tasks. The lowest ranking tasks were S-turns, Chandelles, Steep Spirals, Lazy Eights, and Eights on Pylon. These tasks did not differ significantly from each other in terms of overall importance but were rated as having significantly lower overall importance than all but 2 of the remaining thirty-two tasks.

TABLE 1. Mean ratings of overall task importance (Frequency X Importance).

Task	**Mean Score (Frequency*Importance)**	**Standard Deviation**	**N**
Normal Takeoff/Climb	23.96	2.48	103
Normal Approach/Land	23.96	2.48	103
Traffic Pattern	22.88	3.29	103
Instrument communication, Navigation, and Radar Services	21.20	5.34	103
Navigation and Radar Services	21.10	4.84	103
Pilotage and Dead Reckoning	20.44	5.70	103
Instrument Turns to Heading	19.40	5.89	103
Straight/Level Instrument Flight	18.50	5.91	103
Constant Speed Instrument climb	17.73	6.61	103
Constant Speed Instrument descent	17.05	6.30	103
Go-Around/Rejected Landings	16.41	6.02	103
Slow Flight	14.84	6.54	103
Spin Awareness	13.77	7.42	103
Short-Field Takeoff	13.01	5.59	103
Short-Field Approach/Landing	13.00	5.48	103
Diversion	12.40	6.17	103
Lost Procedures	11.90	6.75	103
Power-Off Stall	11.38	5.97	103
Soft-Field Approach/Land	11.23	5.88	103
Multi-Engine Maneuvering with one engine inoperative	11.10	5.99	103
Forward Slip to Land	11.09	4.94	103
Power-On Stall	11.04	6.07	103
Soft-Field Takeoff	10.98	5.84	103
Recovery from Unusual Attitudes	10.71	5.05	103
Rectangular Course	10.61	7.10	103
Multi-Engine Instrument Approach – One engine inoperative	10.56	5.81	103
Multi-Engine Engine Failure during flight	10.36	5.61	103
Steep Turns	9.65	5.17	103
Emergency Descent	9.34	5.64	103
Emergency Approach and Landing	9.08	5.76	103
Turns Around a Point	6.92	4.61	103
Power-Off 180 degree Accuracy Approach and Landing	6.40	4.72	103
S-turns	6.38	4.09	103
Chandelles	5.93	4.55	103
Steep Spiral	5.45	4.50	103
Lazy Eights	4.42	3.81	103
Eights on Pylons	3.91	3.62	103

Realistic

The mean ratings of how reflective task completion standards are of performance in actual flight are listed in Table 2. The tasks are listed from highest mean to lowest mean.

The mean ratings of realism were analyzed with a one-way within subjects ANOVA. A significant difference did appear with $F(36, 3708) = 36.72$, $p = .000$. Partial eta squared of .263 indicates that 26.3% of the variance in Realism rating depends on the task.

Upon examining the means in Table 2, many differences appear. These apparent differences were analyzed using Tukey HSD post hoc comparisons. The Critical Difference score (.05) = 0.31. Thus, any two means whose difference (absolute value) is equal or greater than 0.31 have a significant difference at the $p = .05$ level.

For example, all tasks listed on Table 2 between Normal Takeoff/Climb and Short Field Takeoff have a Critical Difference less than 0.31. All others have a Critical Difference greater than 0.31.

The tasks whose completion standards received the lowest scores for realism were Power-Off 180 degree accuracy approach and landing, steep spiral, lazy eights, and eights on pylons. Again, these four tasks did not differ significantly from each other but were rated as having significantly lower realism than most of the other tasks. Chandelles was a very close 5th, and though it did differ significantly from some of the other four, the 2 tenths of a point difference offers little practical significance.

Table 2 . Mean ratings of "reflective of performance in actual flight"

