

AD-A071 106

CANYON RESEARCH GROUP INC WESTLAKE VILLAGE CALIF  
DEVELOPMENT OF AN OBJECTIVE GRADING SYSTEM ALONG WITH PROCEDURE--ETC(U)  
MAY 79 J M CHILDS  
FTR-09-79

F/G 5/9

DAH19-77-C-0008

NL

UNCLASSIFIED

ARI-RN-79-18

| OF |  
AD  
A071106



END  
DATE  
FILMED  
8-79  
DDC

ADAC71106

4

LEVEL II



See 1473 in back

DDC FILE COPY,

DDC  
RECEIVED  
JUL 13 1979  
A

DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited

Canyon Research Group, Inc.

79 07 12 043

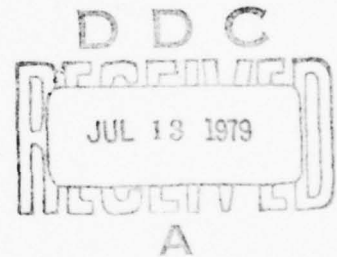
Human Factors Research in Aircrew  
Training Performance Enhancement  
DAHCR9-77-C-0008

RESEARCH NOTE 79-18

DEVELOPMENT OF AN OBJECTIVE GRADING SYSTEM  
ALONG WITH PROCEDURES AND AIDS FOR ITS  
EFFECTIVE IMPLEMENTATION IN FLIGHT

Jerry M. Childs

CANYON RESEARCH GROUP, INC.  
741 Lakefield Road, Suite B  
Westlake Village, California



US ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES  
5001 Eisenhower Avenue  
Alexandria, Virginia 22333

May 1979

*Approved for public release; distribution unlimited*

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

<b>19</b> REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER RN-79-18	2. GOVT ACCESSION NO. <b>18</b> ARI	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Development of an Objective Grading System Along with Procedures and Aids for its Effective Implementation in Flight		5. TYPE OF REPORT & PERIOD COVERED Research Note, 15 Nov 77 - 14 Nov 78	
7. AUTHOR(s) Jerry M. Childs		6. PERFORMING ORG. REPORT NUMBER FTR-09-79	
9. PERFORMING ORGANIZATION NAME AND ADDRESS CANYON RESEARCH GROUP, INC. 741 Lakefield Road, Suite B Westlake Village, CA 91361		8. CONTRACT OR GRANT NUMBER(s) DAHC19-77-C-0008	
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Research Institute for the Behavioral & Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <b>16</b> 2Q763743A772	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) <b>12</b> 37 p		12. REPORT DATE <b>11</b> May 1979	
		13. NUMBER OF PAGES 28	
		15. SECURITY CLASS. (of this report) Unclassified	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) <b>14</b> FTR-09-79			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Objective Grading Procedures      Objective Maneuver Grades Inflight Assessment Procedures      Flying Proficiency Assessment Performance Measurement      Basic Instrument Flight Training Performance Assessment			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report contains a description of the characteristics and test of two alternative inflight scoring procedures. These procedures were designed to meet the requirements of (1) minimal data collection, and (2) objective scores for Initial Entry Rotary Wing (IERW) student performance on Basic Instrument maneuvers. The procedures were criterion-referenced, employing different performance criteria, sampling techniques, and scoring algorithms. They were subjected to tests in the UH-1 simulator to assess the potential			



20. Abstract - Cont'd

value of various characteristics within each procedure for meeting the requirements. Results of the tests provided general indications of those characteristics which best discriminated proficiency within and among students across training days.

Accession For	
NTIS Grant	<input checked="" type="checkbox"/>
DDC TAB	
Unannounced	
Justification	
By _____	
Distribution/ _____	
Availability Codes	
Dist.	Avail and/or special
A	

UNCLASSIFIED

TABLE OF CONTENTS

	<u>Page</u>
INTRODUCTION . . . . .	1
Current IERW Flight Performance Assessment . . . . .	2
Past Attempts at IERW Objective Flight Performance Assessment . . . . .	2
Air Force Flight Performance Assessment . . . . .	4
Requirements and Guidelines . . . . .	4
METHOD . . . . .	6
Definition of Variables for Measurement . . . . .	6
Measurement Criteria . . . . .	7
Inflight Sampling . . . . .	7
Alternatives for Scoring Critical Variables in Flight . . . . .	8
Procedure 1 Characteristics . . . . .	8
Procedure 1 Tests . . . . .	13
Procedure 2 Characteristics . . . . .	14
Procedure 2 Tests . . . . .	15
RESULTS . . . . .	18
Procedure 1 Results . . . . .	18
Procedure 2 Results . . . . .	20
DISCUSSION . . . . .	21
REFERENCES . . . . .	22
APPENDIX - Evaluation Guidelines for Enclosed Gradeslip . . . . .	24

LIST OF FIGURES

<u>FIGURES</u>		<u>PAGE</u>
1	Procedure 1 inflight scoring format (example of climb maneuver shown) . . . . .	11
2	Objective procedure for assessing proficiency on the acceleration and deceleration maneuver (cell values suggested from observation flights) . . . . .	12
3	Objective procedure for assessing proficiency on the unusual attitude recovery maneuver (cell values suggested from observation flights) . . . . .	13
4	Procedure 2 inflight scoring format . . . . .	17
5	Procedure 1 maneuver scores for Basic Instrument students (n = 4). . . . .	19

LIST OF TABLES

<u>TABLES</u>		<u>PAGE</u>
1	Characteristics of the Alternative Scoring Procedures . . . . .	9
2	Procedure 1 Scoring Algorithm . . . . .	10
3	Procedure 1 Data Collection . . . . .	14
4	Procedure 2 Scoring Algorithm . . . . .	16
5	Number of Procedure 1 Subcriterion Scores (1-3) Attributable to Flight Variables at Various Sampling Times . . . . .	18

PREFACE

The work described in this report was performed by Canyon Research Group, Inc. as tasked by, and under contract to, the Army Research Institute for the Behavioral and Social Sciences (ARI), Fort Rucker, Alabama. This effort on objective performance assessment was performed as part of Contract No. DAHC19-77-C-0008, "Human Factors Research in Aircrew Training Performance Enhancement," to conduct research at the Initial Entry Rotary Wing (IERW) level.

DAHC 19-76-C-0018



## ACKNOWLEDGEMENTS

Research projects of this nature cannot be conducted without assistance and support from those who are actively involved in training.

The following individuals were extremely helpful to the conduct of this project. Their assistance is gratefully acknowledged.

Mr. Charles A. Gainer, Chief, and Dr. James A. Bynum, Technical Team Leader, US Army Research Institute Field Unit, both provided helpful comments concerning the measurement approach.

