Federal Aviation Administration

DOT/FAA/AM-08/18
Office of Aerospace Medicine
Washington, DC 20591

# Designing Questionnaires for Controlling and Managing Information Complexity in Visual Displays

Jing Xing

Civil Aerospace Medical Institute
Oklahoma City, OK 73125

August 2008

Final Report

OK-08-1997

# NOTICE

## Technical Report Documentation Page

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| DOT/FAA/AM-08/18 | | |

**4. Title and Subtitle**
Designing Questionnaires for Controlling and Managing Information Complexity in Visual Displays

**5. Report Date**
August 2008

**6. Performing Organization Code**

**7. Author(s)**
Xing J

**8. Performing Organization Report No.**

**9. Performing Organization Name and Address**
FAA Civil Aerospace Medical Institute
P.O. Box 25082
Oklahoma City, OK 73125

**10. Work Unit No. (TRAIS)**

**11. Contract or Grant No.**

**12. Sponsoring Agency name and Address**
Office of Aerospace Medicine
Federal Aviation Administration
800 Independence Ave., S.W.
Washington, DC 20591

**13. Type of Report and Period Covered**

**14. Sponsoring Agency Code**

**15. Supplemental Notes**
Work was accomplished under approved task AM-HRRD522

**16. Abstract**
Information complexity of automation displays has become a bottleneck that limits the usefulness of new technologies in air traffic control (ATC). Previously, we developed a set of metrics to measure information complexity in ATC displays. While these metrics provide measures of display complexity, their use is somewhat limited due to required human factors expertise and understanding of the display design. Technology developers and human factors practitioners often desire quick, easy-to-use tools to assess the display during design and acquisition evaluation. Questionnaires provide a quick and inexpensive means to gather data from a potentially large number of respondents. We developed two questionnaires to evaluate ATC display complexity, based on the metric indices. The first questionnaire employs a multiple-choice format and allows quantitative evaluation of complexity. The second questionnaire uses a Likert rating format and is intended for qualitative assessment of complexity. We conducted an initial assessment of the questionnaires with seven subject matter experts on a radar display (STARS). The results indicate that both questionnaires produced consistent complexity evaluations among the subjects. Thus, we recommend that the multiple-choice questionnaire is more suitable for assessing quantitative complexity control during acquisition evaluations, and the Likert rating questionnaire is more suitable for complexity management during design of new ATC technologies.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| Information Complexity, Usability, Design, Display, Air Traffic Control | Document is available to the public through the Defense Technical Information Center, Ft. Belvior, VA 22060; and the National Technical Information Service, Springfield, VA 22161 |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 21 | |

**Form DOT F 1700.7** (8-72) Reproduction of completed page authorized

# ACKNOWLEDGMENTS

We thank the controller training instructors from the FAA Academy for their serving as the subject matter experts in the development and initial assessment of the two complexity questionnaires. We appreciate all the FAA reviewers and editors whose efforts greatly improved this report. This work was sponsored by the Office of Terminal Operation and the Human Factors Research and Engineering Group of the FAA.

## CONTENTS

# Designing Questionnaires for Controlling and Managing Information Complexity in Visual Displays

## INTRODUCTION

Interactive automation technologies have gained wide acceptance in air traffic control (ATC). These automated technologies typically employ visual displays. Compared to traditional paper-based task aids, automation systems are superior in their visual realism and capability of delivering details of complex operational procedures. However, while advances in sensor development and communication bandwidths allow an automation system to convey an increased amount of information, the capacity of human information processing within a given period of time is limited (Marios & Ivanoff, 2005). Thus, information complexity (IC) becomes a bottleneck that constrains the effectiveness and efficiency of interactive technologies. Assessment and control of IC in visual displays during design are important to prevent such bottlenecks.

Previously, we developed a framework to measure IC in interactive systems Xing & Manning 2005, Xing 2007). The framework is described as follows: a) IC is the combination of three basic factors: quantity, variety, and relation of basic information elements; b) complexity factors need to be evaluated at three stages of mental processing: perception, cognition, and action; and c) complexity metrics can be derived by associating task requirements with the mechanisms of human information processing. Within this framework, we identified nine complexity metrics for ATC displays, each measuring the demand of a complexity factor during perceptual, cognitive, or action information processing. We anticipate that these metrics are used for design and acquisition evaluation. Table 1 presents a brief description of these metrics.

While these metrics provide a means for the objective measurement of display complexity, technology developers and human factors practitioners often desire quick, easy-to-use tools to assess the display during design, prototype, and acquisition evaluation. It is also important to obtain subject matter experts' opinions in the evaluation of new technologies.

A questionnaire is an inexpensive means to acquire such data from a potentially large number of respondents. In fact, questionnaires have become one of the primary methods to assess interface usability (Kirakowski & Corbett, 1993; Lewis, 1995; Schneiderman, 1992). In aviation studies, human factors practitioners often use post-scenario questionnaires for acquisition evaluation of new ATC automation systems (Willems & Truitt, 1999; Willems, Allen, & Stein, 1999). In such studies, subjects first use the system to perform assigned tasks with pre-generated scenarios and subsequently complete questionnaires designed to assess the system. The results allow researchers to assess the usability and collect subjective opinions about users' satisfaction with the system.

In this study, we developed two questionnaires based on the complexity metrics for ATC displays. We intended that the two questionnaires would be used to control complexity during the development and acquisition evaluation of new ATC technologies. While questionnaires are easy to administer, developing an effective questionnaire can be a challenge. Various multi-stage procedures for developing questionnaires have been proposed in the literature. Independent of the exact methodology, the following four steps are essential: a) design a questionnaire based on a task analysis and objectives of the assessment; b) modify the questionnaire by integrating individual criticism or comments; c) test the questionnaire with a small sample of respondents; and d) validate the questionnaire through a large sample of respondents. In this report, we will first describe steps a-c. The validation results of step d) will be reported in the near future.

## DEVELOPING TWO COMPLEXITY QUESTIONNAIRES

### Specify the objectives of the questionnaires

In the literature, the purpose of designing questionnaires falls into two categories--descriptive and analytic. Descriptive studies provide estimates of the parameters of certain system characteristics. Analytic studies provide a systematic comparison of characteristics across several systems or explore the relationship among variables for a single system.