Task	Mean Score (Realism)	Standard Deviation	N
Normal Takeoff/Climb	4.63	.61	104
Multi-Engine Maneuvering with one engine inoperative	4.62	.66	104
Normal Approach/Land	4.62	.61	104
Traffic Pattern	4.57	.59	104
Go-Around/Rejected Landings	4.55	.61	104
Instrument Turns to Heading	4.55	.64	104
Multi-Engine Engine Failure during flight	4.54	.72	104
Instrument communication, Navigation, and Radar Services	4.54	.67	104
Navigation and Radar Services	4.53	.67	104
Recovery from Unusual Attitudes	4.50	.71	104
Multi-Engine Instrument Approach – One engine inoperative	4.49	.79	104
Straight/Level Instrument Flight	4.44	.69	104
Constant Speed Instrument climb	4.40	.70	104
Constant Speed Instrument descent	4.40	.72	104
Lost Procedures	4.39	.79	104
Pilotage and Dead Reckoning	4.38	.79	104
Diversion	4.34	.82	104
Soft-Field Takeoff	4.32	.80	104
Power-Off Stall	4.31	.78	104
Power-On Stall	4.31	.79	104
Emergency Approach and Landing	4.29	.71	104
Spin Awareness	4.29	.90	104
Soft-Field Approach/Land	4.28	.82	104
Short-Field Takeoff	4.28	.79	104
Short-Field Approach/Landing	4.23	.87	104
Slow Flight	4.21	.86	104
Emergency Descent	4.18	.94	104
Forward Slip to Land	4.13	.75	104
Steep Turns	3.99	.96	104
Rectangular Course	3.98	.93	104
Turns Around a Point	3.74	1.06	104
S-turns	3.72	1.04	104
Chandelles	3.46	1.09	104
Power-Off 180 degree Accuracy Approach and Landing	3.41	1.17	104
Steep Spiral	3.34	1.20	104
Lazy Eights	3.21	1.15	104
Eights on Pylons	3.13	1.24	104

Redundancy

The mean ratings of redundancy (i.e., degree to which the skills required for execution of this task are evaluated in other tasks) are listed in Table 3. The tasks are listed from highest mean to lowest mean.

The mean ratings of redundancy were analyzed with a one-way within subjects ANOVA. A significant difference did appear with F(36, 3744) = 27.57, p =.000. Partial eta squared of .210 indicates that 21% of the variance in Redundancy rating depends on the task.

Upon examining the means in Table 2, many differences appear. These apparent differences were analyzed using Tukey HSD post hoc comparisons. The Critical Difference score (.05) = 0.35. Thus, any two means whose difference (absolute value) is equal or greater than 0.35 differ significantly at the p = .05 level. To summarize some of the post-hoc comparisons, the majority of the tasks did not differ significantly from each other. Most tasks were rated "4" or higher meaning that the raters did not see a lot of redundancy issues. For example, all the tasks in Table 3 with a mean of 4.16 or higher (i.e., Multi-Engine Instrument Approach – One engine inoperative through Straight/Level Instrument Flight) do not differ significantly from each other in terms of redundancy rating. However, they were rated as significantly less redundant than Power-Off 180 degree Accuracy Approach and Landing, Chandelles, Lazy Eights, Eights on Pylons, S-Turns, Turns Around a Point, Steep Spiral, Rectangular Course (i.e., those tasks with mean ratings of 3.76 and below).

Finally, Rectangular Course (mean = 2.91) was rated as having the most redundancy with other tasks. This mean was significantly lower (translating as more redundancy with other tasks) than the mean ratings of all other tasks.

Table 3. Mean ratings of degree to which the skills required for this task are evaluated in other tasks. (5 = Low Redundancy, 1 = High Redundancy)

Task	Mean Score (Redundancy)	Standard Deviation	N
Multi-Engine Instrument Approach – One engine inoperative	4.50	.84	105
Multi-Engine Engine Failure during flight	4.49	.84	105
Multi-Engine Maneuvering with one engine inoperative	4.49	.86	105
Recovery from Unusual Attitudes	4.43	.78	105
Navigation and Radar Services	4.40	.83	105
Pilotage and Dead Reckoning	4.36	.94	105
Normal Approach/Land	4.36	.98	105
Spin Awareness	4.36	.91	105
Lost Procedures	4.33	.85	105
Diversion	4.32	.85	105
Go-Around/Rejected Landings	4.32	.91	105
Normal Takeoff/Climb	4.32	1.00	105
Emergency Approach and Landing	4.31	.87	105
Emergency Descent	4.30	.86	105
Short-Field Takeoff	4.27	.84	105
Short-Field Approach/Landing	4.27	.84	105
Instrument Comm., Nav., and Radar	4.26	1.02	105
Power-On Stall	4.25	.89	105
Power-Off Stall	4.24	.84	105
Soft-Field Takeoff	4.19	.89	105
Slow Flight	4.19	.90	105
Soft-Field Approach/Land	4.18	.92	105
Straight/Level Instrument Flight	4.16	1.08	105
Constant Speed Instrument descent	4.11	1.05	105
Constant Speed Instrument climb	4.10	1.04	105
Traffic Pattern	4.09	1.05	105
Forward Slip to Land	4.08	.95	105
Steep Turns	4.06	.90	105
Instrument Turns to Heading	4.03	1.11	105
Power-Off 180 degree Accuracy Approach and Landing	3.76	1.07	105
Chandelles	3.68	1.08	105
Lazy Eights	3.45	1.18	105
Eights on Pylons	3.36	1.22	105
S-turns	3.29	1.12	105
Turns Around a Point	3.24	1.16	105
Steep Spiral	3.16	1.08	105
Rectangular Course	2.91	1.21	105

DISCUSSION

The data presented generally supports the efficacy of the job-analysis and survey method to determine content validity of the PTS tasks. It also supports the notion that a discrepancy exists between the skills required to pass a flight check versus the skills required for actual flight.

