

RESEARCH REPORT SERIES  
(*Survey Methodology* #2007-28)

Accessibility Assessment of Statistical Web Pages

Sid Schneider\*  
Lawrence Malakhoff

Westat\*

Statistical Research Division  
U.S. Census Bureau  
Washington, D.C. 20233

Report Issued: August 7, 2007

*Disclaimer:* This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

# Accessibility Assessment of Statistical Web Pages

By

Sid Schneider (Westat)  
Lawrence Malakhoff (U. S. Census Bureau)

April 24, 2007

*This report is released to inform interested parties of (ongoing) research and to encourage discussion (of work in progress.) The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.*

# Table of Contents

	<u>Page</u>
EXECUTIVE SUMMARY .....	1
1 INTRODUCTION AND BACKGROUND.....	3
1.1 Statistical Web Sites and Section 508 .....	3
1.2 Usability Research .....	5
1.3 Goals .....	6
2 METHOD .....	8
2.1 Participants .....	8
2.2 Forms .....	8
2.3 Tables .....	9
2.4 Questionnaire and Practice Tasks .....	10
2.5 Analysis.....	11
3 RESULTS .....	13
3.1 Forms .....	13
3.1.1 Times.....	13
3.1.2 Ratings .....	13
3.2 Tables .....	13
3.2.1 Accuracy .....	14
3.2.2 Times.....	14
3.2.3 Ratings .....	14
3.3 Automated Tools.....	14
4 DISCUSSION.....	15
4.1 Feasibility of the Method .....	15
4.2 Online Forms.....	17
4.3 Online Tables .....	15
4.4 Recommendations .....	16
4.4.1 Usability Laboratory Methods .....	16
4.4.2 Forms .....	17
4.4.3 Tables .....	17
4.4.4 Caveat .....	17
4.5 Future Research.....	17
5 REFERENCES .....	19

## List of Figures

Figure 1.	Time required by the 19 participants to find the requested item on the forms having two, four, six, and eight drop-down menus.....	25
Figure 2.	Time required by the 19 participants to find the requested item on the forms having two, four, six, and eight drop-down menus, excluding the six tasks in which the times exceeded 200 seconds. ....	26
Figure 3.	Ratings on a 10-point scale, of the difficulty of the tasks when the forms had two, four, six, and eight drop-down menus. ....	27
Figure 4.	Time required by the 10 participants who found the items in the tables correctly in all four tasks. ....	28
Figure 5.	Time required by the 15 participants who found items in the tables within the allowable time in all four tasks, regardless of whether the items were correct or incorrect. ....	29
Figure 6.	Time required by the 15 participants who found items in the tables within the allowable time in all four tasks, regardless of whether the items were correct or incorrect. The 6 tasks in which a participant required more than 200 seconds were excluded from these calculations. ....	30
Figure 7.	Ratings on a 10-point scale, of the difficulty of the tasks with 2-by-2, 2-by-15, 15-by-2, or 15-by-15 tables. These data are for the 10 participants who found the items in the tables correctly in all four tasks. ....	31
Figure 8.	Ratings on a 10-point scale, of the difficulty of the tasks with 2-by-2, 2-by-15, 15-by-2, or 15-by-15 tables. These data are for the 15 participants who found items in the tables within the allowable time in all four tasks, regardless of whether the items were correct or incorrect. ....	32

## List of Tables

	<u>Page</u>
Table 1. Content of the eight drop-down menus in the online forms .....	21
Table 2. An online form with eight drop-down menus. ....	22
Table 3. The online table with 15 rows and 15 columns .....	23
Table 4. Number of errors committed in the tasks involving online tables by all 20 participants, excluding the three who did not complete any of these tasks successfully .....	24

## EXECUTIVE SUMMARY

Section 508 of the Rehabilitation Act of 1973, as amended in 1998 (29 U.S.C. § 794d) requires that all Federal information technology be designed to provide disabled people with a level of access equivalent to that enjoyed by people who have no disability. The United States Access Board has issued guidelines to help ensure that Federal web sites are compatible with screen readers, the assistive technology used by persons who cannot see. However, more research is needed to ensure that web sites are optimally designed for persons who cannot see, to help ensure that they have equal access to online information.

The Federal government offers over 600 web resources that provide statistical data. Users often request particular data by completing an online form. The requestor then receives an online table containing the data. These online forms and tables can be challenging for people who rely on screen readers. The goal of this study was to clarify how participants who cannot see are affected by the number of drop-down menus in online forms and the number of rows and columns in online tables. The results could suggest the optimal designs for online forms and tables.

A total of 20 participants who cannot see carried out four tasks using online forms and four tasks using online tables. The forms had respectively two, four, six or eight drop-down menus. The tables had respectively two rows and two columns, two rows and 15 columns, 15 rows and two columns, or 15 rows and 15 columns.

This study applied usability laboratory methods to accessibility research. Usability researchers often measure participants' speed, accuracy and satisfaction in the controlled conditions of the usability laboratory. In this study, the participants' accuracy and estimated speed were recorded for each task. The participants rated the difficulty of each task on a 1-to-10 scale.

The results from the time estimates suggest that the participants were speediest working with online forms that had only two or four drop-down menus. However, adding drop-down menus beyond four did not consistently increase the amount of time that the test participants required to carry out the tasks. Also, the participants did not consistently perceive that adding drop-down menus to a form makes it more difficult to use.

The results from the time estimates also suggest that adding rows and columns to online tables affects the performance of persons who cannot see. When rows or columns were added to the online table, the participants required more time to complete the task. The participants made more errors, and rated the tasks as more difficult, when the tables had 15 columns, as compared to when they had two columns.

Several recommendations follow from the results:

- Online forms should be limited to no more than four drop-down menus if possible. Larger forms should be broken into smaller ones when feasible.
- However, once the forms contain more than four drop-down menus, the effect of adding still more menus, up to eight, is not apparent in these data. Further research is needed.
- Online tables should contain as few rows and columns as feasible. Large tables should be broken down into smaller ones when possible.
- The number of columns in a table may be especially likely to add to the difficulty of persons who cannot see. Therefore, designers should try to break tables that have a large number of columns down into two or more tables with smaller numbers of columns.
- Methods to help people who cannot see query table data may be desirable. These methods might be modeled upon advanced searches in web browsers.

## 1 INTRODUCTION AND BACKGROUND

Section 508 of the Rehabilitation Act of 1973, as amended in 1998 (29 U.S.C. § 794d) requires Federal agencies to eliminate the barriers facing disabled people who use information technology such as web sites, software products, and computer hardware. Under this law, Federal agencies must provide disabled employees and members of the public with a level of access to information technology that is comparable to the level enjoyed by people who have no disability. Since the summer of 2001, all Federal agencies have been required to design their web sites for compliance with Section 508.

People who cannot see web pages adequately can use assistive technology called screen readers. The most commonly used screen reader in the United States is called *Job Access With Speech*, or JAWS, marketed by Freedom Scientific ([www.freedomscientific.com](http://www.freedomscientific.com)), but several others including WindowEyes ([www.gwmicro.com](http://www.gwmicro.com)) and Home Page Reader ([http://www-03.ibm.com/able/solution\\_offerings/hpr4\\_demo\\_text.html](http://www-03.ibm.com/able/solution_offerings/hpr4_demo_text.html)) are also widely used. Screen readers, as their name implies, read web pages aloud. Persons who cannot see usually have little problem using screen readers to hear the content of web pages that contain only text. However, features like forms and tables must be designed appropriately or they will be incomprehensible to people who use screen readers. Accordingly, the Access Board, a Federal agency devoted to promoting accessible design and compliance with Section 508, has developed standards covering a range of issues, including the design of online forms and tables (General Services Administration, 2004).

### 1.1 Statistical Web Sites and Section 508

In Fiscal Year 2005, the Federal government spent \$5.14 billion collecting and disseminating statistical data. Over 40 percent of this amount was spent by the ten agencies that have statistical activity as their principal mission, such as the Census Bureau. The remainder was spent by the 73 other agencies that maintain statistical programs (Office of Management and Budget, 2004). The web portal [www.fedstats.gov](http://www.fedstats.gov) lists over 600 Federal web sites and sections of Federal web sites that disseminate statistical data about such topics as health, transportation, crime, energy, immigration, housing, employment, trade, demography, and the economy. A recent survey found that the public regards access to such information as one of the most significant benefits of the Federal government's Internet activities (Harper, 2004).

The Census Bureau conducts several well-known data collection projects, including the decennial Census, the American Community Survey, and the Economic Census, which are key sources of data about the population and commerce of the United States. The Census Bureau's web site contains resources such as *Economic Indicators*, *Quarterly Workforce Indicators*, *QuickFacts* and *American FactFinder*, which disseminate the findings of these data collection efforts. A user first completes a brief online data request form to specify the information that she needs. She then receives the data table containing that information.

