Paper to Electronic Questionnaires: Effects on Structured Questionnaire Forms

Anna Trujillo[1]

NASA Langley Research Center
MS 152
Hampton, VA 23681 USA
anna.c.trujillo@nasa.gov

Abstract. With the use of computers, paper questionnaires are being replaced by electronic questionnaires. The formats of traditional paper questionnaires have been found to affect a subject's rating. Consequently, the transition from paper to electronic format can subtly change results. The research presented begins to determine how electronic questionnaire formats change subjective ratings. For formats where subjects used a flow chart to arrive at their rating, starting at the worst and middle ratings of the flow charts were the most accurate but subjects took slightly more time to arrive at their answers. Except for the electronic paper format, starting at the worst rating was the most preferred. The paper and electronic paper versions had the worst accuracy. Therefore, for flowchart type of questionnaires, flowcharts should start at the worst rating and work their way up to better ratings.

Keywords: electronic questionnaires, Cooper-Harper controllability rating, questionnaire formats.

1 Introduction

Paper questionnaires are slowly being replaced by electronic questionnaires. Respondents' ratings, though, may subtly change when using an electronic format [1, 2]. This research begins to determine how electronic questionnaire formats change subjective ratings from the traditional paper formats; in particular, how electronic formats may affect responses to a structured, flowchart type of questionnaire.

1.1 Background

Previous research found that ratings change depending on the electronic format of a traditional paper questionnaire [1, 2]. Furthermore, an electronic version of the NASA-TLX questionnaire had a higher workload rating associated with it [3]. Even

---

with the traditional paper formats of questionnaires, the format may affect a subject's ratings [4, 5].

For this experiment, subjects used the Cooper-Harper (CH) Controllability Rating Scale [6, 7] on a control task that required them to keep a randomly moving target centered. Subjects were told that desired performance was maintaining the target in the inner portion of the screen while adequate performance was maintaining the target in the middle portion of the screen (Fig. 1). Each rating was also described to the subjects with respect to the control task.
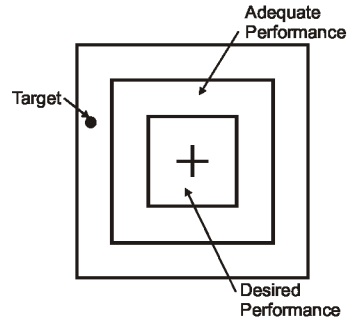

Fig. 1. Target Tracking Task with Indicated Desired and Adequate Performance

1.2 Objective

The objective of this research was to determine whether electronic formats of paper questionnaires change subjects' ratings. In particular, how electronic formats may affect responses to a structured, flowchart type of questionnaire.

2 Experimental Variables

2.1 Subjects' Piloting Experience

Twenty people participated as subjects. Ten were certificated pilots with at least a current Private Pilot license [8]. The rest of the subjects were non-pilots. The average age of the pilots was 48 years and the average age of the non-pilots was 40 years. The pilots averaged 22 years of piloting experience and they had an average of 7314 hrs of total piloting time.

2.2 Cooper-Harper (CH) Controllability Rating Scale Formats

Each subject saw five CH controllability rating scale formats – the standard paper format and 4 electronic formats. The electronic formats were: (1) electronic paper, (2) forced choice bottom, (3) forced choice middle, and (4) forced choice top.

Paper CH Format. The Paper CH format was the standard CH format [6, 7].

