HUMAN FACTORS, 1982, 24(1), 61–73

Pilot Judgment: Training and Evaluation

RICHARD S. JENSEN,[1] *Department of Aviation, The Ohio State University, Columbus, Ohio*

An analysis of accident statistics reveals that over 50% of pilot-caused civil aviation accident fatalities are the result of faulty pilot judgment. Although the FAA requires examiners to evaluate pilot judgment, it provides no definition or criteria against which such an evaluation can be made. In spite of the statistics implicating pilot judgment in many aviation fatalities, attempts to teach it are almost nonexistent. It is but a slight overstatement to say that good pilot judgment is learned by the lucky and the cautious over many years of varied flying experiences. This paper examines some of the decision-research literature in an attempt to provide an operational definition of pilot judgment and to suggest ways that pilot judgment may be taught and evaluated in civil aviation.

INTRODUCTION

From the beginning of aviation history, pilots have been expected to exercise a considerable amount of judgment in the overall task of flying an airplane. However, in recent years, increasing societal demands for safety, dependability, economy, effectiveness, and reduced energy consumption have increased the complexity of civil and military flying operations, thus magnifying the pressures for good pilot judgment. Furthermore, technological advances that have eased much of the pilot's burden for precise aircraft control have not greatly eased the pilot's decision-making workload. In many cases these advances have only created demands for higher levels of skill, knowledge, and judgment for which few pilots have been trained, and the costs to train them to operate effectively in this changing system are becoming prohibitive.

If it were merely a matter of teaching flying skills, the task of training pilots to operate safely in our complex aviation system would be a much smaller one than it is. Unfortunately, because actual conditions are never quite the same as those used to develop aviation regulations, procedures, and performance limitations, the safety of a given flight also depends upon a significant amount of evaluation and interpretation of existing conditions by the pilot.

For example, the conditions used to develop flight performance values for a particular type of airplane may be ideal; e.g., clean airplane surfaces, a new engine, a new propeller, an unrestricted air filter, and a company test pilot. In actual conditions, the pilot must compare these book values obtained in ideal conditions with those actually being encountered. These actual conditions may include a dirty airplane, a slightly used engine, a few marks on the propeller, a somewhat dirty air filter, and a less than perfect pilot. The pilot must then evaluate many other conditions, such as gross weight, center of gravity, wind, temperature, humidity, altitude, etc., for comparison with those used in

[1] Requests for reprints should be sent to Dr. Richard S. Jensen, Department of Aviation, The Ohio State University, Box 3022, Columbus, OH 43210.

the book to determine the expected flight performance. Finally, the present and forecast weather, the terrain, and the expected traffic density must be checked. These are then compared with the pilot's estimate of her or his own capability before determining whether or not the planned flight will be safe.

Examples of decisions being made with imperfect information can be found in all areas of flight activity. Furthermore, every decision that the pilot makes is colored by physiological, psychological, and social pressures that are virtually impossible to weigh properly on the spot. For example, just as persons watching a sporting event may "see" an infraction or foul differently, depending upon their vantage points and which team they support, a pilot may be influenced to view the weather outlook or his or her own abilities differently, depending upon the importance or value he or she assigns to a given flight. Evidence from the social psychology literature (Kogan and Wallach, 1964) suggests that one's self-image and the need to maintain an external image largely determine how much effect the costs and payoffs for a given flight will have on the judgment of one's own ability to make a safe flight.

Some pilots may be susceptible to social pressures that result in less than rational pilot judgment. Irrational pilot judgment has been exhibited by such unsafe practices as flying under bridges, landing on busy highways, attempting to land in football stadiums, and flying "formation" on other unsuspecting pilots. Potential sources of social pressure that may lead to these types of activities include peer reactions, fear of failure, censure from superiors or family members, and many other factors (Janis and Mann, 1977).

The Problem

An analysis of accident statistics by categories of pilot activities reveals the serious nature of the problem. Most analyses of aviation accident statistics have found that from 80 to 85% of accidents can be assigned broadly to "pilot error" and the remainder to mechanical malfunctions. To determine why pilots are making accident-causing errors, it is useful to provide a more molecular examination of pilot activities. One such classification (Roscoe, 1980) divides these activities into three categories as follows:

(1) *Procedural Activities*—the management of the power plant, fuel, vehicle configuration, autopilot, displays, navigation, and communication.
(2) *Perceptual-Motor Activities*—including vehicle control, judgment of distance, speed, altitude, clearance, hazard detection, and geographic orientation.
(3) *Decision Activities*—including the self-assessment of skill, knowledge, physical, and psychological capabilities; the assessment of aircraft and ground-system capabilities; hazard assessment; navigation planning; and flight priority adjustment.

