Testing Pointing Device Performance and User Assessment with the ISO 9241, Part 9 Standard

Sarah A. Douglas and Arthur E. Kirkpatrick
Computer and Information Science Dept.
University of Oregon
Eugene, OR 97403 USA
douglas@cs.uoregon.edu
ted@cs.uoregon.edu

I. Scott MacKenzie
Dept. of Computing and Information Science
University of Guelph
Guelph, Ontario, Canada N1G 2W1
smackenzie@acm.org

ABSTRACT

The ISO 9241, Part 9 Draft International Standard for testing computer pointing devices proposes an evaluation of *performance* and *comfort*. In this paper we evaluate the scientific validity and practicality of these dimensions for two pointing devices for laptop computers, a finger-controlled isometric joystick and a touchpad. Using a between-subjects design, evaluation of performance using the measure of *throughput* was done for one-direction and multi-directional pointing and selecting. Results show a significant difference in throughput for the multi-directional task, with the joystick 27% higher; results for the one-direction task were non-significant. After the experiment, participants rated the device for comfort, including operation, fatigue, and usability. The questionnaire showed no overall difference in the responses, and a significant statistical difference in only the question concerning force required to operate the device—the joystick requiring slightly more force. The paper concludes with a discussion of problems in implementing the ISO standard and recommendations for improvement.

Keywords

Pointing devices, ergonomic evaluation, ISO 9241 standard, isometric joystick, touchpad, Fitts' law

INTRODUCTION

During the past five years the International Standards Organization (ISO) has proposed a standard entitled *ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 9 Non-keyboard Input Device Requirements* [3]. The primary motivation of the standards effort is to influence the design of computer pointing devices to accommodate the user's biomechanical capabilities and limitations, allow adequate safety and comfort, and prevent injury. Secondarily, the standards establish uniform guidelines and testing procedures for evaluating computer pointing devices produced by different manufacturers. Compliance can be demonstrated through testing of user *performance*, *comfort* and *effort* to show that a particular device meets ergonomic standards or that it

meets a de facto standard currently on the market.

Crafting and adopting any set of standards for the evaluation of pointing devices raises a number of questions:

- Are the standards consistent with accepted scientific theory and practice?

- Do the standards allow practical implementation and conformance?

- Are the expected results reliable and ecologically valid in order to predict behavior and evaluate devices?

The goal of our present study is to answer these questions for the ISO proposed standard for the evaluation of *performance* and *comfort*. We first implement the proposed testing procedures using a case study experiment of a finger-controlled isometric joystick and of a touchpad. Secondly, we reflect upon that experience. To that end, we are specifically interested in problems that arise in replication, interpretation and reliability of results.

ISO 9241 - Part 9

ISO standards are written by committees drawn from the research and applied research communities. As of September 1998 the *ISO 9241-Part 9* is in Draft International Standard version and is currently awaiting a vote of member organizations. If adopted, certification of conformance to this standard will be legally required for devices sold in the European Community. The general description of the Standard and the particulars of Part 9 are described in Smith [6].

The proposed standard applies to the following hand-operated devices: mice, trackballs, light-pen & styli, joysticks, touch-sensitive screens, tablet-overlays, thumb-wheels, hand-held scanners, pucks, hand-held bar code readers, and remote-control mice. It does not cover eye-trackers, speech activators, head-mounted controllers, datagloves, devices for disabled users, or foot-controlled devices.

Part 9 specifies general guidelines for physical characteristics of the design including the force required for operating them as well as their feedback, shape, and labeling. In addition to these general guidelines, there are requirements for each covered device.

ISO 9241 defines evaluation procedures for measuring user *performance*, *comfort* and *effort* using an experimental protocol which defines subject samples, stimuli,

experimental design, environmental conditions, furniture adjustments, data collection procedures, and data analysis recommendations.

