

Usability Evaluation: ATUS Data Collection Instrument 1

Note: Permission has been granted by the student to post this work on the website. No part of this work may be copied or used. This applies to this and all sample assignments posted on this site.

–Dr. Dringus

Usability Evaluation:
American Time Use Survey Data Collection Instrument
Diane E. Herz
MMIS 680: Human Computer Interaction
February 10, 2002

Abstract

In this paper, the author evaluates a prototype of the diary portion of the American Time Use Survey's proposed data collection instrument. Three factors are evaluated: consistency, compatibility of data entry fields to data entry required, and efficiency. The text provides summary evaluations, and details are included in the appendices. Recommendations for usability improvements are included.

Table of Contents

	Page
Introduction	4
Evaluation Methods and Factors	5
Evaluation of ATUS Diary Section	7
Recommendations	11
References	14
Tables and Figures	
Table 1. Selected Evaluation Features	6
Table 2. Property Checklist Summary Ratings for Consistency	8
Table 3. Scenario Tests and Summary Ratings for Compatibility	9
Figure 1. Diary Screen Shot—WHO field	8
Figure 2: Error Message 10	11

Introduction

In December 2000, Congress approved funding for the Bureau of Labor Statistics (BLS) to conduct a new survey to measure how people spend their time. The American Time Use Survey, or ATUS, will begin in January 2003, will run continuously, and will be conducted by telephone. Census Bureau interviewers will simultaneously interview respondents and enter data about the respondent's activities during the prior day into a windows-based graphical user interface--the ATUS data collection instrument.¹ This usability evaluation focuses on the portion of the ATUS data collection instrument that is designed to collect information about users' activities during the prior 24 hours (the "diary" portion.) This evaluation can be considered a prototype evaluation—as the software is still being modified prior to introduction in 2003. (Jordan, 1998).

The data collection (and the related activity coding application) is built in [Blaise](#), a windows-based graphical user interface (GUI) developed by [Statistics Netherlands](#) and distributed in the U.S. by [Westat, Inc.](#) The [Census Bureau](#) has recently chosen Blaise as the standard for software development in the agency. The organization is beginning to work with survey sponsors to build new instruments and to convert current DOS-based instruments--used in both Computer-Assisted Telephone Interviews (CATI) and Computer Assisted Personal Interviews (CAPI)--to Blaise.

The ATUS is the second survey that will use Blaise, and is the first CATI survey to do so. Hence, the Census Bureau is still early in the process of developing and refining Blaise user interface guidelines. Because both Blaise and ATUS are new to the CATI interviewers, all system users will be novice users when the survey goes to full production in 2003.

Evaluation Methods and Factors

The evaluation presented here is non-empirical, involving only the author as participant (Shneiderman, 1998; Jordan, 1998). The evaluation focuses on the core part of the ATUS data collection instrument—the “diary” portion. This evaluation was conducted using cognitive walkthrough style. The author knows many of the Census Bureau interviewers, the large majority of whom are high school graduates without any college. Most are familiar with the World Wide Web, but are only familiar with DOS-based survey collection instruments. With their education and experience in mind, the author reviewed software properties against property checklists (based on Census guidelines), and conducted scenario tests and task analyses (Jordan, 1998).

Three features were chosen for evaluation: consistency, compatibility with data collection, and efficiency. (See table 1.) Flexibility was also considered for evaluation, but was left out due to length of study requirements. Learnability and ease of use were deemed out of scope for a non-empirical research project, as they are better examined an empirical method, such as focus groups or user observations (Jordan, 1998).

All ATUS data collection instrument users will begin as novices. They will not have used the application, and none will have used instruments with GUIs to collect survey data; that is—all other surveys these interviewers conduct have DOS-based interfaces. In addition, none have ever conducted the ATUS interview. Eventually, many users will become experts, as the application will be continuously in use during the year-round ATUS. Still, interviewer turnover is high as a rule; thus designing for novice users is important. According to Shneiderman,

¹ After the interview is completed, data will be exported from that instrument into the ATUS activity coding application. Using the coding application, coders will assign an activity code to

applications for novice users should be designed so that the interface is simple and logically organized.” Simplicity requires consistency. Consistency is also the first of Shneiderman’s “Golden Rules” of interface design, and is mentioned by many other authors (Jordan, 1998; Jeffries et. al, 1991; Holzschlag, 2000; Grudin, 1989). For these reasons, consistence was chosen for this evaluation.