There are compelling reasons to make such Federal web resources accessible to persons who cannot see. First, people without any sight or who have low vision constitute a sizable constituency. Of the 33.7 million American adults who are limited in their usual activities due to chronic conditions, 9.3 million are visually challenged (National Center for Health Statistics, 2002). These numbers are likely to increase as the population ages. Second, the Internet is an important information resource for persons who cannot see. Internet users represent only 38 percent of all disabled American adults (Lenhart, Horrigan, Rainie, Allen, Boyce, Madden, and O'Grady, 2003), but 62 percent of American adults who cannot see (Gerber & Kirschner, 2001; National Telecommunications and Information Administration, 2002). Making Federal statistical web resources accessible increases the audience for Federal data and helps people without sight to be knowledgeable citizens who are not shut out of educational and employment opportunities. Thirdly, Section 508 legally requires Federal agencies to make all web resources, including online statistical data, accessible to people without sight.

Two of the Access Board's Section 508 standards apply specifically to data tables. The first states, "Row and column headers shall be identified for data tables" (<http://www.section508.gov/index.cfm?FuseAction=Content&ID=12>; Subpart B, § 1194.22 g). This standard means that the HTML code for data tables must include the "scope" attribute or other markup that causes screen readers to recite each cell's row and column headers along with the data in the cell, so that people who use screen readers can comprehend the meaning of the data (General Services Administration, 2004). The second standard states, "Markup shall be used to associate data cells and header cells for data tables that have two or more logical levels of row or column headers" (<http://www.section508.gov/index.cfm?FuseAction=Content&ID=12>; Subpart B, § 1194.22 h). This standard means that data tables must have HTML markup that causes screen readers to recite all the levels of the row and column headers (General Services Administration, 2004).

Another Section 508 standard (<http://www.section508.gov/index.cfm?FuseAction=Content&ID=12>; Subpart B, § 1194.22 n) applies to data entry forms. It states, "The form shall allow people using assistive technology to access the information, field elements, and functionality required for completion and submission of the form,



including all directions and cues.” This standard is meant to ensure that people who cannot see an online form can rely on screen readers to explain the form to them. All fill-in fields should have <label> tags so that screen readers vocalize the labels associated with the fields. HTML markup should group together related sets of fields, and the response alternatives to each multiple choice question. The tab order of the components of the form should be carefully designed, because persons without sight use the tab key, not a mouse, to navigate (General Services Administration, 2004).

The Access Board’s Section 508 standards are in accord with other accessibility guidelines, such as the guidelines of the Web Accessibility Initiative ([www.w3.org/WAI](http://www.w3.org/WAI)) and the National Center for Accessible Media (<http://ncam.wgbh.org>), Bobby accessibility assessment software ([www.cast.org/bobby](http://www.cast.org/bobby)), and several web accessibility books (Koyani, Bailey, and Nall, 2003; Paciello, 2000; Slatin and Rush, 2003; Thatcher, Bohman, Burks, Lawton Henry, Regan, Swierenga, Urban, and Waddell, 2002). The Section 508 standards ensure that persons who cannot see are not shut out from using online forms and data tables. However, the standards are intended only to “set forth provisions that establish a minimum level of accessibility” (Section 508, Subpart A). The standards by themselves are not meant to specify user-centered designs that provide the least burdensome experience for people who cannot see.

## **1.2 Usability Research**

In several published studies, researchers observed participants who could not see completing online forms or locating information within online tables. Amtmann, Johnson, and Cook (2002) observed nine participants who could not see attempting to extract information from online tables with various designs. This study suggested that persons who cannot see often struggle with tables that sighted people find trivial to use. It also suggested that persons who cannot see are more successful when tables have a descriptive summary that explains the structure of the table, and when the tables are constructed so that each contains as little data as is practical.

Barnicle (2000) and Gerber (2002) observed persons who could not see using the Windows operating system, as well as the web sites for Medicare and the American Foundation for the Blind. These studies suggested that consistency in design from page to page, long known to be a virtue for sighted users (Nielsen, 1993), is especially helpful to people who cannot see because they tend to rely on their memories to determine how to proceed. The studies also suggested that a graphical user interface that is well designed for sighted users may be inappropriate for persons without sight, because screen readers cannot take advantage of visual cues.

Theofanos and Redish (2003) observed 16 participants without sight carrying out a set of tasks on Federal web sites. Based on their observations, these researchers wrote 31 guidelines for accessible design, including some which pertain to online forms: a minimum of text should be on the same page as a form; a form should not be far down or to the right on a page; information should not be located between fields; and, if the user has an option of filling out either of two fields, and they are mutually exclusive, these directions should be explained in the label of the first field.

In all of these studies, researchers observed participants without sight in naturalistic settings such as offices, not in a usability laboratory. Usability laboratories allow researchers to observe participants under controlled conditions. The participants carry out specific tasks on a web site, while the researchers monitor their performance. The researchers debrief the participants, to learn their opinions about the web site. Usability laboratory research has increasingly become an indispensable part of the process of developing web sites (Dumas and Redish, 1999). When web developers have usability laboratory data, they do not have to rely solely on their own best judgments about web design; they can refer to data derived from observations of actual people using design prototypes to complete realistic tasks.

The American National Standards Institute (2001) established that usability laboratory tests should obtain three measures: (1) the participants' accuracy on each task; (2) the amount of time participants require on each task; and (3) the participants' subjective ratings of ease of use. These three usability measures are also relevant to accessibility research. Usability laboratory methods potentially could help web designers ensure that their products are easy for the all people to use, regardless of their ability to see.

### **1.3 Goals**

The present project was conducted with the goal of developing research-based design recommendations for the online forms and data tables that appear on Federal statistical web sites. By applying these recommendations, web developers could go beyond the Section 508 standards, which simply ensure that online forms and tables are compatible with screen readers; developers would be able to optimize the user experience for persons without sight who visit statistical web sites.

To reach this goal, the project's first objective was to develop a methodology that applied usability laboratory methods to assess the performance of people who cannot see. The designs of the online forms and tables were systematically varied while the participants were videotaped in the usability laboratory. The participants' speed, accuracy and subjective ratings were assessed for each design.

The project's second objective was to address the following questions:

- When persons who cannot see access an online form using a screen reader, how does the number of drop-down menus on the form affect the participants' accuracy and speed completing the form, and the participants' subjective rating of the level of difficulty?

- When persons who cannot see access an online table using a screen reader, how does the number of rows and columns in the table affect the participants' accuracy and speed extracting data from the form, and the participants' subjective rating of the level of difficulty?

The answers to these questions could help web developers understand how the number of drop-down menus in online forms and the number of rows and columns in online tables affect people who cannot see. The answers could also suggest the best ways, from the standpoint of accessibility, to present online forms and tables.

## 2 METHOD

### 2.1 Participants

A total of 20 people who cannot see participated. The inclusion criteria were as follows: the participants were 18 to 65 years old, attained a high school diploma or equivalent, were visually impaired from birth, used the Internet at least monthly for the past year, were unable to use a computer without a screen reader, used JAWS more than any other screen reader, and had no non-visual physical or psychiatric disabilities. Random sampling was not used for recruitment. Eighteen participants had received training from Columbia Lighthouse for the Blind in JAWS and MS-Office applications at various times from 1996 to 2005. Two participants were contractors who used JAWS daily and could be considered experts. Ten of the participants were female, and 10 were male. Thirteen were African American.

The participants reported to the Census Bureau's usability laboratory in Suitland, Maryland. The test administrator explained the research procedure, and the participants verbally gave their consent for videotaping. They also signed an informed consent form and a payment voucher, with the aid of a template. Each session lasted no more than 60 minutes. The sessions were videotaped.

The participants were seated at a computer in the testing room of the usability laboratory. The test administrator, sitting close by, provided the instructions for each task individually. The participant carried out eight tasks on web pages that had been created for this project. The first four tasks involved online forms. The next four tasks involved online tables.

If any time remained, participants carried out three additional tasks, one on each of three online resources of the Census Bureau: *American FactFinder*, *Population Finder*, and *Economic Indicators*. These three tasks used the actual online resources, not web pages created for this project.

### 2.2 Forms

In each of the four tasks involving online forms, the participant was presented a web page containing a form with, respectively, two, four, six, and eight drop-down menus. When the form had eight drop-down menus, the menus were titled as follows: county, gender, age, race, ethnicity, home ownership, quarter, and year. The contents of each of these drop-down menus are shown in Table 1. Table 2 shows a form with eight drop-down menus. These eight drop-down menus were similar to those on Census Bureau web resources like *American FactFinder*. When the form contained six drop-down

menus, the six were randomly selected from those eight. When the form contained four drop-down menus, the four were randomly selected from the ones in the form with six drop-down menus. Similarly, when the form contained two drop-down menus, the two were randomly selected from the ones in the form with four drop-down menus. The drop-down menus selected for each form varied across participants.

For each task, the experimenter asked the participant to select a particular item from one of the drop-down menus in the form. The participant then located the correct drop-down menu, and made the selection from the items listed in that drop-down menu. The participants were instructed to say the word “done” when they had completed a task. If they could not complete the task within 15 minutes, they were asked to stop working on the task and to begin the next task.

