Discover problems with instructions before you ask pilots to follow them.

Ready, Set . . . Test

BY HEMANT BHANA

Usability testing is a concept from the software industry. It measures how effectively a product enables the end user to accomplish the goal for which the product is designed.

Usability testing has direct applications in aviation safety. Aviation safety professionals who write standard operating procedures (SOPs), special procedures and operations manuals should be as concerned with usability testing as software designers are. If a manual or procedure is unclear, verbose, poorly formatted or does not efficiently transfer information, its value as a safety tool diminishes.

Presenting information accurately the first time is important. This avoids safety managers having to present multiple revisions to clear up ambiguous data. Unfortunately, issuing hastily conceived instructions and procedures is endemic in the industry and can harm an organization's safety culture.

Safety professionals can and should plan and conduct an aviation usability test. The test will ensure that the product is accurate, unambiguous and easy to use. Most important, it will eliminate the need for costly and time-wasting post-release corrections.

The Test

A basic aviation usability test does not require the level of sophistication used by, for example, Microsoft. The premise, however, is the same — find a sample of test participants representative of the end user, identify what the test intends to address and give the participants tasks to perform in various scenarios (Table 1, p. 27). These actions then form the basis for any changes to a procedure or instruction *prior* to

its formal release. It tests how well the product accomplishes its goals.

Step 1: Identify the relevant issues

The first step in conducting an aviation usability test is to identify what topics or problems the proposed instructions or manual is supposed to address. This step is the backbone of the actual usability test.

Aviation safety officials can derive this information from sources such as safety, training or survey data, or from a detailed analysis of end user tasks. Identifying the major issues first defines the scope of the test, since the goal is not to resolve every problem but to address the major concerns.

As an example, flight managers learn that there is confusion about autopilot usage during nonprecision approaches. The airline decides to issue guidance to pilots clarifying the procedure. Prior to dissemination, the airline tests the impending instructions for usability.

At this point, the issues are broad and consist of questions such as, "Can pilots use the pending guidance to properly use the autopilot during a nonprecision approach?"

Step 2: Define concrete questions

This step breaks down the large issues into specific questions. A good method is to walk through the users' experience and try to identify what is most important for them to grasp.

Step 3: Define tasks and scenarios

The tasks, based on the concrete questions, are the actions the user must perform to answer the questions.

The scenarios are a real-life approximation of how the user interfaces with the task. The problem with just giving the user a task is that all the issues might not be evident unless the user sees the task in context. For example, task one involves finding out when you cannot use the autopilot — relatively straightforward. However, asking a user to perform a task in its proper context could yield additional information. The user might look in a completely different area of the manual to meet his or her expectations of where the information is found. The goal is to eliminate confusion when the user has to use the product outside the artificial setting of a test.

To get the most accurate results, the scenarios should describe situations that the participants are likely to encounter.

Step 4: Determine what data to collect

Usability testing is not academically rigorous. Interpretation of the data is mostly subjective, since the goal is to uncover major problems with the material, not to conduct statistically significant research.

In our continuing example, tasks one through three involve qualitative

Breaking Down an Issue

Issue	Can pilots find the necessary information in the pending guidance that will enable them to properly use the autopilot during a nonprecision approach?			
Concrete questions	Can pilots find the autopilot limitation information in the guidance?	Is the guidance clear on when autopilot usage is mandatory?	Do pilots understand the SOP in the guidance pertaining to autopilot usage during a nonprecision approach?	Can pilots quickly search key portions of the guidance?
Tasks	Task 1. Use the guidance to find when you cannot use the autopilot.	Task 2. Use the guidance to find under what conditions autopilot usage is mandatory.	Task 3. Use the guidance to explain how to use the autopilot during a nonprecision approach.	Task 4. Use the guidance to quickly learn about using the flight-path-angle mode.
Scenarios	You are briefing the approach, and the person you are flying with wants to use the autopilot to attain a particular altitude. You are unsure whether this is permissible. Using this guidance, inform the other pilot when the autopilot may not be used.	You are close to the airport and receive a weather report that is worse than expected. You mention to the other pilot that you are planning to hand-fly the approach to maintain proficiency. The other pilot asks, "Is that allowed?" Using the guidance, inform the other pilot when hand-flying is permissible.	You are the captain. You are flying with a new-hire first officer who is confused about how to use the autopilot during a nonprecision approach. During the approach briefing, the first officer says, "I have no idea how to do this!" Using this guidance, find the portion that describes the SOP, read and explain it to your new-hire copilot.	You are close to the airport and at the last minute decide to use the flight-path-angle mode of the autopilot. Using this guidance, as fast as you can, find the portion of the manual that describes how to use the flight-path-angle mode.