Other features as well as consistency were chosen because of their import in data collection instruments (like ATUS) and their ability to be evaluated using non-empirical methods. Shneiderman identifies the most important features for evaluation for data collection instruments as consistency of data entry transactions, minimal input actions by users, minimal memory load on users, compatibility of data entry with data display and flexibility for user control of data entry. Compatibility with data collection and efficiency were chosen for evaluation. Flexibility was considered, and is defined in Table 1, but is not included in the evaluation.

Table 1. Selected Evaluation Features

Feature	Description
Compatibility	Data entry fields should be compatible with the way that data are collected and provided (Shneiderman. 1998).
Internal consistency	<i>Internal consistency</i> refers to the consistent use of standards throughout the application. These can include graphical layout, action sequences, fonts, headers, naming conventions, and error messages (Shneiderman, 1998; Grudin, 1989). ²
Efficiency	An efficient system is one where minimal input action is required by users. Requiring minimal memory load on users can also improve efficiency. ³
Flexibility	Shneiderman (1998) indicates suggest that users should have some

each activity reported.

² Other types of consistency not evaluated here include consistency with other applications used by the same users and correspondence of interface features to familiar features of the world outside this user group (e.g. the web). (Grudin, 1989).

³ Systems can be too efficient, however. Jordan notes that using the number of keystrokes as a guide can lead to processes that are efficient but remove important steps--such as user confirmation that the task should occur. (Jordan, 1998).

	ability to manipulate the environment to suit their preferences. Adopting flexibility can be difficult to do while maintaining consistency. Sometimes it can be achieved by enabling the same action to be conducted using multiple paths. Some examples of flexible features include enabling users to use shortcuts, change the display, or use either the keyboard or the mouse to enter data or to use keystrokes or pull-down menus to perform an action.
--	--

Evaluation of ATUS “Diary” Section

The ATUS diary section of the interview focuses on filling in a grid of activities for 24 hours. Cognitive research shows that respondents tend to report about activities done yesterday in a largely sequential format (Stinson, 1999). However, recalling yesterday’s activities can be a difficult task, and respondents sometimes forget activities and need to go back in their day to “insert” them (Schwartz, 2002). They also can remember activities that occurred at some time later in the day, but forget those that occurred in between 2 salient events (Schwartz, 2002).

The diary’s main screen was designed to look like paper diaries that have been used successfully in time-use surveys done by other countries (Stinson, 1999). Blaise design enables up to 3 panes to display on each screen: the top pane includes question text; the middle pane, response options; and the 3rd pane, data entry. (See Figure 1.)

Consistency. To evaluate consistency, a property checklist was used and each element was examined separately for consistency with Census Bureau guidelines. (Jordan, 1998; Jeffries et. al, 1991; U.S. Census Bureau, 2001). Screen color, interviewer instructions, font size and color, data entry pane positioning, error message format, and variable name placements were evaluated for consistency. For each element, several specific items were examined (See Appendix A). To examine interviewer instruction consistency, for example, screens were examined to see if instructions were available, if they were written in blue text, if diamonds preceded them, and if they began with a verb.

