Federal Aviation Administration

DOT/FAA/AM-09/14
Office of Aerospace Medicine
Washington, DC 20591

Validating Information Complexity Questionnaires Using Travel Web Sites

Chen Ling[1]
Miguel Lopez[1]
Jing Xing[2]

[1]School of Industrial Engineering
University of Oklahoma
Norman, OK 73019

[2]FAA Civil Aerospace Medical Institute
P.O. Box 25082
Oklahoma City, OK 73125

July 2009

Final Report

OK-09-0434-JAH

NOTICE

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site:
www.faa.gov/library/reports/medical/oamtechreports

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-09/14		

4. Title and Subtitle	5. Report Date
Validating Information Complexity Questionnaires Using Travel Web Sites	July 2009
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Ling C[1], Lopez M[1], Xing J[2]	

9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)
[1]University of Oklahoma School of Industrial Engineering Norman, OK 73019	[2]FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.

12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved project number HRRD522

16. Abstract

With the prevalent use of visual interfaces and the increasing demand to display more information, information complexity in human-computer interfaces becomes a major concern for technology designers. Complex interfaces affect the effectiveness, efficiency, and even the operational safety of a system. Previously, researchers at the Federal Aviation Administration developed two questionnaires to evaluate information complexity of air traffic control displays. This study adapted the questionnaires for commercial computer interfaces and validated them with two types of tasks on three travel Web sites. The results demonstrated that both questionnaires had acceptable reliability, validity, and sensitivity.

17. Key Words	18. Distribution Statement
Interface Evaluation, Usability, Information Complexity, Assessment, Air Traffic Control Displays	Document is available to the public through the Defense Technical Information Center, Ft. Belvoir, 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	20	

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

ACKNOWLEDGMENT

This research was supported by the Federal Aviation Administration Civil Aerospace Medical Institute, Oklahoma City, Oklahoma, with a grant entitled "Investigating Information Complexity in Three Types of Air Traffic Control (ATC) Displays," grant number FAA 06-G-013.

CONTENTS

VALIDATING INFORMATION COMPLEXITY QUESTIONNAIRES USING TRAVEL WEB SITES

INTRODUCTION

With the prevalent use of interactive visual interfaces and increasing demand to display more information, interface complexity became a major concern for users and designers (Maeda, 2006). Complex interfaces may affect system effectiveness, efficiency, and even operational safety. Understanding information complexity and measuring it properly are important for interface design. The objective of the current study was to validate two questionnaires evaluating information complexity.

Complexity Questionnaires

Xing (2004, 2007) proposed a framework to measure information complexity and developed a set of complexity metrics for air traffic control (ATC) displays. According to the framework, information complexity can be assessed by measuring three dimensions: quantity of basic information elements, variety of those elements, and the relationship between them. Each dimension is evaluated by the resource demands for three stages of human information processing: perception, cognition, and action. Each dimension can therefore be described by three *constructs* at each stage of information processing, as shown in Table 1. A combination of all nine constructs gives rise to the complexity of a display.

Based on the complexity metrics, Xing (2008) developed two questionnaires to evaluate information complexity for air traffic control displays. The first questionnaire (referred to as *QA*) has 13 questions, nine corresponding to the complexity constructs described earlier, three corresponding to overall perceptual, cognitive, and action complexity, and one corresponding to the overall complexity of the interface being evaluated. Each question is provided with four statements describing different levels of complexity, ranging from *not complex* to *too complex*. These statements are used as multiple-choice answers to the question. Participants choose one statement that best describes their understanding of the interface being evaluated.

The second questionnaire (referred to as *QB*) contains 13 questions, each accompanied by three to six statements. To eliminate response bias, some statements imply positive answers (corresponding to *not complex*) to the question, whereas others are negative (corresponding to *too complex*). Participants are asked to rate the degree of agreement to every statement using a six-point Likert scale, ranging from *strongly disagree* to *strongly agree*.

These two questionnaires have been tested preliminarily with a small set of subjects but need to be validated with a larger population. Due to the cost and operational difficulties in recruiting a large number of air traffic controllers, this study used college students to validate the questionnaires. Since several questionnaires that evaluate Web site usability have been well validated, we chose several commercial Web sites for this validation study. We slightly adapted the two complexity questionnaires by modifying the words to fit the Web site applications and leaving the overall structure of the questionnaires unchanged.

In this report, we focused the validation of the complexity questionnaires on three aspects: reliability, construct validity, and sensitivity. These are generic aspects regarding the quality of a questionnaire, and they can be generalized to different user populations and different types of displays. On the other hand, since we used naïve

Table 1. Information Complexity Framework and Metrics

	Perception	Cognition	Action
Quantity	No. of fixation groups	No. of functional units	Cost of action (key strokes, mouse movement, etc.)
Variety	No. of distinctive visual features	Unpredictable dynamic changes	Action depth (No. of selective action steps)
Relation	Degree of clutter (Text readability)	No. of variables to be related	No. of Simultaneous action goals

participants examining commercial Web sites, we could not in the current study address the specificity of the questionnaires to professional controllers who examine ATC displays.

