



**Aviation Research Lab
Institute of Aviation**

**University of Illinois
at Urbana-Champaign
1 Airport Road
Savoy, Illinois 61874**

**THE POTENTIAL OF SUBJECTIVE
ESTIMATES OF TRANSFER**

**Gary Bradshaw, Henry Taylor,
Gavan Lintern, and Don Talleur**

Technical Report ARL-99-6

May 1999

Prepared for

**Innovative Technologies Corporation
Dayton, OH**

Contract INNOV TECH 97E-522

INTRODUCTION

To train pilots, employing airplane simulators as training devices represents a promising alternative to training based upon operational equipment. Advantages of the simulator over an aircraft stem from many sources. First we find an economic advantage -- simulators are usually far more economical to operate than airplanes. From a social standpoint, many high-performance training exercises can only be conducted in somewhat confined airspace, limiting the areas where training may occur. From a safety perspective, simulators are forgiving of trainee errors, while in-flight errors may have serious consequences. Finally, from an educational perspective, simulators can preserve a detailed record of student actions for later review and analysis. Collectively, these factors offer a powerful incentive to employ simulators in pilot training wherever possible.

Simulator training is only worthwhile to the extent that the skills obtained in the simulator transfer to aircraft. In the final analysis, the only thing that matters is a pilot's proficiency in an airplane, not how good they happen to be in a simulator. Unfortunately, airplane simulators have a decidedly mixed record of transfer. High levels of transfer are found in the commercial world, where experienced pilots can be transitioned to an unfamiliar cockpit using only a high-fidelity simulator. It is common for pilots to make their first flight in the new plane on a regularly scheduled line flight: Pilots can adjust to a new cockpit given sufficient simulator experience. But when learning a new skill, such as a beginning pilot first learning to fly, or pilots learning air combat maneuvers, simulators are far less effective. In a recent study, Lintern, Taylor, Koonce, Kaiser, and Morrison (1997) trained student pilots to make 60 'landings' in a general aviation training device equipped with an IRIS visual system, but found such training had only the modest effect of reducing the amount of time needed to solo by 10%.

Given the rapid changes in computer hardware and display technology, new simulators are constantly being introduced that have increased capabilities. The new capabilities of these simulators offer promise of improved rates of transfer. From a procurement standpoint, the problem becomes one of identifying which simulators will lead to effective transfer, and to further determine how cost-effective they will be. With increasing pressure on the federal budget, it is necessary to carefully document the cost-effectiveness of new models to justify their purchase.

The traditional method of establishing the quality of a new simulator is to assess the *transfer of training* that results from its use. Two groups are compared in a typical transfer of training study. The control group receives training in the aircraft only. A transfer group first receives training in the simulator, and then training is completed in the aircraft. A common measure of transfer performance is to assess the percent transfer:

$$\text{Eq. 1} \quad \% \text{ Transfer} = \frac{(\text{control-group} - \text{transfer-group})}{\text{control-group}} \times 100$$

The scores used to measure performance can either be dimensioned by time or by trials. For example, if the control group needed twenty hours of training, and the transfer group required ten hours, the percent transfer would be 50%. Similarly, if the control group needed 60 landings and the transfer required 30, the percent transfer would once again be 50%.

The percent transfer measure is insensitive to the amount of prior training needed in the training device. A second measure, known as the transfer effectiveness ratio (TER), measures the efficiency with which transfer is achieved. The formula used to compute TER is:

$$\text{Eq. 2 Transfer Effectiveness Ratio} = \frac{(\text{control-group} - \text{transfer-group})}{\text{training}} \times 100$$

where the control group performance is measured by the time or trials spent in the airplane, the transfer group performance is similarly measured by the time or trials in the airplane, and the training score reflects the time or trials spent in the simulator. Following one of the above examples, the control group needed twenty hours of training, the transfer group required ten hours of training, and spent 12.5 hours learning in the simulator, the transfer effectiveness ratio would be 80%. Knowing the percent transfer, transfer effectiveness ratio, and the operational costs of the simulator and aircraft, it is possible to establish how cost effective the new simulator would be in a training program.

Unfortunately, obtaining the % transfer measure is an expensive proposition. To obtain this number, two groups of students must be trained on a task: the control group and the simulator group. To obtain statistical reliability, each group should include from twenty to thirty students. Training typically involves mastery of a large set of skills to effectively control a complex craft, and so requires weeks or months of training to reach proficiency. The costs inherent in such a study include personnel (instructors and students), operational and maintenance costs, in addition to the effort to collect and analyze the data produced in the study.

Such costs would be acceptable if the simulator proved to be highly cost-effective. But if the simulator cannot be used in a cost-effective manner, the costs of the study are, for all intents and purposes, lost. Here the rapid introduction of new simulators is a decidedly mixed blessing: determining the merits of numerous models is an expensive proposition. For these reasons, finding a cheaper method of assessing the quality of simulators is highly desirable. Even a crude technique that helped weed out inadequate models would be of considerable benefit, as it would allow limited evaluation resources to be concentrated on a few promising alternatives.

Approaches to Transfer

In developing possible alternatives to objective measures of transfer of training, it is worth remembering that the issue is fundamentally one of transfer. Studies of transfer have a long and venerable history in psychology, dating back to the turn-of-the-century research performed by Thorndyke (Thorndyke & Woodworth, 1901). Prior to Thorndyke's seminal research, it was commonly believed that transfer was broad and general, so that learning a difficult topic, like Latin, would benefit students in areas like logic, science, and the like. Thorndyke showed that such general transfer did not occur. Instead, Thorndyke believed that transfer was limited to

activities that shared common situation-response elements. Thorndyke expressed his theory of identical elements in this way:

“One mental function or activity improves others in so far as and because they are in part identical with it, because it contains elements common to them. Addition improves multiplication because multiplication is largely addition; knowledge of Latin gives increased ability to learn French because many of the facts learned in the one case are needed in the other.” (Thorndyke, 1906, p. 243)

Applying Thorndyke's view of identical elements to the world of aviation simulators, we might expect transfer only where the simulator exactly duplicates the experience of pilots in airplanes. To allow transfer of visually-based skills, where pilots look out into a virtual environment, Thorndyke's view suggests we need to have a lifelike visual experience. This would require a very high performance visual display system, including the following parameters:

A stereoscopic, spherical field of view, 3-D display of the virtual environment; fully colored, richly textured, capable of a wide range of illumination intensity (from the sun shining into the pilot's eyes to a landscape illuminated by moonlight), with resolution that matched the best pilot's vision (perhaps 20/15 or 20/12), updated at a high frequency (100 frames per second or more).

In short, the theory of identical elements demands that we make the visual experience of the pilot in a simulator indistinguishable from the cockpit view in order to achieve transfer. Clearly these requirements greatly exceed the abilities of contemporary display technology. Fortunately, later research has shown that Thorndyke's theory of identical elements was overly narrow. For example, Thorndyke predicted that if students learned to solve a geometry problem that included one set of letters to label points in the diagram, students would be unable to solve an otherwise identical problem that incorporated a different set of letters. Yet students can easily solve such a new problem (Anderson, 1995). Some of the situation-response elements may vary, and still lead to effective transfer.

An unsolved issue in psychological research is to determine which elements must remain the same, and which are free to vary. The general belief is that transfer occurs when abstract knowledge structures (such as schemas or propositions) remain constant from one task to another (Singley & Anderson, 1989). Given a limited understanding of abstract knowledge structures employed by people in various tasks, psychologists have not been able to predict transfer on the basis of a formal analysis of the task.

Yet such a prediction lies at the heart of our goal: to identify aircraft simulators that afford high rates of transfer, and so are cost-effective in training. One thing is certain: a simulator that produced a visual experience indistinguishable from the pilot's view out of the cockpit would produce high transfer for visual tasks. Recognizing that such a system greatly exceeds current display technology, we need to answer an important question: "How close to the real thing is good enough?"

Subjective Estimates of Transfer

In the void left by our incomplete understanding of human thinking, the Training Systems Product Group (Brown et. al., 1994) has suggested a promising technique: exploit the experience of instructors to determine the adequacy of a system. This approach, known as *subjective estimates of transfer*, asks instructor pilots to evaluate the adequacy of a system for the purpose of training.

The subjective estimate methodology is simple and straightforward. Instructor pilots are asked to perform a series of missions on a prototype simulator. Each mission consists of a sequence of tasks that resemble training tasks used to teach student pilots. Following each mission, the instructor then judges the adequacy of the simulator for training *that mission* on a 5-point scale (Table 1). If the simulator was judged inadequate, several additional questions were asked that explored specific defects in the visual display system, the focus of the study, in training that mission. Examples of the latter questions are provided in Table 2. In addition to a yes/no judgment about the visual defects, the instructor pilots were given the opportunity to provide additional written detail, and also were later interviewed to solicit additional information about the simulators beyond that requested on the questionnaire.

Table 1. 1994 Vis-Eval rating scale:

1. No similarity between visual simulation and aircraft training. Cannot train requirement with visual system. Provides negative training and has major deficiencies.
2. Little similarity between visual simulator and aircraft training. Only minimal training can be accomplished using visual system. Major modifications would be required to provide adequate training capability.
3. Training capability is acceptable. Essential parts of the task can be taught with this visual system.
4. Visual training capability is nearly equal to that experienced in the aircraft. Negligible, if any, modifications required to train tasks. Most of the task can be trained with this visual system.
5. Training capability is equal to that experienced in the aircraft. Task can be fully trained with the visual system.

Table 2. Questions about specific defects of visual system from 1994 Vis-Eval study.

Did the visual resolution (object detail) affect the performance of the task?

Did the visual field-of-view (viewable area) affect your rating of the task?

Did the visual brightness (object detail) affect your rating of the task?

Did the visual contrast (object detail) affect your rating of the task?

Was the appearance of displayed objects distorted?

In the accomplishment of this task, did light or small objects appear blurry?

The measures are subjective, rather than objective, because no actual transfer-of-training data is collected. Instead, the method relies upon the experience of instructors to determine whether the simulator will be satisfactory or deficient as a training device. In making such judgments, the expert instructors have two different kinds of information they can draw upon: *simulator realism*, and *training experience*. Simulator realism refers to the similarity of experience between the simulator and the corresponding aircraft. Training experience is knowledge an instructor has developed through teaching of the particular skills and abilities students need to acquire during training.

Although the Vis-Eval team did not provide specific guidance about which kind of knowledge expert instructors were to employ in making their overall rating, the secondary questions shown in Table 2 were designed to explore faults in simulator realism. These questions were only asked if the expert instructor pilots gave an 'unacceptable' rating (1 or 2 on the scale shown in Table 1) for a particular simulator on a given mission.

Regrettably, the Vis-Eval team did not have sufficient resources to explore the *validity* of subjective estimates of transfer. Thus, we do not know how well the subjective judgments made by expert instructor pilots would predict the amount of transfer of training as measured through objective techniques, or the transfer effectiveness ratio. To be sure, the measure has a fair amount of surface validity: who better than experts to judge the capabilities of a simulator as a training device? But surface validity does not always coincide with true validity. The only way to determine the accuracy of subjective estimates of transfer is to compare them with traditional objective estimates (such as % transfer of training) to see how well experts are able to judge the effectiveness of simulators. Before considering this matter in greater detail, we will explore further different kinds of estimates we might ask our experts to make.

ALTERNATIVE FORMS OF SUBJECTIVE ESTIMATES

As we noted, the Vis-Eval study did not provide any guidance to expert pilots about whether they should focus on simulator realism or their training experience in determining their overall rating. But the additional questions asked of the experts all explored defects in simulator

realism, so the experts may have developed an implicit bias to favor realism in making their judgments.

Realism is, of course, another way of expressing similarity. Similarity is often helpful in transfer studies, but we have already noted that perfect similarity is not essential for perfect transfer. Indeed, in many complex skills, perfect similarity can be too much of a good thing. It would be dangerous to teach a young child how to ride a bicycle on the busy streets of downtown Manhattan. Instead, we teach children the basic skills on empty driveways or lightly traveled sidewalks. A further violation of realism occurs if parents add auxiliary training wheels to the bicycles. After learning the basics in a spartan, elementary, and forgiving situation, skills may be honed in a more demanding environment. Realism, the essence of perfect similarity, may lead to less transfer than training in a simplified, "unrealistic" situation.

An alternative to judgments based on realism is to ask expert instructors to use their knowledge of teaching to judge whether a simulator preserves the *essential features* needed for training. Such judgments must be based on the individual theories that instructors have about what is needed for training, and what is unessential detail. Of course, there is no reason to believe such judgments have more validity as estimates of subjective transfer than judgments of simulator realism, but they may tap more directly into an expert instructor's unique skills and abilities as a teacher: anyone who can perform a mission can judge how realistic a simulator is at duplicating the experience of the mission, but teachers presumably have particular expertise in knowing what makes for effective practice. Accordingly, we conducted a small study to determine how to collect judgments based on training experience, to determine if such judgments could compliment those based on simulator realism.

Overview of Study

There are four essential elements of the subjective estimates technique:

- 1) A panel of experts must be selected
- 2) A simulator platform or platforms are chosen
- 3) A mission or series of missions are developed for the platforms
- 4) A set of questions is developed to sample the expert opinions on simulator performance.