To determine the relative importance of each of these activities in civil aviation accidents, statistics from the National Transportation Safety Board (NTSB) Automated Aircraft Accident and Incident Information System from 1970 through 1974 were classified into the three behavioral categories given above (Jensen and Benel, 1977). Then the total numbers of both fatal and nonfatal accidents during the 5-year period were determined for each of these behavioral categories. The results of these analyses are shown in Table 1.

Examinations of these data provide valu-

TABLE 1

Number and Percentage of the Total General Aviation Accidents in Which the Pilot Is Listed as a Cause or Factor (1970-1974)

	Fatal	*Nonfatal*
Procedural	264 (4.6%)	2 230 (8.6%)
Perceptual-Motor	2 496 (43.8%)	14 561 (56.3%)
Decisional	2 940 (51.6%)	9 087 (35.1%)

able indications of possible weaknesses in current civil aviation programs. For example, a majority of the nonfatal pilot-caused accidents (56.3%) were the result of faulty perceptual-motor behavior. The most significant factors here ("failure to maintain flying speed" and "misjudgment of distance, speed, altitude, or clearance") represent one type of pilot judgment. On the other hand, a majority of the fatal pilot-caused accidents (51.6%) were the result of faulty decision-making behavior, another type of pilot judgment. The most significant factors in this area were the familiar "continued VFR (visual flight rules) into known adverse weather" and "inadequate preflight planning or preparation."

It is apparent from these accident statistics that both aspects of the decision-making function are important to safe flight and may be neglected in the present training and testing process. However, because it suffers from greater misunderstanding in aviation circles, pilot judgment as represented by general decisional activities is the topic of concern in this paper. Although a significant amount of research has been done on the decisional aspect of judgment in recent years (Janis and Mann, 1977), no one has specifically examined the judgment problem faced by the pilot, the flight instructor, and the pilot examiner.

Three major tasks must be accomplished before major improvements in pilot training and evaluation can be realized. The first task is to establish a common definition of judgment as it applies to flying. At present, even though the term is used repeatedly in aviation circles, and although FAA examiners are required to evaluate candidates on the basis of judgment, no such definition exists.

The second major task is to determine whether or not pilot judgment can be taught and, if so, how best to teach it. Since some aspects of pilot judgment are closely akin to personality characteristics, they may be difficult to modify. It might be necessary to use testing and selection procedures to improve this aspect of aviation safety and effectiveness. Other aspects of pilot judgment are more easily modified through systematic training procedures.

The third major task is to determine whether or not pilot judgment can be evaluated reliably, meaningfully, and objectively. Because judgment is primarily a mental process, it may be difficult to evaluate it reliably. On the other hand, behavioral events frequently have been used to indicate mental activity. Although personality tests have proved to be somewhat unreliable, research results using these instruments may be useful in the development of instruments for evaluating and predicting judgment-making behavior (Fishbein and Ajzen, 1975).

Judgment Definition

As indicated above, the word "judgment" has been used to describe two somewhat different mental processes in aviation. Perhaps its most common usage has been to describe the mental activity that takes place at the perceptual-motor level. The other describes the mental activity involved in choosing a course of action from among several alternatives. Obviously, this second usage of the term is similar to the first in that both involve making choices. However, there is a basic difference.

The first refers to highly learned perceptual responses that must be made in a very short time, in some cases continuously. The second refers to cognitive decisions for which set procedures either have not been established or may have been forgotten. Flight instructors have used various terms referring to this type of judgment, including "headwork," "thinking ahead," and "staying ahead of the aircraft." In a more general sense, pilots have referred to it as professionalism or commandability. In contrast to the perceptual-motor judgments, selection of alternatives usually

occurs when more time is available to evaluate the situation, a larger number of possible courses of action must be considered, and there is a greater degree of uncertainty concerning the existing situation and possible outcomes. For these reasons, cognitive judgments have been the source of greater misunderstanding in pilot training and evaluation.

These two aspects of judgment may be considered as two ends of a continuum based on cognitive complexity and decision time. At one end of the continuum are the common perceptual judgments of distance, altitude, speed, and clearance. These perceptual judgments are less complex in that they involve fewer pieces (frequently one) of fairly accurate information from which responses are determined with highly learned motor behavior. They may require simple responses, but frequently call for immediate control movement. At the other end are what might be called cognitive judgments. As described above, these judgments are complex in that they usually involve a large number of relevant pieces of highly probabilistic information, they usually require the specification of and choice from among several alternatives, and they are frequently affected by emotions, values, and social pressures. In addition, cognitive judgments usually permit some deliberation before a control response is required.