The forms were presented in random order. The item that the participant was asked to find and the drop-down menu on which that item appeared were randomly selected. This randomization procedure was intended to avoid confounding the effect of practice with the effect of the number of drop-down menus. It was also intended to eliminate unintended biases, such as the potential effect of the location of the drop-down menu. The first drop-down menu in the form was never the one that contained the requested item. The participant therefore always had to hear the titles of at least two drop-down menus to find the correct one.

The instructions for each task were in this format: “On this form, select ‘males’ from one of the menus” or “On this form, select ‘people 19 to 21 years old’ from one of the menus.” The instructions did not reveal which drop-down menu was the appropriate one. After each task, the experimenter asked the participant to rate the difficulty of that particular task, on a scale from 1 (very easy) to 10 (very difficult).

### **2.3 Tables**

In each of the four tasks involving tables, the participant was presented a web page containing a table in which data were presented in respectively two rows and two columns, two rows and 15 columns, 15 rows and two columns, or 15 rows and 15 columns. Table 3 shows the 15-row by 15-column table. The tables were created specially for this study. They contained simulated data about the number of people in various Maryland counties employed in various economic sectors. In this way, they resembled the tables on the Census Bureau’s online *Quarterly Workforce Indicators*.

The rows represented Maryland counties. A row header showed the names of the counties. Table 3 shows the counties that were included when the tables had 15 rows. When the tables had only 2 rows, the counties included were Baltimore and Prince George’s.

The columns represented economic sectors. A column header showed the names of the sectors. Table 3 shows the sectors that were included when the tables had 15 columns. When the tables had only two columns, the sectors included were agriculture and mining.

For each task, the experimenter asked the participant to find the number of people employed in a particular economic sector in a particular Maryland county. The instructions might be “Find the number of people employed in agriculture in Prince George’s County” or “Find the number of people employed in the wholesale trade in Harford County.” The participants were instructed to say the word “done” when they had completed a task. If they could not complete the task within 15 minutes, they were asked to stop working on the task and to begin the next task.

The 2-by-2, 2-by-15, 15-by-2, and 15-by-15 tables were presented in random order. The county and the economic sector for the requested information were randomly selected for each task. When the table contained 15 rows or 15 columns, the requested information was never in the first five or last five rows or columns; it was always centrally located, between the sixth and the tenth row or column. The purpose of positioning the needed information this way was to minimize the effect of the location of the information upon the participants’ performance.

The instructions for each task were in this format: “Find the number of people employed in agriculture in Prince George’s County” or “Find the number of people employed in the wholesale trade in Harford County.” After each task, the experimenter asked the participant to rate the difficulty of that particular task, on a scale from 1 (very easy) to 10 (very difficult).

## **2.4 Questionnaire and Practice Tasks**

At the start of each testing session, the researcher administered a questionnaire to the participants, asking about their demographic and computer backgrounds. Then, each participant carried out two practice tasks. The experimenter began with an introduction similar to the following:

Thank you for coming to the Census Bureau today to participate in this project. We are trying to learn the best way to design forms and tables for the Internet so that persons who cannot see can use them easily. I am going to ask you to select items from the menus on a form. Then I am going to ask you to find information on a table. Please try to be both as accurate and as speedy as you can. Do you have any questions?

The experimenter then administered the two practice tasks. In the first one, the participant was presented a form with four drop-down menus. The instructions were as follows:

Let's try a practice form right now. I am going to show you a form on your computer screen. Your screen reader will read it to you. The form will have several menus. Your task this time will be to select "males" from one of the menus. First you will find the menu that is labeled "gender." There are two items on the "gender" menu. Those items are "males" and "females." You will select "males." When you have selected "males," please let me know by saying "All done." Do you understand? Let's try it.

These instructions deliberately identified the correct drop-down menu, even though the instructions that the participants would receive in the actual tasks did not. The intent of the practice instructions was to familiarize the participant with the forms.

In the second practice task, the participant was presented a table with two columns and 15 rows. The instructions were as follows:

Now let's try using a table. The table that I am going to present to you shows the number of people employed in different industries in different counties in Maryland. Please find the number of people employed in agriculture in Dorchester county. You must find the entry in the column labeled "agriculture" and the row "Dorchester." When you have found the number that is in the column labeled "agriculture" and the row "Dorchester," tell me that number.

Again, the instructions were more extensive than the instructions that the participant would receive in the actual test. The intent of the test instructions was to familiarize the participant with the tables.

## **2.5 Analysis**

Because of a technical problem at the Census Bureau's usability laboratory, no sound was recorded on the videotapes of the first seven participants. As a result, Census Bureau staff could not precisely time these participants' tasks. Census Bureau staff obtained the total time on a task screen by computing the difference between the time the screen appeared and changed to the next screen on the videotape. They therefore estimated the times by deducting 18.6 seconds from the total time required for each task screen. The figure 18.6 seconds was obtained by a usability lab intern who calculated the

average time required to obtain the participants' ratings of the difficulty of the tasks from eight observations of the eighth through the twentieth participants.

Census Bureau staff analyzed the videotapes using Noldus Observer Video Pro software. For each task, the following was recorded:

1. The outcome of the task, which could be one of the following: (a) the participant found the requested item on the form or table, (b) the participant found an incorrect item on the form or table, or (c) the participant worked until the maximum time without reaching any solution.

2. The amount of time that the participant required to find the item on the form or table, whether or not that item was the requested item. Of course, this time could not be calculated if the participant worked until the 15 minute maximum time elapsed.

3. The participant's rating of the difficulty of the task, as number from 1 to 10.



## 3 RESULTS

### 3.1 Forms

All but one of the 20 participants were able to find the correct item on the correct drop-down menu for each of the four tasks successfully. The one exception found the correct items for the first two forms but then discontinued the session. That participant had to be excluded from the analyses.

**3.1.1 Times.** Figure 1 shows the estimated or approximate amount of time that the participants required in the four tasks. The participants' performance shown on Figures 1 is in the shape of an inverted U. A total of 12 of the 19 participants required more time with four drop-down menus than with two; 11 of the 19 participants required more time with six drop-down menus than with four; however, only four of the 19 participants required more time with eight drop-down menus than with six.

The times shown in Figure 1 were influenced by four participants who needed unusually large amounts of time for at least one of the four tasks. Two participants required more than 200 seconds for two of the four tasks, and two participants required more than 200 seconds for one of the four tasks. The mean times were recalculated, excluding these six tasks in which the time exceeded 200 seconds. The results are shown in Figure 2. The inverted U shape is no longer apparent. The results depicted in Figure 2 suggest that the participants may have required more time when the form had six or eight drop-down menus, as compared with two or four drop-down menus.

**3.1.2 Ratings.** Figure 3 shows the participants' ratings of the difficulty of the four tasks, on a scale from 1 to 10, where 1 is "very easy" and 10 is "very hard."

### 3.2 Tables

Only 10 of the 20 participants were able to find the requested items correctly in all four tasks involving online tables. Five participants found incorrect items on at least one task. The remaining five participants worked until the maximum time allowed for at least one task.

**3.2.1 Accuracy.** Of the 20 participants, three did not complete any of the tasks involving the online tables successfully. Table 4 shows the number of errors committed by the remaining 17 participants.

**3.2.2 Times.** Figures 4 and 5 show the estimated amount of time that the participants needed to find items in the tables. Figure 4 shows these data for the 10 participants who found the items correctly in all four tasks, while Figure 5 shows these data for the 15 participants who completed the four tasks within the maximum allowed time, regardless of whether they found the correct item or an incorrect item.

Inspection of the data depicted in Figure 5 revealed that five participants required unusually long amounts of time – more than 200 seconds – to find the item in the table in at least one of the four tasks. One participant required more than 200 seconds for two tasks, while another four participants required more than 200 seconds for one task. Two of these tasks occurred in the two rows by 15 columns condition; four of these tasks occurred in the 15 rows by 15 columns condition. Figure 6 shows the data for the 15 participants, excluding the data for these six tasks.

**3.2.3 Ratings.** Figures 7 and 8 show the participants' ratings of the difficulty of the tasks, on a 10-point scale. Figure 7 shows these data for the 10 participants who found the correct items in the tables on all four tasks. Figure 8 shows the ratings for the 15 participants who found an item in all four tasks, regardless of whether that item was correct or not.

### **3.3 Automated Tools**

Only eight participants had enough time at the end of the session to carry out the three tasks on the Census Bureau web site. No statistical analyses were possible with this small number of participants, nor were comparison data collected from sighted users. The participants who did carry out the tasks required a mean of 454.87 seconds ( $SD=272.12$ ) to find an item on *American FactFinder*, 199.95 seconds ( $SD=116.55$ ) to find an item on *Economic Indicators*, and 305.18 seconds ( $SD=116.86$ ) to find an item on *Population Finder*.

## **4 DISCUSSION**

### **4.1 Feasibility of the Method**

The results suggest that usability laboratory methods can yield data about the performance of people without sight when they work with various user interfaces. However, accurate task timing is needed for reliable analysis. Some participants reported that they did not have access to JAWS at home and did not use it daily. In this study, all of the participants but one were able to complete the tasks involving the online forms, but only half could complete the tasks involving the online tables.