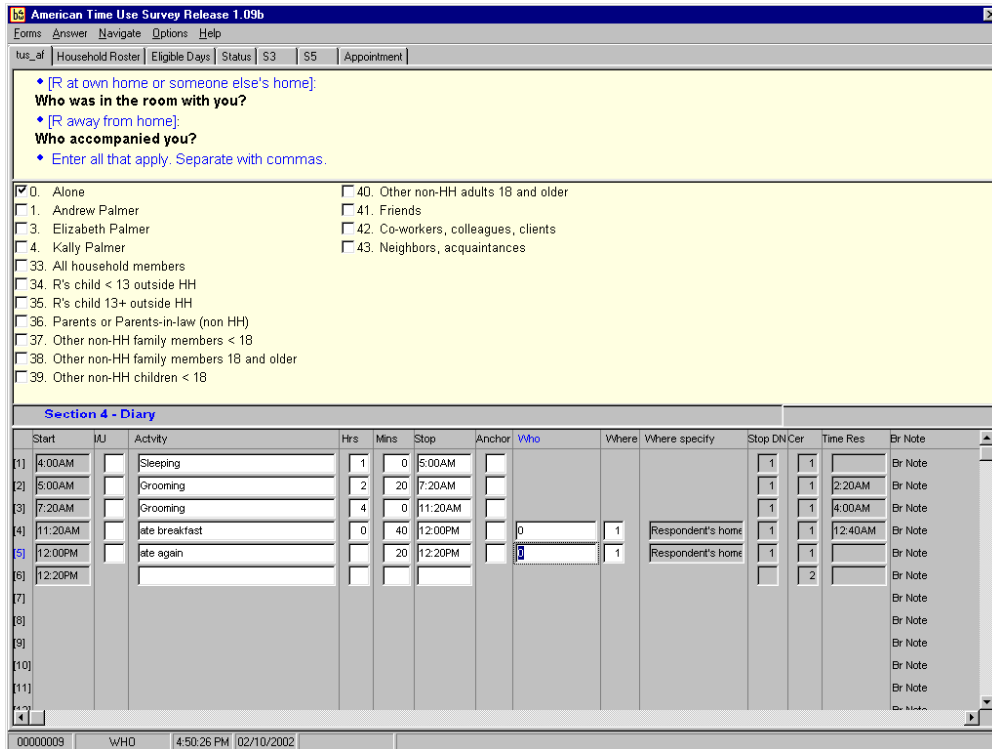


Figure 1. Diary Screen Shot—WHO field

Overall, the consistency rating for the diary section is 3.3 out of 5 (total score 20 divided by 6 elements). Screen colors, variable formats, and placements were consistent from screen to screen. Data entry pane placements were problematic, as response options were in the top pane on some screens and the middle pane on others. Error messages for out-of-range entries varied in text, format, and response. See Appendix A for specific reports.

Table 2. Property Checklist Summary Ratings for Consistency

Property Checklist	
Element	Rating
Screen color	☺☺☺☺☺
Interviewer instructions	☺☺☺
Font size and color	☺☺☺☺
Data entry pane	☺☺
Error message format for out-of-range variable	☺
Variable name available on bottom left	☺☺☺☺☺

Compatibility. To evaluate whether the data entry fields were compatible with the data collected, 5 scenarios were tested (Jordan, 1998). In each scenario, in addition to the activity entry changes, the author varied how activity duration was entered—sometimes providing hours and/or minutes and other times providing time of day. Who and where options were varied across possible selections. The overall rating of the ATUS data collection diary for compatibility was 2 out of 5 (12 divided by 6).

The diary fields were laid out well and seemed to have good “guessability” when the activities were entered sequentially or when a “forgotten” activity was inserted (Jordan, 1998). When the respondent changed reporting style for an activity from times (e.g. “I stopped watching TV at 7:00.”) to durations (“I don’t actually recall the time. I watched for an hour”), the instrument refused to accept the changes. The diary was least compatible with data collection when the interviewer tried to insert 2 activities at once, when there was a gap of time between the activity being reported and the next activity the respondent remembered, and when durations did not successfully overwrite stop times. (See Appendix B for detailed results.)

Table 3. Scenario Tests and Summary Ratings for Compatibility.

Scenario	Rating
1. Respondent reports activities sequentially, and remembers all of them, including the contextual information about who was with them and where they were during the activity.	☺☺☺☺☺
1. Respondent first reports activity times in hours and minutes. Then alters times by changing duration. Then same test reversed.	☺
Respondent forgets an activity and interviewer has to go back and insert it.	☺
Respondent remembers an activity he did in the middle of another one. Interviewer has to “sandwich” insert it.	☺☺☺
Respondent remembers something done later in the day, and the interviewer has to fill in the information.	☺
Respondent remembers that an activity occurred 2 days ago rather than yesterday, and interviewer must delete it.	☺

While error messages were not specifically targeted for compatibility testing, a notable problem emerged. An error screen appeared with options that did not correspond to the question text. The text in the message, initiated when a 3 was entered to undo insertion of a line, asks if the interviewer would like to “Proceed.” The response options are S for Suppress, C for Close and G for Goto. It’s not clear what any of these options will do. None of them actually executed the deletion.