The experts used in our study were drawn from our pool of certified flight instructors at the University of Illinois Institute of Aviation, who routinely teach students at all skill levels. Several of our most experienced instructors were able to participate, along with a few more junior instructors.

To obtain a crude sense of the quality of subjective estimates, we employed three different simulator platforms in our study that deliberately covered a wide range of quality. At the low end, we employed a PCATD simulator, which consisted of a PC-based flight trainer with flight controls by Precision Flight Controls, a 20-inch monitor, and hood. In terms of visual

fidelity and performance characteristics, this system is the least realistic of our simulators. The second simulator is a Frasca 141 with a single-channel PC-based visual system. This simulator uses a 17-inch monitor to provide a forward-looking view updated at 20 frames per second, and can depict a realistic airport scene. This simulator was designed to be the “midrange” device in terms of its capabilities. The third simulator is a Frasca 142 simulator with an Evans and Sutherland SP-X 2-channel image generator that can project both a front view and one side view. This visual system incorporates atmospheric-effects, full texturing capabilities, and a 60 Hz update rate. The 142 simulator is our most sophisticated system, both in terms of the visual capabilities and in terms of modeling the flight dynamics. Each of the simulators was programmed to simulate a Beechcraft Sports/Sundowner, a single-engine, fixed pitch propeller plane with fixed landing gear. By choosing simulators with such a wide range of capabilities, we should be able to determine how sensitive the experts are to differences in simulators.

Before constructing a mission and our questionnaires, we conducted a task analysis of basic flight maneuvers and skills needed in order for students to make their first solo flight. This analysis enabled us to ask more detailed questions of the strengths and weaknesses of simulators than a global estimate of transfer. In essence, we were able to develop questions that paralleled those in Table 2, but instead of exploring detail about simulator realism, our questions explore detail about the skills students might acquire in a simulator, and the maneuvers they could practice using those skills.

To begin the task analysis, we selected maneuvers with a visual component from the full set of maneuvers listed in the Private Pilot Practical Test Standards (PTS) (US Department of Transportation, 1995). The PTS prescribes an acceptable level of performance on all visual maneuvers required for pilots to obtain their Private Certificate. Four of these maneuvers, straight and level flight, turns, climbs, and descents, are basic maneuvers that form the building blocks for more complicated maneuvers (Flight Training Handbook, US Department of Transportation, 1980): All other maneuvers are made up of these in combination of one sort or another. To complete our task analysis, we needed to extend the analysis in two directions: first, to identify the task elements that constitute control in these four basic maneuvers, and second, to show how more complicated maneuvers are built from this basic set.

Task elements can be easily examined by looking at the controls a pilot manipulates, and the instruments or environmental cues that are consulted in such manipulation. We initially examined a simple maneuver, taxiing, to refine our ability to conduct this analysis. Taxiing was chosen because it involves so few control manipulations compared to in-flight maneuvers, and so represented an unusually easy case. After identifying the relevant set of controls and their effects, we similarly analyzed the four basic maneuvers. Having decomposed the four basic flight maneuvers into task elements, we identified commonalties between them. Kershner's (1993) Flight Instructors Manual was also referenced for teaching techniques that revealed the elements students must know to fly, and how they are used in various maneuvers.

We then selected crosswind landings as a representative example of a higher-level maneuver that, if we were successful in our earlier analyses, could be decomposed using the existing more primitive elements in our system. Task elements that appeared in the four basic

maneuvers constituted the bulk of the decomposition, though a few additional elements were added to complete the analysis.

Our final task analysis (see Appendix 1) decomposes the elements of flight into two levels. At the first level we find various maneuvers that are practiced in learning to fly. These include basic maneuvers such as straight and level flight, shallow turns, climbs, and descents. To perform these maneuvers, pilots must possess various skills. Essential elements in our analysis are shown in Table 3.

Table 3. Maneuvers and skills needed to solo.

Maneuvers	Skills
taxi	basic aircraft control
normal takeoff	aircraft trimming
climbs	yaw control
straight and level flight	roll control
descents	pitch control
shallow turns	effects of angle of bank on lift
steep turns	effects of airspeed on lift
power-off stalls	effects of pitch on airspeed
power-on stalls	effects of flaps on airspeed
slowflight	visually detect drift with respect to ground track
landing approach	make pitch changes with respect to horizon
normal landing	make bank-angle changes with respect to horizon
engine failure emergency	judge ground proximity during landing
gliding descent	

The mission we developed from our task analysis requires our expert flight instructors to make a simple flight, including taxiing onto the runway, takeoff, climb to altitude, then make a series of simple maneuvers, including climbs, descents, 360° turns, stalls, slowflight, and make a full-stop landing. Details of the mission are included in Appendix 2.

The task analysis was then employed to construct two different questionnaires for our expert pilots. The first questionnaire (Appendix 3) explored the merits of simulators in teaching maneuvers (e.g., shallow turns) to students. The second questionnaire (Appendix 4) was designed to focus our expert instructors on the merits a simulator might have in training skills (e.g., yaw control). A preliminary questionnaire (Appendix 5) was also developed to collect biographical information from our expert instructors, and tap into their experience with and opinions of airplane simulators employed as training devices.

Method

Subjects. Six flight instructors were asked to participate in the study, four males and two females. Their average age was 36, with a range in ages from 24 to 60. In addition to their Commercial certificates, all instructors were multi-engine and single-engine instrument instructors; four instructors also held an Advanced Ground Instructor certificate. Their total flight time experience ranged from 1160 to 16400 hours, with an average of just over 4500 hours. All of the subjects had prior experience with at least one Frasca airplane simulator equipped with an external display, and most were familiar with several different models of simulators, including motion-based and PC-based systems. Their general attitude toward simulators was positive: In response to the question "How effectively do you believe practice on private pilot maneuvers in a ground trainer can substitute for training in the aircraft?" 4 out of our 6 subjects chose the answer "substitutes very effectively" (5 out of 5) while two subjects chose 4 out of 5.

Equipment. Three different flight simulators were used: a PCATD system, a Frasca 141 with a PC-based visual system, and a Frasca 142 with a more sophisticated Evans and Sutherland visual system.

Procedure. Subjects were brought into the lab individually. On their first session, we asked them to fill out a biographical questionnaire (Appendix 5). Then they were asked to make a simple flight in one of the three simulators. Following that flight, they filled out a questionnaire about the adequacy of the simulator in training various maneuvers (Appendix 3). Subjects returned for additional sessions with the remaining two simulators. Across subjects, the order of exposure to different simulators was balanced to avoid order effects. Following the completion of their first three sessions, subjects were brought back for three additional sessions. These sessions were largely the same as the first three -- they flew a short mission then filled out a questionnaire. But during this block of sessions, subjects answered questionnaires that explored their opinions about the effectiveness of each simulator in teaching various skills (Appendix 4).

Results

Somewhat to our surprise, the Frasca 142, with its greater fidelity both visually and dynamically, was not judged overall as the best system to use as a training device. Instead, the more modest 141 scored a little higher on average. Accordingly, this system will be used as the standard for comparison in our analyses of the other systems, the PCATD and the 142, against this system. Figure 1 plots the average responses our expert pilots gave for questions about whether a system could be used effectively to train various simple maneuvers needed to solo. A

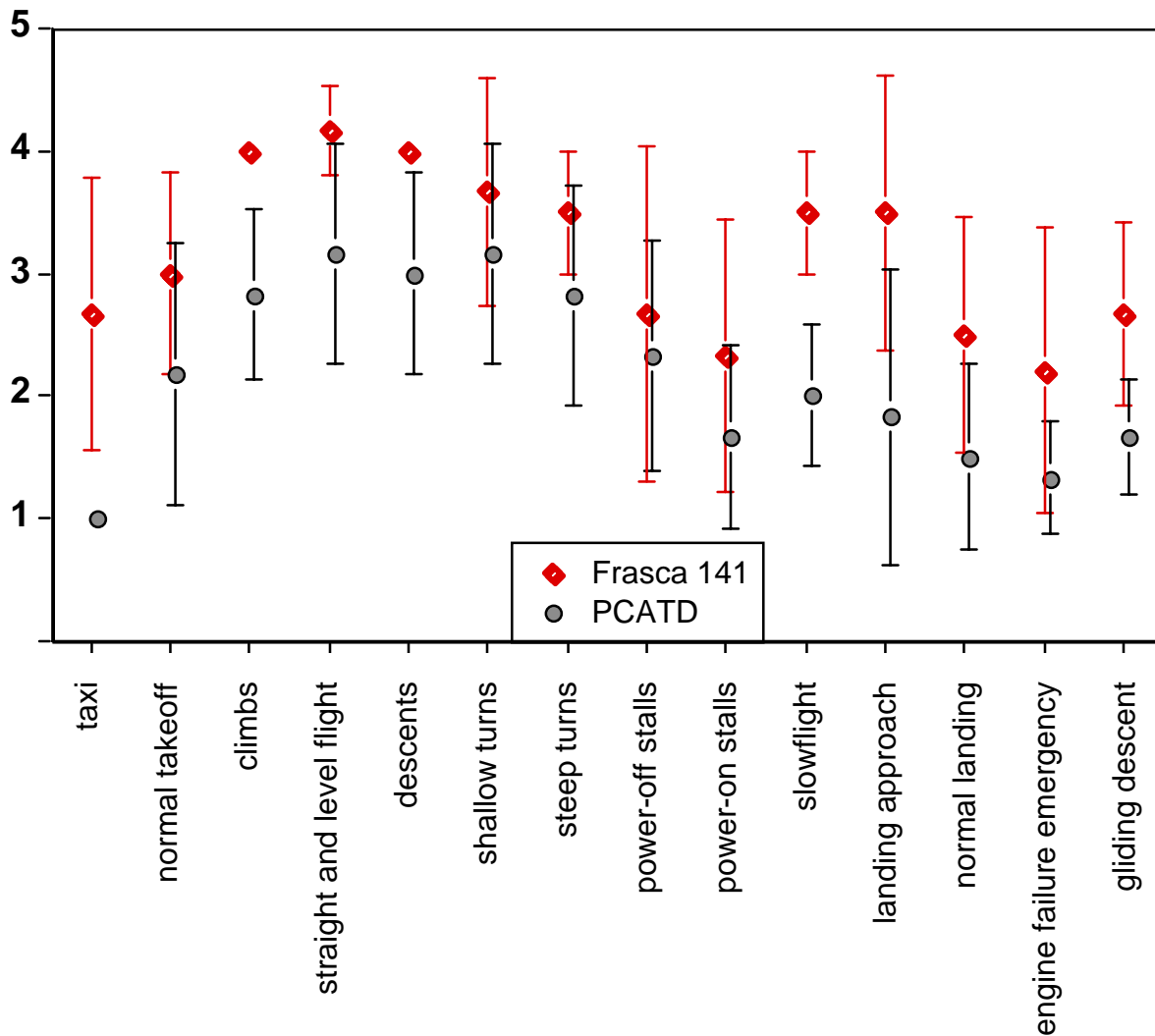


Figure 1. Task ratings of Frasca 141 vs. PCATD.

rating of 1 meant the experts predicted that students could learn “very little” while a rating of 5 meant they predicted students could “learn a lot.”

Error bars represent the standard deviation on our measures. For a few judgments, such as whether the PCATD trainer could train students to taxi, the experts were unanimous in their prediction, so no variability was observed. But on many judgments, experts were far from unanimous in their ratings: We see considerable variability in the ratings given to the systems. This variability is characteristic of most judgments we collected. Unfortunately, we were not able to explain the variation by examining the proficiency of our experts, their experience with simulators, or any other potential factors. In spite of that variability, we see that the 141 is uniformly rated as a superior training device than the PCATD (Figure 2), a result we anticipated.