### **4.2 Online Forms**

The results for the participants' ratings do not suggest that the participants generally perceived that adding drop-down menus to a form makes the form more difficult to use. However, a number of participants did rate the forms with six or eight drop-down menus as more difficult than the forms with two or four drop-down menus.

In summary, the results provide some support for a hypothesis to be validated in future testing that participants may be somewhat quicker working with forms that have two or four drop-down boxes than they are working with forms that have more drop-down boxes. Some participants may also rate forms with two or four drop-down boxes as easier, as well.

### **4.3 Online Tables**

The results provide some support for a hypothesis to be validated in future testing that adding rows or adding columns to an online table cause the participants to work for a longer absolute amount of time, and to rate the task as being difficult. Also, the results provided some evidence that the participants committed more errors, and found the tables to be more difficult, when the tables that had 15 columns than when the tables had 2 columns.

## 4.4 Recommendations

The results of this study suggest the following recommendations:

**4.4.1 Usability Laboratory Methods.** Web designers can employ usability laboratory methods to collect data to inform their decisions on accessible web design. They must be sure to use a reliable method for collecting task-completion times. Baseline measures on sighted users will be useful for comparison.

Some participants require relatively long periods of time to carry out tasks in the usability laboratory. Their slow speed may be attributable to factors unrelated to the web design. Some forgot the task question and needed reminding, possibly due to the amount of data they were hearing from JAWS.

Apparent limitations in the test participants JAWS expertise and problems with the terminology contributed to the long periods of time needed to carry out tasks in the usability lab. As observed by the test administrator, most test participants lacked familiarity with JAWS table commands or usage of context-specific help. Several test participants appeared to be impatient with JAWS and kept pressing the arrow keys and went past the table data value they were seeking, and then needed to navigate the table again.

Some test participants did not seem to be familiar with opening and navigating within drop-down menus or the fact that options are selected, not typed in. Several test participants remarked that they expected the JAWS virtual find (control-F) function to detect drop-down menu options or give data relationships in a table but had to find another method when virtual find did not function as they expected. When they could not access a control, most tried repeating the same behavior over and over instead of bringing up JAWS help for that item.

The terminology used for drop-down menus caused unintended problems in this study. A number of test participants seemed to be confused by the term “Ethnicity” when searching for the Hispanic or Latino category because they looked for this category in the Race drop-down menu. One person could not locate the time period, First Quarter, because he expected it in a menu called “Quarters”. Several test participants could not find “male” or “female” because these options were in a drop-down menu called “gender” and they did not know that gender was a synonym for sex.

Two participants who work as contractors used table commands to access data in the test tables. The remaining test participants who performed table tasks recognized that a county would have sub-categories, but when the cursor focus was on a data item they had difficulty relating it back to the county name. One test participant did remember how to use JAWS help and found and used table commands. Besides the contractors, test participants who successfully found data values memorized the number of

the column they were seeking and used the arrow keys when the cursor focus was on the correct county. These participants counted keystrokes until they reached the correct data value.

In accessibility tests, measures should be taken of both the length of time that participants need to complete the tasks and the participants' subjective ratings of the difficulty of the tasks.

**4.4.2 Forms.** The preliminary results from this study suggest online forms should be limited to no more than four drop-down menus if possible.

**4.4.3 Tables.** Online tables should contain as few rows and columns as feasible. Large tables should be broken down into smaller ones when possible.

Both the number of rows and the number of columns affect people without sight. When there is a choice about the way to make a table smaller, greater priority should be given to eliminating columns.

**4.4.4 Caveat.** The results of this study must be approached with caution. The technical problem that occurred in the usability laboratory added uncertainty to the measures taken from the first seven participants. The tendency of some participants to use unusually long times for some tasks suggests that factors other than the screen design may have influenced the dependent measures. For example, some respondents may have simply forgotten the task as they were working.

Because only 20 persons who cannot see participated in this study, small effects may have been missed. Also, fully half of the 20 participants could not complete at least one of the four tasks involving the online tables. Because the sample was not a random, probability sample, the results cannot be generalized to any larger population. A much larger study, using a nationally representative sample of web users without sight, might yield different results.

## **4.5 Future Research**

Future accessibility research can better clarify how usability laboratory methods can be applied to accessibility issues, allowing researchers to collect data about the accessibility of various web designs quickly and at low cost. For example, future work might employ a panel of geographically dispersed people without sight. Researchers might run remote usability tests using WebEx or Placeware, saving the expense of transporting persons who cannot see to the usability laboratory.

Future accessibility research could address a number of issues to help ensure that Federal statistical web sites are designed so that people who cannot see can use them easily. The data collected in usability research—the participants' speed, accuracy, and satisfaction—appear to be applicable to

accessibility research as well. Researchers can continue to explore how the size of online forms and tables affects the performance of people who cannot see. Researchers may also explore how people who use screen readers can be helped by various notations and descriptions of online forms and tables, hidden to sighted users but read aloud by screen readers. Research could be done on the location of the data item to be found; the item could be, for example, in a random position of a group of drop-down menus instead of always in the middle as it was in this study. The drop-down menus used in the research could be varied to explore how the nature of the menus affects the results. Research could also explore how interactive web pages might help users without sight locate the information that they seek. Statistical web sites increasingly employ Online Analytical Processing (OLAP). Research is needed to ensure that people without sight can use OLAP resources easily.

Future research might involve a nationally representative sample of web users who require assistive technology. The challenge is identifying a suitable frame for such a sample. We are aware of no exhaustive list of such web users, and their low incidence in the population makes traditional screening approaches, such as through an area probability or random-digit-dial frame prohibitively expensive. One possible source for a nationally representative sample would be an existing large Federal survey such as the National Health Interview Survey (NHIS), which the Census Bureau conducts for the National Center for Health Statistics. The 2005 NHIS contained the item “What condition or health problem causes you to have difficulty?” with the response alternative “vision problem or problem seeing.” The NHIS respondents who indicated that they had a problem seeing could be contacted and asked whether they use the web, and if so, whether they need assistive technology. The NHIS employed a nationally representative sample of 98,649 in 2005, with a response rate of 86.5 percent. The survey, therefore, could yield a relatively large number of Internet users who require a screen reader. There is precedent for using NHIS respondents in additional surveys: respondents for the Medical Expenditure Panel Survey are recruited from NHIS respondents. Future research on web users without sight might also use a dual-frame design, which is commonly used for enhancing sample sizes for rare populations. For example, the NHIS respondents with visual problems could serve as one frame and a list of people without sight could serve as the second frame. The American Foundation for the Blind has expressed an interest in working with Westat and the Census Bureau in future research, and might be able to provide such a list.

The overarching goal of future research should be to establish a cost-efficient, valid technique for web designers to collect objective data quickly. With these data, designers could confidently decide upon the optimal accessible designs for their web pages.

## 5 REFERENCES

- American National Standards Institute (2001) *Common industry format for usability test reports*. New York: Author.
- Amtmann, D., Johnson, K., & Cook, D. (2002) Making web-based tables accessible for users of screen readers. *Library Hi Tech*, 20, 221-231.
- Barnicle, K. (2000) *Usability testing with screen reading technology in a Windows environment*. Paper presented to the Technology and Persons with Disabilities conference, Los Angeles.
- Dumas, J.S. & Redish, J. (1999) *A practical guide to usability testing*. Portland, OR: Intellect.
- General Services Administration, Center for IT Accommodation (2004) *Designing accessible websites*. Online training available at [www.section508.gov](http://www.section508.gov).
- Gerber, E. (2002) *Conducting usability testing with computer users who are blind or visually impaired*. Paper presented to the Technology and Persons with Disabilities conference, Los Angeles.
- Gerber, E. & Kirchner, C. (2001) Who's Surfing? Internet access and computer use by visually impaired youths and adults. *Journal of Visual Impairment & Blindness*, 95, 176-181.
- Harper, F.M. (2004) Data warehousing and the organization of governmental databases. In Pavlichev, A. & Garson, G.D. *Digital government: Principles and best practices*. Hershey, PA: IRM Press, pp. 236-247.
- Koyani, S.J., Bailey, R.W., & Nall, J.R. (2003) *Research-based web design and usability guidelines*. Washington, DC: Communication Technologies Branch, National Cancer Institute.
- Lenhart, A., Horrigan, J., Rainie, L., Allen, K., Boyce, A., Madden, M., O'Grady, E. (2003) *The ever-shifting Internet population: A new look at Internet access and the digital divide*. Washington, DC: Pew Internet and American Life Project.
- National Center for Health Statistics (2002) *Summary health statistics for the U.S. population: National Health Interview Survey, 2002*. Washington, DC: Author.
- National Telecommunications and Information Administration (2002) *A nation online: How Americans are expanding their use of the Internet*. Washington, DC: Author.
- Neilsen, J. (1993) *Usability engineering*. Boston: AP Professional.
- Office of Management and Budget (2004) *Statistical programs of the United States government*. Washington, DC: Author.
- Paciello, M.G. (2000) *Web accessibility for people with disabilities*. Lawrence, Kansas: CMP Books.

Slatin, J. & Rush, S. (2003) *Maximum accessibility: Making your web site more usable for everyone*. Boston: Addison Wesley.

