Human Testing

- Ethics and protocols for human testing
- Test protocols
- Mental workload assessment techniques
- Specific instructions for testing
- Download this information from https://spacecraft.ssl.umd.edu/academics/academics.html

Testing Humans

- Not like testing hardware
  - Individual variability
  - Alertness and fatigue
  - Motivational differences
  - Skill differences
- Focus on obtaining statistically significant sample size
- Investigator responsible for ethical treatment of test subjects

Ethics of Human Testing - Safety

- Institutional Review Board (IRB) approval required for experiments using human subjects
- Legal and ethical responsibility to ensure informed consent is obtained and documented
- Subjects must be volunteers without being coerced (overt or covert) and free to withdraw at any time
- Subjects must be drawn from the most diverse pool possible within limitations of test requirements

Ethics of Human Testing - Diversity

- Federally supported research mandates the use of the most diverse test population possible
  - Gender
  - Race
  - Age
- Most space human testing is severely limited in feasible subject pool, number of available test subjects
- Ideal test pool ~40-60 test subjects; aerospace testing frequently limited to ~8-10 or even less

Ethics of Human Testing - Privacy

- Subjects must never be identified by name, initials, or other means which would allow identification of individuals
- Subjects typically identified by numbers, letters, or other random labels
- Lead investigators maintain index from code to test subject names under lock and key
- Characteristics potentially relevant to performance (size, gender, experience, etc.) must be correlated to identifiers

Collection of Subject Metrics

- Some parameters are potentially relevant to performance of human tasks
  - Gender
  - Physical size
  - Age
  - Relevant experience
  - etc.
- Data collected via questionnaire filled out prior to testing, investigator measurements, qualifying tests, etc.

Control of Variables - Motivation

- Performance is generally strongly driven by individual motivation (perfection, competition, boredom, etc.)
- Test subjects should not be exposed to other test subjects to avoid competition (particularly powerful in college-age males)
- Investigators generally create a set of instructions provided to test subject (often read aloud) with very specific instructions on desired parameters of testing (e.g., “minimum time”, “maximum precision”)

Control of Variables - Performance

- Test subject access to performance metrics is very highly restricted
  - Never provide comparative performance with other test subjects
  - Frequently do not tell subjects their own performance metrics after individual runs
- Subjects should seldom be allowed to observe other subjects performing task
  - Avoid adopting techniques from other subjects
  - Have to be concerned about subject-subject "hints"

Control of Variables - Learning

- People tend to improve with repetition
- Tends to follow exponential heuristic:

$$t_n = t_1 n^p$$

- Typically quantify learning as reduction following doubling of number of trials - "80%" where t_2=0.8t_1, t_4=0.8t_2, etc.

$$LC\% = \frac{t_2}{t_1} = 2^p \Longrightarrow p = \frac{\ln(LC\%)}{\ln(2)}$$

- Have to ensure test subjects have essentially leveled off on learning curve for task performance

Control of Variables - Trial Order

- People tend to prefer system they first train on
- Greatest number of repetitions will be on first trial configuration
- Where possible, have sufficient test population to have multiple subjects in each of all possible permutations of order for test matrix
- Where fully populated test matrix is not possible, at least have distribution in starting mode

Research Focus

- Current spacesuits limit wearers to visual and audio inputs only
- Planetary exploration and orbital operations would be augmented by visual displays compatible with spacesuits
- Step 1: investigate the influence of visual overlays/ augmented reality on simple task performance

Test Taskboard

Your Task

- Each test subject will perform a series of 50 simple tasks (e.g., “push grey button 4”, “flip switch B3”) in response to three types of prompts
  - Spoken commands
  - Graphic/test overlays (Microsoft HoloLens)
  - Graphics plus spoken commands
- Performance metrics:
  - Time to complete tasks
  - Cooper-Harper rating
  - NASA Task Load Index
  - Incidence of errors