Usability Questionnaire

A concept related to complexity is *usability*. Usability is defined in ISO 9241 as "the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments." Many questionnaires have been developed for interface usability evaluation. Among the most widely used ones is the Post Study System Usability Questionnaire (PSSUQ) for scenario-based usability evaluation (Lewis, 1995). It evaluates three dimensions of usability: system usefulness, information quality, and interface quality. An overall usability score can be derived by averaging the answers to all the items. PSSUQ has been validated and demonstrated high reliability, with an overall Cronbach alpha of 0.97. Its construct validity has been established through factor analysis. After ten years of use, the PSSUQ is considered reliable and valid (Lewis, 2002).

While a few questions in PSSUQ capture some aspects of complexity, it does not systematically evaluate information complexity. Neither does it yield much information about the underlying structure of complexity. On the other hand, Xing's complexity questionnaires (Xing, 2008) are diagnostic in terms of information complexity issues. They elucidate the underlying complexity structure of a visual display because they were based on a well-structured framework. The two complexity questionnaires could be valuable additions to the usability community. Therefore, another purpose of this study was to adapt Xing's questionnaires for commercial interfaces, validate them, and compare them to the established usability questionnaire PSSUQ.

METHODOLOGY

Psychometric Theories in Questionnaire Validation

To validate the complexity questionnaires QA and QB, several criterions of the psychometric instruments (including reliability, validity, and sensitivity) needed to be established (Nunnally, 1978).

Reliability

Reliability refers to the "consistency," or "repeatability" of the measures. The most common way to establish reliability for summative scales is with internal consistency by calculating Cronbach alpha coefficients (Nunnally, 1978). For the instrument to be considered reliable, the alpha coefficient should be at least 0.70 (Nunnally, 1978). For both QA and QB, the complexity dimensions of quantity, variety, and relation contribute to the perceptual, cognitive, and action complexity. The internal consistency among these dimensions needs to be established for each type of complexity. Because QB uses summated scales from several items for each complexity construct, we can also calculate the internal consistency of each complexity construct for QB.

Validity

A questionnaire's validity refers to the extent to which it measures what it claims to measure (Nunnally, 1978). Researchers usually use the Pearson correlation coefficient to assess criterion-related validity. The correlation is computed between the measure of interest and other concurrent or predictive measures. The correlation coefficient does not have to be large to provide evidence of validity. A value as small as 0.30 to 0.40 is large enough to justify the use of psychometric instruments (Nunnally, 1978). If the predictor and the criterion measure occur at the same time, then the calculated validity is considered concurrent. On the other hand, if the predictor precedes the criterion measure, then the calculated validity is considered predictive. The validity of the QA and QB can be investigated through their correlation with the usability questionnaire PSSUQ results. QA and QB can also be cross-validated by finding correlations between their ratings.

Sensitivity

The sensitivity of the questionnaire is concerned with the question: "Are the questionnaires sensitive to experimental manipulation?" (Nunnally, 1978). A sensitive questionnaire is able to capture differences resulting from experimental manipulations. ANOVA is commonly used on questionnaire responses in different experimental manipulations to establish a questionnaire's sensitivity.

Participants

Fifty-one university students (18 females and 33 males) participated in the study. The average age of the participants was 22.8. All are experienced users of computers and commercial Web sites.

Apparatus

Microsoft Internet Explorer® version 6.0 was used as the Web browser. Three Web sites were studied: www.expedia.com, www.travelocity.com, and www.orbitz.com. The task performance time was recorded with a Casio stopwatch (continuous) and an Ipod Nano® (in stopwatch mode). Three sets of questionnaires were used, including two complexity questionnaires (adapted from Xing's complexity questionnaires QA and QB) and the usability questionnaire PSSUQ (Lewis, 1995).

Tasks

Before the experiment, we conducted a task analysis of each Web site to identify the experimental tasks. The experiment consisted of two types of tasks: *directed* and *exploratory* tasks. Users performed the directed tasks using the standard toolbox on the top of each Web site's homepage. The toolbox helped users find the optimal results. The experiment employed three directed tasks: 1) buy airline tickets for two adults and two children from Dallas, TX, to Yellowstone National Park on particular dates; 2) buy an airline ticket for one person from Oklahoma City, OK, to Chicago, IL, on particular dates; 3) buy cruise tickets for two adults for the *Western Caribbean Sea,* which sails from Miami, FL, for seven days, requiring ocean view rooms and to cost no more than $900 per person.

In the exploratory tasks, participants were asked not to use the standard toolbox but to search through the multi-layered Web site structure to accomplish the task goals. The experiment employed three exploratory tasks: 1) buy the cheapest trip to Paris, France, for four nights for one adult and one child from Miami, FL; 2) plan a seven-day honeymoon trip from San Francisco, CA, to Hawaii with a $5,000 budget for a couple in mid-August of next year; 3) find the best travel deal to go to Las Vegas this weekend.

Many computer interfaces support these two types of tasks. Users performed directed tasks more frequently than exploratory ones. Because exploratory tasks demand more mental effort, the complexity ratings for them are expected to be higher. Therefore, we used both types of tasks to test the sensitivity of the two complexity questionnaires. A questionnaire with enough sensitivity should be able to capture the differences in complexity between these two types of tasks.