Candidate judgment definition. Considering these factors, a candidate definition of cognitive judgment in flying airplanes is

(1) The ability to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action, and to determine expected outcomes from each alternative.
(2) The motivation to choose and authoritatively execute a suitable course of action within the time frame permitted by the situation, where: (a) "Suitable" is an alternative consistent with societal norms; (b) "Action" includes no action, some action, or action to seek more information.

The first part of the definition refers to intellectual or discriminating abilities. This part depends upon human capabilities to sense, store, retrieve, and integrate information. This function is what Van Dam (in Jensen and Benel, 1977) calls the "discriminating ability" in professional pilots. It is analogous to detectability (d') in signal detection theory. It is purely rational and could be stated mathematically. If it were possible to separate this part of human judgment from the second part (which it is not), people would solve problems in much the same way as a computer does. This is not to say that the decisions would be error free. Probabilistic information is used, and the performance is dependent upon the amount, type, and accuracy of information stored, as well as on inherent and learned capabilities to process information.

The second part of the definition refers to motivational tendencies. The emphasis in this case is on the directional aspects of motivation rather than on intensity. It says that a part of human judgment is based upon bias factors (costs and payoffs) or tendencies to use less than rational information (defined by the society) in choosing courses of action. The society would probably consider the use of any information other than that required to define the safety risk (e.g., convenience, monetary gain, gain in self-esteem, adventure seeking, etc.) as less than rational. This part of human judgment is analogous to the response criterion (β) in signal detection theory. It is what Van Dam (in Jensen and Benel, 1977) has called the "response pattern" of the professional pilot. If properly developed, this part of human judgment would tend to halt the use of information not directly related to the safety of the flight and to direct the pilot's decision toward the use of rational processes.

JUDGMENT TRAINING

One of the most important questions to be addressed after establishing the definition is

whether or not pilot judgment, as defined, can be modified through training. The paucity of judgment-training guidelines in pilot-training and training-research literature might lead one to conclude that pilot judgment cannot be taught. Literature and syllabi commonly used in flight-instructor courses contain large sections on how to teach the motor skills of flying but offer very little on how to teach pilot judgment (see Federal Aviation Administration, 1977). The typical private-pilot course offers a scattering of judgment-making instruction in the areas of weather avoidance and power-plant emergencies, but does not offer systematic judgment training.

However, there is evidence in aviation that at least one form of judgment training, i.e., assigning procedures for every conceivable situation that might arise, may be effective. In the military these are referred to as "Boldface" training procedures (Thorpe, Martin, Edwards, and Eddowes, 1976). Similar training procedures used by the airlines have even been more successful than those used by the military (Gibson, 1969; Trans World Airlines, 1969).

Looking outside the field of aviation one finds other evidence indicating that judgment making may be taught. For example, although signal detection theory (SDT) was not designed specifically to handle cognitive judgments, many of its methods can be used to explain and perhaps even to modify pilot judgment behavior. SDT divides an individual's decision behavior into two components representing his or her sensitivity (d') and response criterion or bias (β), roughly corresponding to the two aspects of the judgment definition previously stated.

The sensitivity is affected both by the inequality of the stimulus in the situation (signal vs. background noise) and the quality of the sensory apparatus of the observer. In cognitive judgment this is the intellective component. It refers both to the strength of the evidence favoring the selection of one alternative over another and to the knowledge and ability of the pilot to use effective strategies in gathering and processing the evidence.

On the other hand, the response criterion represents the level of excitation, above which the observer is willing to say "signal." It is the amount of relevant information in the presence of irrelevant information or noise needed to tip the decision one way or the other. The criterion level (i.e., the amount of relevant information required) is influenced by motivation, knowledge of the signal's probability of occurrence, and the costs and payoffs attendant with a given response. In cognitive judgment the response criterion is the motivative component.

The response criterion can be manipulated through a wide range of values by adjusting probabilities, costs, and payoffs (Birdsall, 1955). One can infer from the vast amount of psychophysical decision research that cognitive judgments can be modified in a similar way. The flight instructor communicates decision biases, attitudes, risk tendencies, consideration for passenger safety, and pilot motivation by example, if not by design, at all levels of pilot training. The method by which the instructor teaches these values is influenced by the instructor's assignment of probabilities, costs, and payoffs to actions of the student.