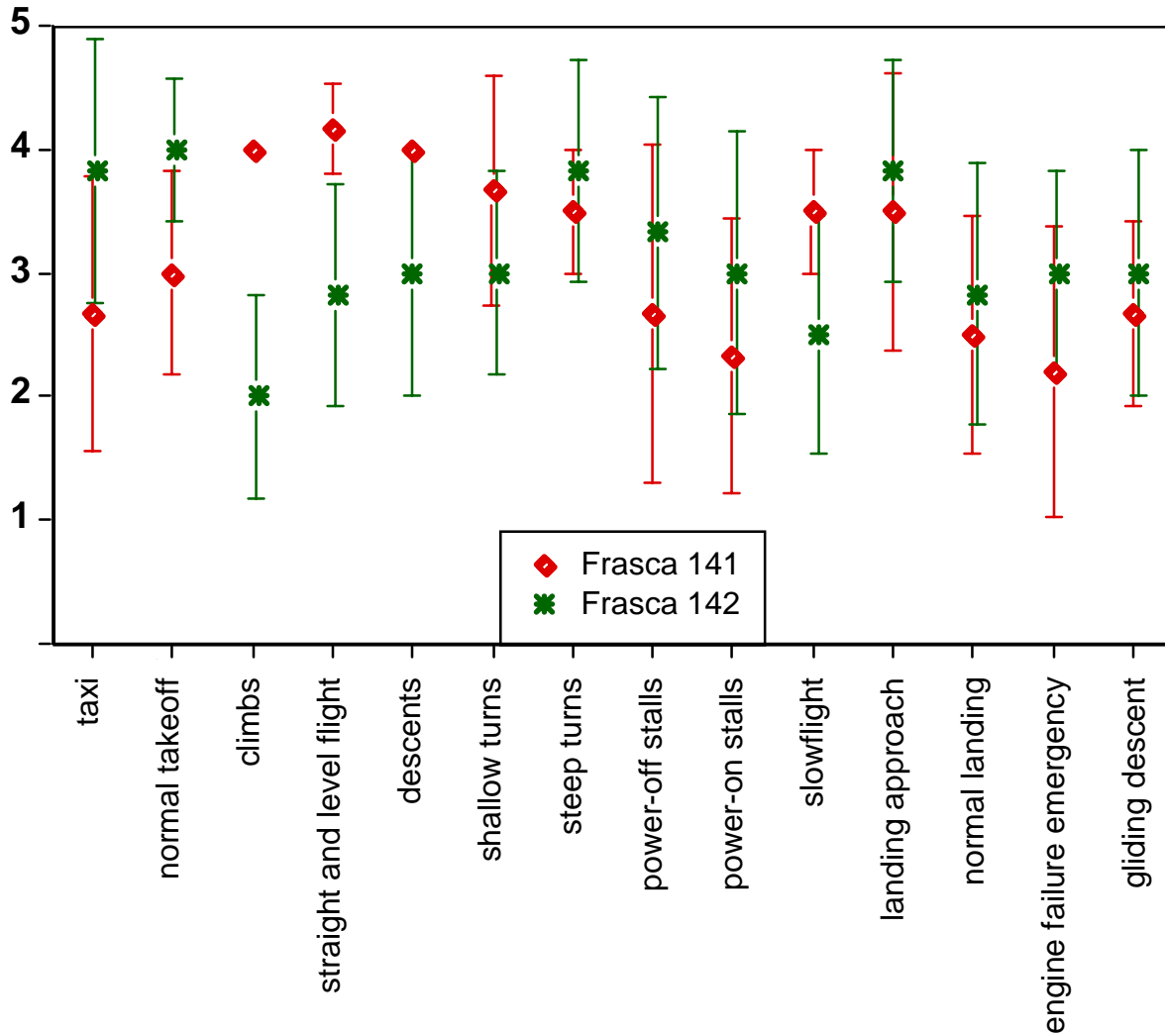


Figure 2. Task ratings of Frasca 141 vs. Frasca 142.

In comparing the Frasca 141 and the Frasca 142 (Figure 3), we find on some maneuvers, such as climbs and straight and level flight, the 141 is rated as clearly superior to the 142. On other maneuvers, such as taxi and takeoff, the 142 seems to be superior, while on other measures, the performance is approximately equivalent. On 9 out of 14 maneuvers, the 142 was rated more highly than the 141, while on the other 6, the 141 received a higher rating. We had anticipated that the 142 would be more highly rated on all scales, so this judgment came as something of a surprise to us.

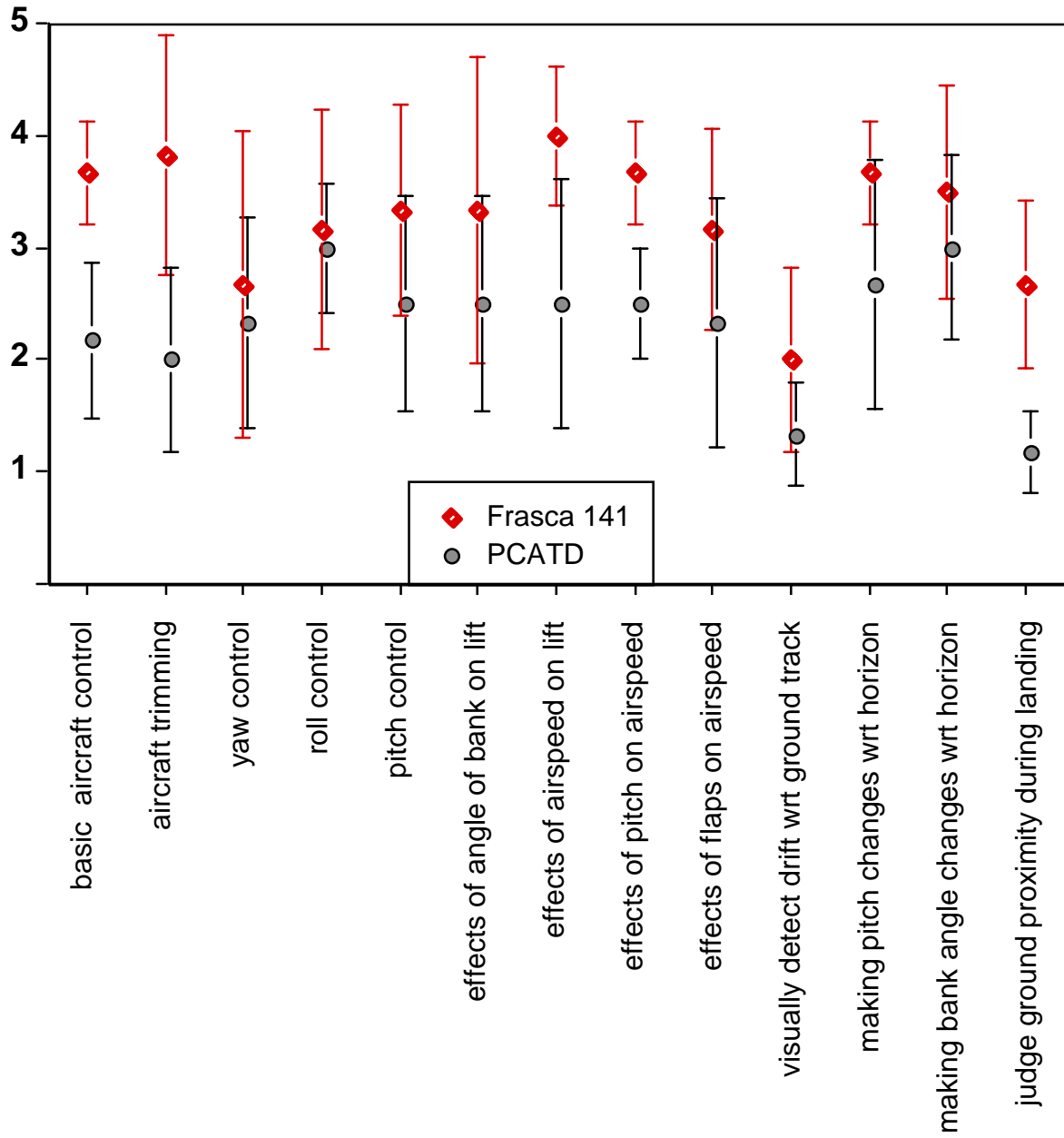


Figure 3. Skill ratings of Frasca 141 vs. PCATD.

A partial explanation of this result was obtained through an analysis of the skill-based judgments made by the experts. Once again, the ratings were made on a 5-point scale, where a rating of 1 meant that our expert pilots thought students would “learn very little” and a 5 meant the students would “learn quite a lot.” The first graph compares the 141 with the PCATD system, and once again, we find the 141 is rated more highly on each basic flight element than the PC-based system. Many of these differences are quite modest, especially given the large variability in judgments by different experts, but some differences, such as basic aircraft control and judgments of ground proximity during landing, represent significant differences in ratings.

The last four items all heavily depend on the external scene depicted by the simulator. The PCATD simulator was, as expected, rated lower on these factors than the Frasca 141. Figure 4 compares the 141 with the 142 on our basic flight elements. For many elements, the two simulators are nearly identical, though curiously, the 141 was rated superior to the 142 on making judgments of drift with respect to the horizon. From additional information we collected from our subjects, we were able to determine that subjects felt the 142 display, set to 15-mile visibility, produced a “hazy” horizon that our expert pilots felt would be problematic for students. They instead preferred the artificially clear horizon of the 141, which has, in effect, infinite visibility conditions. Although the ability to produce atmospheric effects makes the 142 more similar to actual flying conditions, especially in the hazy Midwest, our expert pilots believed that an artificial scene rendering was superior to a more natural system. Apparently the presence of atmospheric haze was important in judging the quality of the simulator in training many of the flight maneuvers, shown earlier in Figure 2.

To summarize, we find evidence of sensitivity on the part of our instructors: the PCATD received lower ratings than did the more capable Frasca simulators. But we also observed considerable variability in the ratings of all three systems for some judgments: several questions generated responses that ranged from 1 to 4 -- almost the full scale of our questionnaire. Attempts to account for this variability by means of biographical covariates (years of flight time, experience as a flight instructor) were not successful.

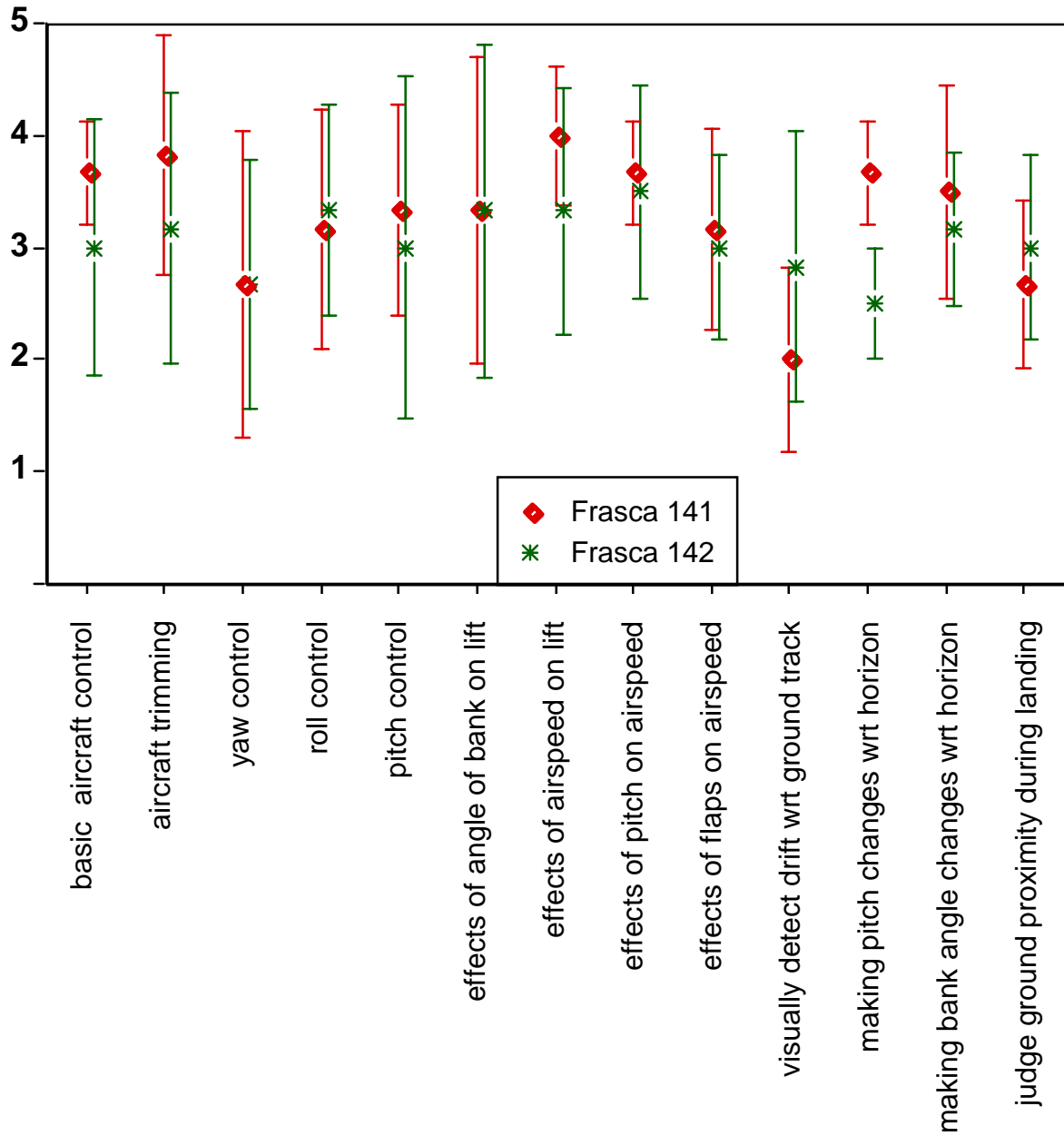


Figure 4. Skill ratings of Frasca 141 vs. Frasca 142.

Our study cannot illuminate the central question: how well do subjective measures of transfer predict objective measures of transfer. Given the variability we observed, we further suggest that our questions, which ask experts to make judgments based on their training experience, should be used in conjunction with realism-based questions of the kind asked in the Vis-Eval studies. Next we discuss how subjective estimates of transfer might be validated in a formal study.

A STUDY IN VALIDATION

We have repeatedly mentioned that subjective estimates of transfer is an unproven method. Unfortunately, there is only one way of validating the subjective estimates technique: by correlating subjective estimates of transfer with objective transfer measures. This is unfortunate for precisely the reasons mentioned at the beginning of this report: objective measures of transfer are expensive and take a long time to collect. Any correlation between objective and subjective measures requires *multiple samples* of each kind of measure, further compounding the cost and extent of a formal study.