Your Test Protocol - Testing Organization

- You will be given a questionnaire for test subjects listing critical parameters which could affect performance (e.g., gender, size, age, relevant prior experience) - https://forms.gle/8FzooKqc1ca2egU38
- You will be able to run about two test subjects per hour through the full set of tests, including documenting assessments after each test case
- Keep subjects from observing other subjects for technique or performance metrics

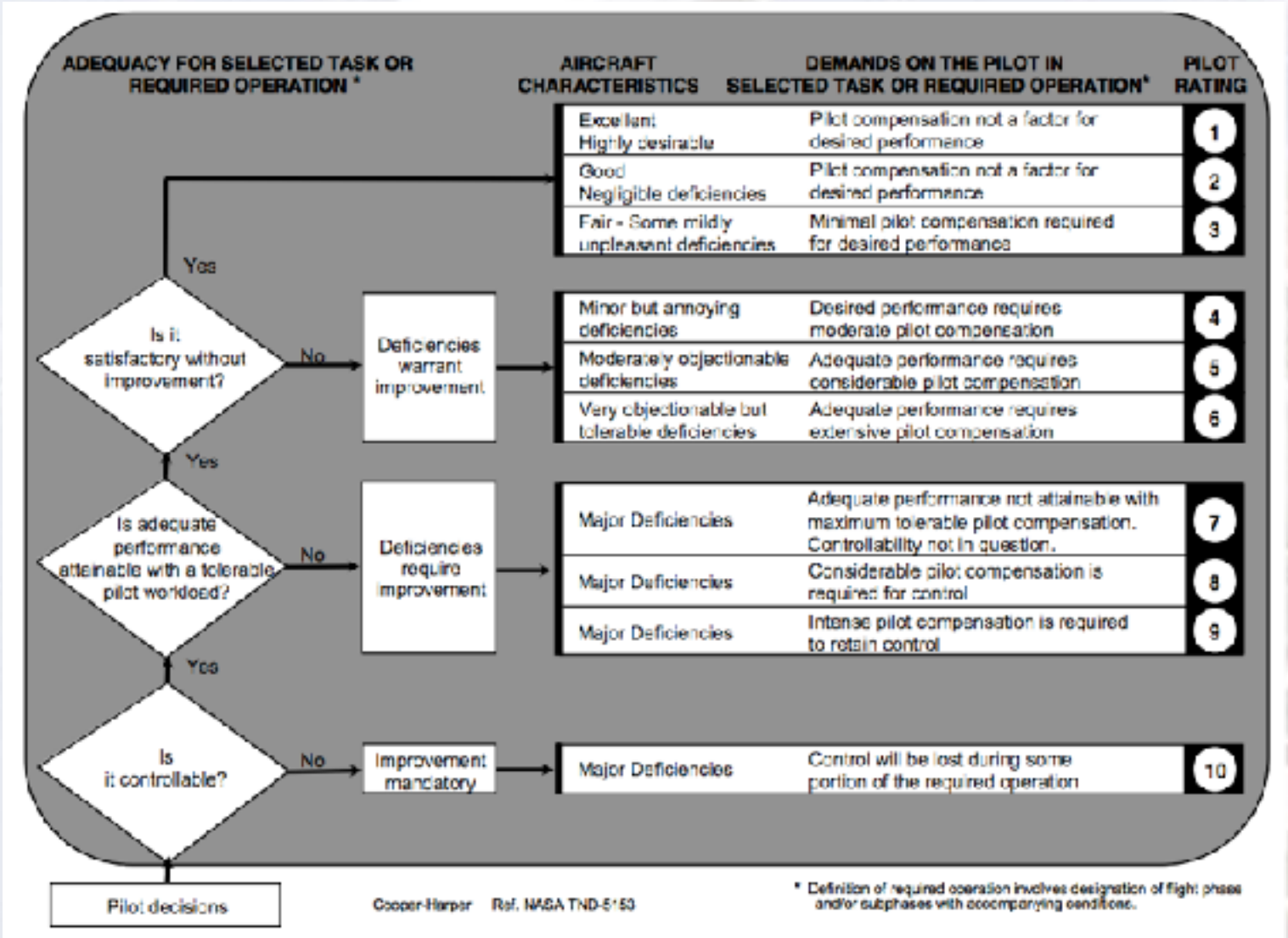
Mental Workload Assessment Techniques

- Primary Task Performance Measures
- Secondary Task Performance Measures
- Bedford Scale
- Defense Research Agency Workload Scale
- Instantaneous Self Assessment Workload
- Malvern Capacity Estimate
- Modified Cooper-Harper Rating
- NASA Task Load Index
- Subjective Workload Assessment Technique
- Workload Profile Technique
- Cognitive Task Load Analysis