The objective of this study was to develop questionnaires to evaluate information complexity of visual displays during acquisition and to provide a way for developers to manage information complexity during the design and prototyping of new ATC systems. Information gained from the application of the questionnaires can be used to determine when the complexity of a display is beyond an operator's capacity limits of information processing; thus, the display is unacceptable for efficient and safe

**Table 1.** Metrics of information complexity for ATC displays

| | Metric | Definition | Potential consequences of complexity |
|---|---|---|---|
| **Perceptual complexity** | Number of fixation groups | A fixation group is a set of visual stimuli that can be perceived with a single fixation for detail analysis. | Increased time and difficulty in visual search. |
| | Variety of visual features | The number of distinctive colors, texture, luminance contrast, spatial frequency, and motion signals. | Increased difficulty in visually organizing information and detecting salient targets. |
| | Degree of clutter | The effect of visual perception of a stimulus being masked by the presence of other stimuli in the visual field. | Increased visual search time and reduced target detection as well as text readability. |
| **Cognitive complexity** | Number of functional units | Functional units are the independent elements or dimensions of information that are maintained in an active, quickly retrievable mental state. | Increased working memory load and reduced situation awareness. |
| | Dynamic complexity | The amount or frequency of unpredictable information onset that demands a change in the contents of the mental representation of a display. | Increased memory load and reduce situation awareness; deteriorated mental representation of a display. |
| | Relational complexity | The number of independent elements or dimensions of information that must be simultaneously combined to use the information. | Increased memory load and cognitive computational cost. |
| **Action complexity** | Action cost | The minimal amount of keystrokes, mouse movements, and transitions of action modes needed to use displayed information. | Increased task performance time. Takes users away from perceptual and cognitive tasks. |
| | Action depth | The number of serial steps needed to plan (or select from a number of action options) to acquire information. | Reduced situation awareness; increased chances of performance errors. |
| | Number of simultaneous action goals | The number of simultaneous action goals needed to use displayed information. | The brain has to switch back and forth between goals; errors may occur when switches of action goals occur at a fast pace. |

operation. The objective can be best met by obtaining quantitative indices of complexity. The second objective for complexity management of a display requires a tool to assess the effects of display design variables on complexity so that they can modify those variables that cause high complexity and manage the complexity to meet users' requirements. Differences in the applications will lead to the development of two independent questionnaires.

**Format Considerations**

Questions may be either open-ended or close-ended. Open-format questions ask for unprompted opinions, and respondents are free to answer in their own words. Open-format questions are good for soliciting subjective opinions. However, the diversity of the responses reduces standardization and greatly increases the time required for systematic analysis. In contrast, close-ended questions require respondents to select one or more answers from those provided. Available response choices can vary in format, such as checklists, ranking scales, Likert scales, and multiple-choice.

The Likert scale is an ordinal, one-dimensional scaling method in which values have an inherent order or sequence but do not correspond to a precise mathematical value. A traditional Likert scale item includes a statement to which respondents make a judgment on a five-point or seven-point scale. For example, a Likert agreement response scale can be formatted as:

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

The multiple-choice format generally involves the use of questions with predetermined responses from which the respondent is requested to choose the most appropriate response. Respondents are asked to select a single response or multiple responses to the question (e.g., select all that apply). The multiple-choice format allows one to obtain gradations of opinions and combinations of reasons or actions. It draws attention to possible alternatives instead of requiring the respondent to generate them, provided that the designer of the questionnaire is sufficiently knowledgeable to identify all measurable alternatives. One advantage of this format is that responses are standardized, preventing a respondent from introducing personal bias and reducing the likelihood of item misinterpretation. Another advantage is that responses are relatively independent of the respondent's ability to express opinions (e.g., writing skills, handwriting). A disadvantage of the multiple-choice format is that questions tend to be more complex and require more care in design. It also requires that the developer have extensive knowledge of the area being investigated. We chose this format for the purpose of complexity control because the advantages stated above will allow it to yield more objective and quantitative evaluation results.

**Item Authoring**

*Multiple-choice questionnaire.* We used the previously developed metrics (Xing, 2007) as a guide for item generation. We designed the questions to assess the metrics in terms of their effects on task performance. For example, a metric for perceptual complexity is the number of fixation groups. A fixation group is defined as the visual stimuli that can be perceived with one fixation for detailed analysis. The average time to search for a particular target on a display increases with the number of fixation groups. From that information, we derived the question "Ease of finding information: How easily can you find the information you need on the display?" We developed nine questions in a similar fashion for each metric. In addition, while the evaluation with individual metrics provides information about specific aspects that make a display complex, evaluators often also want to know about the overall complexity of a display. Hence, we also developed four questions to assess the overall perceptual, cognitive, action complexity, and overall display complexity. These questions are listed in Table 2.

The biggest challenge in designing multiple-choice questions was to define the given choices logically. Obviously, there need to be sufficient choices to cover the range of answers but not so many that the distinction between them becomes blurred. Moreover, multiple-choice response categories should be mutually exclusive so that clear choices can be made. Non-exclusive answers frustrate the respondent and make interpretation difficult, at best. We tried to define the choices from the perspective of controllers' experiences with displays.

Previously we conducted ATC facility observations to understand how controllers use color displays (Xing, 2006). During the observations, we informally collected controllers' opinions about display complexity by asking questions such as "How would you describe the complexity of this system in terms of its effect on your task performance?" A classification of the answers indicated that controllers tended to describe display complexity using four levels (quoting controllers' words):

- "It is not complex at all, very easy to use, I like it."
- "It is moderately complex, yet it helps me a lot so I use it most of the time."
- "It is complex; I only use it when I am not busy with the traffic."
- "It is too complex to use. I do not have time to use it. I figure out my own ways."