Performance is measured by task performance on any of six tasks: one-direction (horizontal) tapping, multi-directional tapping, dragging, free-hand tracing (drawing), free-hand input (hand-written characters or pictures) and grasp and park (homing/device switching). The tasks selected for testing should be determined by the intended use of the device with a particular user population.

For the tapping tasks which are essentially basic point-select tasks, the ISO recommends collection of the following performance data. The primary ISO dependent measure is *Throughput* (*TP*).

$$Throughput = ID_e / MT \qquad (1)$$

where

MT is the mean movement time, in seconds, for all trials within the same condition,

and

$$ID_e = \log_2(D/W_e + 1) \qquad (2)$$

ID_e is the effective index of difficulty, in bits, and is calculated from D, the distance to the target and W_e, the effective width of the target. W_e is computed from the observed distribution of selection coordinates in participants' trials:

$$W_e = 4.133 \, SD \qquad (3)$$

where SD is the standard deviation of the selection coordinates. Throughput has units bits per second (bps).

Readers will note that Equation 1 for throughput is the usual Fitts' Index of Performance (*IP*) except that effective width (W_e) replaces actual measured size of the target (W). Using effective width incorporates the variability observed of human performance and includes both speed and accuracy [4]. Thus, throughput precludes a separate computation of error rate.

The ISO 9241 standard also argues that evaluating user performance using a short-term test is not enough for a complete evaluation of a device. Consequently, the ISO 9241 requires assessment of *effort* as a biomechanical measurement of muscle load and fatigue during performance testing. Finally, *comfort* is ascertained after performance testing by having participants subjectively rate the device using a questionnaire form which assesses aspects of operation, fatigue, comfort, and overall usability.

METHOD

An experiment was designed to implement the performance and comfort elements of the ISO testing. The third element, effort, was not tested due to our inability to obtain the sophisticated equipment and technician for measuring biomechanical load.

Performance testing was limited to pointing and selecting using both a one-direction test (1D Fitts serial task) and a multi-directional test (2D Fitts discrete task). The testing environment was modeled on the ISO proposal as described in Annex B [3]. Comfort was evaluated using the ISO

"Independent Questionnaire for assessment of comfort". The design attempted to follow as reasonably as possible the proposed description in Annex C [3].

Testing was conducted for two different pointing devices, a finger-controlled isometric joystick and a touchpad, both connected to the same computer.

Participants

Twenty-four persons participated in this experiment, twelve for each device. For the touchpad, all participants were right-handed. For the joystick, eleven participants were right-handed and one left-handed.

Participants were unpaid volunteers recruited through posters and personal contact. They were offered the opportunity to win a dinner for two selected randomly from among the participants. All participants were screened using a questionnaire which assessed their prior experience with computers and pointing devices. All participants had prior computer experience and extensive experience with the mouse pointing device. Participants were assigned to the device for which they had no prior experience. If they had no experience on either, they were randomly assigned to one. They all signed an informed consent document informing them of the goals and activities of the study, their rights to terminate, and the confidentiality of their performance.

Apparatus

An IBM Thinkpad® laptop computer was fitted with a separate 21 inch color display monitor. The tested device for the joystick was the installed Trackpoint® III located on the keyboard between the "G" and the "H" key. For the second device, a Cirque Glidepoint® Touchpad 2 Model 400 was connected through the PS/2 port. For both devices, the "gain" was set to the middle value in the standard NT driver software for setting pointing device speed.