Although the SDT model assumes that the sensitivity component is quite stable for a given individual, there is a growing field of research indicating that, when dealing with the intellective component of cognitive judgment, sensitivity can be modified as well. For example, attempts have been made to discover the mental processes that are used by expert judges, such as stockbrokers, livestock judges, and medical diagnosticians, in making their decisions (Shanteau and Phelps, 1977; Slovic, 1969; Anderson, 1969; Hoffman,

Slovic, and Rorer, 1968). The thesis is that if models of the mental processes used by these experts in decision making were available, they could be used in training others to use similar processes. In each of the areas studied, judgment training traditionally occurs over a fairly long apprenticeship program, in which the trainee observes the expert making decisions and learns by this observation. However, as in aviation, because of the complexity of the information used to make decisions, observation or even trial and error are inefficient training methods.

Research on the motivational aspect of cognitive judgment also indicates that training can have a beneficial effect. The major research efforts in this area are reported by Janis and Mann (1977). These authors, speaking from a clinical perspective, begin with the assumption that psychological stress is a frequent cause of errors in decision making. They say that stress arises from at least two sources. First, the decision maker is concerned about the material and social losses that may be suffered from whichever course of action is chosen, including the costs of failing to live up to prior commitments. Second, the pilot's reputation and self-esteem as a competent decision maker are at stake. The more severe the anticipated losses, the greater the stress.

Janis and Mann have constructed a "conflict-theory" model of decision making, postulating that the way difficult choices are resolved is determined by the presence or absence of three conditions: "awareness of risks involved," "hope of finding a better solution," and "time available in which to make the decision." They have developed several clinical procedures to improve decision making under the titles, "awareness of rationalizations," "emotional role playing," "balance sheet," and "outcome psychodrama." They report that these procedures have demonstrated effectiveness in changing decision-making tendencies and in modifying attitudes.

Some Learning Principles

Because some well-established learning principles are misapplied in many training programs, a discussion of these principles as applied to pilot judgment training is needed. Perhaps the most popular "principle" is that the best way to learn an activity is to practice that activity (Gagne, 1962). The assumption is rooted in much of the educational literature and is often identified by the catch-phrase "learning by doing." Gagne points out that it may also be a generalization from research on the conditioned response, in which learning, particularly in animals, appears to have occurred only after a response (practice) has been made.

However, Gagne argues that practice is not an effective training method by itself, even for the acquisition of such motor skills as field gunnery. He says that "instruction about the correct sighting picture for ranging is more effective in bringing about improved performance than is practice on the task" (Gagne, 1962, p. 85). The point is that training should emphasize the principles and procedures (thought processes) involved, and practice should be directed to take advantage of these principles. If this is the proper emphasis for teaching motor skills, it is even more important in the teaching of judgment-making skills, which are more deeply rooted in thought processes.

A second learning principle that is frequently misapplied in training situations is variously called reinforcement, feedback, or knowledge of results during practice. This principle has been found to be most effective in choice behavior. However, Gagne points out that some manipulations that artificially improve feedback during practice failed to show reliably better transfer to the opera-

tional environment, and that others showed negative transfer, perhaps because of a learned dependence on the artificial cues. Apparently the form of the feedback is important.

Any beginning flight student will say that such usual feedback information as "you did it right" or "you did it wrong" is almost useless. The time interval between execution and feedback may be long; the feedback is often cluttered with interfering information; and the trials themselves are often so complex that the student may learn very little from such a response by the flight instructor. A student really needs to know *why* the response was right or wrong. He or she needs to know what rules should have been followed and if and how those rules were violated. Although practice and right/wrong types of feedback may be useful in some training situations, they should be deemphasized in favor of these thought-oriented teaching principles in all types of pilot training, but especially in judgment training.

A Systematic Approach to Training

The need for pilot judgment training has been established for all levels of flight instruction. Without a systematic judgment-training program, good pilot judgment is acquired by the cautions and the lucky over years of flying experience in many varied situations. The primary task of aviation educators using systematic judgment-training techniques is to compress a lifetime of flying experience into a relatively short training program to instill good pilot judgment into the emerging pilot.

In addition, a ground school might include instruction on information integration and subjective probability estimation (Goldberg, 1968). In expert judges, judgment-making behavior is characterized by chunking, or the formation of clusters of stimulus attributes and response alternatives, for economy in the thought process. Ground school students could be taught to use these processes in their judgment-making activity. The instructor would show how various types of probabilistic information, such as weather forecasts, predicted aircraft system malfunctions, and predicted air traffic control problems, should be combined when making flight decisions.