CW2 Charles E. Carr, US Army Research Institute Field Unit, provided much of the information concerning maneuver execution requirements and demonstrated many of the maneuvers in both the UH-1 aircraft and simulator. His instrument flying expertise proved invaluable in many ways.

DAC Instructor Pilots Bobby Hogan, George Welch, Frank McGaffey, and Connie Crane, DOFT, USAAVNC, assisted by discussing and demonstrating instrument maneuver execution requirements.

Doss Aviation Instructor Pilots, Bill Sexton, Chuck Seitz, Rick Palsulich, and Gary Slanga participated in tests of the scoring procedures and provided expert comment concerning their potential utility for IERW training.

Majors G. D. Robnett, J. C. Hare, and J. M. Smith, DOFT, IQD, USAAVNC provided useful comments concerning current IERW assessment procedures and underlying philosophy.

Major W. C. White, Jr., Research and Development Coordinator, US Army Research Institute Field Unit, enabled many of the preliminary tests of the scoring system to be conducted by reliably scheduling both aircraft and simulator and by participating in the initial tests.

Dr. George E. Long, Technical Director, Canyon Research Group, Inc., Fort Rucker, Alabama provided invaluable assistance with the development of the overall measurement methods included in the scoring procedures described herein. His direction of this effort was supportive and helpful in every way.

## SUMMARY

This report contains a description of the characteristics and tests of two alternative inflight scoring procedures. These procedures were designed to meet the requirements of minimal data collection and objective scores for Initial Entry Rotary Wing (IERW) student performance on Basic Instrument maneuvers. The procedures were criterion-referenced, employing different performance criteria, sampling techniques, and scoring algorithms. They were subjected to tests in the UH-1 simulator to assess the potential value of various characteristics within each procedure for meeting the requirements. Results of the tests provided general indications of those characteristics which best discriminated proficiency within and among students across training days.

## INTRODUCTION

Objective performance assessment is based upon explicitly defined, published performance criteria. Subjective assessment employs unpublished, personal criteria comprised of defined and undefined variables. These variables may or may not be performance based, confined to currently observed performance, or related to environmental and task constraints. Performance assessment is objective to the extent that personal criteria resulting from undefined or ill-defined variables are withheld or minimized.

The need for objective flight performance assessment procedures has long been recognized. Of the extensive aviation research efforts to develop such procedures, some of the more successful have identified and described the problems involved and provided viable solutions to many of them (e.g., Williams, 1971;<sup>1</sup> Obermayer and Vreuls, 1972;<sup>2</sup> Povenmire and Roscoe, 1971;<sup>3</sup> Carter and Semple, 1974;<sup>4</sup> Carter, 1977<sup>5</sup>). However, many flight training programs continue to employ assessment procedures that are primarily subjective. Initial Entry Rotary Wing (IERW) flight training is such a program.

The purpose of this research was to develop and test methods for objectively evaluating IERW pilot flight performance. The scoring procedures of such methods should enable instructors to assess student performance reliably and objectively without interfering with either instruction or safety.

Potential benefits of objective performance assessment for IERW pilot training include more efficient use of training resources, better training management, a valid data base, a framework for evaluating current training guidelines, a means of generating maneuver performance profiles as training aids, and more accurate determination of proficiency attainment and loss rates. Further, objective performance assessment is

---

<sup>1</sup>Williams, A. Discrimination and Manipulation on Goal-Directed Instrument Flight. *Aviation Research Monographs*, 1971, 1 (1).

<sup>2</sup>Obermayer, R. W. and Vreuls, D. Measurement of Flight Training Research, *Proceedings of the 16th Annual Meeting of the Human Factors Society*, Beverly Hills, California, October 1972.

<sup>3</sup>Povenmire, H. K. and Roscoe, S. N. An Evaluation of Ground-Based Flight Trainers in Routine Primary Flight Training. *Human Factors*, 1971, 13, 109-116.

<sup>4</sup>Carter, V. E. and Semple, C. A. *Specific Behavioral Objectives for VF-121 Training in Basic Air Combat Maneuvering (Confidential)*. Contract N62269-74-C-0314 (Volume 3), Northrop Corporation Aircraft Division, October 1974.

<sup>5</sup>Carter, V. E. Development of Automated Performance Measures for Introductory Air Combat Maneuvers. *Proceedings of the 21st Annual Meeting of the Human Factors Society*, San Francisco, California, October 1977.

desirable, and perhaps necessary, for use in training programs that advance students on the basis of observable performance because proficiency gain and loss can be accurately determined. The Army's interest in proficiency-based progression, and in proficiency loss and retraining requirements underscores the relevance of the present research.

#### Current IERW Flight Performance Assessment

IERW flight performance currently is assessed on the basis of subjective ratings by instructor pilots (IPs). Letter grades, as defined below, indicate the proficiency of students on both a maneuver and daily basis, as judged by the IP.<sup>6</sup>

*A (Excellent proficiency). The best accuracy of performance that can be expected of students.*

*B (Good proficiency). The accuracy of student performance which is better than fair but below excellent proficiency.*

*C (Fair proficiency). Above that minimum performance required of all students.*

*U (Unsatisfactory). Below required minimum performance standards.*

Students also are evaluated on the basis of mental, emotional, and physical factors as judged by their instructor. These factors are evaluated by using twelve "basic qualities," (e.g., motivation) as indices. Each basic quality is assigned one of the four letter grades. Downgrades (*C* or *U*) on maneuvers require that one or more basic qualities be cited as the cause of the weak performance. Additionally, downgrades require written explanations on the gradeslip describing the specific problems observed in the maneuver execution.

Training effects are difficult to discern due to the infrequent assignment of grades other than *B*. This appears to be influenced by the wide range of acceptable performance comprising the *B* category for any given point in the training program. Since no specific performance standards are published for grade categories, assessment of proficiency is largely contingent upon the amount of training time received relative to this range of *B* category performance. If a student progresses according to his instructor's expectations, his grades are predominantly *B*'s. However, the instructor's expectation level may be influenced by many factors related to the student's personality and needs as well as observed performance. Thus, IERW flight performance assessment currently is not made solely on the basis of observable performance relative to explicitly defined standards.