Thatcher, J., Bohman, P., Burks, M., Lawton Henry, S., Regan, B., Swierenga, S., Urban, M.D., & Waddell, C.D. (2002) *Accessible web sites*. Birmingham, U.K.: Glasshaus.

Theofanos, M.F., & Redish, J. (2003, November/December) Guidelines for accessible—and usable—web sites: Observing users who work with screenreaders. *Interactions*, pp. 38-51.



**Table 1. Content of the eight drop-down menus in the online forms**

<b>County</b>	<b>Gender</b>	<b>Age</b>	<b>Race</b>	<b>Ethnicity</b>	<b>Ownership</b>	<b>Quarters</b>	<b>Year</b>
All	All	All	All	All	All	All	2005
Allegany	Male	14-18	White	Hispanic or Latino	Owner-occupied units	First	2004
Anne Arundel	Female	19-21	Black or African American	Not Hispanic or Latino	Renter-occupied units	Second	2003
Baltimore		22-24	American Indian or Alaska Native			Third	2002
Calvert		25-34	Asian			Fourth	2001
Carroll		35-44	Native Hawaiian or Pacific Islander				2000
Charles		45-54	Some other race				
Dorchester		55-64					
Frederick		65-74					
Harford		75 & over					
Howard							
Kent							
Montgomery							
Prince George's							
Queen Anne's							
St. Mary's							

Table 2. An online form with eight drop-down menus.

## Select individual characteristics

County

Ethnicity

Gender

Ownership

Age

Quarters

Race

Years

**Table 3. The online table with 15 rows and 15 columns**

Number of people employed by occupation and by county in Maryland															
	Agriculture	Mining	Utilities	Construction	Manufacturing	Wholesale Trade	Retail Trade	Transportation	Information	Finance	Real Estate	Professional Services	Management	Administrative Services	Education Services
Allegany	4230	16	84833	2037	543	832	53	1541	57983	9999	6261	813	7320	66	28339
Anne Arundel	988	147	24638	54616	1173	964	114	27	51	6740	1496	190	647	655	192
Baltimore	6477	939	33	48	2644	6647	341	34	87	17	23668	3655	8009	45	7858
Calvert	229	66000	87028	79	91025	25623	52436	27	644	80	499	86	775	9119	7111
Carroll	925	8049	5731	36	43154	640	7747	17408	5726	1507	31847	171	5799	508	77
Charles	997	878	148	640	7837	8415	6082	2730	991	133	6877	5575	568	966	492
Dorchester	6468	88	44	98	2145	822	92	839	8042	7564	56	48	87	37	934
Frederick	75	77	870	337	6477	449	95	503	46	821	8454	7614	468	63	94
Harford	2866	42	519	514	348	960	69	75	26	41	5780	836	53	667	9894
Howard	993	625	4290	13	2989	28	154	311	708	906	7063	3269	31	585	2105
Kent	3822	96	854	796	7556	70	84	9540	44	386	44	47	32	63	715
Montgomery	9253	86	469	77	655	85	406	946	6903	452	2692	479	656	81	5091
Prince George's	7296	1401	9092	858	198	93	6794	78	38	290	585	83	810	3679	26
Queen Anne's	39	4300	58	7500	81	71	75	482	51	747	682	197	762	34	540
St. Mary's	707	769	52	48	646	83	34	16	516	97	958	99	44	529	680

**Table 4. Number of errors committed by 17 participants in the tasks involving online tables. (Note: The data exclude three participants who did not complete any of these tasks successfully.)**

<b>Rows</b>	<b>Columns</b>	<b>Number of Errors</b>
2	2	1
2	15	5
15	2	3
15	15	4

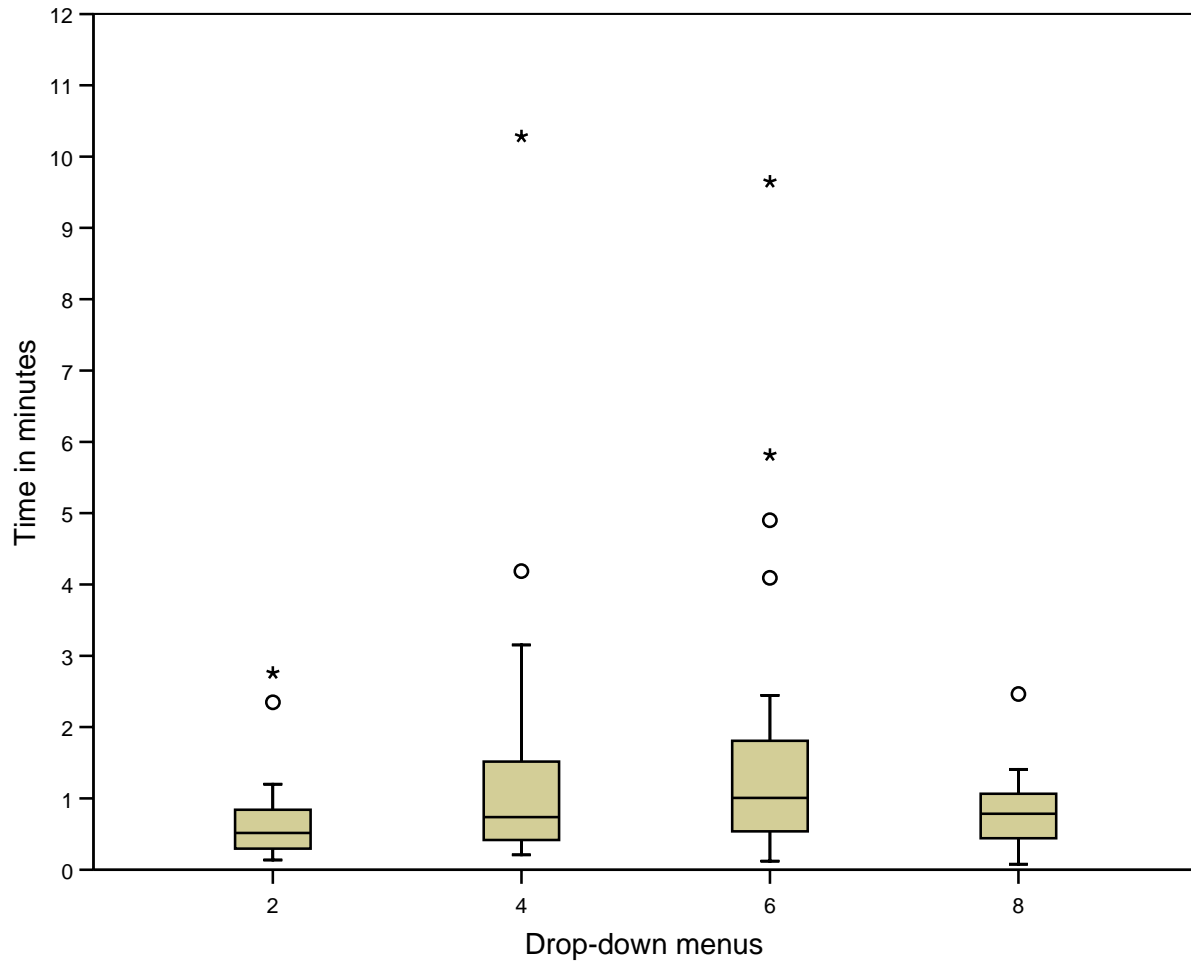


Figure 1. Box plots for the estimated time in minutes required by the 19 participants to find the requested item on the forms having two, four, six, and eight drop-down menus.

Note—In box plots, the black line indicates the second quartile (i.e., the median). The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.

