DOCUMENT RESUME

ED 395 023 TM 025 040

AUTHOR Oltman, Philip K.
TITLE The Effect of Complexity of Mouse Manipulation on Performance in Computerized Testing.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-94-22
PUB DATE May 94
NOTE 16p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Computer Literacy; *Difficulty Level; Individual Differences; *Minority Groups; *Performance Factors; Sex Differences; Teacher Evaluation; *Test Bias; Whites
IDENTIFIERS *Computer Mice; Praxis Series

ABSTRACT

The possibility exists that some people might be at a disadvantage in registering their responses to a computerized test if mouse manipulation were complicated. This would be of particular concern if some groups were differentially affected by the complexity of manipulation. This study analyzed data from a pilot test conducted in 1991 for the Computer-Based Academic Skills Assessments of the Praxis Series of professional assessments for beginning teachers. Computerized test items in reading and mathematics that required simple or complex mouse manipulation were administered to 337 (244 females and 93 males) white examinees and 148 minority group examinees (104 females and 44 males). Examinees were asked to describe their own experiences with mouse use. Very little evidence of differential effects was observed for sex or ethnicity. Whatever differential effect may exist is difficult to disentangle from other test behavior differences, but there would seem to be little cause for concern about mouse use. (Contains two figures and two references.) (Author/SLD)

RESEARCH REPORT

THE EFFECT OF COMPLEXITY OF MOUSE MANIPULATION ON PERFORMANCE IN COMPUTERIZED TESTING

Philip K. Oltman

Educational Testing Service
Princeton, New Jersey
May 1994

The Effect of Complexity of Mouse Manipulation on Performance in Computerized Testing

Philip K. Oltman

Abstract

The purpose of the study was to explore the possibility that requirements for complex mouse manipulation in computer-based testing might make more of a negative impact on certain groups than on others. Very little evidence of such differential effects was observed in this study.

Rationale

Use of the mouse in responding to computer-administered test items may be difficult for some people, particularly those with little or no experience in using a mouse. With simple four-alternative multiple-choice items, using the mouse to select the desired answer would be unlikely to pose a problem even for the mouse neophyte. However, the possibility exists that with relatively more complex requirements for mouse response, such as setting dials, selecting or scrolling text, or setting the lengths of bars in graphs, some people may be at a disadvantage in registering their responses. This would be of particular concern if identifiable focal groups such as women or minorities experienced these problems at a higher rate than members of the reference group.

The present study addressed this question by providing an analysis of the data collected in a pilot test that was conducted beginning in the Fall of 1991 for the Computer-Based Academic Skills Assessments of The Praxis Series: Professional Assessments for Beginning Teachers™. The test forms that were used contained a range of mouse-response modes, from simple endorsement of multiple choices to relatively more complex highlighting and selecting of stimuli. One would expect little or no difference between focal and reference groups in difficulty with the mouse when the complexity of mouse manipulation is low. The question of interest was whether, as the complexity of the required mouse manipulation increases, some focal groups are put at a relative disadvantage.

Method

Briefly, the pilot study plan included three 40-item forms for each of three content areas (reading, mathematics, and writing). The present analysis was concerned only with the reading and mathematics items. Additional minority students were recruited to oversample beyond those who would otherwise have been represented. The study was conducted at four sites: Arizona State University at Tempe (high Hispanic and Native American population), Norfolk State University (an Historically Black University), University of Wisconsin at Stevens Point, and Rider College.

Various kinds of data were collected throughout the pilot testing. Of interest in this context were indicators of difficulty negotiating the screens with the mouse, such as making many clicks or using excessive time to complete items. In addition, questionnaires were administered that elicited subjective responses to the computerized test in general and the use of the mouse in particular, as well as self-reports of experience with computer and mouse.

Prior to collecting data for the study, items were classified into two groups based upon judgments of the complexity of the mouse manipulations they appeared to require. Items put in the "simple" group were those that required no more than

the selection of an answer from four or five alternatives. Items put in the "complex" group were those requiring more, e.g., highlighting or inserting blocks of text in reading items, and setting dials or filling in tables in mathematics items. Complex items generally required more clicks and more mouse movement to respond, especially when the examinee wished to change an answer. Two persons coded the items and achieved perfect agreement, probably because the coding scheme was quite unambiguous. Items were either single-click multiple choice, or they were something beyond that; the former were coded simple, and the latter were coded complex.