#### Past Attempts at IERW Objective Flight Performance Assessment

The Pilot Performance Description Record (PPDR) developed by Greer

<sup>6</sup>USAAVNC Reg. 350-16, *The Uniform Flight Grading System*, US Army Aviation Center, Fort Rucker, AL, February 1973.



and his colleagues represents one effort to attain a higher level of objectivity in IERW flight performance assessment. The PPDR includes a number of objective scales for the assessment of each flight maneuver. The instructor's task is to record the direction and magnitude of error on each of those scales. The overall maneuver score is determined on the basis of error frequency and magnitude. For example, a normal approach contains eight assessed components: entry, sight picture, RPM, line of descent, rate of closure, lane alignment, pedal application, and termination. The extent to which these components approach the performance standards prescribed by training regulations determines the maneuver grade (A, B, C, or U).

The PPDR currently is employed for assessing checkrides in the Primary Phase of IERW training. Although rather extensive training in the use of this grading system is required, it is reported to be an effective assessment technique. However, because of its length and recording demands, the PPDR requires much of the instructor's time and attention in flight. If used on a daily basis, it might interfere with instruction and it otherwise creates a potential hazard. Hence, its greater degree of objectivity has, to some extent, been obtained at the expense of inflight recording ease.

In 1972, a program of research addressing the future use of turbine aircraft for certain IERW training phases was conducted.<sup>8</sup> The objective was to develop a more efficient and effective training concept encompassing all aspects of training to offset the spiraling costs of aircraft operation. To accomplish this, an objective grading concept to handle self-pacing was developed and employed. Criteria were determined and standardized for all maneuvers of the instrument/contact stage of IERW instruction. Students' scores on these maneuvers consisted of the number of errors committed as defined by deviations from acceptable tolerances. For example, if a maneuver contained twelve criteria and two of them resulted in tolerance deviations, the resulting score would be 2. The student's objective was, of course, to obtain scores of 0 as soon as possible. Maneuver proficiency achieved in fewer flight hours resulted in higher grades. When all criteria were attained by the student for a given maneuver, acceptable proficiency was declared and more advanced maneuvers were introduced.

When all maneuvers had been performed to criterion, the student received a *Stage Completion Time* as a proficiency measure. A checkride was then administered and, if passed, the time score was graphically converted into a numerical grade (70-100). The lower the time score coupled with a passing checkride, the higher the numerical grade.

This objective grading format placed additional responsibilities on instructors, but provided standardized assessment as well as more efficient

---

<sup>7</sup> Greer, G. D., Smith, W. D., and Hatfield, J. L. *Improving Flight Proficiency Evaluation in Army Helicopter Training*, HumRRO Technical Report 77, Human Resources Research Office, Alexandria, VA, May 1962.

<sup>8</sup> USAAVNC, *Turbine Trainer Test*, Human Resources Research Organization, US Army Aviation Center, Fort Rucker, AL, 1972.

flight time management. It yielded progression rates and average proficiency levels for students, provided instructors with greater quality control for training deficiencies, and provided a gross index of their instructional skill.

Apparent disadvantages of the *Turbine Trainer* grading system were the lack of daily grades, the judgmental nature of some of the performance criteria (e.g., "light on skids" in an instrument takeoff), and the lack of validation tests on the time-based proficiency criteria. No followup tests to determine the efficacy of the time-based proficiency criteria for discriminating performance were conducted. The need for such tests provided much of the impetus for the present investigation.

Related disadvantages of the self-paced concept were the lack of standardization in introducing and demonstrating maneuvers, scheduling problems resulting from unsatisfactory checkrides, and possible over-emphasis upon the amount of training time required to attain proficiency and hence, to determine numerical grades.

#### Air Force Flight Performance Assessment

The grading system currently employed by the US Air Force<sup>9</sup> includes two scoring categories: *Maneuver grades* and *lesson grades*. *Maneuver grades* are assigned relative to published training criteria. Proficiency is determined by the extent to which student performance approaches a standardized description of an "ideal" maneuver execution. *Lesson grades*, in which students receive overall performance ratings relative to the estimated average performance level for a given training stage, also are given.

The scale for assigning both *maneuver* and overall *lesson* grades consists of four letter grades--*excellent (E)*, *good (G)*, *fair (F)*, and *unable to accomplish (U)*. This scale, coupled with phase training standards, provides a means for assessing maneuver proficiency without knowledge of the type or amount of training the student has received. This grading system has proven to be quite successful because it documents proficiency change based on published standards, allows problem areas to be determined, and identifies outstanding, as well as weak, student performance throughout training. Like all operational assessment systems, however, there exists some degree of subjectivity.

#### Requirements and Guidelines

This research was directed at the development and test of alternative inflight scoring procedures to assess student performance objectively in IERW training. The requirements of such scoring procedures for meeting IERW training needs were *minimal data collection* and *objective scoring*. The scoring procedures were not to interfere with instructors' other responsibilities in flight; i.e., instruction and safety. The primary task of the instructor is to teach the necessary skills to students to

---

<sup>9</sup>HQATC, *Syllabus of Instruction for Undergraduate Pilot Training (T-37/T-38)*, Department of the Air Force, Randolph Air Force Base, Texas, June 1978.

enable them to control the aircraft skillfully in three dimensions. Additionally, instructors must monitor surrounding airspace for other traffic and obstacles. The element of subjectivity had to be minimized. Evaluation was to be made on the basis of observable performance and explicitly defined criteria.

The following guidelines for accomplishing the above requirements were derived:

1. *Instruction was separated from assessment, and only the latter was addressed.*
2. *To the extent possible, existing IERW training standards were used in the initial development of objective performance measures and criteria.*
3. *The objective scoring procedures were designed for use in either the aircraft or simulator, although initial tests were carried out in the simulator only.*
4. *Only the Basic Instrument phase of IERW training was addressed.*

## METHOD

This research was conducted as a follow-on effort to previous research aimed at the identification of critical IERW performance variables<sup>10</sup> and at the development of scoring procedures for the objective assessment of these variables in flight.<sup>11</sup> The overall research approach is described below.

1. In the initial phase of this effort, the requirements to be satisfied by the research product were defined. They included objective inflight grading capabilities for IERW training, and guidelines for achieving those capabilities, as derived from an analysis of the requirements.

2. The variables to be measured by the grading system were defined. Candidate measures were generated from a Basic Instrument (BI) maneuver analysis resulting from consultation with instructors, students, and other subject matter experts, as well as literature review and observation flights. Maneuvers were segmented, and measures for assessing proficiency within segments were developed.

3. Performance criteria for segments were defined as a result of the maneuver analysis, as well as existing IERW training guidelines. These criteria were applied to the objective performance measures.

4. The problem of when and what to sample in flight relative to safety constraints was considered. Alternative sampling procedures were developed and tested in the UH-1 flight simulator.

5. Alternative criterion-referenced grading procedures were developed. These grading procedures were demonstrated to BI instructors and training managers, then subjected to initial testing in the simulator.