The instructor could teach the student how to anticipate decisions that might have to be made later as a result of immediate choices of action. Such anticipation permits the gathering of relevant information under lower levels of stress when errors are less frequent, rather than later in the flight when the amount of time available to decide may become an error-causing factor. This section of ground school could also include decision-making training using procedures suggested by Janis and Mann (1977) such as "balance sheet" and "emotional role playing."

Computer-assisted instruction. An instruction technique that holds exceptional promise for pilot judgment training and evaluation is computer-assisted instruction (CAI). The great advantage of CAI systems is that they can be used to teach principles and then to permit the student to participate in decision-making processes, a highly effective learning technique (Fishbein and Ajzen, 1975). The disadvantage of these systems in the past has been their limited availability and high cost. However, recent advances in technology are making them available at a relatively low cost (Trollip and Ortony, 1977).

Although CAI programs are available in several forms, the dialogue systems that permit student-computer interaction unrestricted by preset response alternatives (Alpert and Bitzer, 1970) show the greatest potential for application to instruction in judgment making. These systems depend upon a set of stored algorithms that are used by the computer to construct a great variety of responses to student questions. In addition,

student responses are not limited to exact duplicates of prestored expected responses. The program recognizes a variety of student responses and is able to proceed accordingly.

Although practice and feedback principles are frequently used concepts in CAI programs, they could be augmented by presenting principles and reasons for taking certain courses of action. In judgment training the student could be presented with a flight situation requiring judgment. The response could require listing all of the alternatives and the factors affecting each. Student pilots could even be asked to estimate the probability of success for each alternative.

The computer, after examining the student's flight experience data (entered previously) and the stored accident statistics from similar circumstances, would respond with comments on the appropriateness of the student's responses, the alternatives that may have been omitted, and the principles that should have been followed in making the decision. The program could then branch to another problem, the difficulty of which would be based on the level of judgment-making capability evidenced by the student's responses to the previous problem.

Complexity, realism, and time constraints might be included in the judgment-making task by the addition of a simple hand-controller and an airplane symbol with a map on the screen. The controller's purpose would be to provide indications of progress toward a destination and the time available for the decision, not to provide flight control instruction.

CAI has many advantages not commonly associated with other instructional systems. The most important of these is individualization of instruction. CAI can adapt to the specific needs of the individual and interact at his or her current level of ability (Goldstein, 1974). Second, the unencumbered reinforcement capabilities of CAI are a real benefit to the student. There are no ulterior motives to clash with those of the student. Third, CAI systems do not require the presence of a teacher, although it may be beneficial to have one present for occasional consultation. Fourth, they permit standardization of instruction across a wide area. One central computer could potentially support terminals at every pilot instructional center in the United States at a relatively low cost. No student would be handicapped by a poor instructor who underscores weaknesses in the simulation. Fifth, data gathered from student responses could be stored for as long as necessary or could be used in updating instructional programs or evaluating individual pilot judgment-making capabilities.

Flight simulation. In some ways judgment training in a simulator environment would be more cumbersome than in ground school or with CAI because, at least in current practice, it is up to the instructor to create the simulated flight situation primarily through verbal communication. However, if properly structured, the simulated flight environment provides an additional opportunity to teach the principles of making judgments in a somewhat more realistic environment than that provided by ground school or CAI.

Probably the best way to begin judgment training in the simulator is to use the airline approach, i.e., the teaching of procedures that are to be followed in each situation that departs from normal flight. This includes detecting system failure as well as establishing courses of action to correct or counter system failures. Principles as well as corrective procedures would be taught according to this method, and appropriate judgment performance measures could be developed.

The simulator instruction could also include the creation, by the flight instructor, of judgment-demanding situations that do not involve the failure of systems. These situations would demand decisions such as whether to continue a flight into deteriorat-

ing weather, decisions about passenger demands for landing at an unfamiliar alternate airport, decisions about weight and balance considering field conditions, density altitude, etc. In all cases the instructor would ask the student to state several available alternatives and also to state which would be chosen. These situations could be developed from NTSB accident briefs, and they could be a part of the flight instructor's simulator judgment-instruction package.

The airplane. Of all the media available, the airplane is probably the most difficult to use for direct, systematic training in judgment making, because for the sake of safety, convenience, and cost, most judgment problems must be halted before the student sees the final consequences of a decision. The student frequently must take the instructor's word that a decision would have resulted in a safe or unsafe situation. However, the airplane offers special opportunites for instruction in judgment making because the environment is more realistic, it is more meaningful, and, therefore, it is more likely to cause a permanent behavioral change in the student than are other training media.