Paper and Electronic Paper CH Format. The Electronic Paper CH format mimicked the paper version but on a touch screen (Fig. 2). In order to choose a rating, subjects had to touch the appropriate rectangle (*e.g.*, Major deficiencies … 8).
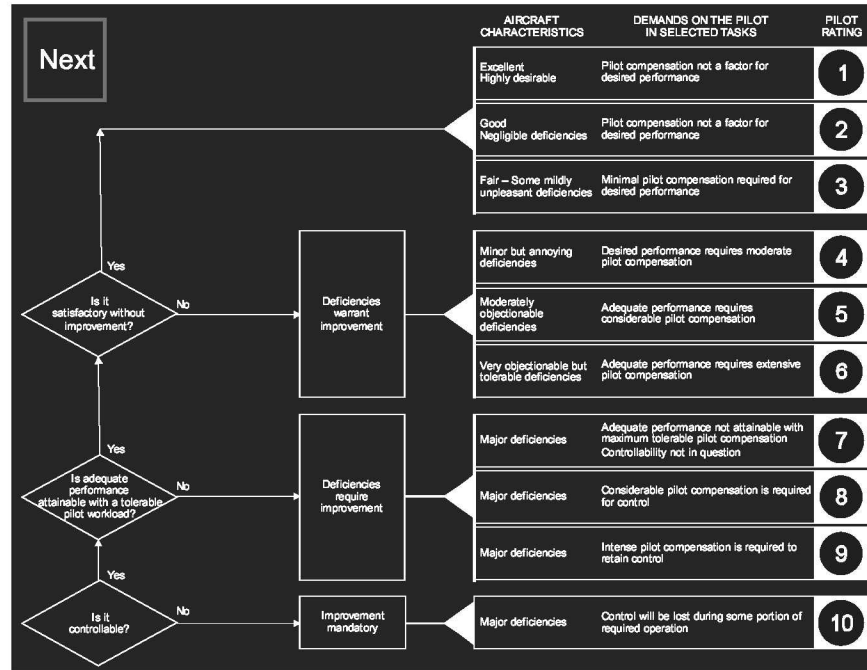
Fig. 2. Electronic Paper CH Format

Forced Choice Bottom CH Format. The Forced Choice Bottom CH format expanded depending on the choices selected by the subject. The flow chart started from the bottom (Is it controllable?) and worked its way up in ratings (Fig. 3). When the subject reached the ratings, only the ratings of the path taken were available. The path the subject had taken to get to those ratings was visible.

Forced Choice Middle CH Format. The Forced Choice Middle CH format also expanded depending on the choices selected by the subject. The flow chart started from the middle (Is adequate performance attainable with a tolerable pilot workload?) and worked its way up or down in ratings. As before, when the subject reached the ratings, only the ratings and their associated path were visible.

Forced Choice Top CH Format. The Forced Choice Top CH format expanded depending on the choices selected by the subject but the flow chart started from the top (Is it satisfactory without improvement?) and worked its way down in ratings. As with the other two forced choice CH formats, when the subject reached the ratings, only the ratings of that path and the path itself were visible.
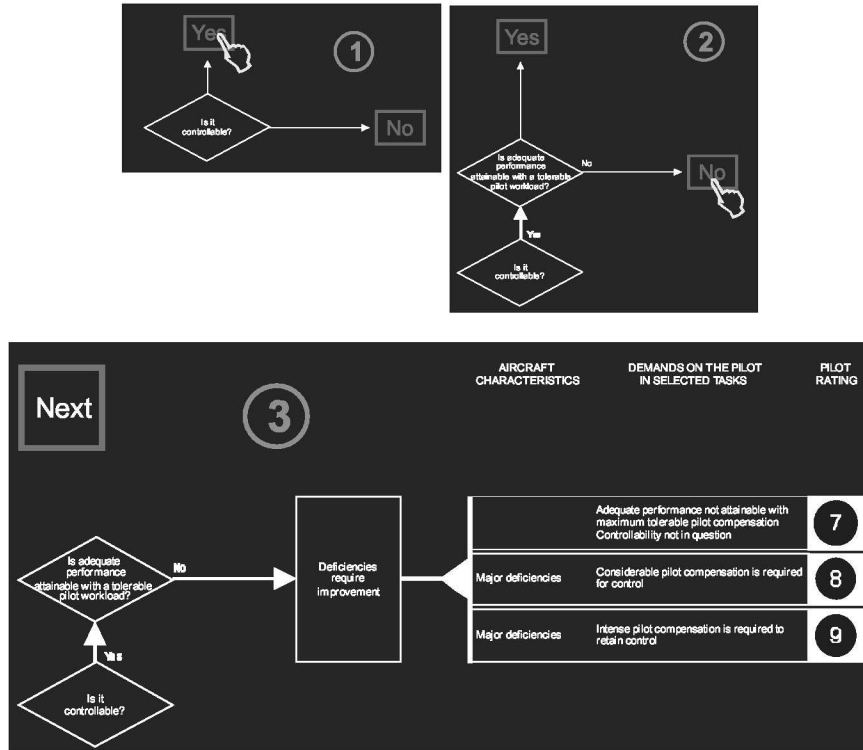
Fig. 3. Forced Choice Bottom CH Format