### Definition of Variables for Measurement

It is clearly possible to define each maneuver in terms of a series of *segments* in which some flight variables are held constant while others are caused to change at a relatively constant rate. When this is done, performance may be assessed by observing how well each of those defined variables is maintained within its defined limits for that segment.

Measures may be differentiated into segments for which one or more *desired values* of critical flight variables--airspeed, altitude, heading, and trim--remain constant. Where critical flight variables are supposed to change, they are marked by observable and verifiable transitions in

---

<sup>10</sup> Childs, J. M. *The Identification and Measurement of Critical IERW Performance Variables*. Research Memorandum, Canyon Research Group, Inc., Fort Rucker, AL, March 1979.

<sup>11</sup> Childs, J. M. *Development of Procedures and Techniques for Inflight Performance Assessment*. Research Memorandum, Canyon Research Group, Inc., Fort Rucker, AL, April 1979.



magnitude of control input on the cyclic, collective, or pedals.<sup>12</sup> In this sense, maneuver segments are relatively easy to observe. These segments can then be used for objectively assessing performance on critical variables throughout all IERW training phases, from the most basic to the most advanced.

In early phases of training, maneuvers usually are designed to teach general control skills. In advanced flight training, students frequently learn maneuvers that later become segments of operational flights. The implication of this is that small elements of performance can, for assessment purposes, be integrated into more comprehensive performance. Basic maneuvers (e.g., climbing turn) may serve as segments of more advanced maneuvers (e.g., instrument takeoff) which then become the segments of even more advanced maneuvers (e.g., tactical instrument flight).

This process can, in fact, be extended to the assessment of entire operational missions in which the maneuvers are considered segments of the mission. Performance assessment variables can thus be aircraft-referenced flight variables assessed within segments of maneuvers or missions. The aircraft's inflight attitude as characterized by observed airspeed, altitude, turn rate, etc., is compared to its desired attitude as defined by prescribed IERW rates and tolerances. Proficiency can then be assessed on the basis of the degree of correspondence between the aircraft's actual and desired attitudes for those segments.

#### Measurement Criteria

The measurement criteria consisted of the *desired values of critical performance variables* that were stipulated on the basis of prescribed IERW rates and tolerances. By observing critical variables at specified times, and determining their degree of deviation from IERW *tolerance limits*, the criteria for assessing performance quality were derived. Where desired values could not be calculated from existing training material, criteria were generated from observation flights.

Where a critical variable is caused to change, its rate of change should remain relatively constant. For example, a standard rate turn maneuver requires the student to maintain airspeed and altitude while turning 180 degrees in approximately one minute. Although the aircraft's heading changes, turn rate and bank angle should remain relatively constant. Objective assessment is accomplished by observing how well each critical variable is maintained within its defined tolerance limits for each segment of the maneuver.

#### Inflight Sampling

An effort was made to standardize inflight sampling procedures. Instructors may sample several variables in assessing student performance, may do so at various times within the maneuver, and may allow many unspecified aircraft parameters and student performance variables to

---

<sup>12</sup>Childs, J. M. *op. cit.*, ref. 10.

affect their final decision concerning the maneuver score. An indication of aircraft control accuracy is derived by standardized sampling of critical variable values over time and comparing those values to desired values. The development of standardized inflight sampling procedures and valid performance criteria was therefore critical.

### Alternatives for Scoring Critical Variables In Flight

Two inflight scoring procedures were developed and tested in the UH-1 simulator. The maneuvers and characteristics of the two procedures are shown in Table 1. The simulator tests were not intended to provide measures using the two procedures that could be compared statistically. Rather, the tests were conducted to provide preliminary, mainly qualitative, information on the potential value of various characteristics within the procedures for meeting the previously stated requirements. The necessity to minimize disruption of ongoing Instrument Phase training prevented rigorous experimental control.

### Procedure 1 Characteristics

Procedure 1 employed *time-based sampling* in which one or more critical variables were sampled at standardized times within the maneuver. Two sampling intervals were considered--ten seconds and fifteen seconds. The latter was selected because it produced less of a recording constraint on the instructors. With this procedure, the evaluator observed and recorded values of at least two specified flight variables (e.g., airspeed and heading) at fifteen second intervals, and also at the completion of the maneuver as defined by its prescribed execution time. This resulted in a total of five to nine data points for each BI maneuver.

This time-based sampling procedure required the evaluator to record values of airspeed, altitude, and heading at the onset of the maneuver. This initial recording provided a baseline against which to evaluate the subsequent portions of the maneuver. These baseline values formed the *initial conditions*. As the maneuver was executed, the evaluator recorded values of one or more of the three variables at fifteen second intervals, and at the completion of the maneuver. These recorded values were defined as *observed values*. By computing the difference between observed values and *desired values*--those values which should be attained as a function of time when standardized IERW rates are followed--an objective evaluation of maneuver performance was obtained. For two of the BI maneuvers--accelerations and decelerations and unusual attitude recoveries--tolerance category criteria were generated from observation flights. Sampling was not time-based for these two maneuvers because no IERW rates were specified.

Procedure 1 used a six-point, criterion-referenced *scoring algorithm* (Table 2). Scores were determined by performance descriptions defined by various tolerance categories for the measures. Criterion level performance (4) was the maintenance of all sampled measures within standard IERW tolerance limits (+ 10 knots airspeed, + 100 feet altitude, + 10 degrees heading). Higher scores (5 and 6) represented proficiency above criterion level by indicating smaller deviations from *desired values* throughout the maneuver. Subcriterion scores (1, 2, and 3)

Table 1  
 Characteristics of the Alternative Scoring Procedures

CHARACTERISTICS

MANEUVERS	MEASURES	CRITERIA	SAMPLING	SCORING ALGORITHM	CRITERION LEVEL MANEUVER SCORE	ASSESS PROFICIENCY ABOVE CRITERION?
1. Straight and Level Turns	Critical Variables a. Airspeed b. Heading c. Altitude	Four Tolerance Categories a. Within 1/2 tolerance b. Within std. tolerance c. Within twice tolerance d. Exceeds twice tolerance	Four Time-Based Points a. 15 b. 30 c. 45 d. 60 Seconds	6-point criterion referenced	4	Yes
2. Standard Rate Turns						
3. Climbs and Descents						
4. Steep Turns						
5. Climbing and Descending Turns	Critical Variables a. Airspeed b. Heading c. Altitude d. Trim	Three Tolerance Categories a. Within std. tolerance b. Within twice tolerance c. Exceeds twice tolerance	Three Segments a. initiate b. maintain c. terminate	6-point criterion referenced	6	No
6. Accelerations-Decelerations						
7. Unusual Attitude Recovery						

S  
C  
O  
R  
I  
N  
G

P  
R  
O  
C  
E  
D  
U  
R  
E  
S

Table 2  
Procedure 1 Scoring Algorithm

PERFORMANCE DESCRIPTION	SCORE
All critical variables within 1/2 standard tolerance	6
Some variables within 1/2 standard tolerance; Some variables within standard tolerance	5
All variables within standard tolerance ( <i>criterion</i> )	4
Some variables within standard tolerance; Some variables within twice standard tolerance	3
All variables within twice standard tolerance	2
One or more variables exceeds twice standard tolerance	1

indicated various levels of unacceptable performance marked by deviations exceeding standard tolerance limits.