Experimental tasks were presented by two different programs. For the one-direction test (1D Fitts task), the *Generalized Fitts Law Model Builder* was used [7]. This program, written in C, runs under Windows 95® in MSDOS mode. Figure 1 illustrates the screen as presented by the software. For each block of trials, the software presented a pair of rectangular targets of width W and
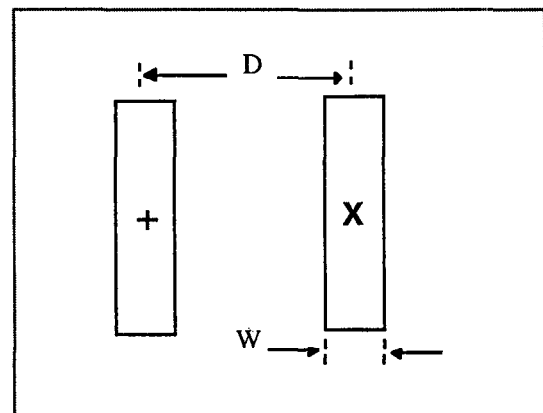

Figure 1. One-direction task.

distance *D*. For this experiment, the target rectangle was varied by three different widths, and three different distances. There were 30 trials in each block.

At the beginning of a trial, a crosshair pointer appeared in the left rectangle and a red X appeared in the opposite rectangle denoting it as the current target. For the next trial the location of crosshair and X were reversed. This allows the participant to move quickly back and forth between the two targets.

The multi-directional test was implemented by software written at the University of Oregon HCI Lab. The basic task environment has been used for prior evaluation work on pointing device performance assessment [1, 2]. It is written in C++ and runs under Windows NT.

Figure 2 illustrates the screen as presented by the software. A trial starts when the participant clicks (selects) in the home square, and ends when the participant clicks in the target circle. The time between these clicks is recorded as the trial time. The cursor is automatically repositioned in the center of the home square at the end of each trial. Combinations of width, distance and angle are presented randomly.

For this experiment, the target circle was varied by three different widths, three different distances, and eight different angles.
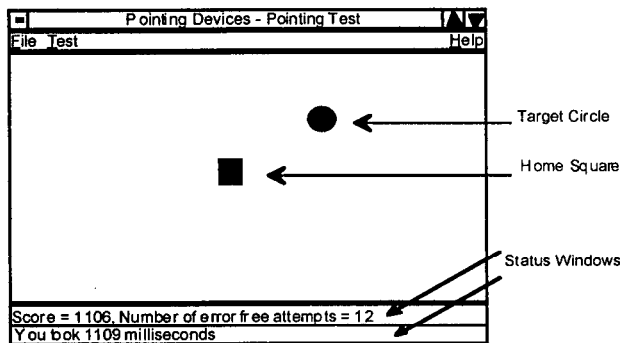

Figure 2. Multi-directional task.

All software is available from the authors.

Procedure

Participants were given the multi-directional task first. The task was explained and demonstrated to the participant. They were instructed to work as fast as possible while still maintaining high accuracy. Participants were instructed to continue without trying to correct errors. Participants performed ten blocks of multiple combinations of target width, distance and angle trials, and were informed that they could rest at any time between trials. Participants using the touchpad were instructed to use the button for selection rather than multiple taps.

After completion of the multi-directional task, participants rested for a few minutes before receiving instruction on the one-direction task. The one-direction task was run for blocks of multiple combinations of target width and distance. Participants were allowed to rest briefly between blocks, but not between trials.

At the conclusion of the performance portion of the experiment, participants were asked to respond to a written questionnaire asking them to rate their experience in using the device. The questionnaire consisted of thirteen questions covering issues of physical operation, fatigue and comfort, speed and accuracy, and overall usability. Participants were asked to respond to each question with a rating from low to high. Figure 3 illustrates this device assessment questionnaire.

DEVICE ASSESSMENT

Please circle the x that is most appropriate as an answer to the given comment.