Multiple samples may be produced in two ways: we can train pilots to perform different tasks or maneuvers on one simulator, and ask our experts to judge the potential of the simulator to train each of those maneuvers, or we may train different groups of pilots to perform the same tasks or maneuvers on different simulators, and ask our experts to judge the potential of various simulators to train the corresponding maneuvers or tasks.

A correlational analysis works best when the paired scores exhibit at least some degree of variability: the worst possible outcome from an analytical standpoint would be for one or both measures to exhibit little variability. Given our experience, obtaining variability on subjective estimates will not be a problem. A greater area of potential concern is that the system or systems employed to produce the objective measures of transfer might lead to nearly equivalent levels of transfer. To promote the chances of obtaining a correlation, it is essential to test conditions where the amount of transfer differs considerably from one condition to another. For this reason, we believe the greatest prospect of validating subjective estimates of transfer will be to employ simulators with a wide range of capabilities. Hopefully, low-end simulators will produce little or no transfer, while high-end simulators will produce better transfer.

In the proposed study, we plan to employ the same range of simulators, from the PCATD up to the Frasca 142, to accommodate a full range of capabilities. (We have adjusted the atmospheric model of the 142, so that the horizon line it creates is much clearer, and hopefully will lead to high levels of transfer as well.) In the past, we have found better transfer with a low degree of scene detail than with a rich display that presents more scene detail to students, so we propose to test the 142 in two different modes: both with low scene detail and with high scene detail. We will collect corresponding subjective estimates from experienced flight instructors on both systems. This will allow us to explore how sensitive the instructors are to simulator realism, and, if our past experience is replicated, to strengthen the case that simulator realism is not always a good thing for students.

Given the relatively low costs of obtaining subjective estimates of transfer, we suggest collecting a fairly-large number of such estimates (from 20 or more experts) for two reasons: to obtain stable estimates of the population means, and to provide a sufficiently-large sample so that we may study the influence of background factors (such as the experience of flight instructors, past exposure to flight simulator, etc.) in making these judgments. This should help to improve the technique should it prove valid and useful.

CONCLUSION

At the core of this report is a serious problem: how to judge the merits of simulators as training devices. Although there is an existing solution, measuring the transfer of training for each simulator, that solution is awkward and expensive. Using subjective estimates produced by experts is a promising and inexpensive alternative to objective estimates. If this technique works, it will represent a major savings in time and money in evaluating new simulators. But subjective estimates of transfer are based on the theories that expert pilots have about what makes a good simulator. Such theories are often implicit, rather than being based on objective information. So, the jury is still out on how much reliance can be placed on these estimates. Only by conducting a comparison study will we be able to determine whether these fast and inexpensive estimates are sufficiently accurate to be helpful in judging the quality of new simulators. A correlational study, where subjective estimates are compared with objective estimates, provides the best means of validation.

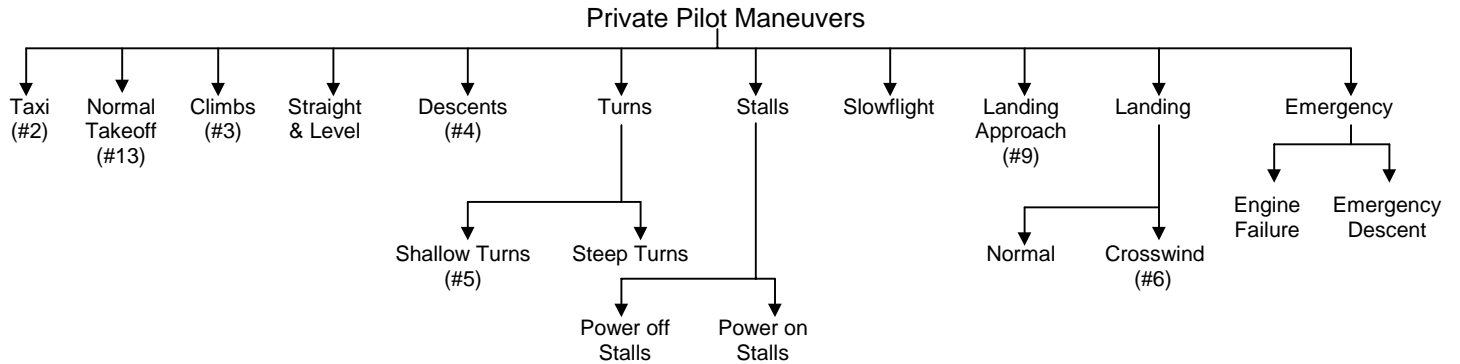
REFERENCES

- Anderson, J. R. (1995). *Cognitive psychology and its implications*. New York: W. H. Freeman and Co.
- Brown, J. E., Lincourt, T. J., Leos, M. J., Poe, D. R., Allard, G. S., & Charpiloz, K. C. (1994). *Visual system operational evaluation* (Technical Report ASC-TR-94-5030).
- Kershner, W. K. (1993). *The flight instructors manual* (3rd ed.). Ames, IA: Iowa State University Press.
- Lintern, G., Taylor, H. L., Koonce, J. M., Kaiser, R. H., & Morrison, G. A. (1997). Transfer and quasi-transfer effects of scene detail and visual augmentation in landing training. *The International Journal of Aviation Psychology*, 7, 149-169.
- Ryder, J. M., & Redding, R. E. (1993). Integrating cognitive task analysis into instructional systems development. *Educational Technology, Research, and Development*, 41, 75-96.
- Singley, K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Thorndyke, E. L. (1906). *Principles of teaching*. New York: A. G. Seiler.
- Thorndyke, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 9, 374-382.
- US Department of Transportation, Federal Aviation Administration, Office of Flight Operations. (1995). *Private pilot for airplane single-engine land practical test standards* (FAA-S-8081-14S). Washington DC.
- US Department of Transportation, Federal Aviation Administration, Flight Standards Service. (1980). *Flight training handbook* (AC 61-21A). Washington, DC.