Results

In the description of the results to follow, differences between groups or between item types were judged statistically significant if they exceeded the .05 level. To help in judging the practical significance of any differences found, the percent of variance accounted for was also computed. Items referred to as complex were judged *a priori* to require somewhat more complex mouse manipulation than those referred to as simple. In this paper, the terms *simple* and *complex* do not refer to anything other than the mouse manipulation demands of the items.

Simple and Complex Items

The numbers of items judged simple and complex in each of the forms were as follows: Reading, Form 1 (30, 9), Reading , Form 2 (28, 11), Reading, Form 3 (27, 13); Mathematics, Form 1 (29, 11), Mathematics, Form 2 (30, 10), Mathematics, Form 3 (28, 12). (One item had to be dropped from Reading, Forms 1 and 2, due to technical problems with the display.) Thus, in each test form, about one quarter of the items were judged to be complex.

Examinees

Because of the limited numbers of examinees in each of the separate minority groups, they were combined. All ethnic group analyses were based on comparisons of the 337 white examinees (244 females, 93 males) with the combined group of 148 minority examinees (104 females, 44 males), deleting those examinees who did not report their ethnic group identification.

Self-Reports of Experience Using a Mouse

To the question, "How often have you used a mouse on a personal computer?," examinees responded on a scale running from 1 ("Never") to 4 ("Most of the time"). The sexes and ethnic groups did not differ significantly on this question. The mean over all was near the scale point labelled 2 ("A few times"), with minority examinees not significantly below this (1.96) and white examinees

just above (2.13). Over 76 percent of the total group reported that they had used a mouse "A few times" or "Never." In general, this group of examinees could not be said to be experienced mouse users. In a study by O'Neill and Kubiak (1992), some groups of minority examinees reported less mouse experience, and in that study those differences were statistically significant.

Objective Indications of Mouse Experience

Although in this study the sexes and ethnic groups did not differ in their self-reports of mouse experience, minority and white examinees did differ significantly in the time they took to complete a tutorial on using the mouse given just before the test administration. Minority examinees took an average of 5.1 minutes to complete the mouse tutorial, while white examinees averaged 3.3 minutes (this difference accounted for 12 percent of the variance, which is considered a medium-to-large effect [Cohen, 1988]). If we make the assumption that mouse tutorial time, measured by the computer during its administration, is a more valid reflection of mouse skill and experience than is a self-report questionnaire response, then we might conclude that minority examinees probably had somewhat less mouse skill/experience, even though as noted above, they did not differ in their self-reports of experience.

Sex Effects, Reading Items

For the reading items, means for percent correct, for number of mouse clicks per item, and for time (in seconds) per item are plotted by sex in the left column of Figure 1. Error bars depicting plus or minus one standard error of the mean are shown around each mean in the figures.

Percent correct. Analysis of variance indicated that the complex reading items were significantly more difficult than the simple reading items across all examinees ($F[1, 482] = 116.9, p < .001$; complex percent correct=57, simple percent correct=65). The sexes (347 females, 137 males) did not differ in percent correct, and the impact of complexity on percent correct was the same for the two sexes (i.e., there was no sex by complexity interaction). Removing the effects of test form, self-reported mouse experience, and mouse tutorial time by analysis of covariance did not appreciably change this pattern of results.

Clicks. Complex reading items elicited significantly more mouse clicks than simple reading items ($F[1, 482] = 51.2, p < .001$; mean clicks, complex=12; mean clicks, simple=11). There was no differential effect by sex. To provide an indication of whether the greater number of clicks on complex reading items might be due to differences in the rate of success for simple and complex item sets rather than mouse complexity, the analysis of clicks was done with percent correct as a covariate. Removing the effect of percent correct made little difference in the results. Similarly, removing the effects of test form, mouse experience, and mouse tutorial time had little effect.

Time per item. Complex reading items also took significantly more time per item ($F[1, 482] = 17.6, p < .001$; mean time, complex=96; mean time, simple=91). Both sexes showed this effect to a similar extent. Covarying on percent correct, test form, mouse experience, and mouse tutorial time did not appreciably change this pattern.

To summarize the results comparing the two sexes on the reading items, on the complex set of items examinees got fewer correct, made more clicks per item, and took more time (in seconds) per item, as compared with the simple items. These effects did not differ for the two sexes.