**Table 2.** The questions in the multiple-choice questionnaire

| Metric | Question |
|---|---|
| Number of fixation groups | Ease of finding information: How easy is it for you to find the information you need on the display? |
| Variety of visual features | Information Organization: How well is the information organized on the display? |
| Degree of clutter | Display clutter: How easy is it for you to read the displayed text? |
| Number of functional units | Awareness of displayed information: How well are you aware of the information provided by the display? |
| Dynamic complexity | Display dynamics: How do the dynamic changes of displayed information affect your using the display? |
| Relational complexity | Relating displayed information: How easy is it for you to understand /comprehend displayed information? |
| Action cost | Performing tasks and retrieving information: How would you evaluate the amount of actions you have to take to perform tasks or acquire information? |
| Action depth | Number of steps to complete an action: How does the number of steps needed to acquire information affect your using the display? |
| Number of simultaneous action goals | Number of action sequences to perform a task: How does the number of parallel action sequences needed to perform a task or acquire information affect your performance with the display? |
| Overall perceptual complexity | Perceptual complexity of the system: How does the perceptual complexity of the display affect your task performance with the display? |
| Overall cognitive complexity | Cognitive complexity: How cognitively demanding is the displayed information? |
| Overall action complexity | Manually using the interface: How easy is it for you to use the display? |
| Overall display complexity | How do you rate the overall complexity of the display from the perspectives of its efficiency, safety, and usefulness? |

Based on controllers' opinions about complexity, we derived four generic complexity levels as the one-dimensional choices for the questionnaire. Figure 1 illustrates the levels. Next, for each metric, we developed four objective descriptions of the metric, each corresponding to one of the complexity levels. For example, for the metric of fixation group, we developed the following four descriptions:

1. I can find the information effortlessly.
2. I can find the information with a few quick glances.
3. I can find the information by searching in a local area of the display.
4. I have to search through the display to find the information.

These descriptions serve as the choices for the given question for a metric. These descriptions forced respondents to make choices between relatively distinct facts rather than come up with their own complexity categories. In this case, each description acts like an anchor. One concern with this approach is whether the four anchors indeed correspond to the four complexity levels. We iteratively modified the descriptions until the descriptions matched to their intended levels. The details will be described later in the subsection "Modify the questionnaires." In total, we developed 13x4 descriptions for the 13 complexity questions in the multiple-choice questionnaire. The latest version of the questionnaire, along with instructions for use, is presented in Appendix A.

*Six-point Likert questionnaire.* We converted each metric into a question, then provided several statements to answer the question. A statement may describe the complexity from the perspectives of "not complex" (a positive statement) or "too complex" (a negative statement). This is to balance the responses to the questionnaire. For example, the metric of fixation groups was converted into the following question-statements, in which statement 1, 2, and 4 were positive, while statement 3 was negative. Notice that we did not attempt to balance the numbers of positive and negative statements.

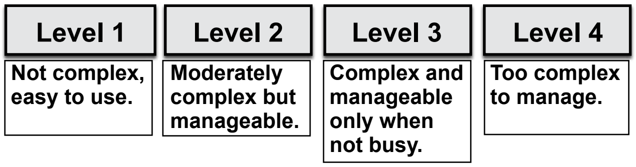


**Figure 1.** Four complexity levels defined in the multiple-choice questionnaire.

**How easily can you find the information you need on the display?**

1. I know where to look to find the information I need.
2. I can find the information I need without searching.
3. I have to search through the display to find the information I need.
4. I can find the information I need with one or a few quick glances.

For each statement, respondents selected one of six response alternatives "strongly agree," "agree," "somewhat agree," "somewhat disagree," "disagree," and "strongly disagree." Notice that each statement depicts one aspect of display design from the perspective of task performance. Therefore, these statements serve as guidelines for complexity management during display development. For instance, if most subjects' ratings for the statement 2 or 4 in the example above were "disagree/strongly disagree," then the display developers should reorganize or modify the display design to reduce the effort required to locate information.

We developed nine question-statement sets for each individual metric. The number of statements for each metric varied from three to six. We also developed four additional question-statement sets to assess the overall perceptual, cognitive, action complexity, and overall display complexity. In total, 13 questions were included in the questionnaires. Table 3 lists the questions corresponding to each metric and overall complexity. The complete questionnaire, along with instructions for its use, is presented in Appendix B.

The author worked with two experts in survey development and linguistics to create these statements and questions. They first reviewed the definitions of the metrics and the statements/questions in the multiple-choice questionnaire, then developed and iteratively revised the new statements/questions for the Likert questionnaire. Therefore, while the statements in Table 3 essentially describe the same information as those in Table 2 (for the multiple-choice questionnaire), the sentences and

**Table 3.** The questions in the Likert rating questionnaire

| Metric | Question |
|---|---|
| Number of fixation groups | How easily can you find the information you need on the display? |
| Variety of visual features | Does the variety of visual features (e.g., size, color, font, and icons) assist you in acquiring information? |
| Degree of clutter | How does display clutter affect reading text and icons? |
| Number of functional units | How does the amount of information provided on the display affect information management? |
| Dynamic complexity | How do information changes on the display affect the way you process information? |
| Relational complexity | Does the way in which information is presented affect your understanding of it? |
| Action cost | How does the action cost (such as keyboard strokes, mouse drags, and transitions between keyboard and mouse) affect you? |
| Action depth | How does the action depth (e.g., number of pull-down menus and/or pop-out windows you have to go through ) required for a task affect you? |
| Number of simultaneous action goals | How do action sequences needed to acquire information affect you? |
| Overall perceptual complexity | How would you evaluate the perceptual complexity of the display from the perspective of perceiving the information? |
| Overall cognitive complexity | How would you evaluate the cognitive complexity of the display from the perspective of understanding the information? |
| Overall action complexity | How would you evaluate action complexity of the display from the perspective of interacting with the display? |
| Overall display complexity | How would you evaluate the overall complexity of the display (from the perspectives of its effectiveness, efficiency, and safety)? |

wording in Table 3 are user-oriented; they can be easily understood without the background knowledge of human factors and display design.