First, the mean overall importance ratings provide a ranking of each task in terms of a combined score of both how frequent the task is performed during actual flight and how important the task is for actual flight. In terms of overall importance ratings, the ground reference and commercial maneuver tasks ranked at the bottom of the list and were well below 30 of the remaining other tasks. This rank ordering of tasks may provide one indication of how to divide training time as well as how to best focus examination time. As an example, Eights on Pylons was ranked at the bottom of the list in importance, yet instructor comments indicate that they often require the most training time. (See Appendix II) Future research should inspect the degree to which this rank ordering of tasks correlates with time spent during training as well as time spent and emphasis of the task during examinations. The findings of such a study could assess if a mismatch occurs between amount of time spent training certain maneuver tasks (e.g., Eights on Pylons) versus the overall importance of mastering this task for actual flight.

Next, consider the mean "realism" ratings, or how reflective each task was of performance in actual flight. As was the case with overall importance ratings, the tasks receiving the lowest realism scores were maneuver-based tasks: steep spiral, lazy eights, and eights on pylons. Maneuver-based tasks are those that are designed for evaluation purposes only and do not realistically reflect any maneuver required in actual flight. Again, these three tasks were definitively below most of the other tasks, and Chandelles placed beneath many others as well. It should be noted, however, that even the tasks who received the lowest mean ratings of realism achieved a 3.1 mean rating on the scale of 1-5. Thus, no tasks were rated completely unrealistic.

Third, in terms of redundancy of evaluation in other tasks, the maneuver-based tasks (i.e., Chandelles, Lazy Eights, Eights on Pylons, S-Turns, Turns Around a Point, Steep Spiral, Rectangular Course) were rated as having more redundancy in evaluation than 23 of the other tasks. The actual mean redundancy ratings, however, were extremely low. The lowest was rectangular course at just under 3 on the 5-point scale.

Despite the maneuvers tasks generally being rated as lower in terms of overall importance and realism as well as having higher redundancy in evaluation, the mean scores were not extreme (i.e., not rated 1s or 5s on the various dimensions) . One explanation for this occurrence is rater error or bias. Human beings are not perfect and when making judgments on Likert type scales, errors can occur. For example, one type of error is "central tendency error." This occurs when raters rated every item in the middle of the scale, and may be the case in our data. Another type of error is "leniency error." This occurs when raters rate every item at the favorable end of the scale. Leniency error may have come into play on the redundancy ratings, as raters could have had difficulty remembering what task evaluations overlap each other.

In addition to the possible leniency error on ratings of redundancy, the lack of extreme scores may indicate a realization by the raters that while some of the skills needed to perform maneuvers (e.g., rectangular course) are also assessed during assessment of other tasks (e.g., traffic patterns), it is not a perfect match. Thus, the raters may have responded according to the logic that if a pilot is trained via a traditional, maneuvers based training, then the best test of mastering the skills taught in training is a maneuvers based test. While this is true, it fails to recognize the important distinction between a test of the skills taught in training versus a test of skills needed in the actual performance environment. That is, depending on how the training was conducted, subtle differences can appear between what is taught in training and how it is assessed (i.e., what criterion is used to assess the skill) and what is needed in actual performance and how it is assessed (i.e., again, the criterion used in actual performance). Consider the following.

1. Maneuver-type tasks (e.g., rectangular course) require many of the same knowledge and skills that are required in more realistic tasks for actual flight (e.g., traffic patterns).
2. Teaching maneuver-type tasks is the traditional approach to fostering basic pilot skills: maintaining aircraft control, clearing for traffic, recognition of wind drift, multi-tasking, etc.
3. Pilots then apply the same knowledge and skills in actual flight that were learned in practicing maneuvers. However, after completing training, pilots rarely (if ever) perform the maneuver-type tasks in actual flight.

Since many of the same knowledge and skills used in accomplishing maneuver-type tasks are used to accomplish tasks important to actual flight, including these maneuvers as part of a proficiency test is *content valid* for the training and actual performance. The problem with including maneuvers as test items is a *criterion validity* issue.