It may be noted (Table 1) that four levels of *performance criteria* were included in Procedure 1: within 1/2 tolerance, within standard tolerance, within twice tolerance, and exceeds twice tolerance. The 1/2 tolerance category was included to enable greater aircraft control precision to be assessed. The principal determinant of the overall maneuver score was the greatest absolute deviation of all assessed variables from their *desired values* at any sampling point. Stated differently, the maneuver score could be no higher than the poorest performance level sampled therein.

The rationale for this is twofold. First, it has been shown that measures of flying proficiency are highly interdependent. Tolerance deviations for given measures are likely to be accompanied by deviations of certain other measures. An example is the inverse relationship between airspeed and altitude in a turn. A maneuver downgrade resulting from only one of several performance measures is therefore rather unlikely. Secondly, maneuver scores reflect the greatest observed deviation from *desired values* and can be used as an incentive for students to maintain control precision throughout the entire maneuver. This control precision would appear to be desirable in learning the more advanced flying skills necessary for operational effectiveness.

An example of how a maneuver was scored using Procedure 1 is shown in Figure 1. In this case, the student was required to climb from 3000



INITIAL CONDITIONS		STUDENT				RANK			
AS	90	INSTRUCTOR				DATE			
HDG	090	MANEUVER CLIMB				SCORE 3			
ALT	3000								
TIME-BASED SAMPLING POINTS (SECS)		15		30		45		60	
CRITICAL VARIABLES	AS	ALT	ALT	HDG	HDG	AS	AS	ALT	HDG
OBSERVED VALUES	90	3100	3120	080	080	90	95	3400	085
WITHIN 1/2 TOLERANCE	X	X				X	X		X
WITHIN STANDARD TOLERANCE (CRITERION)				X	X			X	
WITHIN TWICE TOLERANCE			X						
EXCEEDS TWICE TOLERANCE									

Figure 1. Procedure 1 inflight scoring format (example of climb maneuver shown).

to 3500 feet at 500 feet per minute (IERW rate) in 60 seconds while maintaining airspeed (90 knots) and heading (090). At 30 seconds, the *desired value* of altitude was 3250 and the *observed value* was 3120. The difference between these two values (130) represents the extent to which standard IERW tolerance was exceeded. For this sampling point, the deviation exceeded standard tolerance (100 feet), but remained within twice standard tolerance (200 feet). Thus, the *within twice tolerance* cell was marked. The configuration of marks for a given maneuver determined the score (Table 2) for that maneuver. In the present example, the score was 3. It should be noted that *all* sampled measures were required to be maintained within standard IERW tolerance limits across *all* sampling points to result in a *criterion level score* (4). This was consistent with published IERW grading regulations, but not necessarily in keeping with the majority of the instructors' grading policy.

No time or rate specifications were included in IERW documentation for the acceleration and deceleration or unusual attitudes recovery maneuvers. Five measures were generated for these maneuvers on the basis of observation flights in the UH-1 simulator. These measures were incorporated into procedures (Figures 2 and 3) containing the same four

CONTROL LEVELS	AIRSPEED REVERSAL	TARGET AIR-SPEEDS MISSED	HDG (DEGREES)	ALT (FEET)	TOTAL TIME (SECS)
WITHIN 1/2 TOLERANCE	0	0	<u>+5</u>	<u>+50</u>	<60
WITHIN STANDARD TOLERANCE (CRITERION)	1	1	<u>+10</u>	<u>+100</u>	61-90
WITHIN TWICE TOLERANCE	2	2	<u>+20</u>	<u>+200</u>	91-120
EXCEEDS TWICE TOLERANCE	>2	>2	>20	>200	>120

Figure 2. Objective procedure for assessing proficiency on the acceleration and deceleration maneuver (cell values suggested from observation flights).

CONTROL LEVELS	CONTROL REVER-SALS	BANK RECOVERY TIME (SECS)	PITCH RECOVERY TIME (SECS)	AIRSPEED RECOVERY TIME (SECS)	TOTAL RECOVERY TIME (SECS)
WITHIN 1/2 TOLERANCE	0	1-5	1-5	1-5	3-9
WITHIN STANDARD TOLERANCE (CRITERION)	1	5-9	5-9	5-9	10-14
WITHIN TWICE TOLERANCE	2	10-14	10-14	10-14	15-19
EXCEEDS TWICE TOLERANCE	>2	>14	>14	>14	>19

Figure 3. Objective procedure for assessing proficiency on the unusual attitude recovery maneuver (cell values suggested from observation flights).

levels of performance criteria mentioned above. The values included in each cell were also derived from observations of students performing the maneuver during BI training in the simulator. The reader is referred to an earlier report<sup>13</sup> for a description of the measures used in these scoring procedures.

#### Procedure 1 Tests

Tests on Procedure 1 scoring were conducted in the UH-1 simulator. Four BI students were evaluated on various of the seven BI training days. It was necessary to collect data in several simulators during each training session. Because all data were collected by the same investigator, it was physically impossible to obtain data on all four students for each day of BI training. The data collection schedule is shown in Table 3. Blank cells indicate those training days in which no data were obtained for given students. The investigator was situated either in the jumpseat of the simulator, or at the console of the UH-1 training device. Instructors employed the present IERW gradeslip, training and evaluating their students in the usual manner. Six debriefing sessions were observed by the investigator.

<sup>13</sup> Childs, J. M., *op. cit.*, ref. 11.

Table 3  
Procedure 1 Data Collection

		BI TRAINING DAYS						
		1	2	3	4	5	6	7
STUDENTS	1	X	X	X	X	X		X
	2			X	X	X		X
	3	X	X			X		X
	4	X	X					X

Procedure 2 Characteristics

Procedure 2 was designed to evaluate *segments* of maneuvers without strict adherence to time. A score was assigned to each of three maneuver segments as defined below.