2.3 Control Task Difficulty

Each subject attempted to keep a moving target centered for 1 minute using a right-handed side stick. The control task difficulty levels ranged from a CH rating of 1 to a CH rating of 10. Each scenario had a preset control task difficulty level that was accomplished by linearly changing the speed of the target and inceptor gain.

A pretest to verify that the control task difficulty levels matched an operator's CH rating was conducted. The average difference between the control task difficulty level and the three subjects' CH ratings was -0.07±1.4 with a median of 0. A linear regression of the data was significant ($F_{(1,59)}$=1161.58; p≤0.01). The slope was 0.94 with an R^2=0.95.

2.4 Dependent Variables

The primary dependent variable was the subjects' CH ratings compared to the control task difficulty. The time taken to complete the CH ratings and the workload incurred to complete the CH ratings were also analyzed.

At the end of the experiment, subjects completed a final questionnaire. This questionnaire asked subjects to rate on a continuous scale how easy the CH formats were for rating the control task difficulty and the associated workload to complete the various CH formats. The questionnaire also asked for subject preferences, and likes and dislikes by display type.

3 Procedure

When subjects first arrived, they signed a consent form before being given a verbal briefing on the experiment tasks. Subjects then moved to the simulator where they completed two practice runs with the first CH format. After the practice runs, subjects completed 10 data runs. During each run, subjects had to keep a randomly moving target centered for 1 minute using a right-handed side stick. They also had to indicate when a frequency changed and answer a question that required basic multiplication skills. At the end of each run, subjects completed the CH controllability rating scale and the workload of determining a CH controllability rating. At the end of the 10 data runs with the first CH format, subjects completed at least one practice run with the next CH format and then the 10 data runs with that CH format. This was repeated until subjects had seen all five CH formats. At the end of the simulation runs and questions, subjects completed the final questionnaire.

3.1 Apparatus

The simulations ran on two PCs running Windows™ XP Professional[2]. These had a redraw refresh rate of 60Hz and a graphics update rate of 30Hz. The target tracking task was displayed on a 30-inch LCD screen in front of and slightly above the subject's eye level. The information indicating the frequency change and to answer the multiplication question was on a screen to the right of the subject. The questions were answered using a touch screen to the subject's left. The CH questionnaire was also presented on this left screen at the end of the run. These two touch screens were 19-inch LCD screens with an Elo Touchsystems IntelliTouch overlay for touch-screen capability. The side stick used was a Saitek Cyborg evo joystick. Subjects used their right hand to manipulate the side stick.

3.2 Data Analysis

Data was analyzed using SPSS® for Windows v16. Most of the time, the data was analyzed using a 3-way ANOVA with CH format, control task difficulty, and pilot status (pilot vs. non-pilot) as the independent variables.

---

[2] The use of trademarks or names of manufacturers in the report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

To determine the accuracy of the CH formats, the control task difficulty level was subtracted from the subjects' CH ratings. Therefore, a subject was the most accurate when this difference was 0 and the least accurate when the absolute value of this difference was 9. Furthermore, the CH ratings were on an integer scale. In the ANOVA analysis, the CH rating was treated as a continuous scale even though it is ordinal [9]. The final questionnaire responses were on continuous 100-point scales.