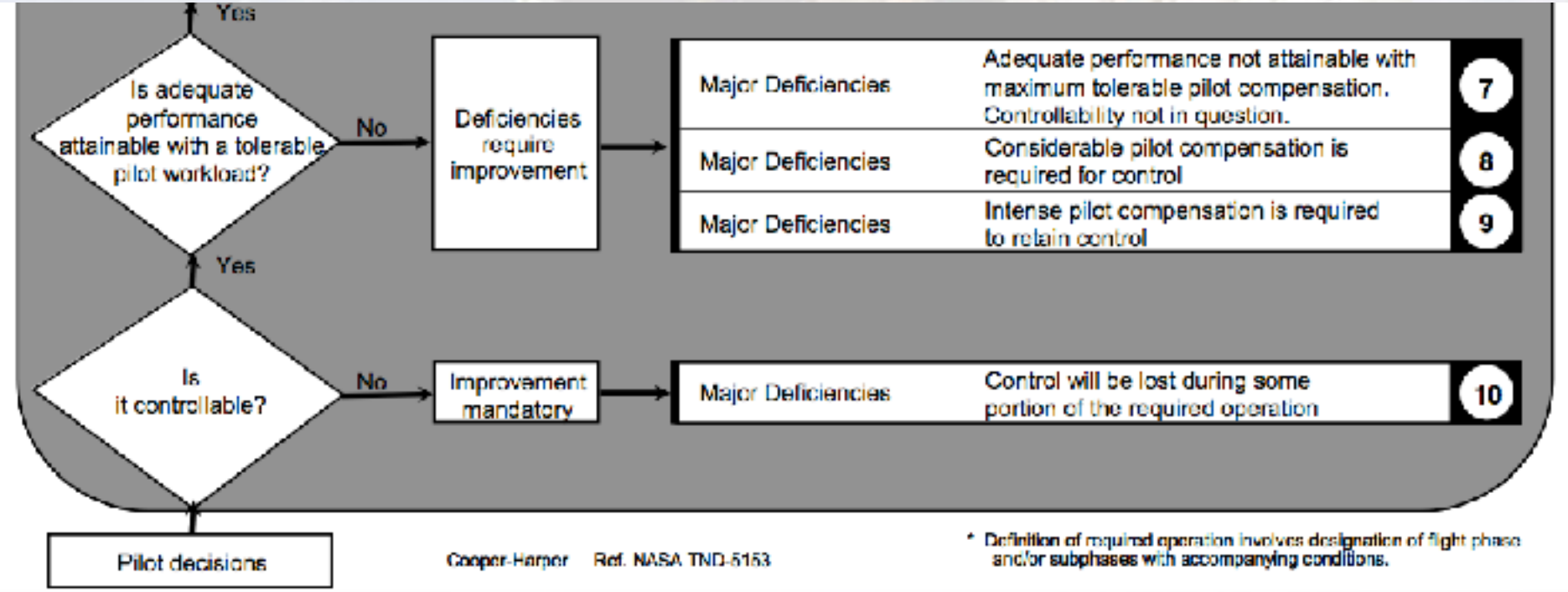
Mental Workload Assessment Techniques

- Primary Task Performance Measures
- Secondary Task Performance Measures
- Bedford Scale
- Defense Research Agency Workload Scale
- Instantaneous Self Assessment Workload
- Malvern Capacity Estimate
- Modified Cooper-Harper Rating
- NASA Task Load Index
- Subjective Workload Assessment Technique
- Workload Profile Technique
- Cognitive Task Load Analysis