1. The force required for actuation was
       x   x   x   x   x
     too low   too high

2. Smoothness during operation was
       x   x   x   x   x
     very rough   very smooth

3. The mental effort required for operation was
       x   x   x   x   x
     too low   too high

4. The physical effort required for operation was
       x   x   x   x   x
     too low   too high

5. Accurate pointing was
       x   x   x   x   x
     easy   difficult

6. Operation speed was
       x   x   x   x   x
     too fast   too slow

7. Finger fatigue:
       x   x   x   x   x
     none   very high

8. Wrist fatigue:
       x   x   x   x   x
     none   very high

9. Arm fatigue:
       x   x   x   x   x
     none   very high

10. Shoulder fatigue:
       x   x   x   x   x
     none   very high

11. Neck fatigue:
       x   x   x   x   x
     none   very high

12. General comfort:
       x   x   x   x   x
     very   very
     uncomfortable   comfortable

13. Overall, the input device was
       x   x   x   x   x
     very difficult   very easy
     to use   to use

Figure 3. Device Assessment Questionnaire.

The total time spent by each participant ranged from slightly less than an hour to one hour and 30 minutes. The performance section took between 45 minutes to one hour to complete.

Design

Pointing Performance

The design for the experiment used a mixed design, with device as a between-subjects factor, and task-type (one-direction or multi-directional tapping) as a within-subjects factor. For the one-direction task, we have the following independent variables:

Target Width (2 mm, 5 mm, 10 mm)

Target Distance (40 mm, 80 mm, 160 mm)

Trial (1 to 30)

Block (1 to 9)

The three target sizes approximated the width of a "o" in 8 pt. Helvetica, the height of a character or word, and the width of an icon in the Microsoft Windows environment. These widths and distances represent Fitts' Index of Difficulty values from 2.3 to 6.3 bits. (We have given the widths in actual physical distance rather than pixels since pixel sizes vary from monitor to monitor.)

In the one-direction case, a block consists of 30 trials of the same width-distance combination. A total of 270 trials were run (30 trials per block x [3 widths x 3 distance] blocks).

For the multi-directional task we have the following independent variables:

Target Width (2 mm, 5 mm, 10 mm)

Target Distance (40 mm, 80 mm, 160 mm)

Target Angle (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°)

Trial (1 to 72)

Block (1 to 10)

Note that a block of the multi-directional task is defined as the 72 fully crossed combinations of target distance, width and angular location from the starting position (3 widths x 3 distances x 8 angles). A total of 720 trials were run (72 trials per block x 10 blocks). Since the multi-directional task was more complex, it was performed before the one-direction task to bring the participants to a criterion level of practice in a much more realistic and ecologically valid task environment for computer users.

Dependent variables are movement time (*MT*) for each trial, throughput (*TP*), and error rate (*ER*). Error rate is the percentage of targets selected when the pointer is outside the target and is not an ISO recommended measure. However, we have found it useful piece of information when assessing performance.

Device Assessment Questionnaire

The Device Assessment questionnaire consisted of thirteen questions taken from the ISO standard. Participants were asked to give a response to each question as a rating on a five point scale from low to high. The data were considered ordinal.

ANALYSIS

The data for movement time (*MT*), and selection point (*x*, *y*) were collected directly by the software which presented the experimental tasks. The data were then prepared for further statistical analysis by computing values for throughput and error rate in addition to *MT* for each trial. Finally, basic statistics and ANOVA were performed using commercial software.

Adjustments to Data

We did not make any adjustments to the data and excluded none of the trials.

Computed Formulas

For the one-direction task, the computation for *throughput* begins by computing W_e according to Equation 3. To achieve this, for each trial the *x* coordinate of the participant's final selection point is recorded. For all participants in a *D* x *W* condition, these constitute a distribution of points. The sample mean can be computed in the usual manner. Then the difference between the participant's selection point and the mean is computed and squared; this can also be interpreted as the square of the distance between the selection point and the mean. For all subjects and all trials, the standard deviation (*SD*) is then computed as

$$SD = \sqrt{\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right)} \tag{4}$$

W_e, ID_e and *throughput* are then computed by Equations 3, 2 and 1, respectively.