Experimental Procedure

We first adapted Xing's complexity questionnaires QA and QB to fit the commercial Web sites. We modified the wording of the questionnaires based on inputs from a language professor and several subject matter experts. The original questionnaires can be found in Xing (2008). The adapted questionnaires are shown in Appendixes A and B.

Each participant filled out a consent form and a demographic survey at the beginning of the experiment. Based on the reported familiarity with the three Web sites, the participants were assigned to use the Web site with which they were least familiar. None of the participants was familiar with all three Web sites. This was to reduce the effect of prior experience. As the result, 18 participants used Expedia, 15 used Travelocity, and 18 used Orbitz. We presented the participants with the tasks and explained the purpose of the experiment, emphasizing the importance of carefully completing the questionnaires. The participants first performed the three directed tasks. The order of the tasks was randomly assigned. Next, they were asked to complete the two complexity questionnaires and PSSUQ usability questionnaire. Subsequently, the participants performed the three exploratory tasks and filled out the three questionnaires again. The orders of these three questionnaires were counterbalanced to eliminate potential ordering effects. Throughout the experiment, we took notes on task performance and recorded comments from the participants. The total time to complete the experiment was about one and one-half hours.

RESULTS

Standardized Complexity and Usability Ratings

Because each participant completed the questionnaires twice, we had 102 sets of responses in total from 51 participants. The answers from QA ranged from 1 to 4, with 1 representing the complexity level "*not complex and easy to use;*" 2 for "*moderate complexity but manageable*;" 3 for "*highly complex and hard to manage*;" and 4 for "*too complex to manage*."

In QB, multiple statements were used for the same construct, corresponding to either *"not complex"* or *"too complex."* We transformed all the answers so that higher indices implied higher complexity levels. The indices ranged from one to six, as we used a six-point Likert scale.

The usability questionnaire PSSUQ (Lewis, 1995) is composed of 19 questions. The overall usability rating was derived by averaging the responses to the 19 questions. A seven-point Likert scale was used to derive the degree of agreement to the questions. A higher number indicated lower usability.

We standardized the average complexity ratings of all items from QA, QB, and PSSUQ to range between 0 to 1 for quantitative comparison. The standardized ratings were derived by dividing the rating received from the questionnaires by the number of maximum points in the scales used, which was four for QA, six for questionnaire B, and seven for PSSUQ. Shown in Figure 1 are the standardized overall complexity rating from QA and QB and usability rating from PSSUQ for the directed and exploratory tasks. After standardizing, all the ratings were closely comparable to each other.

Reliability

Reliability of the questionnaire was established by answering the question: "Are the measured indices from different questionnaires consistent?" (Nunnally, 1978).

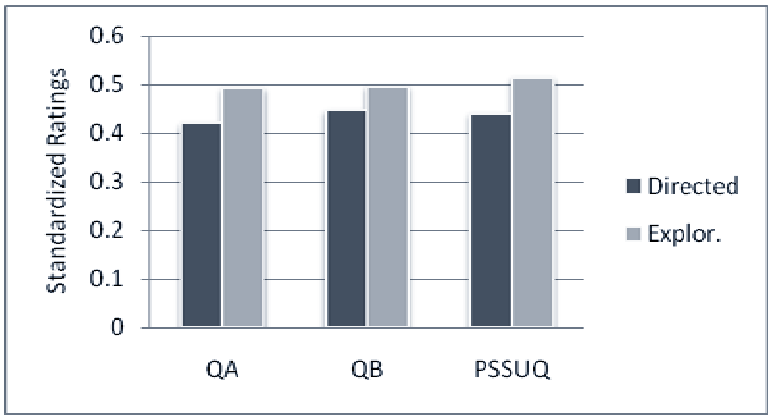

Figure 1. Standardized Complexity and Usability Ratings for Three Questionnaires

The internal consistency of QA and QB was calculated to ensure all statements measuring the same construct were internally consistent.

Internal Consistency of QA

In QA, three dimensions of the complexity — quantity, variety, relation—contributed to the perceptual, cognitive, and action complexity. The ratings on these sub-dimensions associated with the same complexity construct were expected to be consistent. The internal consistency was calculated for the perceptual, cognitive, and action complexity scales, respectively, by computing the Cronbach alpha coefficients. The internal consistency was 0.71 for perceptual complexity (based on questions 2, 3, and 4), 0.71 for cognitive complexity (based on questions 6, 7, and 8) and 0.79 for action complexity (based on questions 10, 11, and 12). These values are acceptable and demonstrate the reliability of QA.

Internal Consistency of QB

In QB, multiple statements were associated with each complexity scale. The responses to the statements associated with the same construct were expected to be consistent. The internal consistency indices among these statements were calculated for every question in QB, and the resultant Cronbach alpha values are shown in Table 2. During the calculation, three statements in QB showed negative correlation with other relevant statements. These statements were referred to as "inappropriate." Cronbach alpha values before and after the removal of the "inappropriate" statements are listed in Table 2. The Cronbach alpha values associated with the "inappropriate" statements were lower (shown in bold in Table 2). However, after removal of those statements, all the resultant Cronbach alpha values were equal or above 0.7, which is considered acceptable for internal consistency.