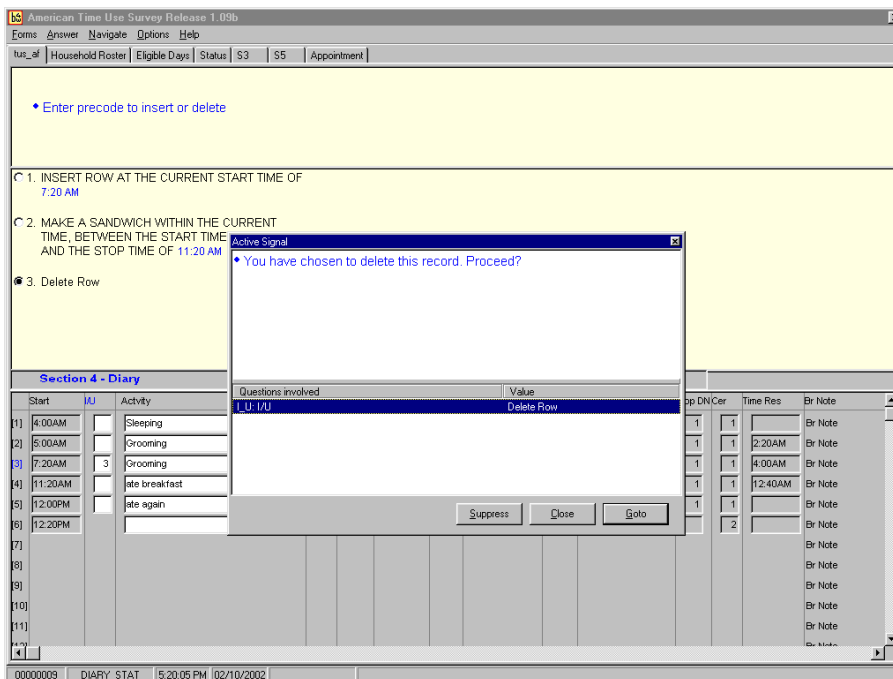


Figure 2: Error Message

Efficiency. To evaluate efficiency, the following tasks analyses were performed: entering a hard-coded activity, entering the duration of an activity, inserting a new activity, changing the duration of an activity, and completing a full 24-hour diary. The overall rating for efficiency was 3.6 (18 divided by 5).

The ATUS data collection instrument has some features that enhance efficiency. The cursor automatically moves to the right when the Enter key is hit, and STOP times are automatically computed when activity durations are entered (and vice versa in most cases). A

copy feature (F-12) enables interviewers to copy the entry in the field directly above the cursor to the current field; this is useful when respondents are with the same people or in the same location for extended periods of time.

The instrument also has some room for efficiency improvements. Fields that enable the interviewer to insert a line or “anchor” the stop time of a particular activity will be used in a small minority of activity rows. Currently, both fields are “on path” during a trip through the instrument. As a result, the interviewer needs to hit <Enter> 2 extra times per line. This could translate to about 50 extra keystrokes per interview, since the average number of activities reported is 27! In addition, the instrument unnecessarily lands on the Stop time of an activity even though the duration of the activity was already reported in hours and minutes, auto-filling the Stop time field.

Table 4. Tasks Analyzed and Ratings for Efficiency

Task	Rating
Enter a hard-coded activity	☺☺☺☺☺
Enter duration of activity, in hours and minutes	☺☺
Insert an activity	☺☺☺☺☺
Change the duration of an activity	☺☺☺
Complete a full diary of 24, 1-hour activities.	☺☺☺

Recommendations

The ATUS instrument performed moderately well in the areas of consistency, compatibility, and efficiency. A number of recommendations emerge from the tests conducted. In order to improve both user trust (Shneiderman, 2000), and interface consistency in the ATUS instrument, the following changes are recommended:

- Ensure that every screen includes interviewers instructions (unless irrelevant).

Specifically, add instructions to screens for the HRS, STOP, WHERE variables.

- Standardize range error formats. They should have the same text (e.g. “Entry is invalid”) and should provide the range of acceptable values.
- Standardize response options. They should be placed in the middle pane and have the same text format and radio button format, consistent with Census guidelines. (Census, 2001).
- Rename the I/U (insert/uninsert) column I/D (insert/delete) to make title consistent with field options.

In order to make the instrument design more compatible with data collection, it is recommended that observable tests be conducted with interviewers and respondents to test compatibility with various types of reporting styles. Specific recommendations for improved compatibility include:

- Enable an hours/minutes duration entry to recompute a new stop time and overwrite the previous STOP time for the same activity.
- Enable entry of start time of interrupting activities, or instruct interviewer how to enter stop time of previous activity, during a sandwich insert.
- Enable entry of a future activity that occurs later in the day but remains out of the time computation until the activities being reported “catch up” to it.
- Change the edit screen when an activity is deleted (and also other similar screens) to provide options consistent with the question (preferably Yes/No options).

Finally, recommendations to improve efficiency include the following:

- Take the I/U and Anchor paths “off path.” Develop a simple method for invoking them when needed.
- Take the STOP time off path when hours and minutes have been entered for an activity. Provide a functionality to overrule this, in case the respondents change the way they report duration for the activity.