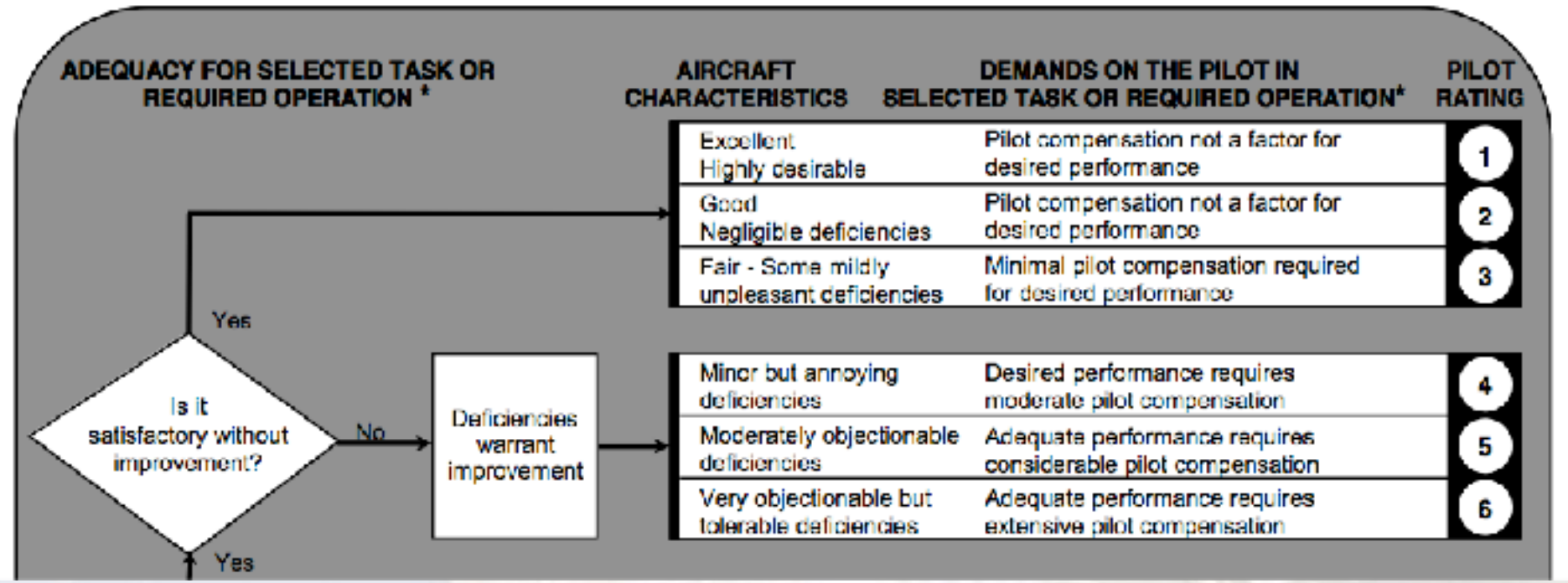
Cooper-Harper Rating



Cooper-Harper Rating (close-up 1)



Cooper-Harper Rating (close-up 2)



Cooper-Harper Definition of Terms

DEFINITIONS FROM TN-D-5153

COMPENSATION

The measure of additional pilot effort and attention required to maintain a given level of performance in the face of deficient vehicle characteristics.

HANDLING QUALITIES

Those qualities or charcteristics of an aircraft that govern the ease and precision with which a pilot is able to perform the tasks required in support of an aircraft role.

MISSION

The composite of pilot-vehicle functions that must be performed to fulfil operational requirements. May be specified for a role, complete flight, flight phase, or flight subphase.

PERFORMANCE

The precision of control with respect to aircraft movement that a pilot is able to achieve in performing a task. (Pilot vehicle performance is a measure of handling performance. Pilot performance is a measure of the manner or efficiency with which a pilot moves the principal controls in performing a task.)

ROLE

The function or purpose that defines the primary use of an aircraft.