For the multi-directional task, the computation of W_e must be modified because selection points are now located in a two-dimensional plane. The "difference" between each actual selection point and the mean is computed now as the Euclidean distance between the selection point and the mean point ($\bar{x}, \bar{y}$).

$$Dist = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \tag{5}$$

In the *SD* computation, we square the distance and hence obtain

$$SD = \sqrt{\left(\frac{\sum_{i=1}^{n}\left[(x_i - \bar{x})^2 + (y_i - \bar{y})^2\right]}{n-1}\right)} \tag{6}$$

W_e, ID_e and *throughput* are then computed by Equations 3, 2 and 1, respectively.

Device Assessment Questionnaire

The mean and standard deviation of the ratings for each of the thirteen questions was computed. Given the ordinal nature of the data, the Mann-Whitney non-parametric statistic was computed to test for significant differences between participants in the two device groups.

RESULTS

Pointing Performance

Multi-directional Task

The mean movement time for the joystick was 1.975 seconds with a standard deviation of .601 seconds. For the touchpad, the mean movement time was 2.382 seconds and a standard deviation of .802 seconds. These differences were

statistically significant ($F_{1,22}$ = 11.223, p= .0029). From this we can conclude that the pointing time for the joystick is 17% faster on average.

Error rates were 2.1% (sd = 5.68) for the joystick and 5.4% (sd = 10.5) for the touchpad. These differences were statistically significant ($F_{1,22}$ = 7.604, p= .0115), with the joystick having a 61% lower error rate.

Throughput was computed for the joystick at 2.15 bps (sd = .40). For the touchpad, it was 1.70 bps (sd = .53). These differences are statistically significant ($F_{1,22}$ = 20.458, p = .0002), and indicate that throughput for the joystick is 27% higher than for the touchpad.

Although the ISO standard does not discuss learning effects, this obviously must be considered when designing and evaluating performance data. In this experiment, participants performed the multi-directional task for ten blocks of 72 trials each. We hypothesized, based on other similar experiments, that they would have achieved a criterion level of practice by block ten. In other words, no significant improvement in performance would be shown in the final blocks.

This was confirmed by the data analysis. Helmhert contrasts show the differences between blocks become non-significant at block 6. For throughput, the effect of Block*Device was significant ($F_{9,198}$ = 2.726, p = .0051). Graphing the Block*Device effect shows only a mild effect (see Figure 4). The main contributor is that the difference between Block 1 and 2 is greater for the pad than the joystick.
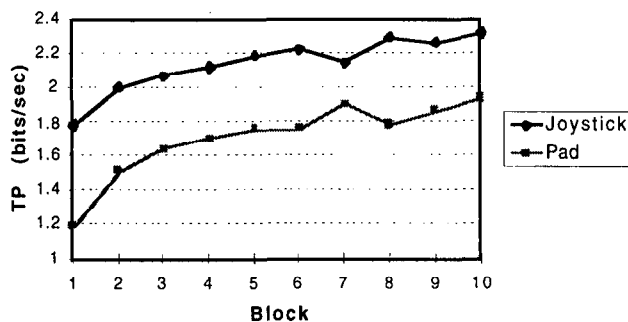

Figure 4. Learning shown for Throughput by Device and Block.

The effect of learning on the task suggests that examining Block 10 alone will give us a good measure of practiced performance.

For Block 10 trials the mean movement time for the joystick was 1.770 seconds (sd = .458). For the touchpad, the mean movement time was 2.132 seconds (sd = .583). These differences are statistically significant ($F_{1,22}$ = 14.462, p = .010). From this we can conclude that the pointing time for the joystick is 17% faster on average.

Error rates were 3.4% (sd = 6.8) for the joystick and 3.8% (sd = 7.7) for the touchpad. These differences are not statistically significant ($F_{1,22}$ = .123, p > .05).

Throughput was computed for the joystick at 2.33 bps (sd = .32). For the touchpad, it was 1.94 bps (sd = .46). These differences are statistically significant ($F_{1,22}$ = 15.873, p < .0006), and demonstrate slightly higher throughput due to practice for both devices—the joystick now 20% higher. (The change in relative performance is due to the steeper slope of the learning curve for the touchpad.)