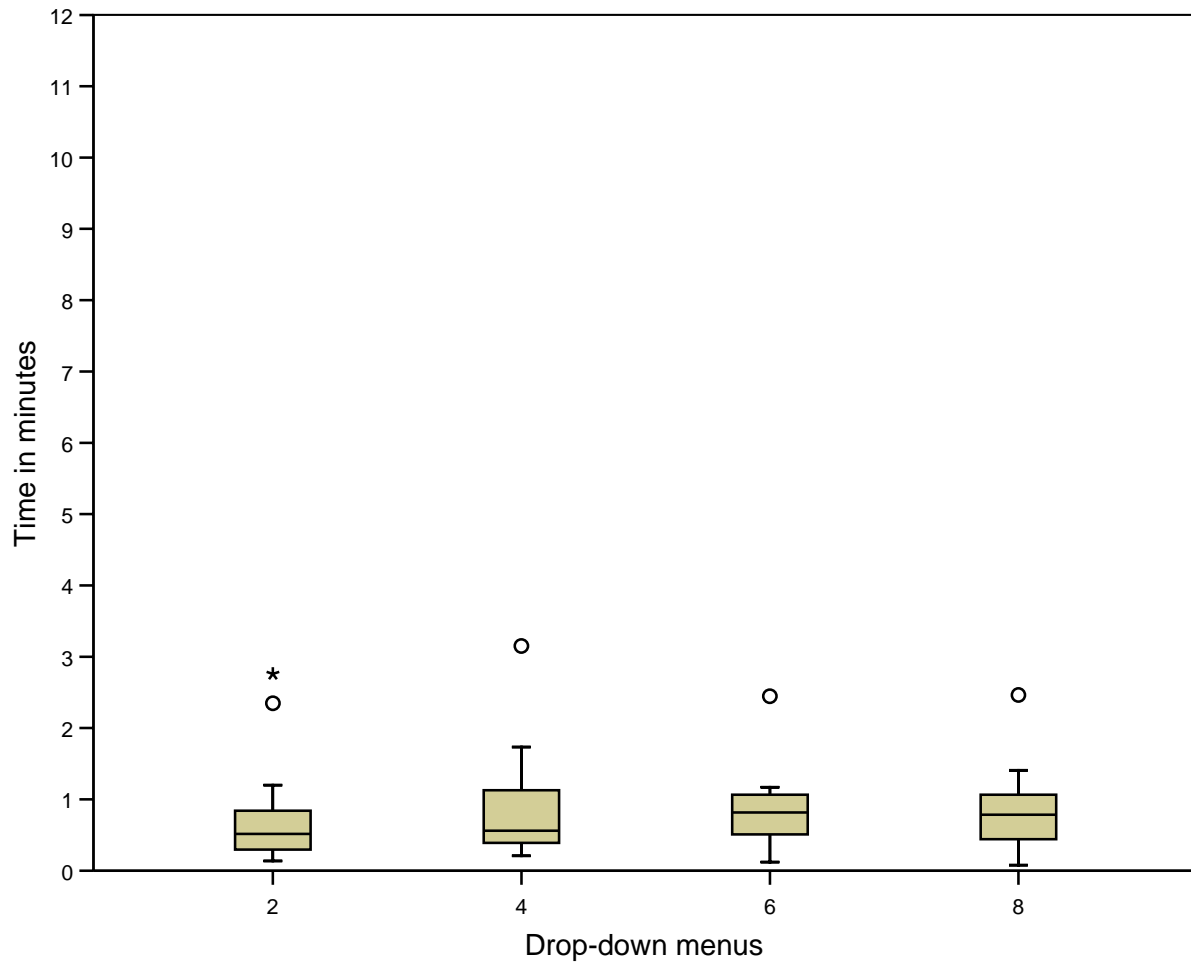


Figure 2. Box plots of the estimated time in minutes required by the 19 participants to find the requested item on the forms having two, four, six, and eight drop-down menus, excluding the 6 tasks in which the times exceeded 200 seconds.

Note—In box plots, the black line in the box indicates the second quartile (i.e., the median). The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.

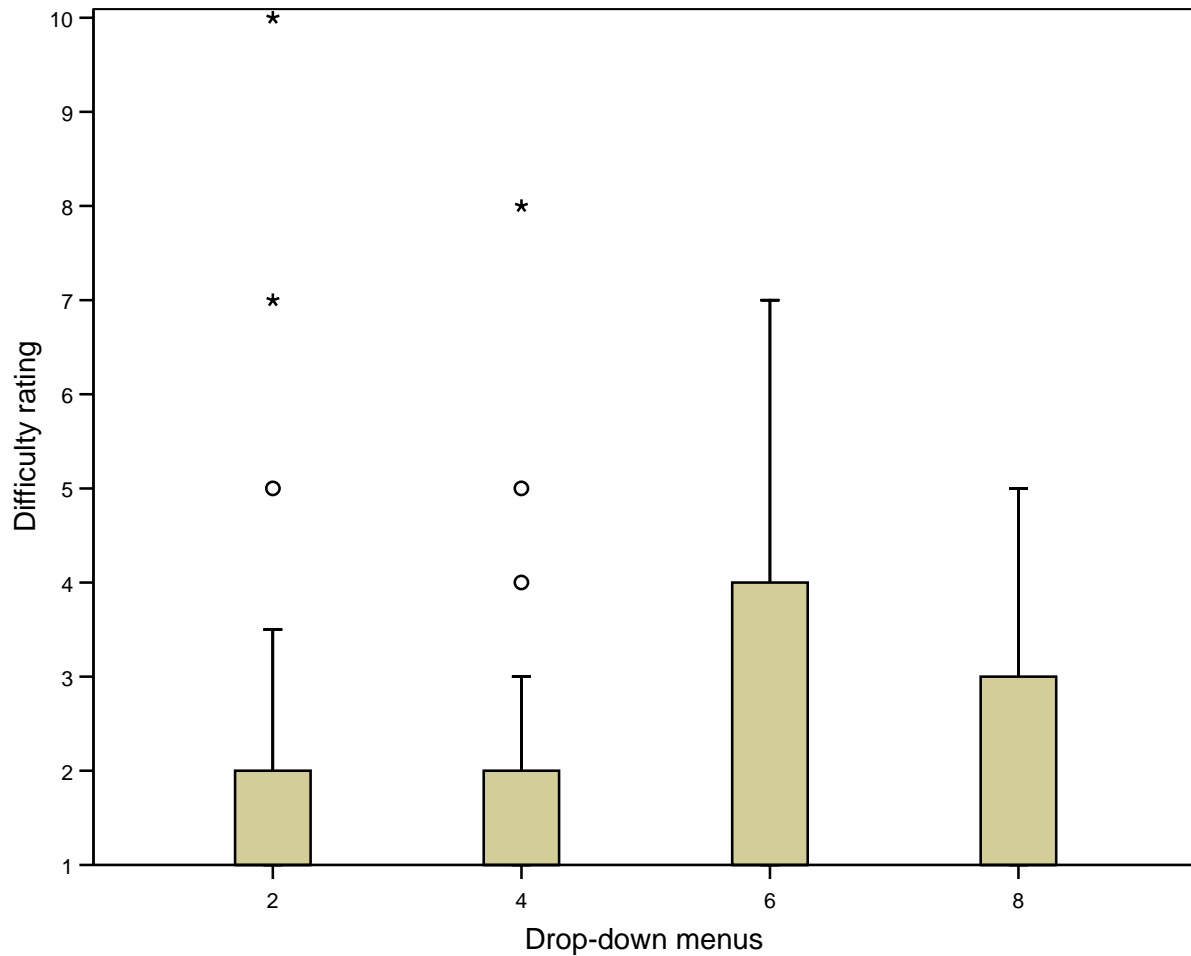


Figure 3. Box plots of the 19 participants' mean ratings, on a 10-point scale, of the difficulty of the tasks when the forms had two, four, six, and eight drop-down menus.

Note— A rating of 1 signifies “very easy,” while a rating of 10 signifies “very hard.” In box plots, the black line in the box indicates the second quartile (i.e., the median). In all four boxes in this figure, that line coincides with the bottom of the box. The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.

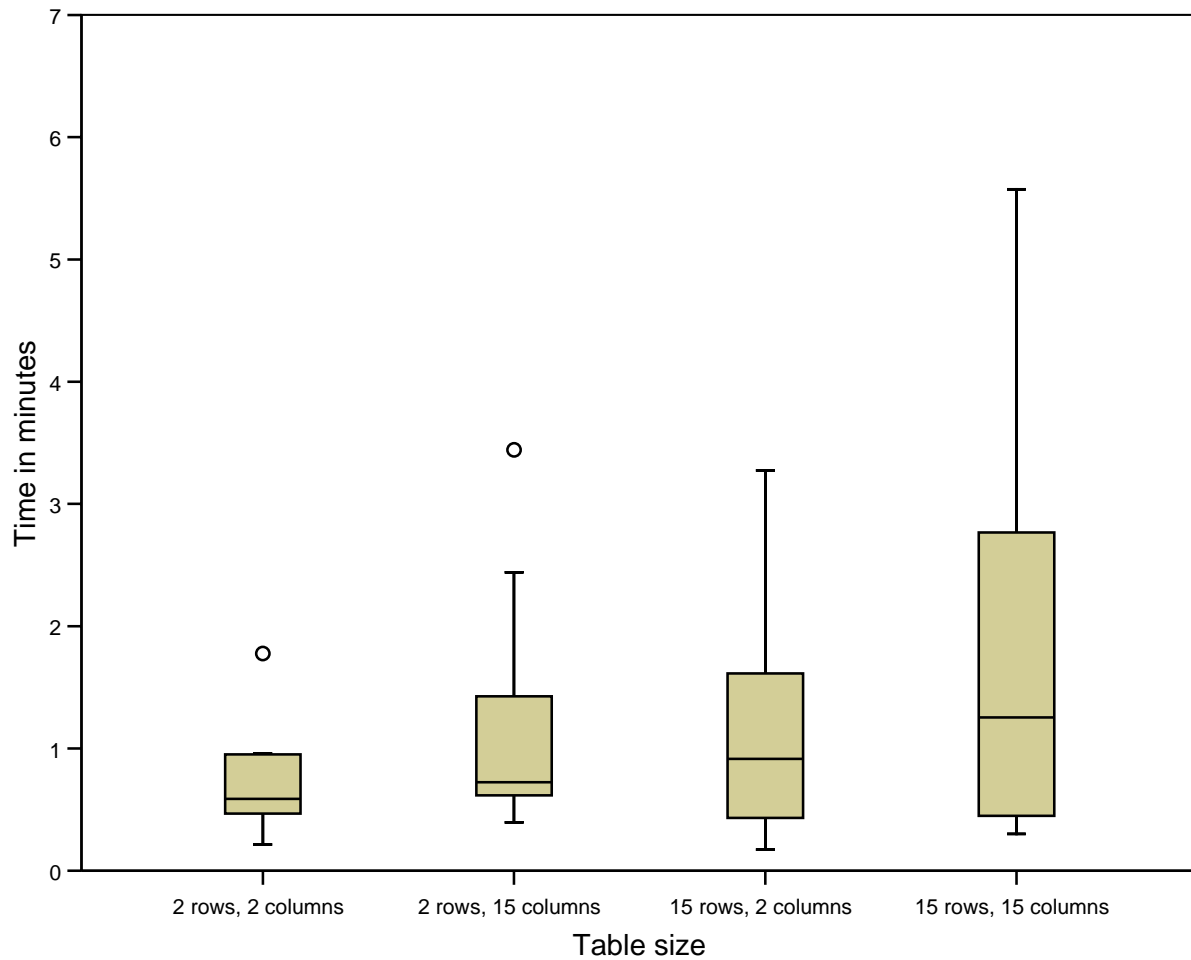


Figure 4. Box plots of the estimated time in minutes required by the 10 participants who found the items in the tables correctly in all four tasks.