We further studied the three "inappropriate" statements. One of them was associated with the perceptual-variety construct. The statement was "I can see information better if I ignore some of the colors, fonts, and text formats." Participants felt that they could not comprehend the sentence. The "inappropriate" statement associated with the overall cognitive complexity was "Using this Web site takes moderate mental efforts." The statement associated with the overall action complexity was, "I can interact with the Web site to accomplish my

Table 2. Internal Consistency for All Complexity Constructs in QB

Complexity Constructs		Cronbach alpha	
		Before Removal of Inappropriate Statement	After Removal of Inappropriate Statement
Perception	Overall	0.791	0.791
	Quantity	0.826	0.826
	Variety	**0.411**	0.836
	Relation	0.817	0.817
Cognition	Overall	**0.350**	0.727
	Quantity	0.860	0.860
	Variety	0.872	0.872
	Relation	0.858	0.858
Action	Overall	**0.616**	0.810
	Quantity	0.798	0.798
	Variety	0.738	0.738
	Relation	0.750	0.750
Grand Overall		0.894	0.894

tasks but with some effort." Both statements described a mediocre level of complexity instead of "not complex" or "too complex." The degree of agreement with those statements was different from strictly positive or negative statements. Because there were other positive and negative statements measuring the same construct, these three "inappropriate" statements were removed from QB without affecting the overall effectiveness of the questionnaire. The improved version of QB had 51 statements, as listed in Appendix B. Please note that all the statistics reported hereafter were based on the responses without the "inappropriate" statements.

After removing the three "inappropriate" statements, another internal consistency analysis on QB was performed on the scores of individual complexity constructs derived by summated item responses for overall perceptual, cognitive, and action complexity. This procedure was similar to that performed for QA. The Cronbach alpha values for perceptual, cognitive, and action complexities were 0.82, 0.83, and 0.72, respectively, which were large enough to demonstrate reliability.

Validity

The validity of the questionnaire was established by answering the question: "Does the instrument measure the intended attribute?" (Nunnally, 1978). The concurrent validity and construct validity of the QA and QB were investigated, as described below. Because complexity was believed to be related to usability, the responses to the PSSUQ were used as a criterion measure for the concurrent validity calculation of QA and QB.

Concurrent Validity With PSSUQ

The complexity ratings measured by QA and QB were correlated with the overall usability value of the PSSUQ usability questionnaire (Lewis, 1995). The results indicated that usability and complexity were negatively correlated with each other. Web sites with lower complexity were easier to deal with and therefore considered to have higher usability. The correlation coefficients of the nine complexity constructs and the overall complexity ratings are shown in Table 3. All correlation relationships, except those bolded in Table 3, were significant. In general, the coefficients were higher for QB than QA. If we use a range of 0.3 to 0.4 as the criterion for acceptable validity (Nunnally, 1978), we can see that nine of the 13 constructs for QA and ten for QB had acceptable validity. For both QA and QB, the overall measure of complexity was moderately correlated with the overall usability rating (r_A=-0.45, r_B=-0.47), which demonstrates reasonably good evidence of validity.

Concurrent Validity Between QA and QB

Because both QA and QB measured the same complexity constructs, establishing the correlation between these two questionnaires could provide evidence for the purpose of cross-validation. For each complexity construct, the correlation coefficient and the p-value were computed (as shown in Table 3). All correlation coefficients were

Table 3. Pearson Correlation Coefficients among QA, QB, and PSSUQ

Complexity Constructs		Btw. QA and PSSUQ	Btw. QB and PSSUQ	Btw. QA and QB
Perception	Overall	-0.26	-0.45	0.55
	Quantity	-0.32	-0.42	0.68
	Variety	-0.32	-0.37	0.49
	Relation	-0.4	-0.3	0.61
Cognition	Overall	-0.2	-0.33	0.42
	Quantity	-0.31	-0.27	0.6
	Variety	**-0.18**	-0.29	0.58
	Relation	-0.27	-0.43	0.5
Action	Overall	-0.42	-0.41	0.6
	Quantity	-0.27	**-0.19**	**0.19**
	Variety	-0.31	-0.39	0.26
	Relation	-0.3	-0.38	0.52
Grand Overall		-0.45	-0.47	0.61
Average Coefficient		-0.31	-0.36	0.51

positive. All correlations were significant except for the action-quantity construct. Eleven of 13 complexity constructs had correlation coefficients larger than 0.30. The average correlation coefficient of the 13 pairs was 0.51, which provided evidence for the similarity between QA and QB. But because neither questionnaire has been validated, similarity between them is not sufficient to prove their validity.