Case 1: Traditional Training

A pilot taught via traditional training acquires knowledge and skills essential for flight via mastering the maneuvers. The notion is that the pilot will later apply the skills and knowledge acquired via learning the maneuvers to successfully perform traffic patterns. When tested, performing a maneuver is an exact replica of training. So, how well that pilot did on the maneuver during training will be an excellent predictor of his/her performance on the maneuver during the test. Thus, a maneuver as a test has both high content validity and high criterion validity *for the training*. However, this pilot will never be required to perform the maneuver again during actual flight. Thus, the degree to which a maneuver has *criterion validity with actual flight* is not as high as it is for the actual training. Consider this: In accomplishing actual traffic patterns, the pilot applies the skills and knowledge s/he developed during maneuver training but in a slightly different manner than before. Thus, the effectiveness of a pilot performing a maneuver will be related but will not be a perfect predictor of performance in a traffic pattern. *However, the pilot's performance of a traffic pattern during training will be an excellent predictor of the pilot's performance of a traffic pattern during the proficiency examination, which will be an excellent predictor of pilot performance of a traffic pattern during actual flight.*

Case 2: FITS training

A pilot taught via FITS methodology develops the knowledge and skills essential for flight via scenario based training. The notion is to teach the knowledge and skills in the context of tasks inherent to actual flight. Thus, pilots taught via FITS learn to maintain aircraft control, clear for traffic, recognize wind drift, and multi-task all in the context of tasks they will continue to perform as licensed pilots (e.g., traffic patterns).

A FITS trained pilot has acquired (through scenario based training) most of the knowledge and skills required to accomplish the maneuver-based tasks. However, never having done one before, and since there is not a perfect correlation between performing maneuvers and succeeding in actual flight, it would take the FITS trained pilot additional practice to do the maneuver as well as a pilot fresh out of traditional pilot training.

Thus, using a maneuver to test a FITS trained pilot is content valid (the knowledge and skills required to do a maneuver are largely the same as those required to perform a traffic pattern). However, using a maneuver to test a FITS trained pilot does not have high criterion validity. That is, this pilot’s performance during training (e.g., on a task such as traffic patterns) would NOT be a highly predictive of the pilot’s performance of a maneuver-based task during an examination. Furthermore, that pilot’s performance of a certain maneuver during an examination would NOT be highly predictive of the pilot’s performance in traffic patterns during post-training, actual flight. On the other hand, *the FITS trained pilot’s performance of traffic patterns during training would be highly predictive of that pilot’s performance of traffic patterns during an examination which, in turn, would be highly predictive of that pilot’s performance of traffic patterns during actual flight.*

In summary, using maneuver-based tasks to predict performance in traffic patterns is not likely to be the best predictor of actual flight performance for anyone (FITS or Traditionally trained). The FITS trained group is really at a loss, as the maneuvers do not have criterion validity with the training or actual task performance. For the traditional group, the maneuvers are a good test of the training, but because we are licensing pilots to fly in non-training settings, shouldn’t the test be optimized to predict performance in actual flight and not training maneuvers?

CONCLUSION

The data from this survey indicates that a variety of maneuvers seem to have low overall importance for actual flight (i.e., are not content valid), seem unrealistic of actual flight (i.e., are not criterion valid*), and may be evaluated in other tasks during pilot examination (i.e., have skill evaluation redundancies).

Flight instructors always maintain the right to train the necessary knowledge and skills however they think best (maneuver-based or scenario-based). However, for the highest content and criterion validity of a licensing test, pilots should be evaluated on tasks they will be expected to perform in the real, non-training flight environment.

*Note that while this wasn't a pure test of criterion validity, it provides an estimate based on SME's opinions.

PAGE INTENTIONALLY LEFT BLANK

Appendix I

PTS SURVEY

UND and ERAU have received FAA funding to evaluate the maneuvers included in the FAA Practical Test Standards. Information that you provide in this survey will help establish *Content Validity*--the degree to which a particular task reflects the knowledge, skills, or abilities needed to accomplish the real job or assignment; and *Criterion Validity*—the degree to which completion standards reflect the performance measures expected during actual job completion.

The survey asks several questions regarding individual PTS maneuvers. The first two questions ask you to evaluate how *frequently* individual tasks are required, and than how *important* each is for actual flight. Even though certain maneuvers are not encountered frequently during actual flight (like Emergency Approach and Landing), you may still consider them critical. For survey purposes, *actual flight* is defined as a routine VFR or IFR cross-country in a standard UND or ERAU configured aircraft.

The next question asks you to evaluate the PTS Completion Standards for each maneuver—are they realistic standards for actual flight operations, or unrealistic (either too stringent or too lenient)?