**Modify the questionnaires**

We worked with three subject matter experts (SMEs) to modify the questionnaires. The SMEs were FAA academy training instructors. After being introduced to the purpose of the questionnaires, the SMEs reviewed the questionnaires and made constructive comments. One challenge in developing the four-level multiple-choice questionnaire was to ensure that the four anchors of a question correspond to the four given complexity levels. We iteratively worked on this with the three SMEs using the following procedure:

1) Researchers developed the initial four-level descriptions, A, B, C, and D for each question; intending to have A for level-1 complexity, B for level 2, C for level-3, and D for level-4;
2) SMEs mapped each description to one of the four complexity scales;
3) If a statement was not mapped to its pre-specified level, researchers discussed the issue with the SMEs and modified the description to more clearly relate to the intended complexity level;
4) Steps 2) and 3) were repeated until the descriptions were mapped to their expected complexity levels.

Finally, we asked two researchers who were professionals in ATC technologies and had experience in developing questionnaires for ATC studies to review and critique the questionnaires. We integrated their critiques into further modifications of the questionnaires.

We used a similar procedure to modify the six-point Likert questionnaire. The challenge in this questionnaire was whether each statement we provided was related to the question. Again, we had the three SMEs evaluate the statements and made modifications accordingly. We continued to process the questions until it was agreed that all the statements were related to their given questions, either positively or negatively. Finally, we collected critiques from several researchers who are professionals in ATC technologies, and we further modified the questionnaire according to their comments. After that, the questionnaires were ready to be tested with subjects.

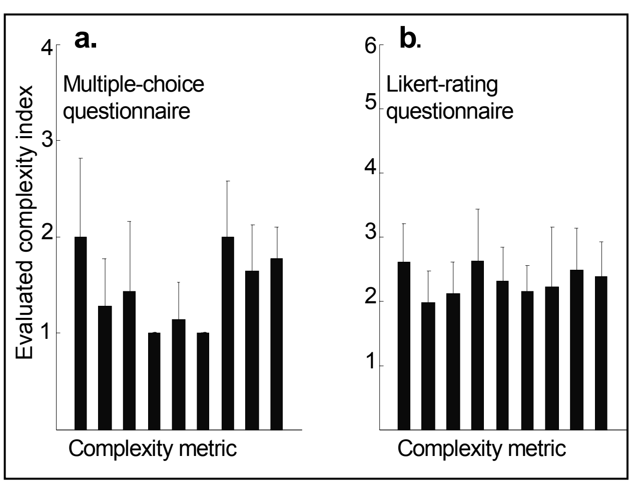


**Figure 2.** Complexity indices for STARS evaluated by individual metrics.

## TESTING THE QUESTIONNAIRES

**Methods**

We tested the questionnaires with seven FAA Academy instructors. The evaluation was made with regard to the Standard Terminal Automation Replacement System (STARS) displays. We first introduced the purpose of the study to the subjects and estimated their familiarity with STARS. The subjects then completed both complexity questionnaires and made critiques. Finally, we discussed with the subjects their responses to the questionnaires.

**Results**

*Complexity evaluation with individual metrics.* We first assessed information complexity using individual metrics. Figures 2a and b show the results produced with the multiple-choice questionnaire (referred to as QA) and the Likert rating questionnaire (referred to as QB). Along the horizontal axis of Figure 2a are the nine metrics in the following order (from left to right): *Number of fixation groups, Variety of visual features, Degree of clutter, Number of functional units, Dynamic complexity, Relational complexity, Action cost, Action depth,* and *Number of simultaneous action goals.* The vertical axis in Figure 2a indicates the four complexity levels in QA, from 1 "not complex" to 4 "too complex." The height of the bars indicates the evaluated metric indices averaged across all subjects; the error bars represent one standard deviation from the mean. The indices for most metrics in Figure 2a are close to level 2, corresponding to the complexity level "information is moderately complex but manageable." The standard deviations range from 0 to 0.79, suggesting that the evaluations was relatively consistent across subjects.

Figure 2b shows the results using QB. The bars along the horizontal axis represent the same metrics as in Figure 2a. The vertical axis of Figure 2b indicates the ratings on the Likert scales, from 1 "strongly agree" or "not complex" to 6 indicating "strongly disagree" or "too complex." For each metric in QB, the complexity index was calculated by averaging the ratings for positive statements and the reversed ratings for negative statements. The height of the bars indicates the assessment indices averaged across subjects; the error bars represent one standard deviation from the mean. The mean values of the metrics vary between 0.19 to 2.6, corresponding to "agree" or "somewhat agree" to the "not complex" statements (suggesting a positive response). The standard deviations range from 0.4 to 0.9, suggesting a relatively consistent rating across subjects.

*Overall complexity.* Both questionnaires contained four questions addressing the overall perceptual, cognitive, action, and display complexity. Figure 3a and b show the overall complexity indices for QA and QB, respectively. The four bars (from left to right) indicate the overall perceptual, cognitive, and action complexity, as well as the overall display complexity. In Figure 3a, the mean indices of perceptual, cognitive, action, and overall display complexity for all subjects are 1.93, 1, 1.14, and 2.21, respectively. Notice that the error bars in Figure 3 are in general larger than those in Figure 2, suggesting that the assessment of overall complexity is less consistent than that of the individual complexity metrics. One reason might be that the statements describing the overall complexity in QA and QB are less specific than those describing individual metrics. The former includes several factors to consider, and respondents may have only focused on some of the factors, or they may have used different rules to pull those factors together to make a single decision on the overall rating scales.

*Relationships between complexity of individual metrics and overall complexity.* While a questionnaire typically consists of multiple questions to assess system characteristics from different perspectives, users often desire a single measure to make their decision. A typical practice is to sum or average the responses to individual questions to generate a single judgment. However, such linear computations may not correspond to users' decision-making strategies. Since we collected users' responses to individual metrics and overall complexity, the data may conceptually elucidate the underlying decision rules. Thus, we studied how the evaluation of overall complexity related to that of the individual metrics.

To compare the overall complexity indices to individual metric ratings, we hypothesized about how participants might combine or integrate their responses to individual metrics together to generate a single number judgment. The most common rules describing how information is combined in the brain are averaging and winner-takes-all. Hence, we tested the averaging and winner-takes-all hypotheses.