Ethnic Group Effects, Reading Items

For the reading items, means for percent correct, for number of mouse clicks per item, and for time per item are plotted by ethnic group in the right column of Figure 1.

Percent correct. Minority examinees got significantly fewer reading items correct than did white examinees ($F[1, 482] = 53.5, p < .001$; minority mean percent correct=52; white mean percent correct=65). As noted earlier, complex reading items were more difficult than simple reading items for everyone, but this was especially true for minority examinees (i.e., the interaction of complexity and ethnic group was significant: $F[1, 482] = 7.8, p < .01$). This interaction can be seen in Figure 1 as the difference in the slopes of the two lines in the top right graph. Removing the effect of test form, experience with using a mouse, and mouse tutorial time by analysis of covariance did not significantly change the pattern of results.

Although *statistically* significant, the difference in the magnitude of the response to complexity between minority and white examinees accounted for only 1.2 percent of the variance. One percent or less is considered a "small" effect, similar in size to e.g., the mean difference in height between 15- and 16- year-old girls (Cohen, 1988). (In Cohen's scheme, an effect would have to be approximately 6 times this large to be considered a "medium" effect, such as the height difference between 14- and 18-year-old girls, or the IQ difference between semiskilled workers and professionals.)

Clicks. Although the ethnic groups did not differ overall in their mean number of clicks per reading item, they did differ significantly in their response to the impact of complexity, in that minority examinees showed a larger increase in clicks per item when going from simple to complex reading items than did white examinees, paralleling the interaction noted in the previous section for percent correct ($F[1, 482] = 4.5, p < .03$). Removing the effect of percent correct and mouse experience did not change this pattern. However, covariance adjustment for test form and for mouse tutorial time reduced the difference to statistical nonsignificance.

Time per item. Minority examinees took significantly longer across all reading items ($F[1,482] = 86.8, p < .001$), and showed a significant interaction in mean time per item ($F[1, 482] = 4.0, p < .05$). That is, they increased their times more when going from simple to complex items than did white examinees. Covarying on mouse experience and on percent correct items did not change these results. However, as with clicks, covariance adjustment for test form and for mouse tutorial time reduced the interaction of complexity with ethnic group to nonsignificance.

Sex Effects, Mathematics Items

The means for the mathematics items are plotted in Figure 2, by sex in the left column, and by ethnic group in the right column.

Percent correct. For all examinees taken as a group, the complex mathematics items were significantly easier (in the sense of percent correct responses) than the simple items ($F[1, 482] = 5.3, p < .02$; complex percent correct=69; simple percent correct=61). Female examinees did slightly but significantly worse than male examinees on the mathematics items ($F[1, 482] = 5.9, p < .02$; female mean percent correct=64; male mean percent correct=68). However, the sexes did not differ in the effect of complexity on performance. Removing the effect of test form, mouse experience, and tutorial time by analysis of covariance did not change these results.

Clicks. Complex mathematics items elicited significantly more mouse clicks than the simple mathematics items ($F[1, 482] = 106.1, p < .01$; mean clicks, complex=10; mean clicks, simple=8). There were no differential effects by sex. Covarying on percent correct, test form, mouse experience, and mouse tutorial time did not change the nature of these results.

Time per item. Complex and simple mathematics items did not differ in time per item, and there were no differential effects by sex. Removing the effects of percent correct, test form, mouse experience, and mouse tutorial time had no effect.

Thus, for the mathematics items as for the reading items, no significant interactions with sex were observed. The mouse-complex items turned out to be easier than the mouse-simple items.

Ethnic Groups, Mathematics Items

Percent correct. As with the reading items, minority examinees got significantly fewer mathematics items correct overall ($F[1, 482] = 110.1, p < .01$; minority mean percent correct=52; white mean percent correct=71). As noted above the complex items had a higher mean percent correct overall, but this effect was less pronounced for minority examinees. That is, a significant ethnic group by

complexity interaction was observed; it accounted for 0.9 percent of the variance($F[1, 482] = 5.2, p < .02$). Covarying on test form and mouse experience did not change this pattern. However, removing the effect of mouse tutorial time reduced the complexity by ethnic group interaction to statistical nonsignificance.