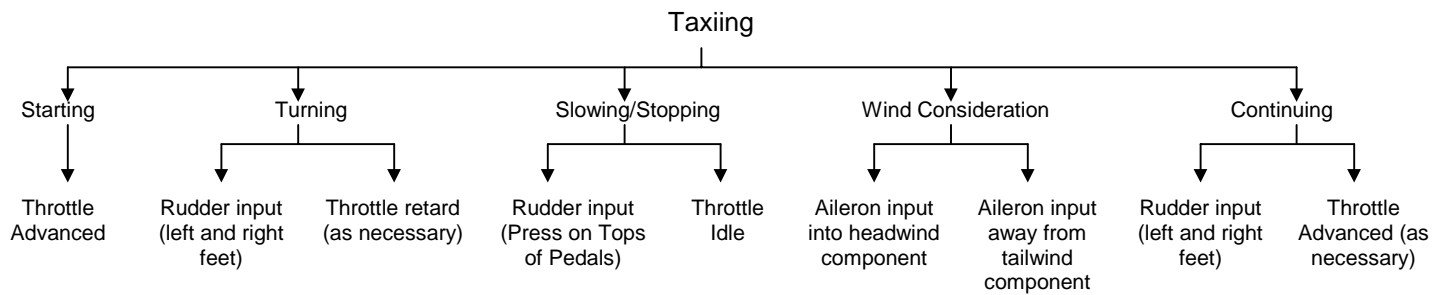
Appendix 1

Task Analysis of Basic Flight Maneuvers and Skills

Maneuver Hierarchy #1

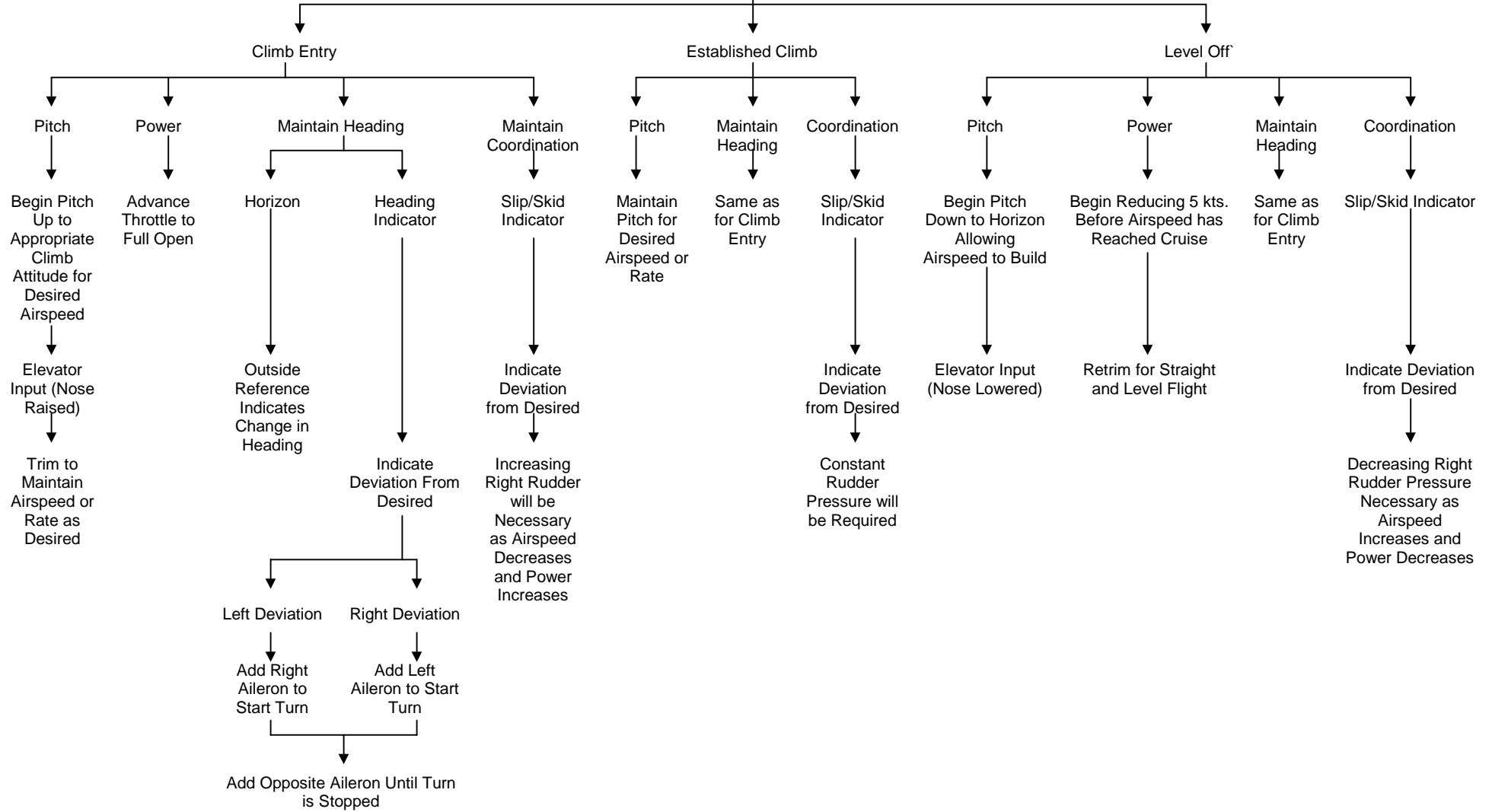


Maneuver Hierarchy #2



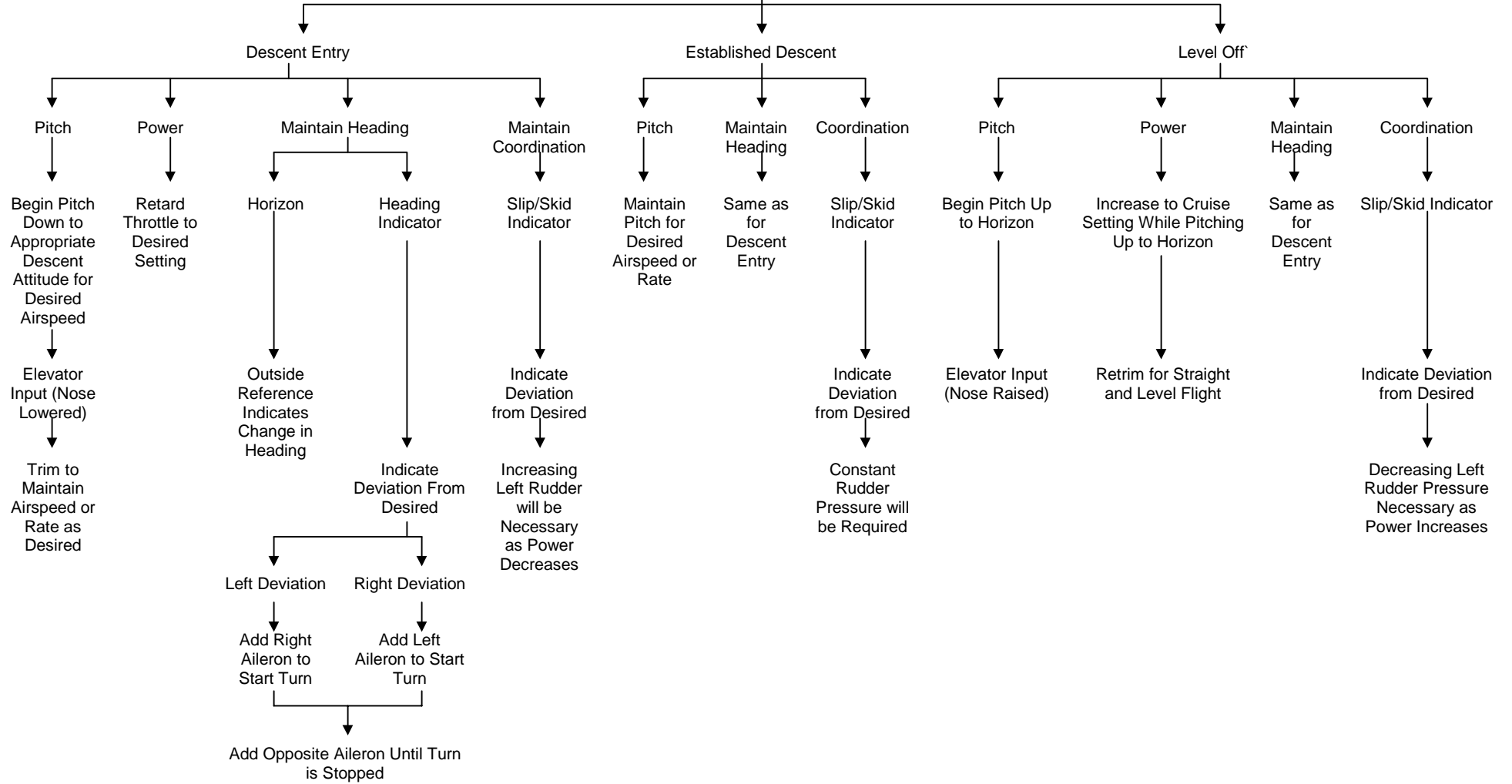
Maneuver Hierarchy #3

Cruise Climb

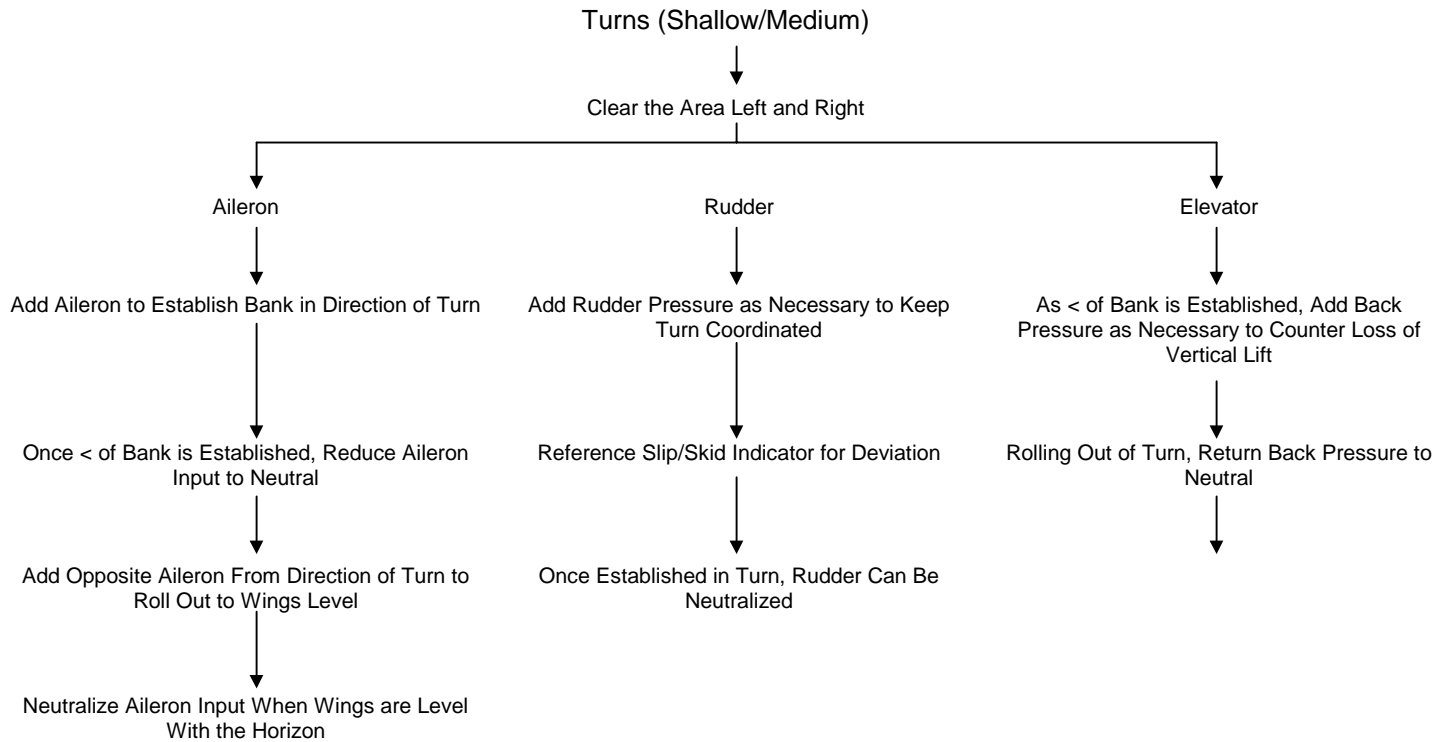


Maneuver Hierarchy #4

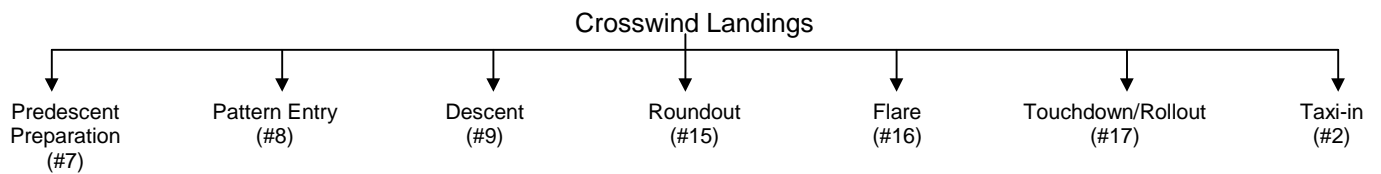
Cruise Descent



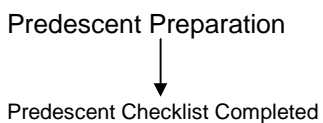
Maneuver Hierarchy #5



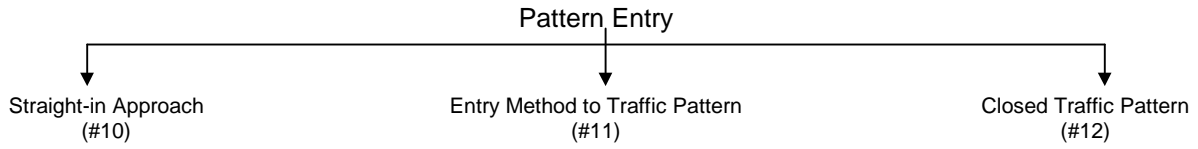
Maneuver Hierarchy #6



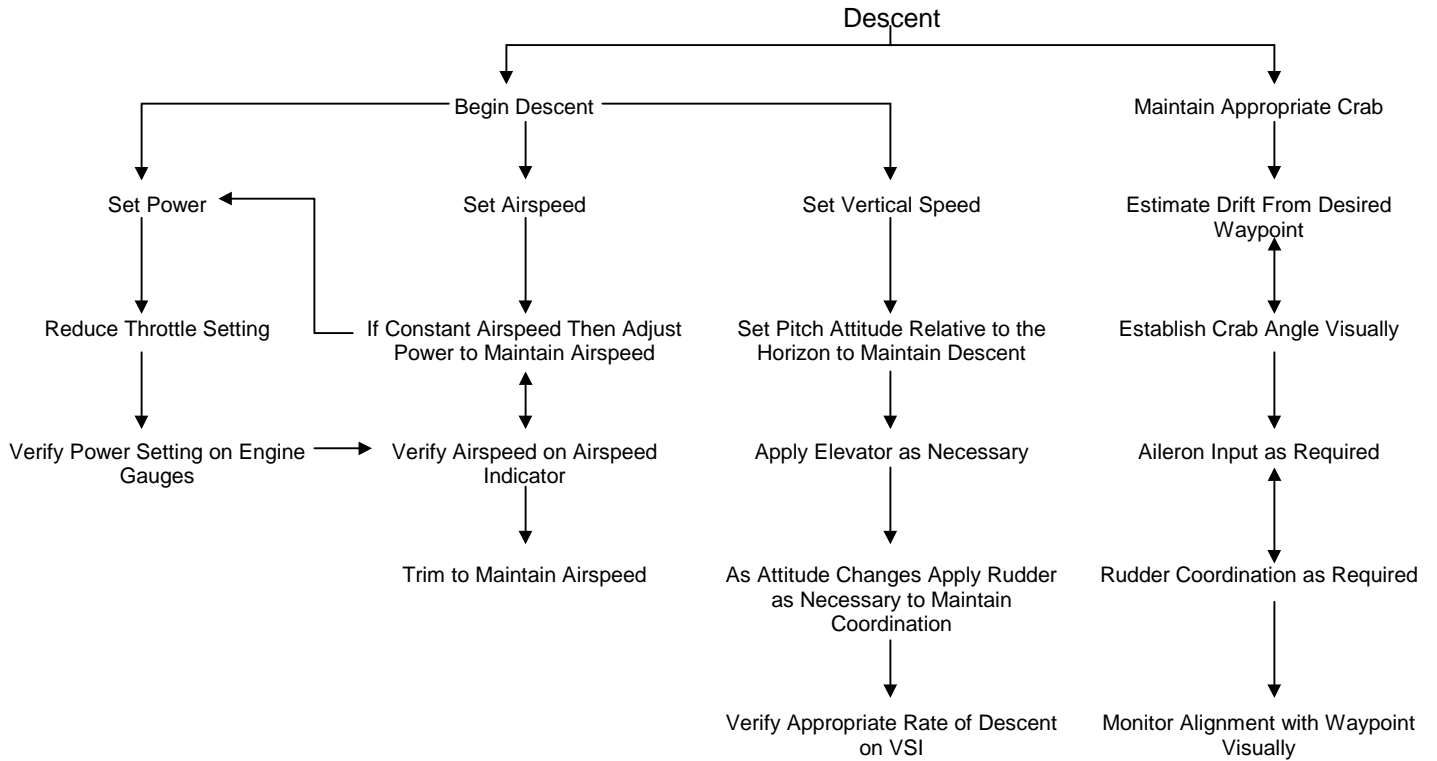
Maneuver Hierarchy #7



Maneuver Hierarchy #8

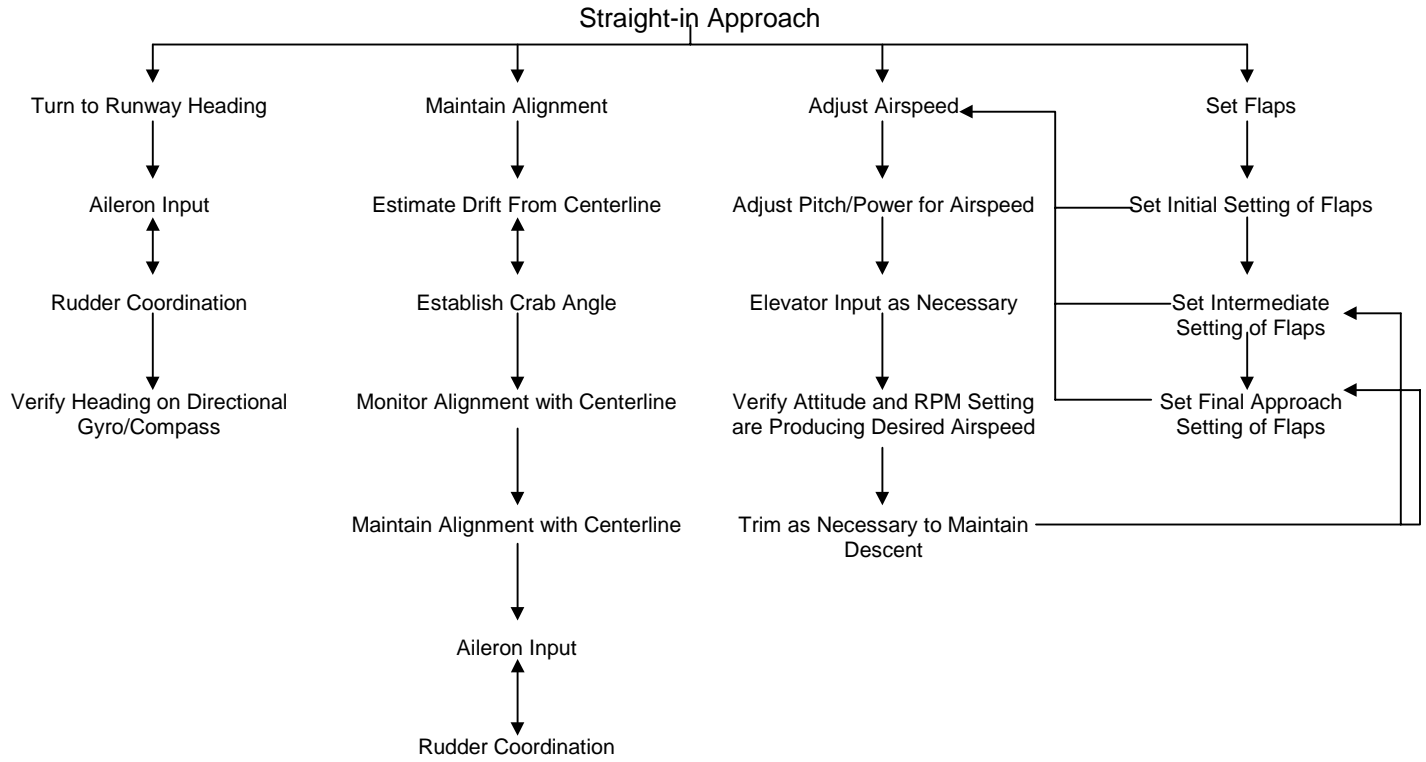


Maneuver Hierarchy #9



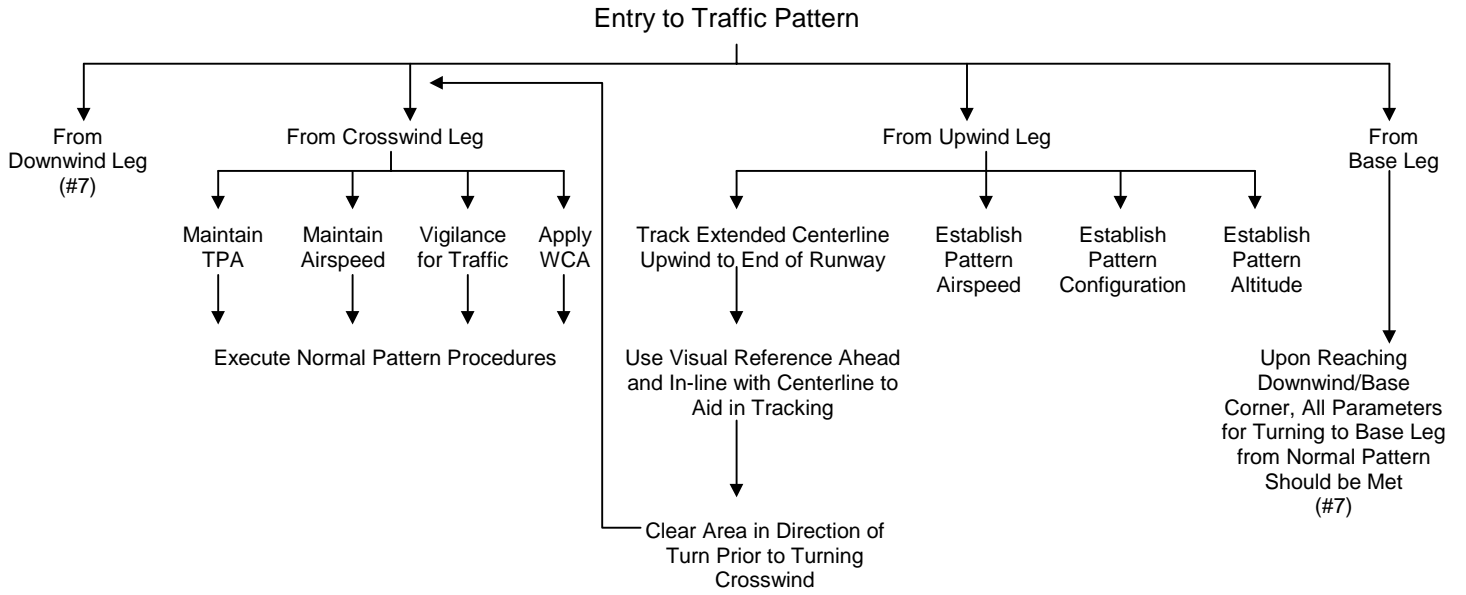
↔ = Loop Activity

Maneuver Hierarchy #10

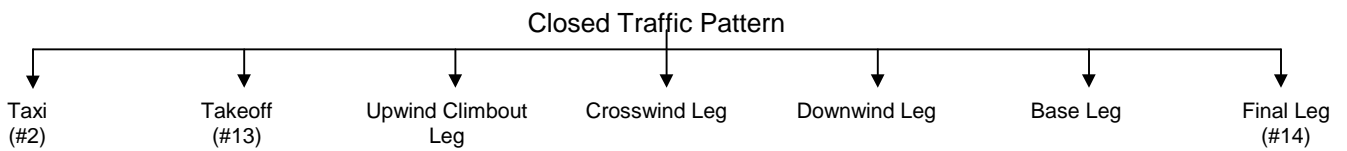


↔ = Loop Activity

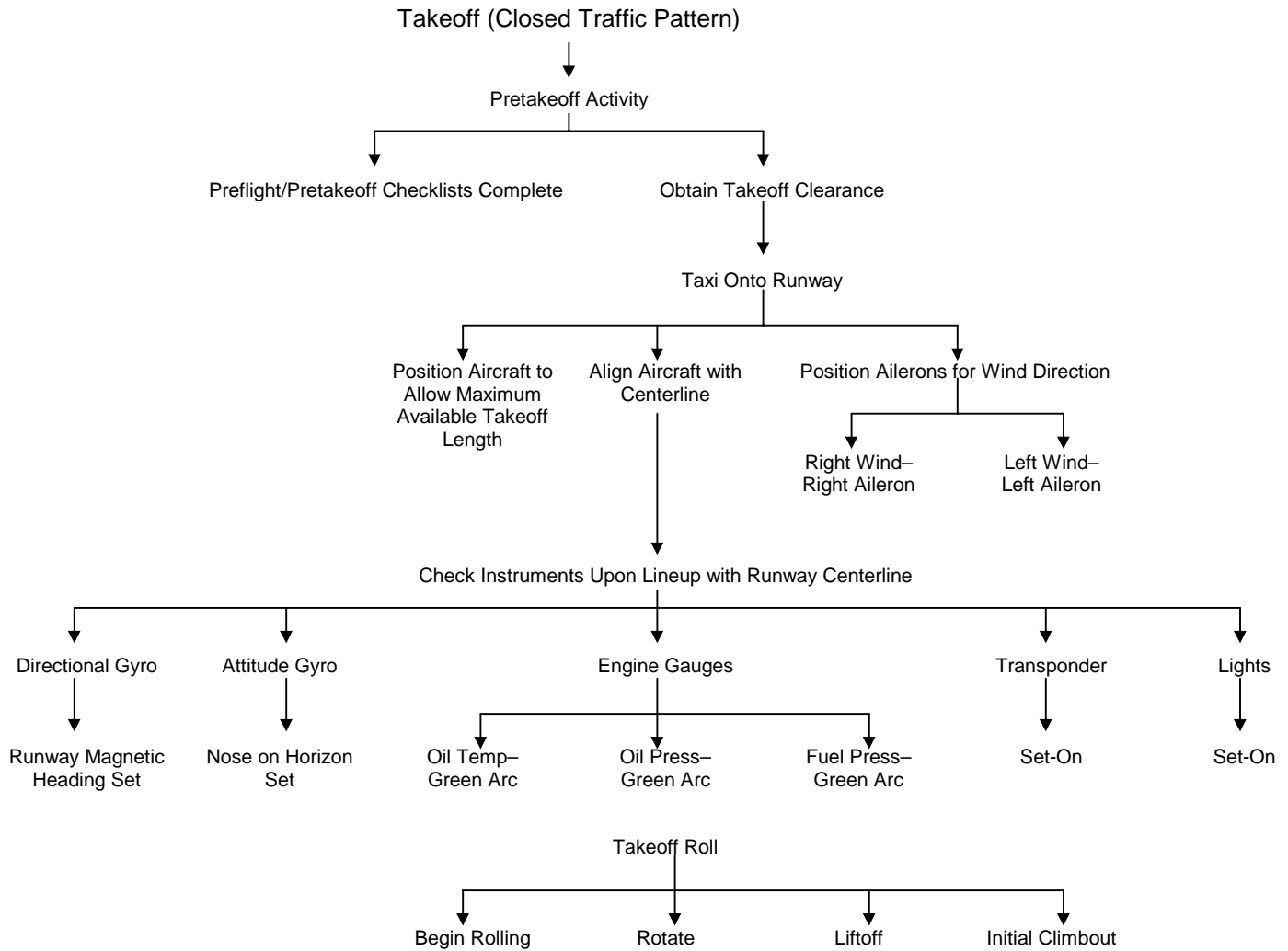
Maneuver Hierarchy #11



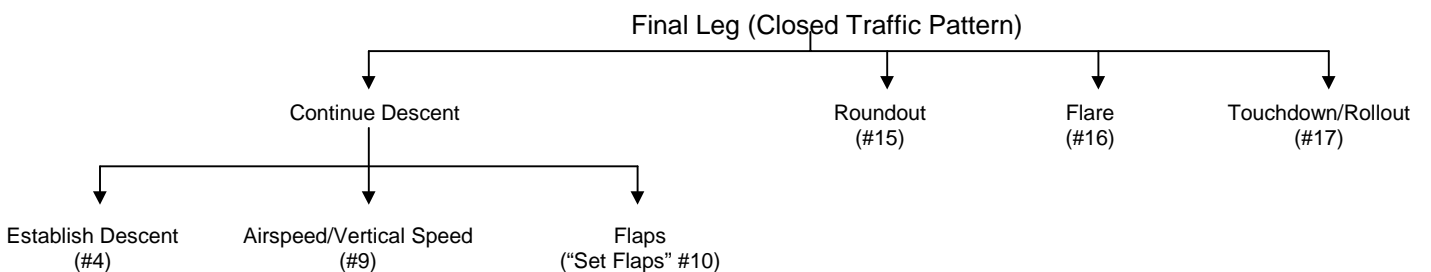
Maneuver Hierarchy #12



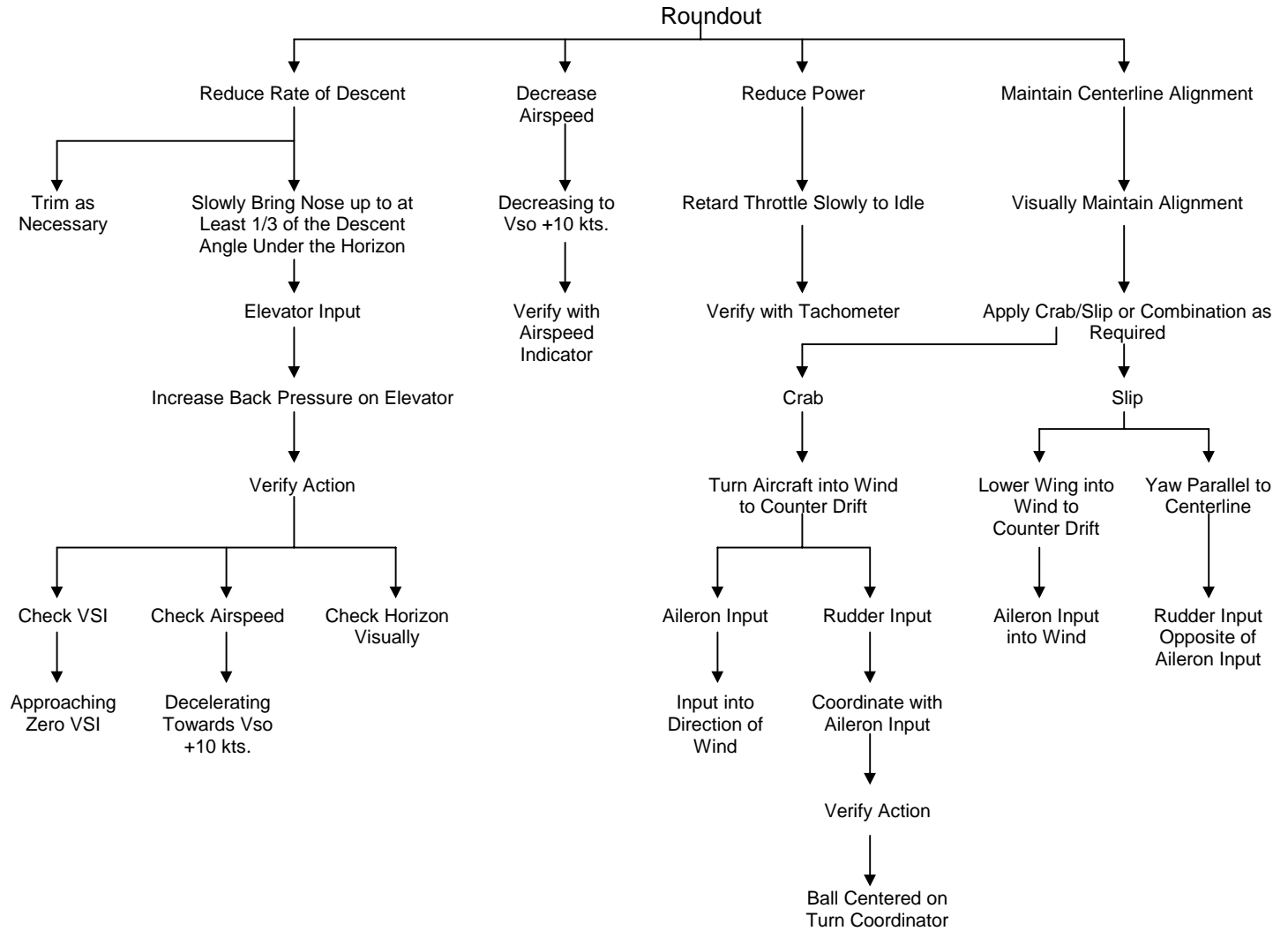
Maneuver Hierarchy #13



Maneuver Hierarchy #14

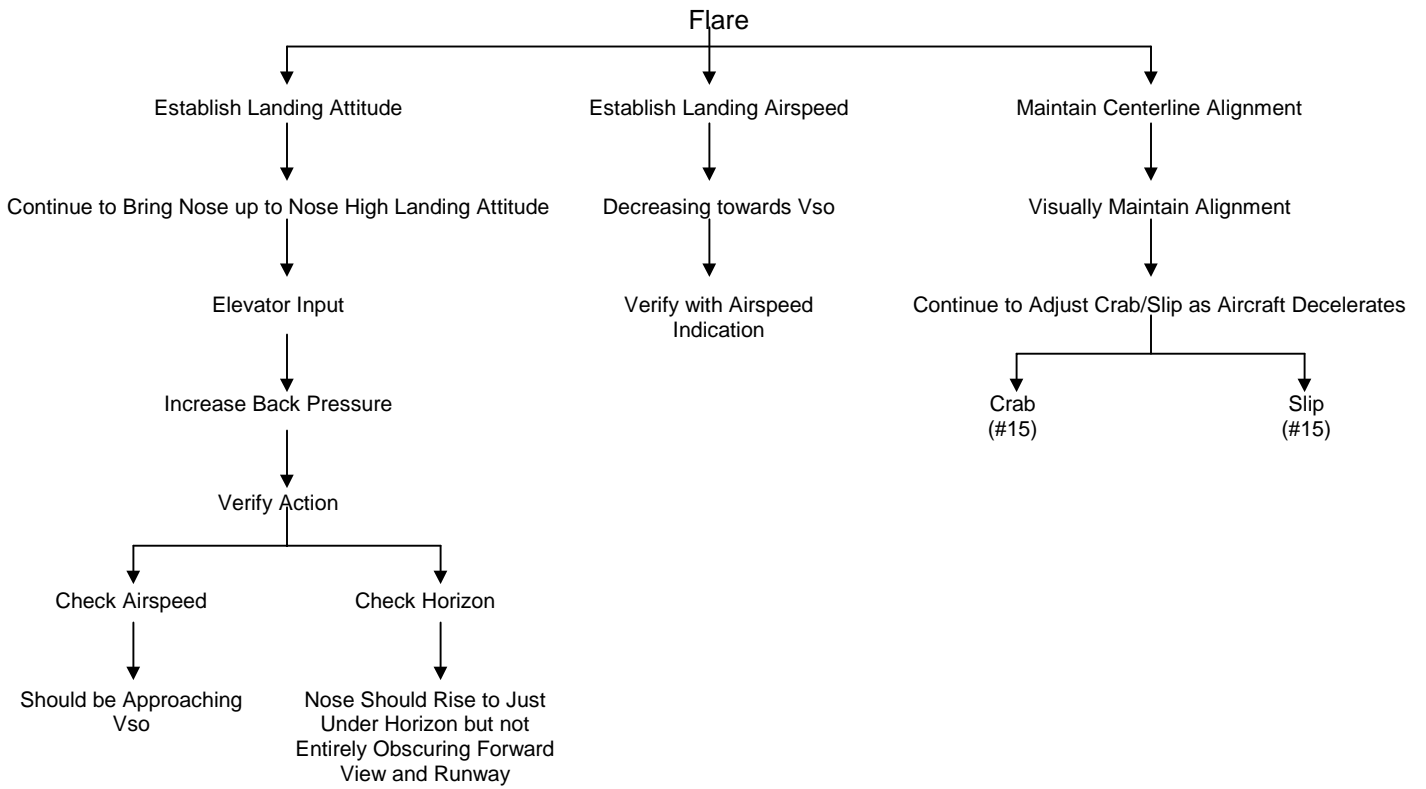


Maneuver Hierarchy #15



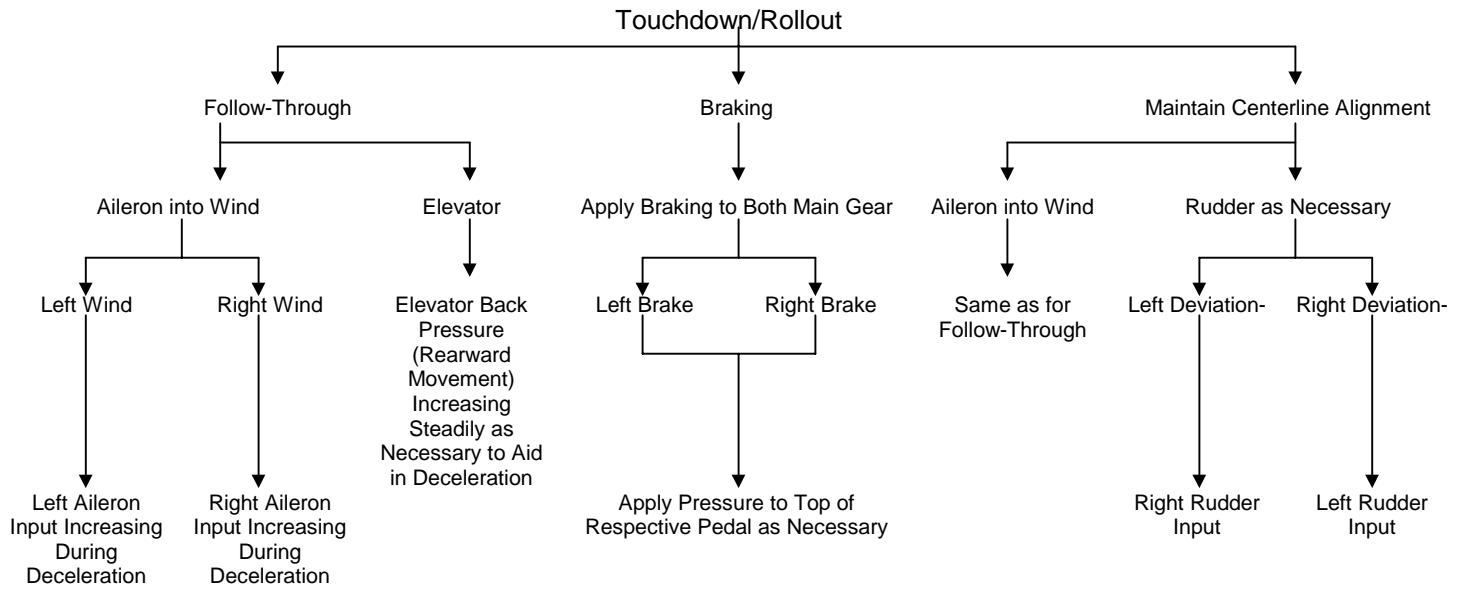
Rate of descent indirectly effects airspeed.
 Power directly effects airspeed.
 Power indirectly effects rate of descent.
 Airspeed directly effects ability to maintain centerline via crab/slip.
 Aileron and rudder inputs in crab or slip directly effect each other.

Maneuver Hierarchy #16



Rate at which attitude is achieved directly effects airspeed.
 Airspeed indirectly effects rate at which attitude is achieved.
 Airspeed directly influences crab/slip requirements.

Maneuver Hierarchy #17



Aileron during follow-through has an indirect effect on centerline alignment and rudder requirements. Braking symmetry has a direct effect on centerline alignment.

Appendix 2

Evaluation Mission Scenario

Introduction:

“During this mission flight you will be flying a mission designed to give you the basic feel of the training device. You will be asked to perform several maneuvers during the mission. None of these maneuvers will be scored and your overall performance is not being recorded. However, you should try to perform each maneuver as well as possible since the goal of the flight is to evaluate the training device that you are flying. After the mission flight, you will be asked to answer a questionnaire about this device and its capabilities. The length of the mission flight depends on how long you take to perform each specified maneuver. You may feel free to ask me questions during the flight if you have questions about the procedure for any particular maneuver.”

“The mission flight will begin shortly. The flight starts on the ground at Campaign and finishes at Champaign.”

Task Narrative:

- 1) *“You are currently on the ground at Champaign on a runway. Release the brake and immediately make a 180 degree taxiing turn. Do not worry about taxiing off of the runway. When you reach your new heading, apply the brakes and come to a complete stop.”*

After subject comes to a complete stop:

- 2) *“Now, taxi back into position on the runway and release the brakes for a normal takeoff. Continue to fly runway heading throughout the takeoff.”*

After subject has taken-off:

- 3) *“Continue to climb on runway heading to 3,000 feet. Upon reaching 3,000 feet, make a normal level-off.”*

After subject has leveled off:

- 4) *“Now make an 82 kt. climb to 4,000 feet. Level out at 4,000 feet.”*

After subject levels out from the climb:

- 5) *“Now fly straight and level for 2 minutes. I will keep the time”*

After two minutes:

- 6) *“Now make a 500 foot per minute climb to 5,000 feet and level out upon reaching 5,000 feet.”*

After subject levels out from the climb:

- 7) *“Now lets make a 500 foot per minute descent to 4,000 feet. Level out at 4,000 feet.”*

After subject levels out from descent:

- 8) *“Now lets make an 82 kt descent to 3,000 feet. Level out at 3,000 feet.”*

After subject levels out from descent:

- 9) *“Now make a 360 degree standard rate turn to the left followed by one to the right.”*

After subject rolls out from turns:

- 10) *“Now make a 360 degree steep turn to the left followed by one to the right.”*

After subject rolls out from turns:

- 11) *“Now perform a power off stall as you would in the aircraft. After performing a full stall, recover the aircraft to normal straight and level flight.”*

After subject has performed stall:

- 12) *“Now perform a power on stall as you would in the aircraft. After performing a full stall, recover the aircraft to normal straight and level flight.”*

After subject has performed stall:

- 13) *“Now lets configure the aircraft for slowflight. The configuration should include full flaps. Also, maintain altitude and heading during this maneuver. Inform me when you are satisfied that you have the device in slowflight.”*

After subject informs that they are established in slowflight:

“Lets maintain slowflight for 1 minute.”

After one minute of slowflight:

“Now resume normal cruise flight. Maintain your heading and altitude during the recovery.”

After subject has resumed normal cruise flight:

- 14) *“Now simulate a engine failure forced landing and emergency approach to a field. Upon reaching minimum safe altitude recover and climb back to 3,000 feet.”*

After subject has climbed back to 3,000 feet:

- 15) *“Now tune the Navigation radios as appropriate for return to Champaign. Once you’re sure about your position, proceed direct to Champaign for a full stop landing on a runway of your choice.”*

After the subject is enroute to the airport:

“You may descend at your discretion. The airport’s elevation is at 754 mean sea-level.”