One-direction Task

The mean movement time for the joystick was 1.544 seconds (sd = .305). For the touchpad, the mean movement time was 1.563 seconds (sd = .285). These differences are not statistically significant ($F_{1,22}$ = .024, p > .05).

Error rates were 17.5% (sd = 13.8) for the joystick and 25.6% (sd = 22.5) touchpad. These differences are not statistically significant ($F_{1,22}$ = 1.136, p > .05).

We were surprised by the high error rates for both devices. In post-experiment interview, many participants commented on the difficulty of the one-direction task with the small targets. A closer examination of the data revealed that many of the errors occurred on the 2 mm target widths. Error rate means for the 2 mm widths were 29.6% (28.5); 5 mm were 20.4% (26.4); and 10 mm were 14.7% (22.1). An ANOVA testing error rate by width shows a significant effect ($F_{2,44}$ = 7.565, p = .0015). We speculate that this difference in error rate might be due to the serial nature of the one-direction task, and the ballistic nature of the initial pointing movement which promotes increased momentum causing overshoot. Further analysis of errors will be needed to confirm that.

Throughput was computed for the joystick at 2.07 bps (sd = .39). For the touchpad, it was 1.81 bps (sd = .62). These differences are not statistically significant ($F_{1,22}$ = 1.469, p > .05).

Overall Pointing Performance

Table 1 illustrates the various computations of throughput. It is clear from these results that the joystick is consistently superior in performance in both overall and practiced analysis of multi-direction pointing. Results for one-direction are non-significant, although the joystick is, again, slightly higher in throughput.

Throughput	Joystick mean (sd)	Touchpad mean (sd)
Multi-direction: All Trials	2.15 bps (.40)*	1.70 bps (.53)*
Multi-direction: Block 10 only	2.33 bps (.32)*	1.94 bps (.46)*
One-direction: All Trials	2.07 bps (.39)	1.81 bps (.62)

*Significant at p < .005

Table 1. Comparison of throughput.

Finally, how do these results compare with other published data of performance for the same or similar devices? On the multi-direction task, the throughput for Block 10 can be compared with published data for the mouse at 4.15 bps and another finger-controlled isometric joystick, the Home-

Row J key, at 1.97 bps [2]. Note, however, that these latter values are throughput computed using W instead of W_e.

Similarly, our results for the one-direction task can be compared with the study of MacKenzie and Oniszczak [5] on selection techniques for a touchpad. Their results indicate a throughput of .99 bps for another button selection touchpad versus our results of 1.81 bps. However, our value is for a one-direction task after the participants received a great deal of practice on the multi-directional task. Their value is an overall value for a task environment in which participants learned the device on one-direction tasks only, and repeated each condition 60 times (20 trials x 3 blocks). Our experiment had participants performing each condition 30 times (30 trials x 1 block). Practice could account for the higher values we observed.

Device Assessment Questionnaire

The means and standard deviations of the responses on the thirteen questions on the questionnaire are shown in Table 2. The results of the questionnaire analysis show that these responses are not statistically significant overall (Mann-Whitney $U = 11653$, $p = .5180$). Individual question analysis showed only Question 1, on the amount of force required for actuation, had significant differences in response between the two devices (Mann-Whitney $U = 33.000$, $p = .0243$). The joystick participants rated the force required slightly higher (3.583) than the touchpad participants (2.833).

Question	Joystick	Touchpad
1	3.583 (.996)	2.833 (.835)
2	2.583 (1.084)	2.000 (.739)
3	3.583 (1.165)	3.333 (.651)
4	3.667 (.888)	3.917 (.669)
5	4.083 (.996)	4.000 (.739)
6	3.583 (.515)	3.500 (.522)
7	3.083 (1.564)	3.167 (1.193)
8	3.500 (1.382)	3.667 (1.371)
9	3.167 (1.337)	2.250 (1.422)
10	2.000 (.953)	2.000 (1.651)
11	1.833 (2.167)	2.167 (1.337)
12	2.333 (.985)	2.667 (.985)
13	2.417 (.793)	2.583 (.900)

Table 2. Results of the Device Assessment Questionnaire.