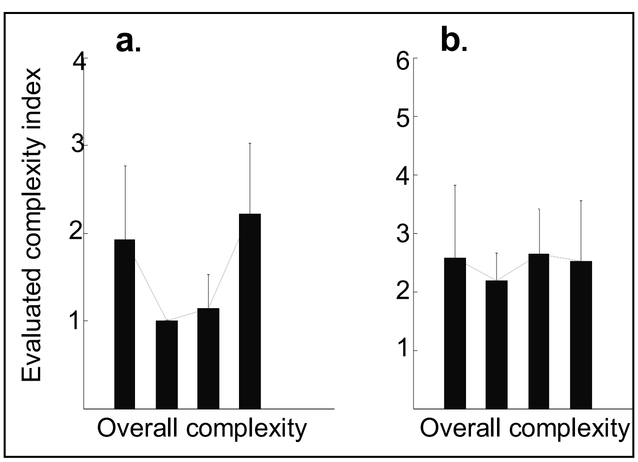


**Figure 3.** Indices of overall complexity for STARS.

For each subject, we calculated the overall complexity predicted by the winner-takes-all hypothesis as follows:

- Predicted perceptual complexity = the maximum of the indices of the number of fixation groups, number of visual features, and degree of clutter;
- Predicted cognitive complexity = the maximum of the indices of functional units, dynamic complexity, and relational complexity;
- Predicted action complexity = the maximum of the indices of action cost, action depth, and simultaneous goals;
- Predicted overall display complexity = the maximum of the indices of overall perceptual, cognitive, and action complexity from the original overall questions.

We then calculated the difference between each predicted and evaluated index. The overall results for QA and QB are illustrated in Figure 4a and b, respectively. The vertical axis indicates the difference between the predicted and evaluated index. Each circle represents the difference of one predicted-evaluated pair for a subject. The 28 circles along the horizontal axis are for all seven subjects. Most circles in Figure 4a clustered along the zero-difference line, while the circles in Figure 4b seem more variable. We calculated the least square error (LSE), which is the root of the summed square of the difference between predicted and evaluated values. The LSE is 0.12 for QA and 0.55 for QB, suggesting that the winner-takes-all hypothesis is the combination rule for QA but not QB. Another way to test the hypothesis is to calculate the correlation between the predicted and evaluated indices. The correlation coefficient for QA is 0.67, suggesting that the predicted and evaluated indices are positively correlated. On the other hand, the coefficient for QB is 0.37, suggesting a very weak correlation. Thus, the winner-takes-all hypothesis fits QA better than QB.

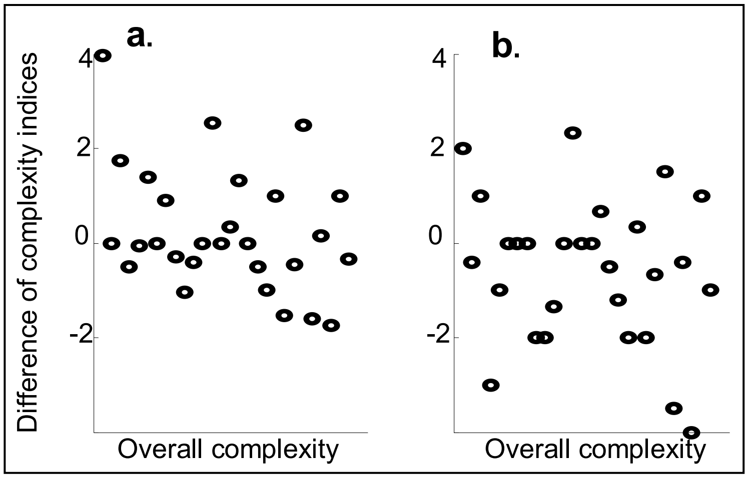


**Figure 4.** Differences between predicted and evaluated overall complexity with the winner-takes-all hypothesis.

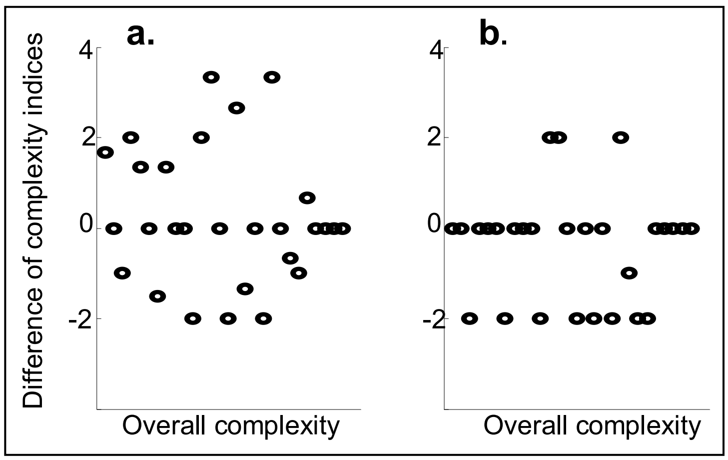


**Figure 5.** Differences between predicted and evaluated overall complexity with the averaging hypothesis.

We repeated the above procedure for the averaging hypothesis by applying a different decision rule – averaging the values of the responses to individual metrics. The results for QA and QB are illustrated in Figure 5a and b in the same format as that of Figure 4. Most circles for QB are clustered around the zero-difference line, and those for QA are more randomly distributed. The LSE is 0.44 for QA and 0.18 for QB, suggesting that the averaging hypothesis seems to fit QB better than QA. Similarly, the correlation coefficient for QA is 0.34, suggesting a very weak correlation between the predicted and evaluated ratings; while the coefficient for QB is 0.51, suggesting a moderate positive correlation. Thus, the averaging hypothesis fits QB better than QA.

This paradox is somewhat surprising because we expected that one hypothesis might work for both QA and QB. It is possible that the paradox was due to the small sample size. With a larger number of subjects, the data may validate one hypothesis and reject the other for both QA and QB. However, we cannot rule out the possibility that the paradox may reveal some intrinsic mechanisms of information processing. That is, subjects may indeed have used more than one rule to integrate the information of individual metrics. QA forced subjects to discriminate the four given complexity levels from the perspective of task performance. The winner-takes-all rule implies that if any one of the metrics reaches a higher complexity level, then the overall effect on task performance is severe. On the other hand, QB assesses a subject's opinions on specific aspects of display design from the perspective of complexity. The averaging hypothesis implies that every individual factor equally contributes to the overall judgment. Therefore, the results suggest that different rules might be used for different approaches to the evaluations.