Note—In box plots, the black line in the box indicates the second quartile (i.e., the median). The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.



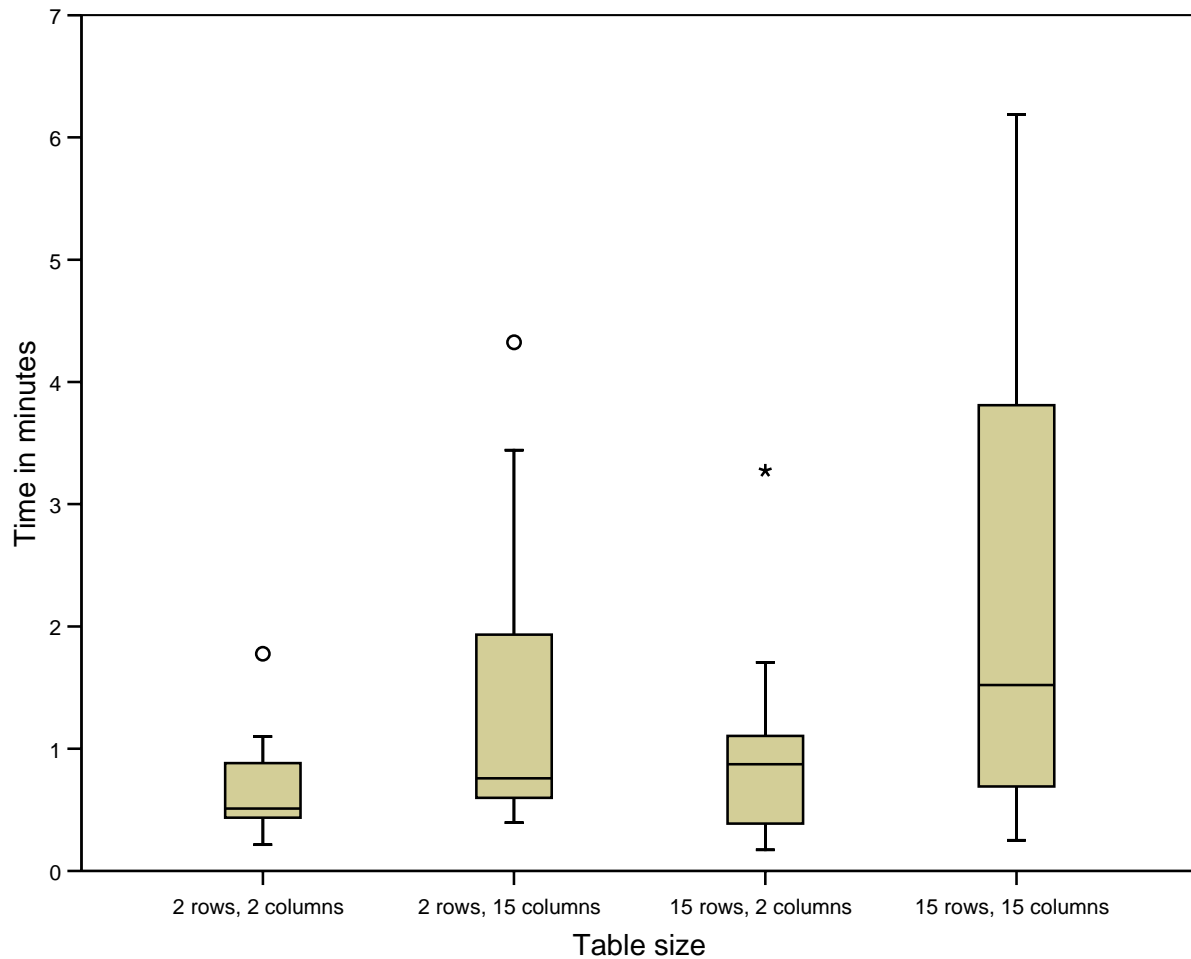


Figure 5. Box plots of the estimated time in minutes required by the 15 participants who found items in the tables within the allowable time in all four tasks, regardless of whether the items were correct or incorrect.

Note—In box plots, the black line in the box indicates the second quartile (i.e., the median). The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.

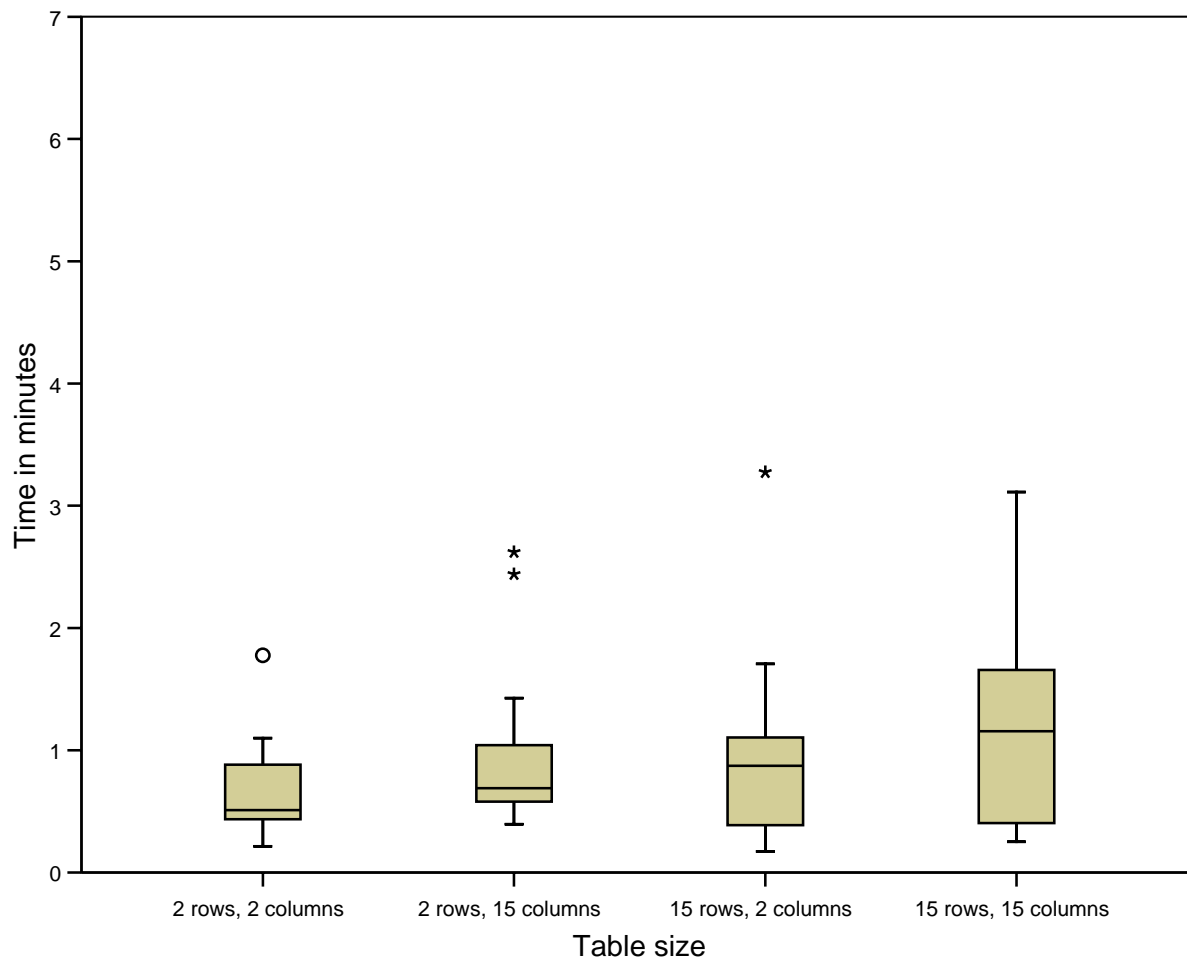


Figure 6. Box plots of the estimated time in minutes required by the 15 participants who found items in the tables within the allowable time in all four tasks, regardless of whether the items were correct or incorrect. The six tasks in which a participant required more than 200 seconds were excluded from these calculations.