Everything that has been said about instructor attitudes and approaches to judgment training is even more important when actually flying the airplane. Effective instruction in the airplane requires a consistent, disciplined flight instructor who always follows the rules that the student is expected to follow, or provides a good explanation for deviating from them. It also requires that the instructor follow the learning principles stated earlier, i.e., that practice and feedback are beneficial only when accompanied by direction and explanation.

Judgment-making instructions in the airplane could take the form of simulated situations created by the instructor that require the use of judgment. Such activities could be interspersed throughout the flight-training program. To some extent, such instruction is already being done through training in simulated engine failures, other system failures, and stalls of all types. This training could be expanded to include many of the hypothetical situations discussed above. Portions of such simulated situations could be a part of every instructional flight.

Through situational training techniques, flight instructors may be able to modify the response tendencies of their students by teaching them that it is not socially demeaning to refuse to fly or to turn around in the face of deteriorating circumstances. Such situations could be made to occur several times during the student's instruction program in the airplane. There is a tendency in many pilots to believe that they are responsible to maintain a virile or courageous image for themselves and for their profession. Pilots have often said that it is most difficult to turn around the first time. Thus, students need to be taught how to avoid the tremendous social pressure that a group of important passengers can exert. In making all important decisions, they need to know how to isolate themselves from flight-naive passengers.

Finally, one of the pilot's most difficult evaluations is the evaluation of one's own skill, knowledge, and judgment-making capability relative to a proposed flight. Although research is needed to determine ways to effectively evaluate one's own capabilities, some guidelines can be suggested for its application. First, pilots should be taught to develop a list of personal limitations on flight procedures based on an estimation of their own capabilities. Second, pilots should be taught that these limitations are applicable to all flights, regardless of passengers' identity or their willingness to pay for a more risky choice. These personal limitations should be invoked during a rational moment, and the pilot's resolution should be strong enough to withstand social pressures to de-

viate from them either before or during the flight.

JUDGMENT EVALUATION

Perhaps the most difficult part of any study of human judgment is the evaluation of performance. This is because much of what must be evaluated cannot be observed directly but must be inferred by observing other related behaviors. From discussions with flight instructors and pilot-examining personnel, it is clear that judgment is not currently being evaluated effectively (Jensen and Benel, 1977).

Although flight test guides published by the FAA specify that civilian pilots are to be evaluated for their "judgment" capabilities, no definition of judgment is provided. For this evaluation, examiners primarily depend upon the assessment of flight instructors who have had the opportunity to examine their student's decision-making capability over a greater variety of circumstances. However, in interviews with flight instructors, Jensen and Benel (1977) found only one who admitted to having failed a student purely on the basis of poor judgment. Although many said that they could recognize poor judgment, students were failed on the basis of a borderline performance of some other more clearly defined flying maneuver involving skilled performance.

Some ideas for judgment evaluation are offered by Van Dam, who directs a flight school in which pilot judgment receives a strong emphasis (Jensen and Benel, 1977). In his approach, the evaluation begins with psychological and intelligence testing prior to admitting students for flight instruction. Initial impressions from these pretraining examinations are augmented with other subjective indicators of judgment, such as "obvious effort and attention to instruction," "relaxation," "division of attention," "response delays," "confidence," "capacity for problem solving," and "initiative." In later pilot training, evidence of judgment development is seen through an "eagerness to learn or high motivation," "teachability," "adaptability and flexibility," "an intuitive quality in thinking or decision making," "a pattern of good choices," and "application of margins and allowances."

The requirements of pilot judgment evaluation are even broader than these. Society expects pilots to make decisions based on the interests of passengers and property owners. Therefore, judgment must also be evaluated in an absolute sense against this poorly defined scale. Thus, there are three major dimensions along which judgment should be evaluated; each presents a unique problem to the evaluator:

(1) The assessment of judgment capabilities and tendencies prior to flight training
(2) The assessment of the effects of training on pilot judgment
(3) The assessment of the amount of training transferred to the operational flying environment

Pretraining Evaluation

It is important from the standpoint of both safety and economics to identify, prior to flight instruction, those persons who may have difficulty with some aspects of flying judgment. If such individuals could be identified, they could either be discouraged from seeking flight training, or their training programs could be modified to offset this deficiency.