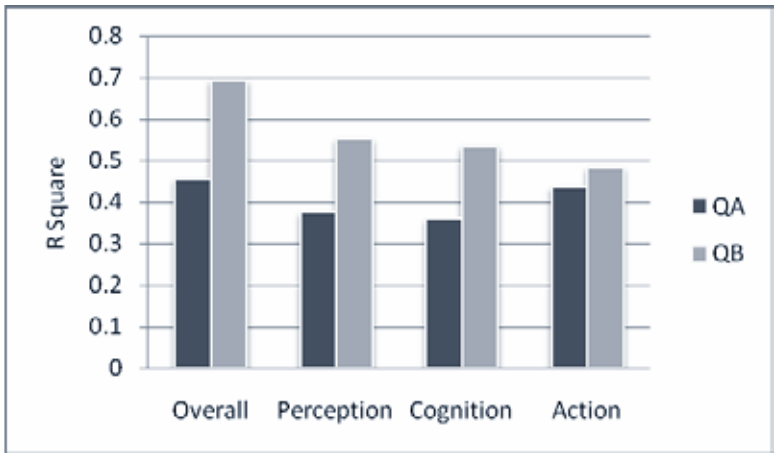

Figure 2. R Square Derived from Multiple Regression for Construct Validity

Construct Validity

We performed multiple regression analyses on responses from both QA and QB to understand how their constructs (see Table 2) were related to each other and, in particular, how the individual complexity constructs contributed to the overall complexity rating. Regression analysis for QA revealed that the perception, cognition, and action complexity together accounted for 46% of the variance in the overall complexity (R^2=0.46). And the quantity, variety, and relation dimensions accounted for 38% of the variance in perceptual complexity, 36% in cognitive complexity, and 44% in action complexity.

Multiple regression analyses for QB revealed that the perception, cognition, and action complexity together accounted for 69% of the variance in the overall complexity (R^2=0.69). Furthermore, the quantity, variety, and relation dimensions accounted for 55% of the variance in perceptual complexity, 53% in cognitive complexity, and 48% in action complexity.

These percentages are shown in Figure 2. The higher adjusted R-square values derived from the responses to QB suggested that ratings of multiple statements for each complexity construct provided a broader coverage of complexity issues than just a single statement for each construct, as in QA.

A common approach for establishing construct validity is through confirmatory factor analysis (Thompson & Daniel, 1996). For a factor analysis to derive reliable factor loadings, the rule of thumb was having at least five responses for each item (Nunnally, 1978). Because QB has 51 items, we needed to obtain at least 255 data points to perform a factor analysis. The available 102 data points were not sufficient for performing a factor analysis to derive reliable factor loadings. More data need to be collected to further validate the construct validity of the two complexity questionnaires.

Sensitivity

The sensitivity of the questionnaire is established by answering the question, "Are the questionnaires sensitive to experimental manipulation?" (Nunnally, 1978). To measure the sensitivity of the two complexity questionnaires, responses from both were compared between the two types of tasks (directed and exploratory), and among the Web sites (Expedia, Travelocity, Orbitz). Because each participant was assigned to use only one travel Web site to perform tasks, *Web site* was a between-subject independent variable. Each participant performed both types of tasks. Therefore, the task type was a within-subject, independent variable. The dependent variables were the responses to questionnaires. Due to the different nature of the tasks, the exploratory tasks were expected to yield higher complexity ratings than the directed tasks.

ANOVA on PSSUQ Results

ANOVA on the overall usability ratings of PSSUQ showed that task type was statistically significant (p=0.011) in affecting the overall usability. The exploratory tasks had lower usability (M=3.08, SD =1.42) than the directed tasks (M=3.60, SD = 1.51). The Web site examined and the interaction between Web site and task type were not significant.

Further analyses of the individual dimensions of PSSUQ (system usefulness, information quality, and interface quality) indicated that task type was a significant factor affecting system usefulness (p=0.009) and interface quality (p=0.005) but not information quality (p=0.107). Figure 3 shows the complexity rating by PSSUQ. System usefulness included aspects of ease of use, learnability, speed, and task performance; interface quality measured whether users felt the system was pleasant and liked it (Lewis, 1995). The significance in both dimensions indicated that there were indeed differences in users' interactive experiences with the system between the two types of tasks. The complexity questionnaires should be able to capture such differences.