References

- Jordan, P (1998). An Introduction to Usability. Philadelphia, PA: Taylor and Francis, Ltd.
- Grudin, J. (1989, October). The Case Against User Interface Consistency. Communications of the ACM, 32 (10), 1164-73.
- Holzschlag, M. (2000, February). In Search of the Elusive Interface. Web Techniques, 5 (2), 24.
- Jeffries, R., Miller, J, Wharton, C. and Uyeda, K. (1991). User Interface Evaluation in the Real World: A Comparison of Four Techniques. Communications of the ACM.
- Schwartz, L. (2002, forthcoming). Response to Winkler: measuring intrahousehold allocation of time. Monthly Labor Review. 125 (3).
- Shneiderman, B. (2000, December). Designing Trust Into Online Experiences. Communications of the ACM, 43 (12), 57.
- Shneiderman, B.(1998). Designing the User Interface. Boston, MA: Addison Wesley Longman, Inc.
- Stinson, L. (1999, August). Measuring how people spend their time: a time-use survey design. Monthly Labor Review, 122 (8), 12-19.
- U.S. Census Bureau. (2001). Guidelines for development of Blaise user interfaces. Internal document.

Appendix A.

Consistency was evaluated using a property checklist/guideline evaluation (Jeffries et. al., 1991). Features were evaluated based on whether the met Census Bureau Blaise interface development guidelines (U.S. Census Bureau, 2001). Table A-1 includes detailed results of that evaluation.

Table A-1. Property Checklist for Consistency Evaluation.

Property Checklist	
Element	Results and Rating
Screen colors 1. Top pane=muted yellow 2. Middle pane=muted yellow 3. Diary pane=Blue with yellow open entry fields	1-3. Colors correct on all screens. RATING: HIGH (5)
Interviewer instructions 1. Always preceded by a diamond 2. Available on every page 3. Always in blue with first word capitalized 4. Always begin with verb	1. Instructions were always preceded by a diamond (8/8) 2. Some (5/8) fields had interviewer instructions. No instructions were provided for entry in HRS, STOP, or WHERE fields 3. Instructions always in blue (5/5) 4. Most (4/5) preceded with verb RATING: MED (3)
Font size and color 1. Arial font all panes, same size 2. Variable names all CAPS 3. Question text in bold black 4. Options in regular black, first word capitalized	1. Arial font on all panes. 2. Variable names all capitalized 3. Question text always in bold black. Question text missing from MINDUR 4. Response options for DAIRY_STAT are capitalized RATING: MED_HIGH (4)
Data entry panes 1. Top pane=question text 2. Middle pane=response options 3. Bottom pane=data entry 4. Response option radio buttons 5. Response options text format consistent	1-2. Top pane for ACTIVITY includes response options. Top pane for WHERE includes titles. Should go on middle pane. Other panes correct. 3. Bottom pane always data entry. 4. ACTIVITY options have no radio buttons; WHO has checkboxes. 5. ACTIVITY includes equal signs after numbers. WHO and WHERE have periods. RATING: LOW-MED (2)
1. Error message format for out-of-range variable	1. 3/8 fields showed message box with same text indicating valid range.

Usability Evaluation: ATUS Data Collection Instrument 16

	<p>ACTIVITY didn't produce error message; entered "Activity X" STOP converted 13AM to 1:03AM</p> <p>DIARY_STAT converted "D" to correct entry "A" Where (text field) accepted numbers RATING: LOW</p>
Variable name available on bottom left	<p>All screens meet</p> <p>RATING: HIGH</p>

Appendix B.

Compatibility of the ATUS instrument to data collection was examined using scenario testing. Each scenario was run by a system designer and evaluator. Scenario descriptions and results are included in Table B-1.