Unfortunately, on the basis of psychological testing research to date, the predicted success of such a pretraining evaluation program is not very good. For example, psychologists and others have made many attempts, with little success, to identify a general personality trait known as risk taking and to link this trait to accident proneness (Shealy, 1974). Shealy found that if one were to limit the scope of the test to specific situa-

tions, such as downhill skiing, its predictive validity would increase greatly. Therefore, efforts to develop pretraining pilot judgment prediction tests should not be discouraged by the limited success of the general tests. Instead, efforts should be made to design an aviation-specific test of judgment with predictive validity.

Pretraining evaluations of judgment ability in pilot-training candidates is a potentially useful adjunct to the entire training and evaluation process. Results from such tests could be used by trainers to adapt their programs to emphasize training in areas identified in these tests. Flight instructors could be alerted to possible weaknesses in individual students and adapt their training accordingly. Tests that could identify risk-taking tendencies (Kogan and Wallach, 1964; Taylor and Dunnette, 1974), tests that identify accident proneness (Shaw and Sichel, 1971), and situation-specific tests are potentially useful in this application.

Training Evaluation

The second major dimension along which pilot judgment must be evaluated is an assessment of the amount of change in the pilot's judgment performance that results from training. This measure would provide an indication of the value of the training program as well as an indication of individual student progress.

The development of clearly defined judgment-evaluation criteria presents the greatest challenge to effective evaluation of pilot judgment in all phases of pilot training. To ensure that evaluations are made along the same dimensions as those along which the training is conducted, the development of these criteria should be based on preestablished behavioral objectives. Judgment criteria should consist of positive statements of acceptable pilot judgment-making behavior for each major area of flight activity. Similar criteria could be developed for every major maneuver taught. These could be graded by the instructor together with evaluations of knowledge and skill each time the maneuvers are attempted.

In pilot training, for each level of pilot experience, certain judgment proficiency levels could be objectively specified. The instructor or examiner who evaluates the judgments would have a range of acceptable performances, also objectively specified. Evaluation of pilot judgment would be a matter of comparing performance against the established criteria in carefully structured situations.

The critical point for evaluating judgment in a national system is the use of the same criteria by all judges as well as by the pilots themselves. One way to ensure standardization of judgment evaluations is to use a nationwide CAI system to administer tests at specific times during each student's training program. Results of such tests could be used to modify the individual student's training or the training program as a whole.

Transfer Evaluation

The final dimension along which pilot judgment must be evaluated is an assessment of the amount of training that is transferred to the operational flight environment. This means that students who have received special judgment training would be compared with those who have not received such training after both groups have moved into the operational flight environment. The results of this evaluation are used to modify both student selection criteria (or pretraining examinations) and program need assessments.

The criteria for such an evaluation are basically the same as those used in training evaluations, except that they would be more highly influenced by societal demands. Indices of the transfer of judgment could include the number of accidents or incidents due to

faulty pilot judgment compared across the two groups.

Operationalizing Judgment Evaluations

The definition of pilot judgment has two components: discrimination among situational dimensions and response selection. Both components must be evaluated. To operationalize these components for use in any specific training or testing situation, the evaluator might ask the following questions:

(1) For discriminative judgment: Did the student consider all of the available alternatives? Was all of the relevant information considered and assigned proper weights? Was the relevant information integrated efficiently before the choice was made?
(2) For response selection tendencies: Did the student exhibit any tendency to consider factors other than safety (such as self-esteem, adventure, or social pressure) in making the response selection? Was the student highly prone to use semirelevant factors, such as financial gain or convenience, in situations where safety should have been the primary consideration?

The initial step toward an operational evaluation of pilot judgment for experienced pilots was taken in recent study by Flathers (1980). In this study conjoint measurement techniques were used to establish the worth functions for four factors affecting a diversion decision during a "paper and pencil" simulation of an alternator failure in instrument flight conditions. The four factors were air traffic control service (radar vs. no radar), the weather at the airport (ceiling of 1,000 or 500 feet), the time to fly to the airport (15 vs. 30 min), and the best approach facilities (ILS vs. ADF) at the alternate airport.

Flathers reports that these techniques effectively discriminate between ATP and instrument rated pilots in terms of the value placed on certain of these four factors. Similar discriminations were found as a function of training backgrounds, type of flying most commonly done, and level of pilot ability to diagnose flight problems as determined from a separate test. Each of these are primarily measures of the discriminative component of judgment.

Flathers also measured the response selection tendencies of the pilot. In this test pilots were asked to state how many of their previously selected top airport choices would be avoided in favor of a more risky choice to take advantage of field maintenance facilities. Although in this case the results were not statistically significant, they do indicate that private and commercial pilots were more likely to take a greater risk than were ATP pilots.