*Compatibility between QA and QB.* Ideally, the two questionnaires would yield consistent evaluation results. However, since QA and QB have different formats and use different complexity scales, it is difficult to quantify the consistency between them. On the other hand, being compatible means that the evaluation results produced

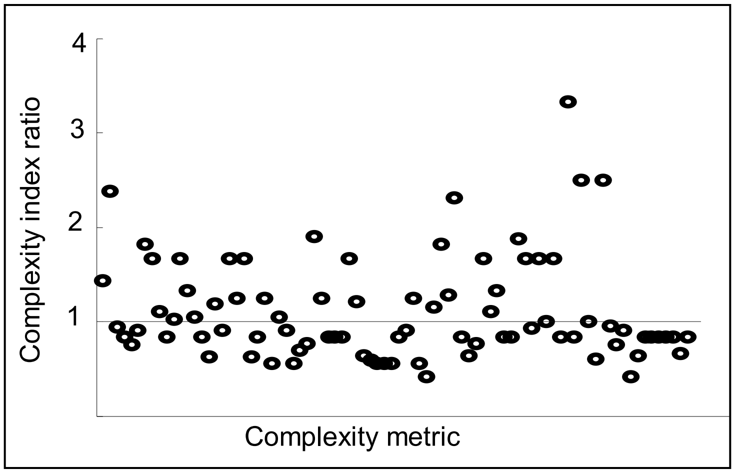


**Figure 6.** Complexity index ratios of the two questionnaires.

using QA are not contradictory with the results produced using QB, and vice versa. We examined the compatibility of QA and QB. For each subject, we calculated the evaluation ratio for every individual metric. The ratio was calculated as the complexity index for a given metric produced with QA divided by the corresponding complexity index produced with QB. Thus, we had 7x13 ratios. We plotted them in Figure 6, with each circle representing one ratio. The data were best fit at the ratio of 0.81, suggesting that the evaluations generated by QA and QB were compatible.

## DISCUSSION

This report describes the development and testing of two questionnaires to evaluate information complexity of ATC displays. We tested the questionnaires by having a small set of subjects evaluate the display complexity of STARS. The results indicated that STARS complexity was rated around the level of "information is moderately complex but manageable." The evaluation data demonstrated considerable consistency across subjects. The results also indicated that, while the two questionnaires are compatible with each other, the subjects may have used different strategies to combine the responses to individual questions in their decision-making.

Both questionnaires were based on information complexity metrics we developed earlier (Xing 2007). The first questionnaire employed a multiple-choice format in which subjects choose one of four levels of complexity for each complexity metric. The data collected with this questionnaire provided a relatively quantitative evaluation of display complexity. Moreover, the responses to the four complexity levels indicate whether the display is too complex to use. Thus, this questionnaire is most appropriate for assessing complexity control in acquisition evaluation of new ATC technologies. The second questionnaire employed a Likert rating format; subjects rated the statements about different aspects of design from the perspective of complexity. The data collected with this questionnaire provided a qualitative evaluation of display complexity. In addition, most statements in this questionnaire describe specific aspects of display design, so the rating of a specific statement can assist developers in managing complexity introduced by the factor described in the statement. Therefore, this questionnaire is better suited to complexity management during design and prototypes of new technologies. Users may choose to use one or both questionnaires as they need. The preliminary testing demonstrated that both questionnaires can be easily and quickly administered, yet provide reliable evaluation of information complexity in ATC displays.

Finally, we would like to emphasize that the main purpose of this report was to describe the questionnaires. We only present the preliminary test results of the questionnaires with a small set of subjects. Thus, the testing reported here was preliminary. Further evaluation is needed across displays with larger numbers of subjects to answer some of the issues raised in this study, such as the decision rules to integrate the assessments of individual dimensions of complexity. We also need to conduct a more complete evaluation of the questionnaires to assess their reliability and overall validity with larger numbers of subjects.

## REFERENCES


Kirakowski J, Corbett M (1993). SUMI: the software usability measurement inventory. British Journal of Education Technology, 24(3), 210–12.

Marios R, Ivanoff J (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, 9(6), 296-305.

Lewis JR (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7,* 57-78.

Schneiderman, B (1992). Designing the User Interface: Strategies for Effective Human-Computer Interaction, (2nd Ed.) Reading, MA: Addison-Wesley.

Willems B, Truitt TR (1999). Implications of reduced involvement in en route air traffic control. Federal Aviation Administration, William J. Hughes Technical Center; Report No: DOT/FAA/CT-TN99/22.

Willems B, Allen RC, Stein ES (1999). Air traffic control specialist visual scanning II: Task load, visual noise, and intrusions into controlled airspace. Federal Aviation Administration, William J. Hughes Technical Center; Report No: DOT/FAA/CT-TN99/23.

Xing J, Manning C (2005). Complexity and automation displays of air traffic control: Literature review and analysis. Washington DC: Federal Aviation Administration, Office of Aerospace Medicine; Report No: DOT/FAA/AM-05/4.

Xing J (2006). Color and visual factors in ATC displays. Washington DC: Federal Aviation Administration, Office of Aerospace Medicine; Report: No. DOT/FAA/AM-06/15.

Xing J (2007). Information Complexity in ATC displays. Washington DC: Federal Aviation Administration, Office of Aerospace Medicine; Report: No. DOT/FAA/AM-07/26.

# APPENDIX A

## A multiple-choice questionnaire to evaluate information complexity of ATC displays

Name of the display you are evaluating ______________
How long have you been using this display? ____________

**Instructions:**

1) The purpose of this study is to evaluate information complexity of the given ATC automation display by completing the questionnaire..
2) The questionnaire consists of 13 questions, each assessing a specific aspect of display complexity. For each question, we have provided you with four statements. Please circle the statement that best describes the complexity of this display.
3) The term "information" in the questionnaire refers to either displayed materials (texts, symbols, etc.) or control functions (action buttons, menus, etc) for users to acquire information.