4 Results

4.1 Accuracy of Subjects' CH Ratings

When subtracting the control task difficulty from subjects' CH rating, pilot status by CH format was significant ($F_{(4, 900)}$=3.21; p≤0.02) (Fig. 4). In general, both pilots and non-pilots underestimated the control task difficulty with non-pilots underestimating the difficulty a bit more than pilots especially for the Forced Choice Middle and Forced Choice Top CH formats. Subjects for these two formats typically underestimated the control task difficulty by a full rating.
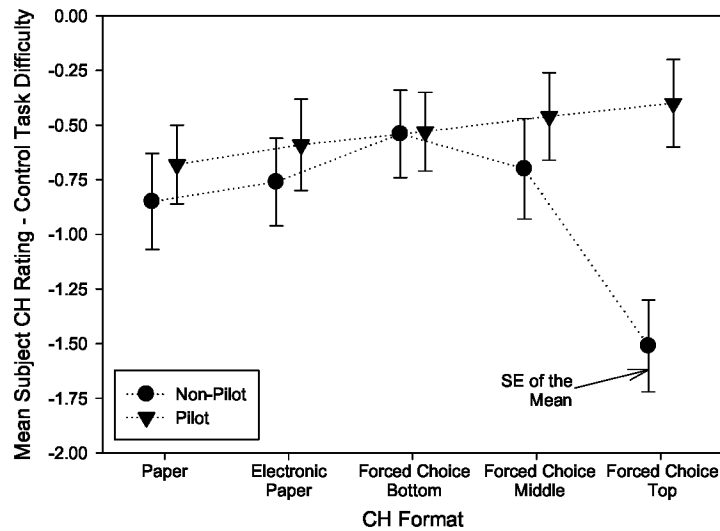

Fig. 4. Mean Subject CH Rating – Control Task Difficulty by Pilot Status and CH Format.

A linear regression estimating the subjects' CH rating by the control task difficulty was done in order to compare the effects of pilot status and CH format. As can be seen in Figure 4 and Table 1, subjects typically underestimated the control task difficulty by 15%. For pilots, the most accurate CH formats were flowcharts while the Forced Choice Bottom CH format was the most accurate format for non-pilots.

Table 1. Linear Regression Statistics of Estimating Subject CH Rating with
Control Task Difficulty by Pilot Status and CH Format.

Pilot Status	CH Format	Slope	R^2
Non-Pilot	Paper	0.80	0.86
	Electronic Paper	0.80	0.89
	Forced Choice Bottom	0.87	0.89
	Forced Choice Middle	0.84	0.87
	Forced Choice Top	0.68	0.86
Pilot	Paper	0.82	0.91
	Electronic Paper	0.82	0.88
	Forced Choice Bottom	0.84	0.91
	Forced Choice Middle	0.85	0.89
	Forced Choice Top	0.85	0.89

4.2 Time to Complete CH Ratings

The CH format was significant for the time it took subjects to complete the CH ratings ($F_{(4, 900)}$=31.98; $p \leq 0.01$) (Table 2). Not surprisingly, the Paper CH format took the longest to complete with the Forced Choice Bottom CH format taking the second longest. This is probably because this format typically requires a greater number of button pushes. The other formats were not significantly different from one another.

Table 2. Time to Complete CH Rating by CH Format.

CH Format	Time to Complete CH Rating (sec)	
	Mean	SE of the Mean
Paper	18.27	0.58
Electronic Paper	10.34	0.71
Forced Choice Bottom	13.16	0.62
Forced Choice Middle	10.99	0.43
Forced Choice Top	10.80	0.46

4.3 Subjective Data

Subjects' preference for the CH formats was dependent on CH format ($F_{(4, 87)}$=2.95; $p \leq 0.03$) and pilot status by CH Format ($F_{(4, 87)}$=4.36; $p \leq 0.01$) (Fig. 5). In general, subjects preferred the Electronic Paper and Forced Choice Bottom CH formats.

Pilot status by CH format was also significant for subjects' reported workload in completing the CH ratings ($F_{(4, 90)}$=2.51; $p \leq 0.05$) (Fig. 6). Workload for the Electronic Paper CH formats was the same for both pilots and non-pilots. But for pilots, the Forced Choice Bottom CH format a slightly higher workload than the Electronic Paper CH format but the workload was on par with the Paper version. The other two flow chart methods had even higher workloads for pilots. For non-pilots, the electronic versions of the CH format did not really affect the workload but they were lower than the Paper CH format.