Table B-1. Compatibility Evaluation Scenarios and Results

Scenario	Result
<p>1. Respondent reports activities sequentially, and remembers all of them, including the contextual information about who was with them and where they were during the activity.</p>	<ul style="list-style-type: none"> Instrument accepted all activities—both “hard-coded” options and text entries without a problem in ACTIVITY field. <p>RATING: HIGH</p>
<p>1. Respondent first reports activity times in hours and minutes. Then alters times by changing duration 2. Same test, reversed</p>	<p>This seemed to work in one direction only. 1. If the respondent first reports in times and then the interviewer changes it using minutes or hours, the row will not accept the data and recompute. 2. If respondent reported in hours and durations, and then interviewer overwrites it using times, the instrument allows the change.</p> <p>RATING: LOW (Showstopper)</p>
<p>Respondent forgets an activity and interviewer has to go back and insert it.</p>	<p>The insert feature works very well when only one activity is inserted at a time. If an interviewer tries to insert 2 activities by entering 1 in the I/U field, the instrument produces an error screen. The error screen does not say “Insert only 1 activity at a time.” Rather, it says, “The valid range is 1 to 3”. It is impossible to get out of this field at that point without hitting 3 (undo), to continue the interview. 3 does not “undo” the first insert. The interviewer has entered 1 originally, so the error screen is a problem.</p> <p>RATING: LOW (Showstopper)</p>
<p>Respondent remembers an activity he did in the middle of another one. Interviewer has to “sandwich” insert it.</p>	<p>The sandwich insert option (I/U=2) does seem to work. However, the interviewer is forced to enter the time the first part of the broken-up activity stopped rather than the time the interrupting activity began. (START is a read-only field). This may not be intuitive for users, but will need to be tested with them to know.</p>

	<p>RATING: MEDIUM</p>
<p>Respondent remembers something done later in the day, and the interviewer has to fill in the information.</p>	<p>There appear to be 2 ways that an interviewer can enter a future activity. (E.g. The respondent forgets what happened 7:20-8:00 but knows what happened at 8.) The interviewer can enter a false duration for the later activity of 1 hour and 40 minutes, then insert an activity before it and call it 40 minutes (or whatever amount). This could produce errors in duration. Or, they could enter the “gap” of 40 minutes from the current to the future activity, then that activity. But, the interviewer would have to be able to compute the length of that time on the fly and that’s unlikely to be possible during a difficult interview and by this group of interviewers.</p> <p>RATING: LOW</p>
<p>Respondent remembers that an activity occurred 2 days ago rather than yesterday, and interviewer must delete it.</p>	<p>This function did not work. It produced a “hard-edit” that said “You have chosen to delete this record. Proceed?” The options are Suppress, Close, and Goto. None of these executed the edit.</p> <p>RATING: LOW</p>

Appendix C

Efficiency of the ATUS instrument was examined using task analyses. Table C-1 describes the tasks analyzed. Efficiency evaluations for each task are also presented in the table.

Table C-1. Tasks Analyzed for Efficiency

Task	Efficiency
Enter an hard-coded activity	<p>The activity is entered efficiently, with only a number and the Enter or Tab key. The instrument automatically advances the cursor to the next field.</p> <p>RATING: HIGH (5)</p>
Enter duration of activity, in hours and minutes	<p>The instrument enters hours with one digit requirement only. It automatically advances to minutes, and accepts up to 120, converting hours automatically if the number is over 90.</p> <p>However, the instrument then lands on STOP time. This does not seem necessary if the duration has already been entered using hours and minutes. It could lead the interviewer to ask a redundant question, and it does force the interviewer to hit enter again to advance.</p> <p>Then, the cursor advances to the Anchor field. The interviewer will use this only rarely, so it should probably not be “on path” for every activity, but rather be optional like insertions. This added keystroke on every activity could lead to a great deal of inefficiency.</p> <p>RATING: LOW-MED (2)</p>
Insert an activity	<p>Inserting an activity requires only 2 extra keystrokes compared to moving sequentially. The interviewer needs to enter 1 for insert and then enter to accept. The <Enter> could be removed from the sequence, but it’s probably a good check on the interviewer’s intention. If it weren’t there, erroneous insertions could occur.</p> <p>RATING: HIGH (5)</p>
Change the duration of an activity	<p>Changing the time of an activity is very fast. One needs only enter a new time on top of the old time, and the instrument advances to the next field.</p> <p>Changing the duration using hours and minutes has the same problems as noted above—once entered, the</p>

Usability Evaluation: ATUS Data Collection Instrument 20

	<p>interviewer lands on STOP time, which seems redundant.</p> <p>RATING MEDIUM. (3)</p>
<p>Complete a full diary of 24, 1-hour activities.</p>	<p>Both the I/U and Anchor fields are “on path” for every activity entry. These will likely be used only sporadically, so it seems very inefficient to require them on every entry. Rating=2</p> <p>A copy feature makes repeated trips through the Who and Where field much more efficient. The interviewer can use the F-12 key to copy the Who and Where information from the prior field. Also, these fields don’t have to be filled until the end of the interview, so it can be done “all at once” enabling flexibility and efficiency. Rating=4.</p> <p>RATING: MEDIUM (3)</p>