SOP = standard operating procedure

Source: Hemant Bhana

Table 1

data, while task four involves quantitative data (time). The data collected should not simply record whether the participant successfully completed the task. As part of the pre-test briefing, test moderators should request that the participants "think out loud" or verbalize their thoughts as they proceed with the tasks. Recording and collecting *these* data are critical, as thoughts and opinions will indicate how well the product accomplishes its goals.

A test participant may successfully complete the tasks, but of vital interest is what obstacles the participant encounters en route. That information is far more valuable, since safety managers can use the information to eliminate these obstacles during the rewrite.

The test moderator may also include several questions at the end of each task that focus on the participant's expectations. For example, the moderator may ask about what terminology the participants were looking for or how the test taker is searching for information. The answers to these questions will bring the material more in line with the expectations of the end users.

In our example, task four is slightly more complicated, as it involves recording time. For this task, having participants find the flight-path-angle information is ancillary because the intent of the test is to measure how searchable the document is. Thus for task four, the data metric is both time to completion *and* thoughts and opinions. The time criterion for a successful test is subjective; the stakeholder determines all the benchmarks for product success.

Since the goal of usability testing is to uncover major problems, test moderators only need five to eight participants per group. Research has determined that five test participants can uncover 80 percent of usability problems.[1]

Each testing group represents a specific category of users. In our example, the testing group is a random selection of captains and first officers. Two groups would be needed to see if captains and first officers interpret the instructions differently.

Step 5: Conducting the test

Test facilitators should conduct the test in a comfortable setting that allows for observation and is free from distraction.

The test facilitator should also work from a script to ensure consistency of participant instructions. The script should emphasize that the usability test is not an evaluation of the participants. This will put the participants at ease and increase the quality of the data.

Step 6: Capturing data

If possible, one person should act as the test moderator, another as the note taker. Alternatively, audio and video recording equipment can capture test participant comments for detailed analysis later. However participant data are captured, the goal is to record the participants' thought processes and observations. The note taker should pay special attention to participants' difficulties. Capturing why the participants stumble or what problems the test taker encounters will yield the most valuable data.

Likewise, the data from the post-test questionnaire should emphasize what the test participants were expecting. Test facilitators can also solicit information with off-script questions if information is not forthcoming from the participants.

Step 7: Interpreting and applying the data

First, the information should be organized according to the task performed. Next, the testing team should look for common themes in the data that would indicate systemic problems. For example, multiple people having trouble finding the flight-path-angle information queried in task four could indicate a problem with information organization. The test team's job is to identify what elements of the guidance structure caused the problems.

The test team should then prioritize the problems and start working on potential fixes. Continuing our example, if the data indicate that the flight-path-angle information was not found where the participants expected it, managers can rewrite the guidance to be more in line with expectations.

Not Only Manuals

The example in this article centered on a proposed SOP or manual change concerning autopilot usage during non-precision approaches. However, aircraft operators can employ usability testing for a variety of products, including emergency procedures.

Keep in mind that the usability test is a measure of how well the product *fits the needs of the user*, not a test of the user or the content of the product. The goal is to identify flaws in how well the final product functions as a tool. Getting this information correct prior to dissemination is vital to prevent confusion and noncompliance, and to uphold high standards of safety.

Hemant Bhana is a lead technical pilot with GE Aviation–PBN Solutions, based in Kent, Washington, U.S.

Note

1. Virzi, R. (1992). "Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough?" *Human Factors*, Volume 34(4), pp. 457–468.