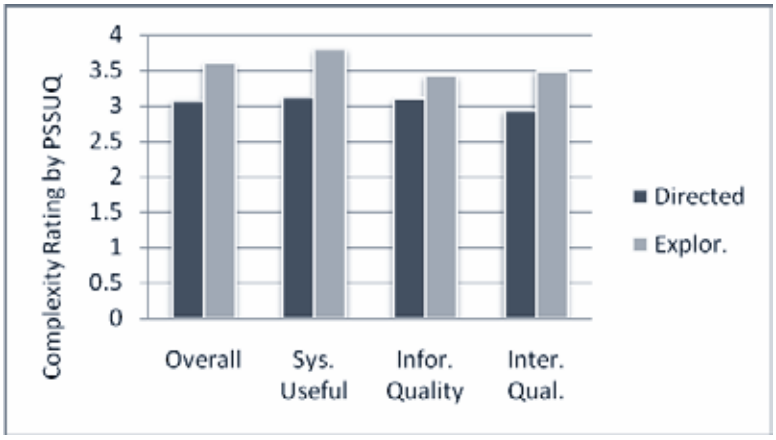

Figure 3. The Usability Ratings Measured by PSSUQ

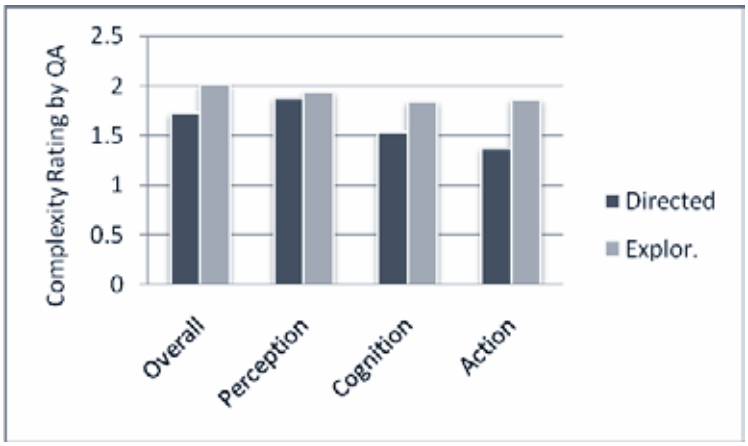

Figure 4. The Complexity Ratings Measured by QA

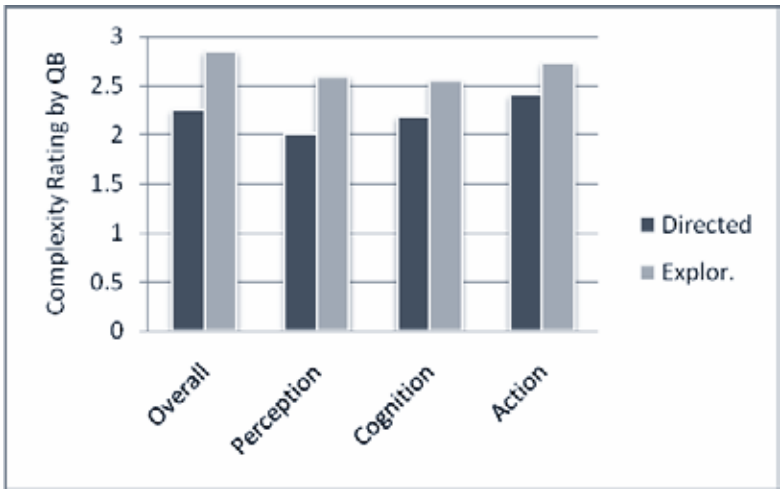

Figure 5. Complexity Ratings Measured by QB

ANOVA on QA Results

The complexity ratings measured by QA are shown in Figure 4. An ANOVA on the overall complexity responses in QA showed that task type was significant (p=0.014). The interfaces were considered as being more complex when performing exploratory tasks (M=2.02, SD=0.73) than directed tasks (M=1.73, SD=0.53). The Web site factor and the interaction between Web site and task type was not significant.

Further analysis of individual complexity constructs (perception, cognition, and action) showed that the task type did not significantly affect perceptual complexity (p=0.68). But task type differences were found for cognitive complexity (p=0.008) and action complexity (p<0.0001). The reasons for the significant differences might be that when the participants performed the exploratory task, they needed to figure out which links to follow to accomplish the tasks. The navigation efforts taken to seek the answer and carry out the tasks were usually more demanding than using the standard tool box. Therefore, both the cognitive and action complexity of the interfaces were higher for the exploratory tasks. It seemed reasonable to think that the perceptual complexity should also be higher for exploratory task than directed task because of the extra browsing through Web pages. The responses to QA did not reveal any significant differences in perceptual complexity. However, such differences were captured by QB.

ANOVA on QB Results

The complexity ratings measured by QB are shown in Figure 5. An ANOVA on overall complexity responses in QB showed that task type was again significant (p=0.007). The exploratory tasks were considered more complex (M=2.84, SD =1.31) than the directed tasks (M=2.25, SD=0.98). The Web sites and the interaction between Web site and task type were not significant.

Further analysis of individual complexity constructs (perception, cognition, and action) showed that the task type was a significant factor affecting all three complexity constructs— perceptual complexity (p=0.008), cognitive complexity (p=0.05), and action complexity (p=0.016). Exploratory tasks resulted in higher complexity ratings in all three

complexity constructs than directed tasks. This result was a bit different from those obtained with QA. In addition to cognitive and action complexity, perceptual complexity measured by QB was significantly affected by task types as well. This may be because participants need to spend more effort searching for information during the exploratory tasks, whereas in the directed task, they only needed to search for information within the framework provided by the search tool. So QB was able to capture differences in perceptual complexity that QA did not. The results suggest that QB was more sensitive than QA.

In summary, both questionnaires were able to detect the differences in complexity of the two types of tasks and demonstrated satisfactory sensitivity. The ANOVA analysis on QB results revealed a significant factor that was not found with QA, the effect of task type on perceptual complexity. The result implies that QB had higher sensitivity than QA. The reason may be that multiple statements associated with each construct in QB better probed a subject's opinions about the complexity of each Web site.

DISCUSSION

In this study, two complexity questionnaires were validated by evaluating subjects' task performance on three travel Web sites. Reliability, validity, and sensitivity of the two questionnaires were calculated and validated. Reliability was validated by calculating the internal consistency for QA and QB. Validity was established by examining the correlation among the two complexity questionnaires and the PSSUQ usability questionnaire, and the correlations between the responses to QA and QB. Sensitivity was established by an ANOVA analysis of results across two types of tasks. The questionnaire was able to capture differences associated with different task types. In general, both QA and QB had acceptable psychometric attributes, including reliability, validity, and sensitivity. Moreover, we found that the reliability, validity, and sensitivity of QB were higher than those of QA. Perhaps this is because QB used multiple items to derive summative scales, while QA used only a single item for each complexity construct. This result is consistent with the findings by Nunnally (1978): Summated scales are more reliable than single-item scales.