Subjects indicated that the CH format affected their ability to arrive at their desired rating ($F_{(4, 83)}$=4.26; p≤0.01) (Table 3). In general, subjects felt that the Paper, Electronic Paper, and Forced Choice Bottom CH formats allowed them to arrive at an accurate CH rating.
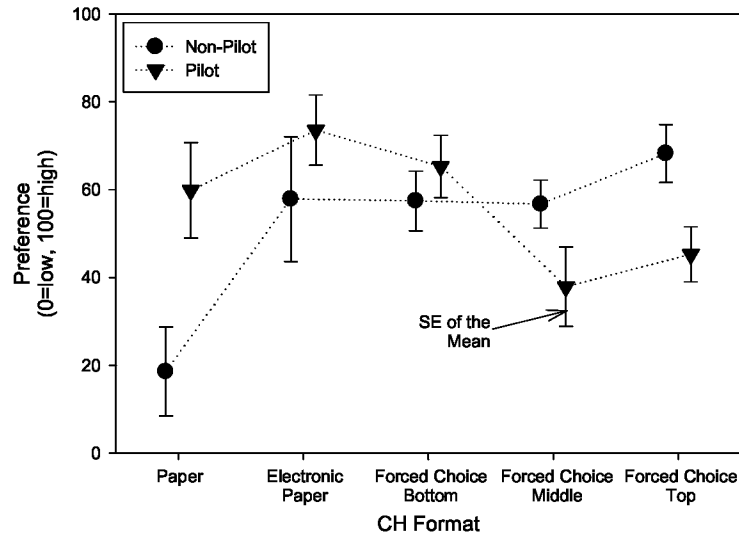

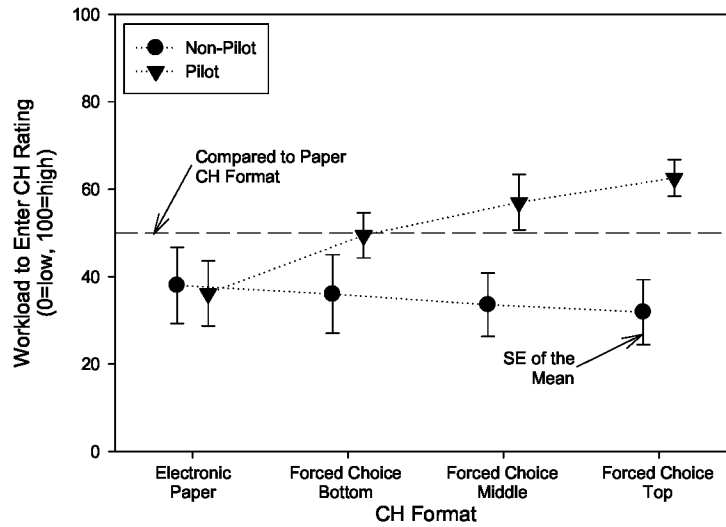**Fig. 5.** CH Format Preference by Pilot Status and CH Format.

Fig. 6. Workload to Enter CH Rating by Pilot Status and CH Format.

Table 3. Ability to Arrive at Desired Rating by CH Format.

CH Format	Ability to Arrive at Desired Rating (0=low, 100=high)	
	Mean	SE of the Mean
Paper	65.32	6.89
Electronic Paper	77.37	6.42
Forced Choice Bottom	65.94	5.09
Forced Choice Middle	48.17	5.21
Forced Choice Top	52.41	4.55

Additionally, subjects indicated that on the Paper version, they specifically went step by step through the flow chart only about half of the time even though they were instructed to arrive at their ratings via sequentially answering the questions in the flow chart: specifically 45% of the time for non-pilots and 64% of the time for pilots. This may be because the Paper and Electronic Paper CH formats allow subjects to "cut to the chase" and choose a number without going through the flow chart (Table 4).