The final question is evaluating the degree of redundancy that exists in the Practical Test Standards. If a particular maneuver tests the same skills and knowledge that are sufficiently evaluated in other maneuvers, than excessive redundancy may exist.

Survey completion will require approximately 15 minutes. Your participation is appreciated and will help to improve the pilot training and evaluation process.

PERSONAL INFORMATION

How many total flight hours do you have? __________

How many instructor hours do you have? ___________

Please circle current licenses:

Certified Flight Instructor	Yes	No
CFI Instrument Airplane	Yes	No
CFI Multi-Engine Airplane	Yes	No
Airline Transport Pilot License	Yes	No

Other: Please list______________________________

PTS SURVEY: Please circle your responses to the following questions regarding Private and Commercial PTS maneuvers.

TASK	**HOW FREQUENTLY IS THIS TASK REQUIRED FOR ACTUAL FLIGHT?** **1 (Seldom) to 5 (Always)**	**HOW IMPORTANT IS THIS TASK FOR ACTUAL FLIGHT?** **1 (Non-essential) to 5 (Critical)**	**HOW REFLECTIVE ARE TASK COMPLETION STANDARDS OF PERFORMANCE REQUIRED IN ACTUAL FLIGHT?** **1 (Not Realistic) to 5 (Realistic)**	**ARE THE SKILLS REQUIRED FOR EXECUTION OF THIS TASK EVALUATED IN OTHER TASKS?** **1 (Much Redundancy) 5 (Zero Redundancy)**
Traffic Patterns	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Normal Takeoff/Climb	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Normal Approach/Land	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Soft-Field Takeoff/Climb	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Soft-Field Approach/Land	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Short-Field Takeoff/Climb	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Short-Field App/Land	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Emergency Approach and Landing	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Power-Off 180° Accuracy Approach and Landing	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Forward Slip to Land	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Go-Around/Rejected Landings	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Steep Turns	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Rectangular Course	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
S-Turns	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Turns Around a Point	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Steep Spiral	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Lazy Eights	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

Eights on Pylons	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Chandelles	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Emergency Descent	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Pilotage and Dead Reckoning	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Nav and Radar Services	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Diversion	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Lost Procedures	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Slow Flight	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Power-Off Stalls	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Power-On Stalls	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Spin Awareness	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Straight/Level Instrument Flight	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Constant Speed Instrument Climb	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Constant Speed Instrument Descent	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

TASK	**HOW FREQUENTLY IS THIS TASK REQUIRED FOR ACTUAL FLIGHT?** **1 (Seldom) to 5 (Always)**	**HOW IMPORTANT IS THIS TASK FOR ACTUAL FLIGHT?** **1 (Non-essential) to 5 (Critical)**	**HOW REFLECTIVE ARE THE TASK COMPLETION STANDARDS OF PERFORMANCE REQUIRED IN ACTUAL FLIGHT?** **1 (Not Realistic) to 5 (Realistic)**	**ARE THE SKILLS REQUIRED FOR EXECUTION OF THIS TASK EVALUATED IN OTHER TASKS?** **1 (Much Redundancy) 5 (Zero Redundancy)**
Instrument Turns to Headings	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Recovery from Unusual Flight Attitudes	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Instrument Comm, Nav, and Radar Services	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
MULTI-ENGINE				
Maneuvering with One Engine Inoperative	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Engine Failure During Flight (by Reference to Instruments)	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Instrument Approach—One Engine Inoperative	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

Please explain your rational for any maneuver rated *2 or less*.

MANEUVER	***EXPLANATION***

Thank you for participating. Please return this survey to your assigned Lead Instructor.

Appendix II

Table 4. Flight Instructor Comments. These comments were taken directly from the remarks section of the PTS survey explaining the rational of a maneuver score of 2 or less.

Maneuvers	**Comments**
Eights on Pylon and Lazy eight	1. Never performed one or needed to other than on the check ride 2. Seldom used 3. Never used for normal flight 4. Teaches patience, how applicable to everyday flight 5. How can this be applied to everyday flight - should be intuition 6. Not used in practical flight 7. Not much point, everyone at UND is doing this maneuver differently 8. Only thing I get out of it is patience 9. Do not understand how they relate to flying 10. The standards are vague, any student will have trouble executing the maneuver, and each instructor (stage pilot) has different expectations for completion 11. Teaches theory (pivotal altitude) used only in this maneuver. Skills required (A/C control, division of attention, etc), are evaluated in other tasks 12. We don't fly 8's generally in real life, we fly cross country 13. Not applicable on everyday flight, time can be spent on s-turns or landings 14. This maneuver in no way relates to any skill required in flight. It's time consuming and frustrating for students to train to standards in this maneuver 15. Completely pointless