The experimental results also contributed to the improvement of the complexity questionnaires. We found that three statements in QB were inappropriate for their associated complexity constructs. Those statements caused confusion and contaminated the responses. Removing them from the data resulted in higher internal consistency.

The two types of tasks used in the experiment were complimentary. Directed tasks are those that users perform every day, while exploratory tasks represent those that are not used on a regular basis and require more effort to perform. Since each task type represents a method of human interaction with interfaces, any evaluation based on just one type of task could not fully account for interface usability and complexity. We recommend that a full evaluation of interfaces be based on the combination of these two types of tasks.

One shortcoming in this study is the relatively small sample size. We had 51 participants and 102 sets of responses. This size was insufficient to run a factor analysis on the results to fully assess the construct validity (including convergent and discriminant validity) of the questionnaires. Further validations should be performed with a larger sample of participants.

Although the two complexity questionnaires were initially developed for ATC displays, our result showed that with minor wording modification, they could also be used to measure the complexity of commercial visual interfaces. On the other hand, validation of the questionnaires with commercial Web sites provided initial evidence that the questionnaires are reliable, valid, and sensitive to interface complexity. Thus, they can be used as a quick assessment tool for evaluating ATC displays. Yet, we need to recognize the differences between travel Web sites and ATC displays. In general, travel Web sites' complexity is not as high as most ATC displays. Also, many ATC displays present dynamic information, while travel Web site interfaces primarily present static information. The current study used college students as naïve users of the commercial Web site to validate the complexity questionnaire. In contrast, controllers using the ATC systems are experts in their task domain. Therefore, for the validation results to be fully generalized to ATC displays, we should perform further validation studies with complex interfaces that present information dynamically.

REFERENCES

ISO 9241 (1998). Ergonomic requirements for office work with visual display terminals (Part 11. Guidance on Usability).

Lewis JR (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7(1),* 57-78.

Lewis JR (2002). Psychometric evaluation of the PSSUQ-using data from five years of usability studies. *International Journal of Human-Computer Interaction,* 14, 463-88.

Maeda J (2006). The laws of simplicity: Design, technology, business, life. Cambridge, MA: MIT Press.

Nunnally JC (1978). Psychometric theory. New York: McGraw-Hill.

Thompson B, Daniel LG (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56 (2), 197-208.

Xing J (2004). Measures of information complexity and the implications for automation design. Washington, DC: FAA Office of Aerospace Medicine, Report No. DOT/FAA/AM-04/17.

Xing J (2007). Information complexity in air traffic control displays. Washington, DC: FAA Office of Aerospace Medicine, Report No. DOT/FAA/AM-07/26.

Xing J (2008). Designing questionnaires for controlling and managing information complexity in visual displays. Washington, DC: FAA Office of Aerospace Medicine, Report No. DOT/FAA/AM-08/18.

APPENDIX A

Information Complexity Questionnaire A

Name of the website you are evaluating

Travelocity.com
Orbitz.com
Expedia.com

How many times have you used this website before today's experiment? _____

Instruction:

1. The purpose of this questionnaire is for users to evaluate the information complexity of web sites. The questionnaire contains 13 questions, each assessing a specific aspect of website complexity. In the questions below we have provided you with four choices A, B, C, and D. You may either circle one of the four choices, or circle two adjacent choices if you feel that both of them describe your feeling towards the website.

2. The term "information" in the questionnaire means either displayed materials (texts, symbols, etc.) that provide information to users or control functions (action buttons, menus, etc) for users to acquire information. The term "tasks" or "primary tasks" means the things that you want to accomplish through using the website.

Thank you for your participation!

1. How do you rate the perceptual complexity of the website?

A. The website looks simple and clear; I can find the needed information easily and quickly.
B. The website looks busy but I can find my 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 website looks too busy for me to quickly find the information.

2. How easy is it for you to find information on the website?

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 website.
D. I have to search through the website to find the information.

3. How well is the information organized on the website?

A. Information organization is very 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 thinking about how the information is organized.
C. The organization of information is confusing; I have to work hard thinking about how the information is organized.
D. The website has too many visual features (colors, symbols, fonts, graphic patterns, etc) for me to recognize how information is organized.

4. 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 stare at them for a longer time).
D. The website has too much clutter; it is hard for me to read the text quickly and correctly.

5. 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 website 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.

6. How well are you aware of the information provided by the website?

A. There are only several chunks of information that I need to be aware of in order to use the website. 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 website, but have difficulties to do so when I have other things in mind.
D. The website provides too many pieces of information for me to be aware of; I cannot mentally build a fixed mental model of the website.

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

A. The website does not present dynamic information or most changes are expected and predictable.
B. I can take care of changes but prefer that the website 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.

8. 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 make certain strategies to use 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 website. It is difficult to hold them all at once.

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

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

10. How would you evaluate the number 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.

11. How do you rate the number of 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.

12. 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. It is highly possible that I may be 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.