TASK

The actual work assigned a pilot to be performed in completion of or as representative of a designated flight segment.

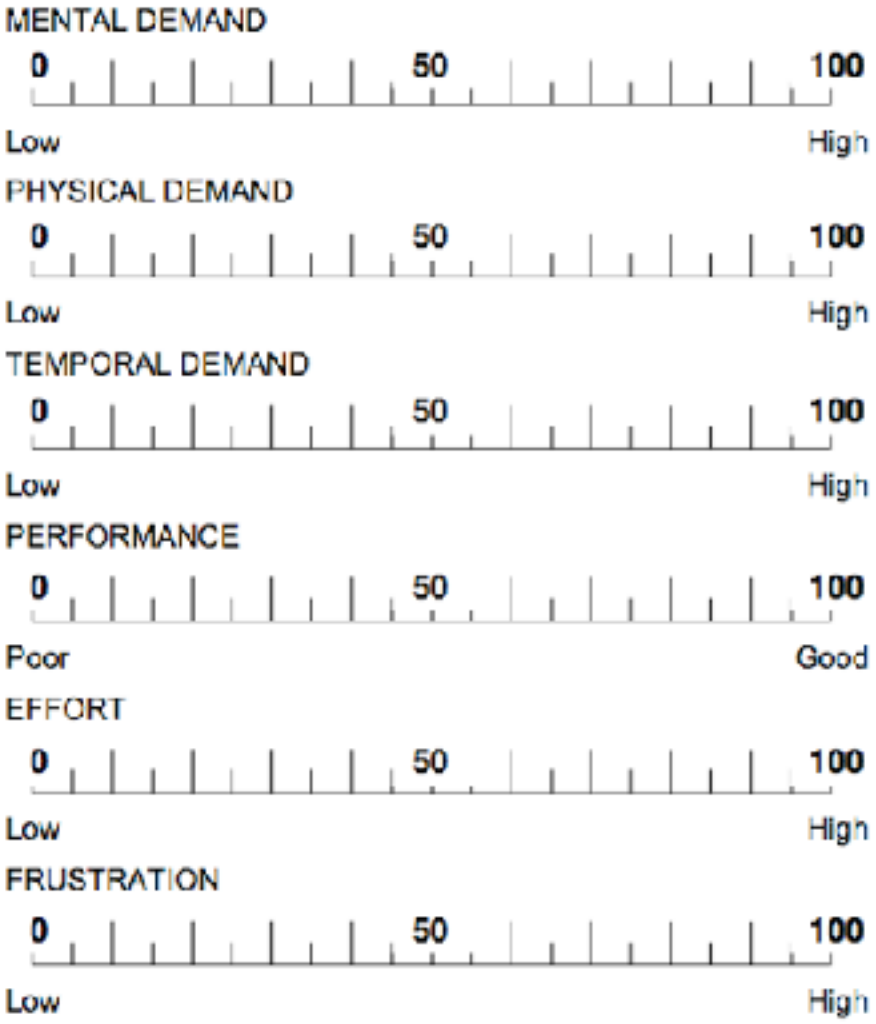
WORKLOAD

The integrated physical and mental effort required to perform a specified piloting task.

NASA Task Load Index (TLX)

Rating Scale Definitions

Title	Descriptions
MENTAL DEMAND	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?



Analysis and Reporting

- Lab reports should include a summary of the test set-up, objectives, and protocols
- Summary presentation of experiential data - times, errors, subjective evaluations
- Perform statistical evaluations of individual data to determine heuristic relationship between performance and workload assessment parameters
  - Perform a two-factor ANOVA to look for statistical significance in time required for each test mode
  - Look for correlation between errors and time
  - Look for correlation between performance and C-H/TLX

Human Testing Lab Sessions

- Thursday and Friday, 2-5 pm
- Advanced Robotics Development Laboratory (Rm. 1309) in the Kim Engineering Building
- Sign up for a 60-minute slot
- We will run as many subjects as possible during that time
- This is experimental research - things fail! Be resilient and resourceful