The Flathers study shows that objective tests can be developed to measure both aspects of pilot judgment. In all likelihood, greater fidelity in the testing procedure could be found using computers or flight simulators. Although the first steps are being taken, a great deal of research still needs to be done to refine these tests and to validate their results in operational environments.

The proper evaluation of pilot judgment requires more of the evaluator than just an occasional passing glance at the instrument panel. It requires the careful structuring of the situation requiring the exercise of judgment and a careful examination of the student's actions. Each evaluation should be considered a training occasion as well, and as such, feedback should be given to the student concerning all aspects of the decisional situation. Evaluations of this sort place high demands on the flight instructor; nevertheless, they seem to be warranted in view of the high number of fatalities caused by faulty judgment.

ACKNOWLEDGMENTS

This research was sponsored by the Systems Research and Development Service of the Federal Aviation Administration. Patrick Russell was the Contracting Officer's Representative. The author wishes to thank Russ Benel, Jim Finnegan, Mike Kelley, and Rob Durst for their assistance in this program.

REFERENCES

Alpert, D., and Bitzer, D. L., Advances in computer-based education. *Science*, 1970, *167*, 1582-1590.

Anderson, N. H. Comment on "analysis-of-variance model

for assignment of configural cue utilization in clinical judgment." *Psychological Bulletin*, 1969, *72*, 63-65.

Birdsall, T. G. The theory of signal detectability. In H. Quastler (Ed.) *Information theory in psychology: Problems and methods*, Glencoe, IL: Free Press, 1955, 391-402.

Federal Aviation Administration. *Aviation instructor's handbook*. Washington, DC: U.S. Government Printing Office, No. AC 61-41, 1977.

Fishbein, M., and Ajzen, I. *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley, 1975.

Flathers, G. W. A study of decision-making behavior of aircraft pilots deviating from a planned flight. Unpublished masters thesis, Ohio State University, Department of Industrial and Systems Engineering, 1980.

Gagne, R. M. Military training and principles of learning. *American Psychologist*, 1962, *17*, 83-91.

Gibson, J. H. Optimized flight crew training: A step toward safer operations. Fort Worth, TX: American Airlines, Flight Training Academy, 1969.

Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, *23*, 483-496.

Goldstein, I. L. *Training: Program development and evaluation*. Belmont, CA: Wadsworth, 1974.

Hoffman, P. J., Slovic, P., and Rorer, L. G. An analysis of variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin*, 1968, *69*, 338-349.

Janis, I., and Mann, L. *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press, 1977.

Jensen, R. S., and Benel, R. A. Judgment evaluation and instruction in civil pilot training. Springfield, VA: National Technical Information Service, Final Report FAA-RD-78-24, 1977.

Kogan, N., and Wallach, M. *Risk taking: A study of cognition and personality*. New York: Holt, Rinehart, and Winston, 1964.

Roscoe, S. N. *Aviation Psychology*. Ames, IA: Iowa State University Press, 1980, p. 175.

Shanteau, J., and Phelps, R. H. Judgment and swine: Approaches and issues in applied judgment analysis. In M. F. Kaplan and S. Schwartz (Eds.) *Human judgment and decision processes: Applications in problem settings*. New York: Academic Press, 1977.

Shaw, L., and Sichel, H. S. *Accident proneness: Research in the occurrence, causation, and prevention of road accidents*. New York: Pergamon, 1971.

Shealey, J. E. Risk-taking in skilled task performance. *Proceedings of the Eighteenth Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors Society, 1974.

Slovic, P. Analyzing the expert judge: A descriptive study of stockbroker's decision processes. *Journal of Applied Psychology*, 1969, *53*, 255-263.

Taylor, R. N., and Dunnette, M. D. Influence of dogmatism, risk-taking propensity, and intelligence on decision-making strategies for a sample of industrial managers. *Journal of Applied Psychology*, 1974, *59*, 420-423.

Thorpe, J. A., Martin, E. L., Edwards, B. J., and Eddowes, E. E. Situational emergency training: F-15 emergency procedures training program. Williams Air Force Base, AZ: Technical Report No. AFHRL-TR-76-47(1), June, 1976.

Trans World Airlines. Flight simulator evaluation. Kansas City, MO: Author, Flight Operations Training Department, 1969.

Trollip, S. R., and Ortony, A. Real-time simulation in computer-assisted instruction. *Instruction Science*, 1977, *6*, 135-149.