13. Overall, how do you rate the complexity of the website in terms of its usefulness?

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

APPENDIX B

Information Complexity QB (improved version)

Instructions: This survey asks you to respond to items designed to measure a specific aspect of a website. For example, Section I asks about how quickly and easily you can find the information you need on the website. When answering an item, think about the lead-in question for that section and indicate your response by circling the choice 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 quickly and easily can you find the information you need on the website?							
1	I know where to look to see the information I need.	*	*	*	*	*	*
2	I can see the information I need without searching.	*	*	*	*	*	*
3	I have to search through the website to find the information I need.	*	*	*	*	*	*
4	I can find the information I need with a few quick glances.	*	*	*	*	*	*
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 website.	*	*	*	*	*	*
6	I can easily see the structure of the displayed information.	*	*	*	*	*	*
7	The variety of visual features on the website is confusing.	*	*	*	*	*	*
8	The website uses too many different sizes, colors, fonts, and icons.	*	*	*	*	*	*
How does the website clutter affect reading text and icons?							
9	The website looks too busy.	*	*	*	*	*	*
10	The text and icons stand out clearly from the background.	*	*	*	*	*	*
11	I have to move closer to the screen to read the text.	*	*	*	*	*	*
12	I have to stare at the website to read the information.	*	*	*	*	*	*
13	I can quickly and correctly read the information presented on the website.	*	*	*	*	*	*
14	Adequate space is provided between pieces of information on the website.	*	*	*	*	*	*

		Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
How does the amount of information provided on the website affect information management?							
15	It is difficult to manage all the necessary information to do a task.	*	*	*	*	*	*
16	It requires a great deal of efforts to manage all the necessary information or control functions.	*	*	*	*	*	*
17	There is too much information and there are too many control functions on the website to remember.	*	*	*	*	*	*
How do information changes on the website affect the way you process information?							
18	Information changes on the website are predictable.	*	*	*	*	*	*
19	Information changes on the website are easy to track.	*	*	*	*	*	*
20	Keeping track of information changes on the website distracts me from performing my primary tasks (makes me too busy).	*	*	*	*	*	*
21	Information changes are too frequent.	*	*	*	*	*	*
22	I would prefer the information on the website to change less frequently.	*	*	*	*	*	*
Does the way in which information is presented affect your understanding of that information?							
23	I think that pertinent information was presented in a direct manner.	*	*	*	*	*	*
24	Interpreting information distracts me from focusing on my tasks.	*	*	*	*	*	*
25	The information is presented so that I can understand it without having to think about it too much.	*	*	*	*	*	*
26	I must relate several pieces of separately displayed information to understand them.	*	*	*	*	*	*
27	The information presented is straightforward.	*	*	*	*	*	*
28	I have to relate (or associate) too many pieces of information at the same time.	*	*	*	*	*	*
How does the action cost (transition between action modes e.g., from keyboard to mouse) affect you?							
29	The website requires too many actions to perform tasks or acquire information.	*	*	*	*	*	*
30	The number of transitions in action modes (i.e., from keyboard to mouse, mouse to keyboard, etc) distracts me.	*	*	*	*	*	*
31	I am comfortable with the number of transitions required to perform actions on the website.	*	*	*	*	*	*
32	Using this website requires me to frequently change action modes, which takes my time away from performing my primary tasks.	*	*	*	*	*	*

		Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
How does the number of action steps (e.g., number of display windows, pull down menus, and pop up windows) affect you?							
33	I have to access too many menu buttons or windows to acquire information/perform tasks.	*	*	*	*	*	*
34	The website's pop-up windows and/or pull-down menus help me to acquire information/perform tasks.	*	*	*	*	*	*
35	I have trouble getting the information and performing tasks because there are so many layers of windows/menus.	*	*	*	*	*	*
How does the number of action sequence to perform tasks or acquire information affect you?							
36	I have to manage multiple action sequences to get a task done. I have a hard time keeping up with all.	*	*	*	*	*	*
37	The number of action sequences is small, and I can perform the tasks easily on the website.	*	*	*	*	*	*
How would you evaluate the overall complexity of the website?							
38	The website is an effective tool for acquiring information.	*	*	*	*	*	*
39	The website is simple and easy to use.	*	*	*	*	*	*
40	I do not like it because it is too complex to use.	*	*	*	*	*	*
How would you evaluate the perceptual complexity of the website?							
41	Only necessary information is presented on the website.	*	*	*	*	*	*
42	I can easily and quickly find the information I need.	*	*	*	*	*	*
43	I can hardly find the information I need.	*	*	*	*	*	*
44	I could not find the information I need.	*	*	*	*	*	*
How would you evaluate the cognitive complexity of the website?							
45	I can easily process the information presented on the website.	*	*	*	*	*	*
46	Using this website takes too much mental effort.	*	*	*	*	*	*
47	I feel overwhelmed by the amount of information presented on the website.	*	*	*	*	*	*
How would you evaluate the action complexity of the website?							
48	The actions required by the website take my attention away from my tasks.	*	*	*	*	*	*
49	The amount of action required to perform tasks or acquire information does not bother me.	*	*	*	*	*	*
50	I can easily interact with the website to accomplish my tasks.	*	*	*	*	*	*
51	I feel overwhelmed by the amount of interaction required by the website.	*	*	*	*	*	*