## Beginning of the questions

**1. How easy is it for you to find information on the display?**

A. I can see the information effortlessly.
B. I can find the information with a few quick glances.
C. I can find the information by searching in a local area of the display.
D. I have to search through the display to find the information.

**2. How well is the information organized on the display?**

A. Information organization is obvious by its visual features (colors, symbols, fonts, graphic patterns, etc); I know how the information is organized at a glance.
B. The organization of information is not obvious by its visual features; I have to spend some effort to figure out how the information is organized.
C. The organization of information is confusing; I have to work hard to figure out how the information is organized.
D. The display has too many visual features (colors, symbols, fonts, graphic patterns, etc) for me to recognize how information is organized.

**3. How easy is it for you to read the displayed text?**

A. Texts and icons stand out clearly from the background; I can read them correctly with a quick glance.
B. Texts and icons can be read easily but the clutter still slows down my reading.
C. Text and icons are cluttered and I have to spend some effort to read them (such as moving closer to the screen or staring at them for a longer time).
D. The display has too much clutter; it is hard for me to read the text quickly and correctly.

**4. How well are you aware of the information provided by the display?**

A. There are only several chunks of information that I need to be aware of in order to use the display. I am aware of the information most of the time.
B. I can manage all the needed information but feel that managing information takes my mental resources away from doing my tasks.
C. I can manage all the displayed information only by fully concentrating on the display, but have difficulties to do so when I have other things in mind.
D. The display provides too many pieces of information for me to be aware of; I cannot mentally build a fixed mental model of the display.

**5. How do you like the dynamic changes of the displayed information?**

A. The display does not present dynamic information or most changes are expected and predictable.
B. I can take care of changes but prefer that the display present information more statically.
C. I have to frequently update my mental model due to the unpredicted changes of displayed information.
D. The displayed information changes too frequently in an unpredictable manner; I have a hard time catching up with the changes.

**6. How easy is it for you to understand /comprehend the displayed information?**

A. The information is very straightforward. I can understand the meaning without thinking.
B. I can integrate the pieces of information and use them properly, but prefer that information be presented in less intermingled manner.
C. I need to use some strategies to manage the displayed information. That takes my mental resources away from my tasks.
D. I have to simultaneously associate (or to relate) multiple pieces of displayed information to use the display. It is difficult to hold them all at once.

**7. How would you evaluate the amount of actions you need to take to perform tasks or acquire information?**

A. It takes only one or a few simple actions to perform tasks or acquire information; the actions can be done nearly subconsciously.
B. It takes me some actions to perform tasks or acquire information, but the amount of actions is manageable.
C. Many actions are needed to perform tasks or acquire information.
D. It takes too many actions (keystrokes, mouse drag/ clicks, etc) to perform tasks or acquire information.

**8. How do you rate the number of action steps needed to perform tasks or acquire information?**

A. It takes one or two steps to perform tasks or acquire information; I can perform them almost automatically without thinking about the steps.
B. I can remember the steps but that distracts me.
C. It takes several steps to perform tasks or acquire information; performing those steps makes navigation difficult.
D. It takes multiple steps to perform tasks or acquire information. I have a hard time remembering all those steps.

**9. How do you rate the number of action sequences needed to perform tasks or acquire information?**

A. Only one sequence of action steps is needed to perform tasks or acquire information; I can perform the action sequence easily and reliably.
B. I can manage the multiple sequences of actions required to perform tasks or acquire information; but that increases task difficulties.
C. I am confused with the action steps in different sequences when I do not fully concentrate on the sequences.
D. It takes too many sequences of steps to perform tasks or acquire information. I have a hard time managing the sequences.

**10. How do you rate the perceptual complexity of the display?**

A. The display looks simple and clear; I can find the needed information easily and quickly.
B. The display looks busy but I can find the information with a little effort.
C. Many pieces of information do not always relate to my tasks; they adversely affect my perception of information.
D. The display looks too busy for me to find the information.

**11. How cognitively demanding is the displayed information?**

A. The information is presented straightforwardly; I can manage all the needed information quickly and correctly.
B. Information is complex but I can manage to use it by focusing on my own tasks.
C. Using this display takes too much attention and disturbs my decision-making in performing my tasks.
D. The information is too overwhelming; it is difficult to interpret the information quickly and correctly.

**12. How easy is it for you to interact with the display?**

A. The display demands very few actions from me.
B. The display is usable but it demands some undesired interactions.
C. The display demands lots of interactions to perform my tasks.
D. The display is too difficult to use. It requires me to do too many things.

**13. Overall, how do you rate the complexity of the display in terms of its usefulness (efficiency, effectiveness, and safety)?**

A. The display is very simple to use and I am fully satisfied with it.
B. The display is moderately complex and I might choose to use it when I need the service.
C. The display is complex and I will use it only when I have to.
D. The display is too complex to use.

# APPENDIX B

## A Likert rating questionnaire to evaluate information complexity of ATC displays

**Instructions:** This questionnaire asks you to respond to items designed to measure a specific aspect of a display. When answering an item, think about the lead-in question and indicate your response by darkening the bubble corresponding to your answer. If you change your response, please make sure your final choice is clear. If the response options do not provide a perfect fit for your unique situation, use your best judgment.

| | | Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|---|---|
| **How easily can you find the information you need on the display?** | | | | | | | |
| 1 | I know where to look for the information I need. | O | O | O | O | O | O |
| 2 | I can see the information I need without searching. | O | O | O | O | O | O |
| 3 | I have to search through the display to find the information I need. | O | O | O | O | O | O |
| 4 | I can find the information I need with one or a few quick glances. | O | O | O | O | O | O |
| **Does the variety of visual features (e.g., size, color, font, and icons) assist you in acquiring information?** | | | | | | | |
| 5 | The variety of visual features, such as size, color, font, and icons, assists me in acquiring the information on the display. | O | O | O | O | O | O |
| 6 | The variety of visual features on the display is confusing. | O | O | O | O | O | O |
| 7 | The display uses too many different sizes, colors, fonts, and icons. | O | O | O | O | O | O |
| 8 | I can see information better if I ignore some of the colors, fonts, and text formats. | O | O | O | O | O | O |
| **How does the display clutter affect reading text and icons?** | | | | | | | |
| 9 | The display looks too busy. | O | O | O | O | O | O |
| 10 | The text and icons stand out clearly from the background. | O | O | O | O | O | O |
| 11 | I have to move closer to the screen to read the text. | O | O | O | O | O | O |
| 12 | I have to stare at the display for a while to read the information. | O | O | O | O | O | O |
| 13 | I can read displayed text or detect icons on a glance. | O | O | O | O | O | O |
| 14 | Adequate spaces between text/icons are provided for on-a-glance reading/detection. | O | O | O | O | O | O |
| **How does the amount of information provided on the display affect information management?** | | | | | | | |
| 15 | It is difficult to manage all the necessary information. | O | O | O | O | O | O |
| 16 | I can manage the displayed information effortlessly. | O | O | O | O | O | O |
| 17 | There is too much information on the display for me to be aware of them. | O | O | O | O | O | O |

| | | Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|---|---|
| **How do information changes on the display affect the way you process information?** | | | | | | | |
| 18 | Most of information changes on the display are predictable. | ○ | ○ | ○ | ○ | ○ | ○ |
| 19 | Most of information changes on the display are easy to track. | ○ | ○ | ○ | ○ | ○ | ○ |
| 20 | Keeping track of information changes on the display distracts me from performing my primary tasks (makes me too busy). | ○ | ○ | ○ | ○ | ○ | ○ |
| 21 | Information changes are too frequent for me to keep up with. | ○ | ○ | ○ | ○ | ○ | ○ |
| 22 | The displayed information should change less frequently. | ○ | ○ | ○ | ○ | ○ | ○ |
| **Does the way in which information is presented affect your understanding of that information?** | | | | | | | |
| 23 | Interpreting information distracts me from focusing on my tasks. | ○ | ○ | ○ | ○ | ○ | ○ |
| 24 | I can use the displayed information without relating it to other pieces of information. . | ○ | ○ | ○ | ○ | ○ | ○ |
| 25 | I have to relate several pieces of separately displayed information to use them. | ○ | ○ | ○ | ○ | ○ | ○ |
| 26 | The information presented is straightforward. | ○ | ○ | ○ | ○ | ○ | ○ |
| **How does the amount of action required to perform tasks or acquire information, such as keyboard strokes or mouse drags affect you?** | | | | | | | |
| 27 | The display requires too many actions to perform tasks or acquire information. | ○ | ○ | ○ | ○ | ○ | ○ |
| 28 | The actions required by the display take my attention away from my tasks. | ○ | ○ | ○ | ○ | ○ | ○ |
| 29 | The amount of action required to perform tasks or acquire information does not bother me. | ○ | ○ | ○ | ○ | ○ | ○ |
| 30 | I feel overwhelmed by the amount of interaction required by the display. | ○ | ○ | ○ | ○ | ○ | ○ |

| | | Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|---|---|
| | **How does the action depth (e.g., number of pull-down menus and/or pop-out windows you have to go through) for a given task affect you?** | | | | | | |
| 31 | I have to access too many menu buttons or windows to acquire information/perform a specific task. | ○ | ○ | ○ | ○ | ○ | ○ |
| 32 | I can effortlessly follow the links of pop-out windows and/or pull-down menus to acquire information/perform tasks. | ○ | ○ | ○ | ○ | ○ | ○ |
| 33 | I have trouble getting the information and performing tasks because there are so many layers of windows/menus. | ○ | ○ | ○ | ○ | ○ | ○ |
| | **How does the number of action sequences required to perform tasks or acquire information affect you?** | | | | | | |
| 34 | I have to manage more than one action sequences to get a task done. I have a hard time keeping up with them. | ○ | ○ | ○ | ○ | ○ | ○ |
| 35 | I can perform most tasks by following a single action sequence. | ○ | ○ | ○ | ○ | ○ | ○ |
| | I might confuse or forget the choices of the actions needed to complete a task when I am busy. | | | | | | |
| | **How would you evaluate the overall complexity of the display (from the perspective of perceiving the displayed information)?** | | | | | | |
| 36 | The display is an effective tool for acquiring information. | ○ | ○ | ○ | ○ | ○ | ○ |
| 37 | The display is simple and easy to use. | ○ | ○ | ○ | ○ | ○ | ○ |
| 38 | I do not like the display because it is too complex to use. | ○ | ○ | ○ | ○ | ○ | ○ |
| | **How would you evaluate the perceptual complexity of the display (from the perspective of perceiving information)?** | | | | | | |
| 39 | Only necessary information is presented on the display. | ○ | ○ | ○ | ○ | ○ | ○ |
| 40 | I can easily and quickly find the information I need. | ○ | ○ | ○ | ○ | ○ | ○ |
| 41 | I don't like the display because it appears to have too much stuff. | ○ | ○ | ○ | ○ | ○ | ○ |
| 42 | I could not find the information I need because the display looks too busy. | ○ | ○ | ○ | ○ | ○ | ○ |
| | **How would you evaluate the cognitive complexity of the display (from the perspective of understanding information)?** | | | | | | |
| 43 | I can effortlessly understand the information presented on the display. | ○ | ○ | ○ | ○ | ○ | ○ |
| 44 | Using this display takes too much mental effort. | ○ | ○ | ○ | ○ | ○ | ○ |
| 45 | I feel overwhelmed by the amount of information presented on the display. | ○ | ○ | ○ | ○ | ○ | ○ |
| | **How would you evaluate the action complexity of the display (from the perspective of interacting with the display)?** | | | | | | |
| 46 | I can easily interact with the display to accomplish my tasks. | ○ | ○ | ○ | ○ | ○ | ○ |
| 47 | I can get confused or even lost by the actions required to accomplish my tasks. | ○ | ○ | ○ | ○ | ○ | ○ |
| 48 | I feel overwhelmed by the amount of interaction required by the display. | ○ | ○ | ○ | ○ | ○ | ○ |