Table 4. Subject Comments on the CH Formats

Subject Comment Categories and Example Comments	Number
All choices are available on Paper and Electronic Paper CH formats "like to see all options"; "easier to compare measures"	18
Too much information on Paper and Electronic Paper CH formats "hard to sort all information"; "information overload"	8
Like the mechanics of flowchart "like flowchart with its logical sequence"	8
Do not like the mechanics of flowcharts "takes longer"	5
Do not like mechanics of Paper CH formats "more cumbersome"; "required most time to answer"	9

Specific comments on where to start in flow chart
    "flow chart pulls you in the direction of where you started"
    "liked starting at the bottom because it was the worst case"
	16

Many subjects commented that they liked having all the information available to them to see at once. Some subjects did say that the Paper and Electronic Paper CH formats induced "information overload" because "there was too much information."

Subjects who liked flowcharts said it was because they had a "logical sequence" which helped "produce a more reasoned rating." As for where to start on the flowchart, most subjects commented that they like to start at the bottom because it was the "most intuitive" and "ask[ed] the most important question first." Other comments relating to other starting points in the flowcharts indicated that the "flow logic was counter intuitive."

Generally, subjects liked having all the information available to them at once but they did feel like the flow chart formats produced a logical thought process. Of the flow chart sequences, the Forced Choice Bottom CH format had the most preferred logic sequence.

5 Discussion

Electronic questionnaires are replacing paper formats. The formats of traditional paper questionnaires have been found to affect a subject's rating. Consequently, the transition from paper to electronic format can subtly change results. This research had subjects use five different formats of the CH Controllability Rating Scale that requires respondents to give their ratings by answering questions posed in a flowchart.

Results indicated that while all formats were reasonably accurate, the Electronic Paper and Forced Choice Bottom CH formats produced the most accurate ratings while being the most preferred. In general, subjects underestimated the difficulty of the control task using all CH formats. Workload in inputting their answers was a bit higher for pilots when using the Forced Choice Bottom CH format but was on par for the workload when using the Paper version. Subjects did indicate that they only went through the Paper flow chart questions about half the time even though they were instructed to arrive at their ratings only after answering the flow chart questions.

Therefore, moving questionnaires from paper to electronic media could change respondents' answers. Specifically, the above results suggest that when using a flow chart type of questionnaire, it is best to have subjects directly answer each decision point while starting at the worst rating. Although this inflicts a slight penalty in time and workload, it does insure that subjects make decisions at each point while minimizing the underestimation of the difficulty of the task.

References

1. Trujillo, A.C., D. Bruneau, and H.N. Press, Predictive Information: Status or Alert Information?, in 27th Digital Avionics Systems Conference. 2008: St. Paul, MN.
2. Trujillo, A.C. and A.T. Pope, Using Simulation Speeds to Differentiate Controller Interface Concepts, in 52nd Annual Meeting of the Human Factors and Ergonomics Society. 2008, HFES: New York, NY.
3. Noyes, J.M. and D.P.J. Bruneau, A Self-Analysis of the NASA-TLX Workload Measure. Ergonomics, 2007. 50(4): p. 514-519.
4. Riley, D.R. and D.J. Wilson. More on Cooper-Harper Pilot Rating Variability. in 8th Atmospheric Flight Mechanics Conference. 1990. Portland, OR.
5. Wilson, D.J. and D.R. Riley. Cooper-Harper Pilot Rating Variability. in AIAA Atmospheric Flight Mechanics Conference. 1989. Boston, MA.
6. Cooper, G.E. and R.P. Harper, The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities. 1969, Technical Report 567, AGARD. p. 52.
7. Harper, R.P. and G.E. Cooper, Handling Qualities and Pilot Evaluation (Wright Brothers Lecture in Aeronautics). Journal of Guidance, Control, and Dynamics, 1986. 9(6): p. 515-529.
8. Federal Aviation Administration. Electronic Code of Federal Regulations - Title 14: Aeronautics and Space Subpart E-Private Pilots Section 61.103. August 28, 2008 [cited 2008 September 2]; Available from: http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=%2Findex.tpl.
9. Bailey, R.E., The Application of Pilot Rating and Evaluation Data for Fly-by-Wire Flight Control System Design, in AIAA Atmospheric Flight Mechanics Conference. 1990, AIAA: Portlan, OR. p. 13.