Since each response was rated on a five point scale, a value of 3 is the mid-point. Indeed, the overall question mean was 3.032 for the joystick and 2.929 for the touchpad, indicating participants rated both devices in the midpoint range. Questions 7-11 regarding fatigue rated both devices in the same range, with finger and wrist fatigue higher than shoulder and arm. Even Question 12 on overall comfort

and Question 13 regarding usability of device were rated near the midpoint for both devices (2.417 for the joystick; 2.583 for the touchpad).

We can conclude from these questionnaire data that the ISO subjective comfort assessment shows little difference between the two devices. Given that the throughput difference was 27% favoring the joystick and that both devices were laptop devices, we might suggest that the difference was not enough to be reflected in differences in subjective evaluation.

DISCUSSION

At the beginning of our paper, we posed three questions concerning the ISO standards. We will now discuss these issues using our experience in implementing this case study.

- Are the standards consistent with accepted scientific theory and practice?

Much of the ISO standard for pointing performance is based on the accepted use of Fitts' law as a basis for the evaluation of pointing performance [1, 4]. Throughput is essentially Fitts' Index of Performance. The human factors literature is filled with studies of pointing device performance based on this scientific theory. In keeping with current practice, the ISO formula for throughput uses W_e (the effective target width) rather than W. Thus throughput incorporates both the speed and accuracy of users' behavior.

Our case study focused on pointing performance evaluation for both one-direction and multi-directional (2D) tasks. The one-direction task has been widely used in human factors work on pointing devices; the multi-directional task less so.

As the standard recommends, in recent years effective width (W_e) has replaced measured width (W) in the computation of *Index of Performance* or the ISO term, *throughput*. While this allows a single measure of both speed and accuracy, we feel that it does not replace separate measures of speed as movement time and accuracy as error rate. Consequently, we recommend computing both movement time and error rate as separate dependent variables.

A serious flaw in the standard is its failure to incorporate learning into the analysis. Existing studies of pointing devices show a significant effect due to learning [2]. The standard does not recommend experimental designs that reach a criterion level of practice nor discusses controls for transfer of training. In our study we applied a repeated measures paradigm and tested for learning effects. We recommend that others do so as well.

Concerning experimental design, the standard recommends a participant sample that is representative of the intended user population, and recommends at least 25 participants. We only used 12 for each between-subjects condition. This is standard practice for pointing device performance experiments, and psychological testing in general.

Finally, we did not agree with the recommended design of the Questionnaire. The standard recommends a 7 point rating scale which it claims is an interval scale. We substituted a 5 point rating scale after pilot tests showed

that participants could not make finer distinctions. We also consider the data ordinal rather than interval.

- Do the standards allow practical implementation and conformance?

In general, the standard is very vague concerning the implementation of the Fitts' concepts into experimental designs for testing environments. No discussion of Index of Difficulty is made, i.e., there are no explanations of how the task conditions should vary as a function of target width and distance to produce a range of difficulties. (We recommend *ID* range from 2 to 6 bits.) Computations for throughput and W_e are not explained in enough detail to implement the data analysis. As our study has shown, extension of the computation of W_e to the multi-direction case is not straight-forward. We propose that our experiment be used as a paradigm.

- Are the expected results reliable and ecologically valid in order to predict behavior and evaluate devices?

In other words, do they really help us evaluate devices? Given the general recommendations of the ISO standard for evaluating *performance* which is based on sound scientific foundations and the proper implementation of an experiment as we have described in this paper, we believe the results are reliable and can successfully predict user behavior. While we implemented both the one-direction and multi-directional tasks, we believe that the multi-directional task is more ecologically valid, presenting a complex task environment closer to what is observed with modern user interface tasks. As we observed, one-direction tasks performed by well-practiced participants gave us non-significant results.