1. *Initiate Segment*--the initial portion of the maneuver of interest. Student proficiency in transitioning the aircraft from straight and level initial conditions to a turn, climb, descent, or acceleration (90 knots to 100 knots) was assessed.

2. *Maintain Segment*--the intermediate portion of the maneuver. Student proficiency in maintaining proper rates of turn, climb, descent or deceleration (100 knots to 70 knots) was assessed. The most critical variable of a maneuver (e.g., turn rate in a turn, climb rate in a climb) was evaluated at approximately the midpoint of a maneuver in the interest of uniformity. If turn rate was slow or fast, for example, it would be reflected by indicated heading at the midpoint. If climb or descent rate was improper, it would be shown by indicated altitude. Other variables (altitude, airspeed, and trim in a turn; heading, airspeed, and trim in a descent; altitude, heading, and trim in an acceleration and deceleration) were monitored closely throughout the segment. Any observed deviation of these variables from standard Instrument Phase tolerance limits reduced the segment score.



3. *Terminate Segment*--the terminal portion of a maneuver. Student proficiency in transitioning from a turn, climb, descent or acceleration (70 knots to 90 knots) back to straight and level flight, was assessed.

With Procedure 2, each segment was assigned a score of 1, 2, or 3, depending on the performance observed for that segment. Segment scores were then used to determine maneuver scores, as shown in Table 4. An example of segment scores and the resulting maneuver score for a climbing turn is shown in Figure 4. Here, the student failed to begin the climb when the turn was initiated. The 500 foot per minute climb rate had not been established during the *initiate* segment; standard tolerance was exceeded, resulting in a segment score of 2. Upon initiating the climb, a compensation to coordinate climb rate and turn rate was made. Both of these variables were maintained within standard tolerance limits during the *segment*. At 30 seconds, a 95 degree turn and a 250 foot climb had occurred (segment score of 3). The student maintained both climb and turn rate and in the *third segment*, rolled out on the assigned heading, and leveled off at the assigned altitude at approximately 60 seconds following initiation of the maneuver. Airspeed did not vary more than 10 knots from initial airspeed and trim was maintained throughout. This maneuver received a subcriterion score of 5 due to a standard tolerance violation for one segment.

#### Procedure 2 Tests

Procedure 2 tests occurred in the UH-1 simulator over six days of BI training. Eight students' performance was assessed using both Procedure 2 and the current IERW gradeslip. All assessment was performed by four Doss Aviation instructors. All Instrument Phase training in IERW is conducted by Doss. Pretest briefings to explain Procedure 2 characteristics were held with the instructors. Evaluation Guidelines (Appendix) were then distributed to each instructor.

Table 4

Procedure 2 Scoring Algorithm

Performance Description	Segment Score
All variables within standard Instrument Phase Tolerance Limits	3
Some variables exceed standard tolerance limits; however, all variables are within twice tolerance limits	2
Any variable exceeds twice tolerance limits	1



	Maneuver Score
All 3's	6 <i>Criterion</i>
2's, 3's	5
All 2's	4
2's, 3's, one 1	3
2's, 3's, two 1's	2
All 1's	1

Student			Rank	
Evaluator				
Date		Tng Day		
Maneuvers Assessed	Segment			Maneuver Score
	1	2	3	
1. Climbing Turn	2	3	3	5
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
Turbulence				
UH-1		UH1FS		

Figure 4. Procedure 2 inflight scoring format

## RESULTS

### Procedure 1 Results

Figure 5 presents performance curves for each of the four students. Each data point represents the mean Procedure 1 score for all maneuvers performed on each training day. Any interpretation of the curves relative to performance improvement was questionable because of the increased difficulty level of the maneuvers trained and assessed during the latter stages of BI training. Because of the nature of the training, maneuver assignments for given training days could not be controlled.

Daily IERW grades as assigned by instructors were all B's with the exception of one C assigned on BI training day 5. It should be noted that the Procedure 1 score exhibited a decrease for that student on that training day. Mean Procedure 1 checkride scores (day 7) were subcriterion for two students. All IERW checkride grades as assigned by BI instructors were well within the passing range, and are shown in parentheses in Figure 5.

A result of interest was the number of Procedure 1 subcriterion scores (1, 2, or 3) attributable to each of the measures at various sampling times. Table 5 presents a summary of these data for all BI

Table 5  
Number of Procedure 1 Subcriterion Scores (1 - 3)  
Attributable to Flight Variables at  
Various Sampling Times (All Maneuvers)

	Times				TOT
	15	30	45	60	
AS	0	1	2	5	8
ALT	10	18	16	18	62
HDG	2	46	52	24	124
TOT	12	65	70	47	194