After touchdown (if subject is successful in making a touchdown) freeze the system and ask them to complete the phase II or III survey.

Appendix 3

Maneuver-Based Questionnaire

In this phase of our experiment you will be asked several questions relating to the use of this particular training device. You will fly a mission similar to that of a private pilot checkflight. However, you will not be graded nor is your performance on any task of interest to us. The goal is for you to evaluate the capability of the device as if you intended to use it to give private pilot instruction. After flying the mission, you will complete the attached questionnaire dealing with the flight maneuvers you've just flown.

Your answers will be coded according to a pre-assigned subject number that will not be linked to your name. None of the information you provide will be disclosed to anyone other than Mr. Talleur or the other collaborators in this study.

So that you understand what you'll be asked to evaluate, please read through the following questions at this time. When finished you may start your evaluation of the training device.

Subject number:

1. Name _____

2. Device being evaluated _____

3. Do you have any experience flying this particular training device or the same model device elsewhere? (Circle one)
Yes No

4. Do you have any experience teaching in this particular training device at this flight school? (Circle one)
Yes No

5. If you answered **NO** to question #4, have you ever given flight instruction in the same model device elsewhere?
Yes No

continue to next page

6. Rate the following VFR maneuvers on how much you feel that a student could learn about the maneuver from practicing in this training device. (Circle a number)

	Will learn very little				Will learn quite a lot
a. taxi					
	1	2	3	4	5
b. normal takeoff					
	1	2	3	4	5
c. climbs					
	1	2	3	4	5
d. straight and level flight					
	1	2	3	4	5
e. descents					
	1	2	3	4	5
f. shallow turns					
	1	2	3	4	5
g. steep turns					
	1	2	3	4	5
h. power-off stalls					
	1	2	3	4	5
i. power-on stalls					
	1	2	3	4	5
j. slowflight					
	1	2	3	4	5
k. landing approach					
	1	2	3	4	5
l. normal landing					
	1	2	3	4	5
m. engine failure					
	1	2	3	4	5
n. gliding descent					
	1	2	3	4	5

Appendix 4

Skill-Based Questionnaire

In this phase of our experiment you will be asked several questions relating to the use of this particular training device. You will fly a mission similar to that of a private pilot checkflight. However, you will not be graded nor is your performance on any task of interest to us. The goal is for you to re-evaluate the capability of the device as if you intended to use it to give private pilot instruction. After flying the mission, you will complete the attached questionnaire dealing with certain aspects of the training device and task elements of the flight maneuvers you've just flown.