As with other Fitts' law results, we adamantly assert caution in comparing results across experiments: It is critical that exactly the same experimental design, task environment, instructions and data analysis be given [1, 4]. The ISO standard does not make this clear. Given these limitations, it is useful to have standardized software such as that used in this experiment for presenting experimental environments, namely the *Generalized Fitts Law Model Builder* [7] available from author MacKenzie which runs both 1D and 2D Fitts' tasks, or the 2D Fitts' task software written at the University of Oregon HCI Lab available from authors Douglas and Kirkpatrick.

We have no means to compare *comfort* which is done through a post-experiment questionnaire. From our interviews with participants after the experiment, the questions were too vague to pick up specific problems with a device. We recommend an additional open-ended questionnaire with the following questions:

- Did you have any trouble moving the cursor to the target? If so, please describe.

- Did you have any trouble selecting (clicking) a target? If so, please describe.

- Do you have any comments in general about using this device for pointing?

- Comparing the tested device to your usual pointing device (which is ____), could you imagine a situation in which you would prefer the tested device?

- Do you have any suggestions how to improve this device?

This will allow the testers to more fully evaluate specific problems with the device.

CONCLUSIONS

Our goal in conducting this study is to assess the ISO 9241, Part 9 standard as a tool to evaluate pointing device performance and comfort by implementing a case study. We have done this by examining the scientific and practical issues. On its scientific merits, the standard appears sound; on practicality, it sorely needs improvement. A major contribution we have made in this paper is to define the experimental design in sufficient detail so as to allow others to replicate it. Finally, we have contributed to the growing evaluation of pointing devices through our study of the joystick and touchpad.

We note that while the ISO standard assesses user performance, comfort and effort, it does not address other issues of interest to users such as footprint, cost or integration of the pointing device with the rest of the hardware and software. These must be evaluated by other means if a broader analysis is needed.

As of the writing of this paper (January 1999) the ISO organization members are in the process of deciding whether the 9241, Part 9 will be adopted or not. Voting began last summer and lasted until October 21, 1998. Members had a choice of four options: approved as written, approved with attached comments, not approved with attached comments, and abstain. To the best of our knowledge and even given the fact that one of the authors (MacKenzie) is an ISO representative for Canada, the results are not known yet. If the Standard is adopted it will have major impact on device manufacturers in terms of cost, time for development, and final marketability of the product.

REFERENCES

1. Douglas, S.A. and Mithal, A.K. *The ergonomics of computer pointing devices.* Springer-Verlag, New York, 1997.

2. Douglas, S.A. and Mithal, A.K. The effect of reducing homing time on the speed of a finger-controlled isometric pointing device. *Human Factors in Computing Systems, CHI '94 Conference Proceedings,* ACM Press, New York, pp. 411-416.

3. ISO. *ISO/DIS 9241-9 Ergonomic Requirements for Office Work with Visual Display Terminals, Non-keyboard Input Device Requirements, Draft International Standard*, International Organization for Standardization, 1998.

4. MacKenzie, I. S. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction, 7*, (1992), pp. 91-139.

5. MacKenzie, I. S. and Oniszczak, A. A comparison of three selection techniques for touchpads. *Human Factors in Computing Systems, CHI '98 Conference Proceedings*, ACM Press, New York, pp. 336-343.

6. Smith, W. *ISO and ANSI ergonomic standards for computer products: A guide to implementation and products*, Prentice Hall, New York, 1996.

7. Soukoreff, W. and MacKenzie, I.S. Generalized Fitts' law model builder. In *Companion Proceedings of the CHI '95 Conference on Human Factors in Computing Systems*, ACM Press, New York, pp. 113-114.