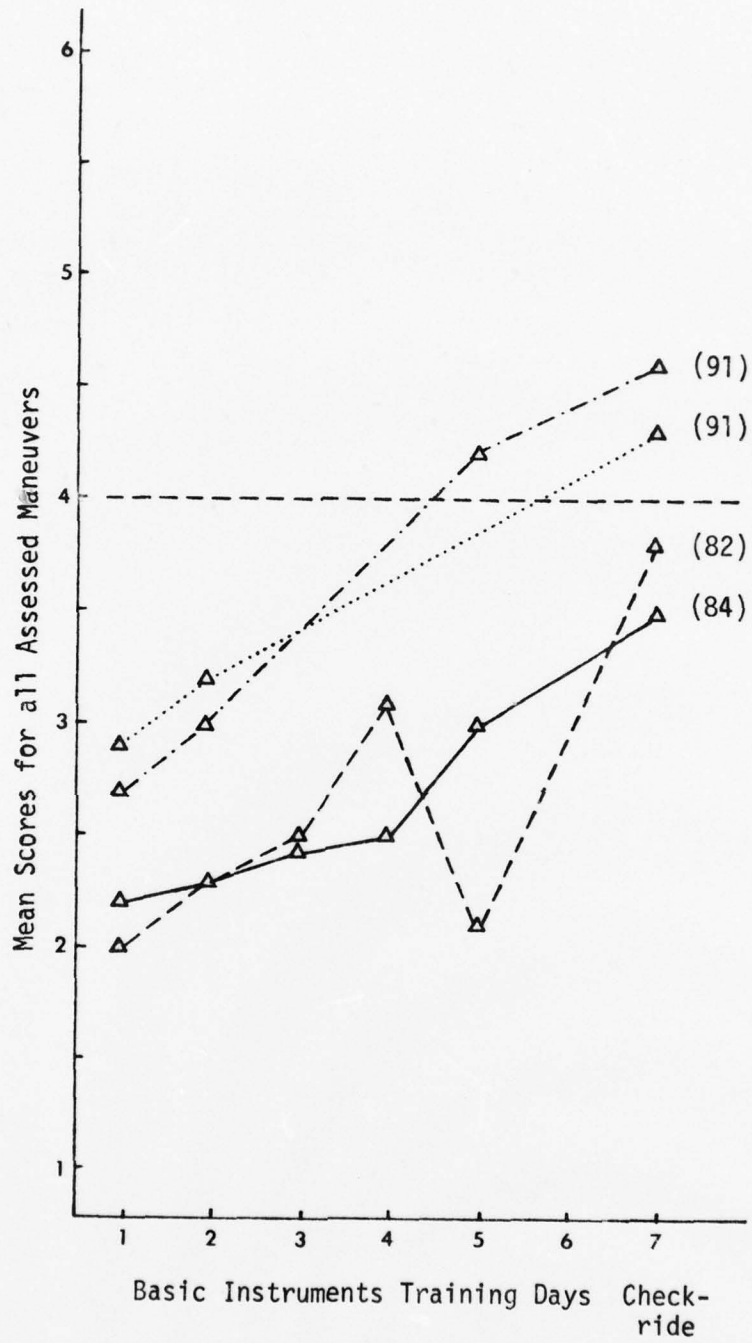


Figure 5. Procedure 1 maneuver scores for Basic Instruments students (n=4).

maneuvers. Heading errors, as measured by deviations from standard tolerance limits, were frequent for the 30- and 45-second sampling times. Approximately 64% of all subcriterion scores were the result of heading errors, while 32% were attributable to altitude deviations, and only 4% attributable to airspeed errors.

#### Procedure 2 Results

In general, mean Procedure 2 scores increased over training days. However, the Evaluation Guidelines were not closely followed, resulting in variations in assessment. For example, the 1/2 tolerance criterion included in Procedure 1, was arbitrarily used on Procedure 2 to assess segments of some maneuvers. For this reason, graphic representation of maneuver scores would be misleading.

Procedure 2 was reported to be highly usable. Instructors were optimistic about the Procedure 2 format for use in the aircraft. However, they indicated that Procedure 2 maneuver scores were, in their judgment, high, relative to students' actual proficiency, especially for training days one and two. Procedure 2 scores did not discriminate proficiency among the eight students across training days.

## DISCUSSION

Basic differences between the two scoring procedures were noted and are now summarized. Because of its fixed-time characteristic, Procedure 1 sampling required the investigator's undivided attention. The time-based sampling procedure was thus assumed to be overly constraining for instructor use because of its observational and recording demands. This assumption was not empirically tested, however, since the time and effort necessary to conduct the tests were not considered to be cost-effective. In general, the Procedure 1 sampling procedure posed an apparent inflight safety hazard.

It should be reiterated that Procedure 1 was employed by the investigator while Procedure 2 was used by instructors. This, in itself, obscures many of the desired relative effectiveness comparisons that might otherwise be made. Based upon instructor reports, however, Procedure 2 sampling would appear to be practical for inflight use. Maneuver segments were easily and objectively assessed although some degree of standardization was likely sacrificed relative to Procedure 1 sampling.

The Procedure 1 scoring algorithm permitted detection and quantification of proficiency differences among and within students across the seven days of Basic Instrument training. Although the Procedure 2 algorithm did not reflect these differences, it is not known whether this was due to overly lenient performance criteria or to a lack of assessment standardization by the instructors. It would appear that each of these factors contributed to the excessively high grades received by all eight Procedure 2 students during early training. The instructors indicated that narrower tolerance limits than those employed in Procedure 2 would seem to be necessary to discriminate proficiency, at least for Basic Instrument training.

The observed differences within flight variables in the determination of Procedure 1 downgrades indicate that certain variables are more critical than others in determining maneuver skill. The predominance of heading errors during intermediate sampling times would seem to have two implications. First, students are not likely encouraged to maintain control precision during the intermediate portions of a maneuver. Rather, the initial and terminal portions tend to be emphasized. This was supported by observations of training sessions, including briefing and debriefing.

Secondly, and perhaps more importantly, the higher overall incidence of heading errors indicates the need of a differential weighting procedure for published IERW tolerance limits. Specifically, these results indicate that  $\pm 10$  degrees of heading is generally more difficult to maintain than  $\pm 10$  knots of airspeed. Further,  $\pm 10$  degrees of heading is likely a more realistic tolerance band for straight and level flight than for a climbing maneuver or an acceleration-deceleration maneuver. These tolerance differences indicate that objective information should be collected using the UH-1FS automated data collection system. This data should provide a means to generate time-critical sampling points and to specify valid weighting procedures. Valid performance criteria could then be identified and subsequently implemented into IERW assessment procedures.

## REFERENCES

1. Williams, A. Discrimination and Manipulation on Goal-Directed Instrument Flight. *Aviation Research Monographs*, 1971, 1 (1).
2. Obermayer, R. W. and Vreuls, D. Measurement of Flight Training Research, *Proceedings of the 16h Annual Meeting of the Human Factors Society*, Beverly Hills, California, October 1972.
3. Povenmire, H. K. and Roscoe, S. N. An Evaluation of Ground-Based Flight Trainers in Routine Primary Flight Training. *Human Factors*, 1971, 13, 109-116.
4. Carter, V. E. and Semple, C. A. *Specific Behavioral Objectives for VF-121 Training in Basic Air Combat Maneuvering (Confidential)*. Contract N62269-74-C-0314 (Volume 3), Northrop Corporation Aircraft Division, October 1974.
5. Carter, V. E. Development of Automated Performance Measures for Introductory Air Combat Maneuvers. *Proceedings of the 21st Annual Meeting of the Human Factors Society*, San Francisco, California, October 1977.
6. USAAVNC Reg. 350-16, *The Uniform Flight Grading System*, US Army Aviation Center, Fort Rucker, AL, February 1973.
7. Greer, G. D., Smith, W. D., and Hatfield, J. L. *Improving Flight Proficiency Evaluation in Army Helicopter Training*, HumRRO Technical Report 77, Human Resources Research Office, Alexandria, VA, May 1962.
8. USAAVNC, *Turbine Trainer Test*, Human Resources Research Organization, US Army Aviation Center, Fort Rucker, AL, 1972.
9. HQATC, *Syllabus of Instruction for Undergraduate Pilot Training (T-37/T-38)*, Department of the Air Force, Randolph Air Force Base, Texas, June 1978.
10. Childs, J. M. *The Identification and Measurement of Critical IERW Performance Variables*. Research Memorandum, Canyon Research Group, Inc., Fort Rucker, AL, March 1979.
11. Childs, J. M. *Development of Techniques for Inflight Performance Assessment*. Research Memorandum, Canyon Research Group, Inc., Fort Rucker, AL, March 1979.
12. Childs, J. M., *op. cit.*, ref. 10.
13. Childs, J. M., *op. cit.*, ref. 11.