Note—In box plots, the black line in the box indicates the second quartile (i.e., the median). The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.

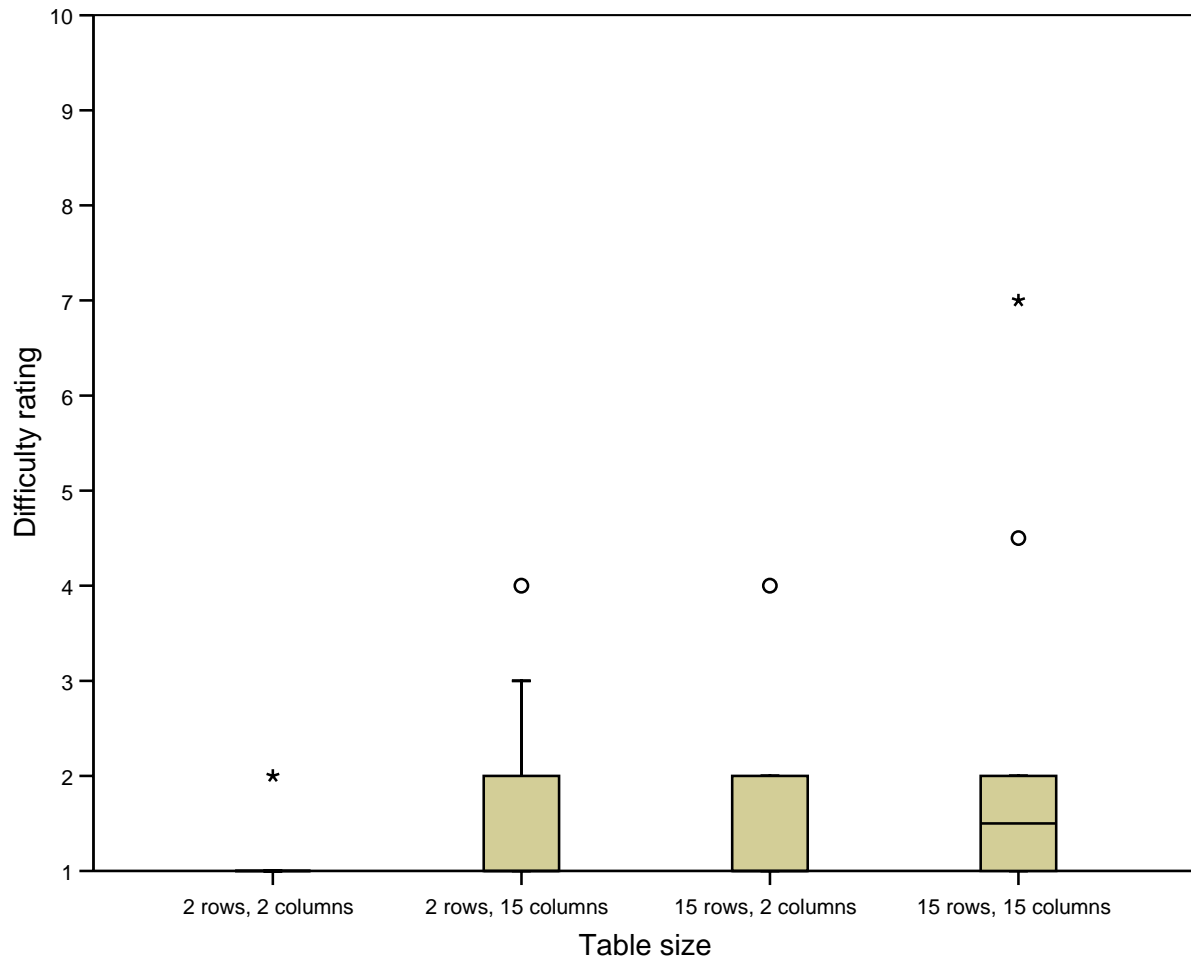


Figure 7. Box plots of the participants' ratings, on a 10-point scale, of the difficulty of the tasks with 2-by-2, 2-by-15, 15-by-2, or 15-by-15 tables. These data are for the 10 participants who found the items in the tables correctly in all four tasks.

Note— A rating of 1 signifies “very easy,” while a rating of 10 signifies “very hard.” In box plots, the black line in the box indicates the second quartile (i.e., the median). In two boxes in this figure, that line coincides with the bottom of the box. One box could not be drawn because all responses except one were “1.” The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.

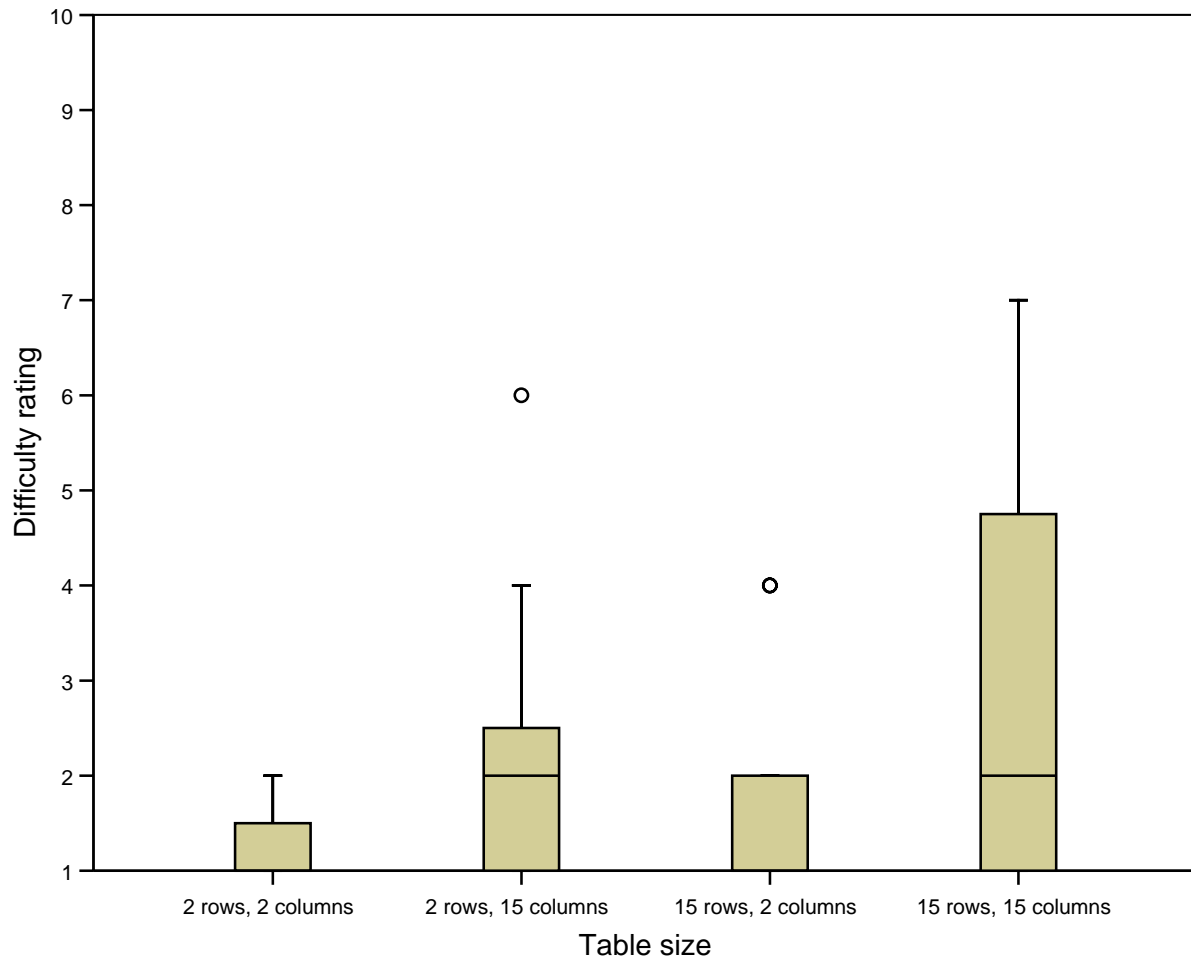


Figure 8. Box plots of the participants' ratings, on a 10-point scale, of the difficulty of the tasks with 2-by-2, 2-by-15, 15-by-2, or 15-by-15 tables. These data are for the 15 participants who found items in the tables within the allowable time in all four tasks, regardless of whether the items were correct or incorrect.

Note— A rating of 1 signifies “very easy,” while a rating of 10 signifies “very hard.” In box plots, the black line in the box indicates the second quartile (i.e., the median). In all two boxes in this figure, that line coincides with the bottom of the box. The inter-quartile range, depicted by the box, is calculated by subtracting the first quartile from the third quartile. The whiskers show the minimum and maximum values that are not outliers. Open circles represent mild outlier values which lie between 1.5 and 3.0 times the inter-quartile range above the third quartile. Stars represent extreme outlier values which lie more than 3.0 times the inter-quartile range above the third quartile.