Your answers will be coded according to a pre-assigned subject number that will not be linked to your name. None of the information you provide will be disclosed to anyone other than Mr. Talleur or the other collaborators in this study.

So that you understand what you'll be asked to evaluate, please read through the following questions at this time. When finished, the mission flight will begin

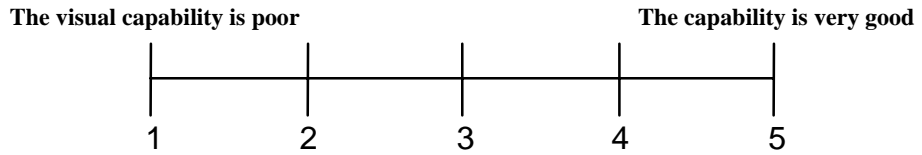
Subject number:

1. Name _____

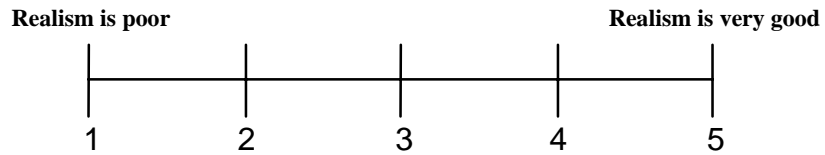
2. Device being evaluated _____

continue to next page

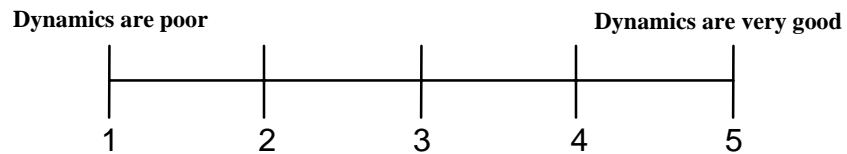
3. How would you rate the visual capability of this training device for VFR maneuvers?
(Circle a number)



4. How would you rate the realism of the cockpit layout of this training device?
(Circle a number)



5. How would you rate the dynamics (control inputs/ performance) of this training device?
(Circle a number)



continue to next page

6. Rate, using the following scale, your impression of how effective this device will be for learning these basic flight elements.

	Will learn very little				Will learn quite a lot
a. basic aircraft control	1	2	3	4	5
b. aircraft trimming	1	2	3	4	5
c. yaw control	1	2	3	4	5
d. roll control	1	2	3	4	5
e. pitch control	1	2	3	4	5
f. effects of angle of bank on lift	1	2	3	4	5
g. effects of airspeed on lift	1	2	3	4	5
h. effects of pitch on airspeed	1	2	3	4	5
i. effects of flaps on airspeed	1	2	3	4	5
j. visually detect drift with respect to ground track	1	2	3	4	5
k. making pitch changes with respect to the horizon	1	2	3	4	5
l. making bank angle changes with respect to the horizon	1	2	3	4	5
m. judge ground proximity during landing phase	1	2	3	4	5

continue to next page

Appendix 5

Biographical Questionnaire

The purpose of this questionnaire is to ascertain your aviation background and inquire about your thoughts on various topics pertaining to simulation. Please answer the following questions as accurately as possible.

Your answers will be coded according to a pre-assigned subject number that will not be linked to your name. None of the information you provide will be disclosed to anyone other than Mr. Talleur or the other collaborators in this study.

Subject number:

1. Name _____

2. Age _____

3. Native Speaking language _____

4. Handed: Right_____

 Left_____

 Both_____

5. Corrective Lenses: None_____

 Glasses_____

 Contacts_____

 Glasses and Contacts_____

continue to next page

6. Circle all flight certificates and ratings held

- a. Private SEL
- b. Private MEL
- c. Private SES
- d. Private MES
- e. Commercial SEL
- f. Commercial MEL
- g. Commercial SES
- h. Commercial MES
- i. Private or Commercial Glider
- j. Flight Instructor SE
- k. Flight Instructor ME
- l. Flight Instructor Glider
- m. Instrument SE
- n. Instrument ME
- o. Flight Engineer
- p. Airline Transport Pilot SE
- q. Airline Transport Pilot ME
- r. Type Ratings (Please list)_____

s. Please list any other pilot certificates or ratings not listed above. _____

7. Please list all non-pilot flight certificates and ratings held_____

continue to next page

8. Please circle all of the following aircraft that you have flown.

<u>Aerospatiale</u>	<u>Beechcraft</u>	<u>Cessna</u>	<u>Commander</u>
TB-9	BE-19	AGPICKUP	114
TB-10	BE-23	AGTRUCK	690
TB-20	BE-A24R	AGWAGON	
	BE-G35	CARAVAN	<u>Piper</u>
<u>Citabria</u>	BE-J35	C-150	DAKOTA
7ECA	BE-V35	C-152	CHEYENNE
	BE-A36	C-170	SARATOGA
<u>Grumman American</u>	BE-B50	C-172	TOMAHAWK
AA-5B	BE-E55	C-172RG	PA23-160
	BE-56TC	C-180	PA23-250
<u>Mooney</u>	BE-B60	C-182	PA24-260
M20	BE-76	C-182RG	PA28-140
M201	BE-B80	C-185	PA28-161
M205	BE-C90SE	C-206	PA28-181
M231	BE-E90	C-207	PA28-200
	BE-B100	C-T207	PA28-201
	BE-200	C-208	PA31-310
		C-210	PA31-350
		C-P210	PA32-301
		C-310	PA32-301T
		C-320	PA34-200
		C-P337	PA44-180
		C-340	PA46-310
		C-402	
		C-414	
		C-421	

9. Please list any other aircraft that you've flown that are not listed above. (including all military types) _____

continue to next page

10. Fill in the appropriate hours

- a. Total time _____
- b. Total dual ground training device and simulator time _____
- c. Total Elite, FS-100, or other desktop PC Simulation Device time _____
- d. Total Simulated Instrument time (Hood) _____
- e. Total Actual Instrument time (IMC) _____
- f. Total Dual received _____
- g. Total Pilot- in- Command (PIC) time _____
- h. Total Second- in- Command (SIC) time _____
- i. Total X-country flight time _____
- j. Total time in last 90 days _____

11. Have you ever received or given flight instruction in a (circle all that apply)

- a. GAT I
- b. GAT II
- c. GAT III
- d. Frasca 141
- e. Frasca 142
- f. Frasca 242
- g. Other Frasca trainer
- h. Illimac trainer
- i. Azuresoft's Elite desktop trainer
- j. MDM's FS-100 desktop trainer
- k. Jeppesen's FS-200 desktop trainer
- l. A full motion (six degrees of motion) simulator
- m. Other non-motion training device
- n. Other PC training device

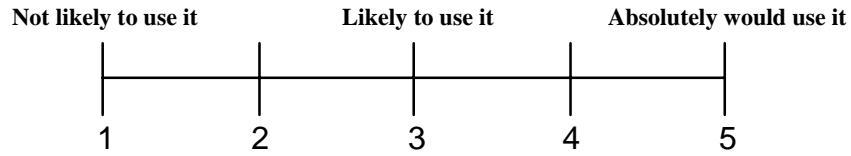
12. Have you ever received or given instruction in a ground training device that has a visual system. (Circle one)

Yes No

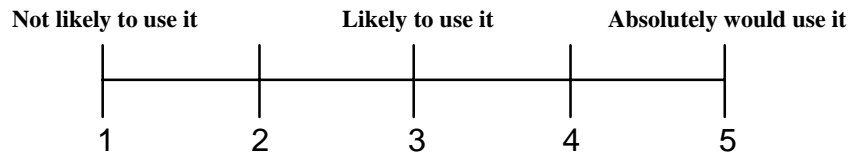
continue to next page

General: Attitudes about simulation

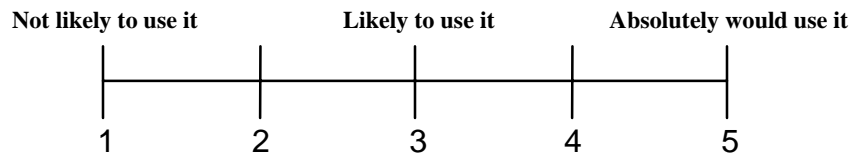
1. If you have the option of using an approved ground training device in training a student for a private pilot's certificate, how likely are you to use it? (Circle a number)



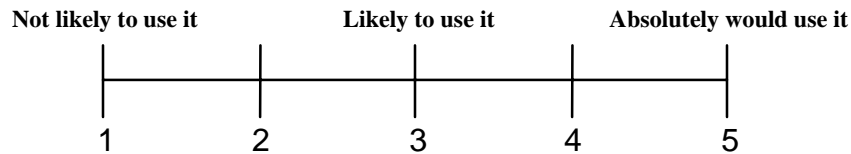
2. If you have the option of using an approved ground training device in training a student for a instrument rating, how likely are you to use it? (Circle a number)



3. If you have the option of using a desktop PC training device in training a student for a private pilot certificate, how likely are you to use it? (Circle a number)

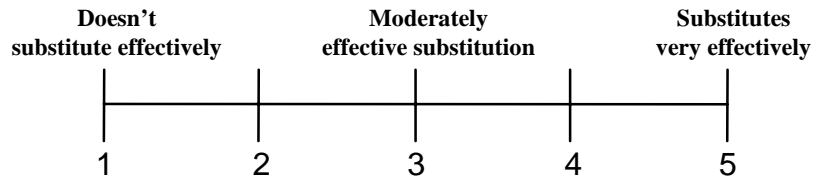


4. If you have the option of using a desktop PC training device in training a student for a instrument rating, how likely are you to use it? (Circle a number)

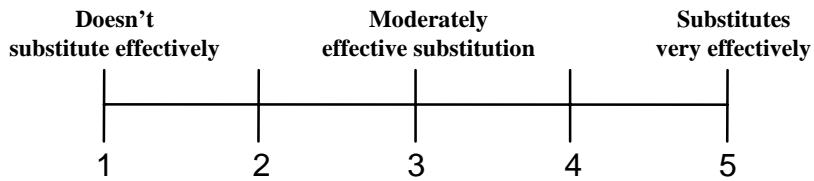


continue to next page

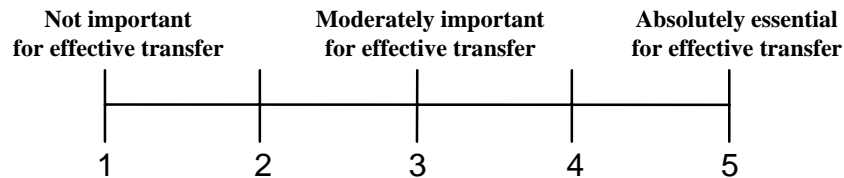
5. How effectively do you believe practice on private pilot maneuvers in a ground trainer with a visual system can substitute for training in the aircraft? (Circle a number)



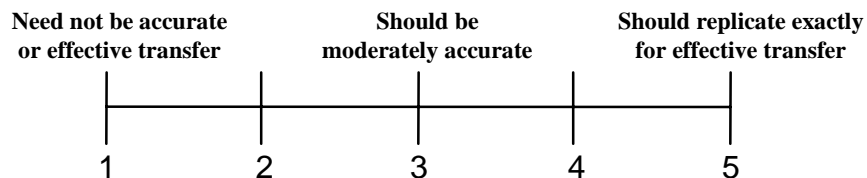
6. How effectively do you believe practice on instrument maneuvers in a ground trainer can substitute for training in the aircraft? (Circle a number)



7. For effective transfer of training from a ground training device to an aircraft, how important is it that the simulator moves to simulate roll, pitch, and yaw? (Circle a number)

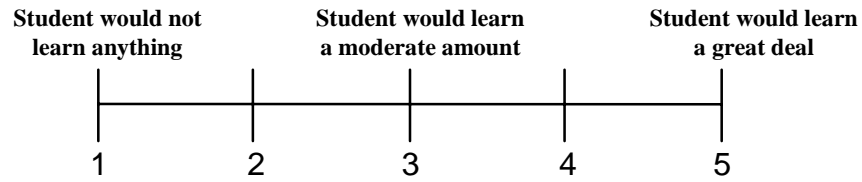


8. How accurately do you believe a ground trainer's instrument panel should replicate the training aircraft's instrument panel layout in order to achieve successful transfer of training? (Circle a number)

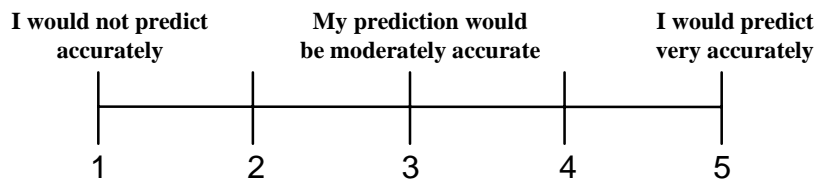


continue to next page

9. How much do you think a student could learn about VFR maneuvers in a non-visual training device that would transfer to VFR flight in the aircraft? (Circle a number)



10. If given the opportunity to test fly a training device, how well do you think you could accurately predict which maneuvers would be worth while to train in that training device? (Circle a number)



stop here