APPENDIX  
EVALUATION GUIDELINES FOR ENCLOSED GRADESLIP

EVALUATION GUIDELINES  
FOR ENCLOSED GRADES LIP

The Army Research Institute (ARI) and Canyon Research Group, Inc. seek your help in developing a performance-based inflight evaluation procedure. One objective in developing such a system is to provide USAAVNC training management with the necessary proficiency data to track students accurately and efficiently across all IERW training phases. We have begun at the Basic Instrument level and intend to extend this work to other IERW phases. We recognize that data gathering of this nature is often tedious and sometimes places excessive recording demands on instructors, jeopardizing inflight safety. Thus, user orientation has been of paramount importance in developing this gradeslip. You, the potential user, are in the best position of helping us achieve this goal. Thus, we solicit all comments - positive and/or negative - you would care to provide regarding its ease of use, measure selection, scoring criteria, etc.

In using this gradeslip, we ask that you evaluate the students' performance solely on the basis of instrument indications of the stated parameters. Any additional information which you feel is pertinent should be included in the Comment section on the back of the gradeslip. This scoring procedure was designed for use in evaluating Basic Instrument maneuvers (except unusual attitude recovery). We recommend that, if possible, all instruction be completed prior to (and withheld during) assessment using this gradeslip.

The following instructions explain the gradeslip components.

1. Student's Name and Rank, Evaluator's Name, Date - self explanatory.
2. Training Day - indicate day number of Basic Instrument training (1-7) for student.
3. Maneuvers Assessed - indicate which maneuver is undergoing assessment (straight and level, standard rate turn, climb, descent, steep turn, climbing turn, descending turn, acceleration/deceleration). Any maneuver may be evaluated more than once.
4. Segment Score - maneuvers have been segmented according to transitions normally occurring in flight parameters over time.
  - o Segment 1 - ("Initiate" Segment) - is intended to assess the student's proficiency in transitioning the aircraft from straight and level initial conditions into some form of turn, climb, or descent, or by increasing airspeed at a relatively constant rate while maintaining heading and altitude. Segment 1 should be assessed during the first 10 seconds of turns, climbs, and descents, and during the initial acceleration (90 knots → 100 knots) of the acceleration/deceleration maneuver. Segment scores are explained in Table 1.

Table 1

Procedure for Assigning Scores to Segments

Observed Control Levels	Segment Score
All assessed parameters are within IERW standard tolerance limits <sup>1</sup>	3
Some assessed parameters exceed standard tolerance limits; however, all parameters are within double tolerance limits <sup>2</sup>	2
Any assessed parameter exceeds double tolerance limits such that an impending safety hazard exists	1

<sup>1</sup>Airspeed  $\pm$  10 knots, altitude  $\pm$  100 feet, heading  $\pm$  10 degrees

<sup>2</sup>Airspeed  $\pm$  20 knots, altitude  $\pm$  200 feet, heading  $\pm$  20 degrees

- o Segment 2 - ("Maintain" Segment) - is designed to evaluate student ability to maintain the proper rates of turn, climb, descent, or airspeed change throughout the intermediate portion of the maneuver. This segment evaluates the deceleration (100 knots—→70 knots) portion of the acceleration/deceleration. The most critical parameter of a maneuver (e.g., turn rate in a turn, climb rate in a climb) should be evaluated approximately 30 seconds after the maneuver has been started in the interest of uniformity. If turn rate ( $3\frac{1}{4}$ /sec) is slow or fast, it will be reflected by heading at 30 seconds. If climb or descent rate (500 fpm) is improper, it will be indicated by altitude. Secondary parameters (altitude, airspeed, and trim in a turn; heading, airspeed, and trim in a climb or descent; altitude, heading, and trim in an acceleration/deceleration) should be monitored as closely as possible throughout the segment (as they presently are). Any deviation of these parameters from standard IERW tolerance limits will reduce the segment score (Table 1).
- o Segment 3 - ("Terminate" Segment) assesses transition from some form of turn, climb, or descent back into straight and level flight, or transition from controlled airspeed change back to constant normal cruise airspeed. This segment should be assessed during the final 10 seconds of turns, climbs, and descents, and during the terminal acceleration (70 knots—→90 knots) phase of the acceleration/deceleration maneuver.

#### Inflight Recording

One of the benefits of this grading system concerns its inflight recording ease and flexibility. Any recording method you care to use is satisfactory as long as control levels are accurately observed for each maneuver segment. Two suggestions are:

1. On a blank sheet of paper, simply record which parameter(s) exceed(s) tolerance or double tolerance within a segment by some coding method such as:

A (altitude exceeds tolerance)

S (airspeed exceeds tolerance)

H (heading exceeds tolerance)

T (trim exceeds tolerance)

AX (altitude exceeds double tolerance)

SX (airspeed exceeds double tolerance)

HX (heading exceeds double tolerance)

or:



2. Draw a matrix such as the one below for recording tolerance violations (slash mark) or double tolerance violations (X).

	Segment 1	Segment 2	Segment 3
HDG		X	
ALT		/	/
AS			
TRIM			

Blank cells indicate that no tolerance violations were observed. In using a recording method like those mentioned above, there is no need to complete the gradeslip until debriefing occurs. However, we believe that, with very little practice, instructors could insert segment scores (1, 2, or 3) on the gradeslip in flight with minimal distraction. We would be interested in knowing your particular recording technique. Maneuver scores can, of course, be assigned later on the basis of segment score configurations. Thus, it is critical that segment scores be accurately assigned. Each segment should be assigned a segment score of 1, 2, or 3 depending on the level of control observed for that segment. In the interest of standardization, all segments should be evaluated as closely to the recommended assessment times as possible. Please evaluate straight and level flight in one-minute blocks of time and include a score for each 20-second segment.

5. Maneuver Score - Assign maneuver scores according to Table 2 guidelines.

Table 2

Assignment of Maneuver Scores on the Basis of Segment Scores

Segment Scores	Maneuver Scores
All 3's	6 (criterion)
2's, 3's	5
All 2's	4
2's, 3's One 1	3
2's, 3's Two 1's	2
All 1's	1

6. Turbulence - indicate turbulence level.

7. UH-1, UH1FS - indicate whether the maneuvers are performed in the aircraft or simulator.

If you have questions regarding these evaluation guidelines, please do not hesitate to call Jerry Childs at 205/598-2453. Your cooperation is greatly appreciated.