Clicks. Minority examinees increased their mouse clicking significantly more than did white examinees when moving from simple to complex mathematics items ($F[1, 482] = 10.7, p < .01$; the interaction accounted for 1.7 percent of the variance). Covarying on percent correct reduced the size of this interaction so that it was no longer significant, but covarying on test form, mouse experience, and mouse tutorial time did not eliminate the significance of the interaction.

Time per item. When moving from simple to complex mathematics items, minority examinees increased their mean time per item, while white examinees actually showed a slight *decrease* in time per item ($F[1, 482] = 12.6, p < .01$). This interaction (2.5 percent of the variance) was reduced to nonsignificance when percent correct was covaried out, but was unaffected by covariance adjustment for test form, mouse experience, and mouse tutorial time.

To summarize the ethnic group results for reading and mathematics items, when moving from simple to complex items, minority examinees showed either a greater decrease in percent correct (reading) or a smaller increase in percent correct (mathematics), and greater increase in mouse clicks and time per item than did white examinees. However, each of these interactions, except that for percent correct on the reading items, was attenuated when covariates were introduced. The size of the latter effect was quite small in comparison to the differences between groups, item types, and individuals.

Summary of Interactions with Sex and Ethnic Group Membership

The main interest in these data is the possibility that complex mouse demands might make more of a negative impact on certain groups than on others. Very little evidence of such differential effects was observed in this study. The tables below summarize the interaction results that bear on this concern, in terms of variance accounted for.

Interactions with Sex	Reading, Complexity by Sex Interaction	Mathematics, Complexity by Sex Interaction
Percent Correct	< 0.1%	< 0.1%
Clicks per Item	< 0.1%	0.2%
Time per Item	0.3%	< 0.1%

Note. None of the above effects was significant.

Interactions with Ethnic Group	Reading, Complexity by Ethnic Group Interaction	Mathematics, Complexity by Ethnic Group Interaction
Percent Correct	1.2%	0.9%[a]
Clicks per Item	0.8%[b]	1.7%[c]
Time per Item	0.8%[b]	2.5%[c]

[a]Not significant when mouse tutorial time was partialled out by analysis of covariance.

[b]Not significant when test form and mouse tutorial time were partialled out by analysis of covariance.

[c]Not significant when percent correct was partialled out by analysis of covariance.

In the first table above, we see that the increase in mouse complexity makes about the same difference for female and male examinees, and therefore there were no significant complexity by sex interactions found. The second table shows some indication that moving from simple to complex mouse demands may cause more problems for minority than for white examinees. Initially, percent correct, clicks per item, and time per item each showed significant interactions with minority/majority status. However, after covariance adjustments, only the minority group's differential performance decrement on the complex reading items remained significant.

The results suggest that for the most part, the differential effects of mouse complexity on the sexes and on ethnic groups are either non-existent or minor. Minority examinees made differentially more mouse clicks and took more time on the complex items than white examinees. However, these differences were removed after covariance adjustment (for mouse tutorial time on the reading items, and percent correct on the mathematics items). Whatever differential effect there may be is difficult to disentangle from other test behavior differences between minorities and whites, including performance level and mouse proficiency.

To the extent that there remains a concern about the effects of differential experience with mice, one approach may be to introduce a criterion-based tutorial tc bring all examinees up to a predetermined level of skill in using the mouse. The criterion might even include click counts and timing on sample items administered at various points in the mouse tutorial. Examinees could be required to reduce their mouse click rates and times per item to some criterion level during the mouse tutorial, prior to proceeding to the test proper. Data could be analyzed on an ongoing basis to determine if mouse complexity effects dissipate as higher levels of mouse skill and experience develop.

mouse-2\mousrep6.wpd

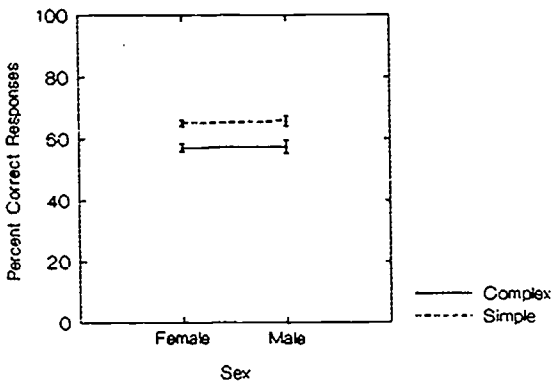
References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

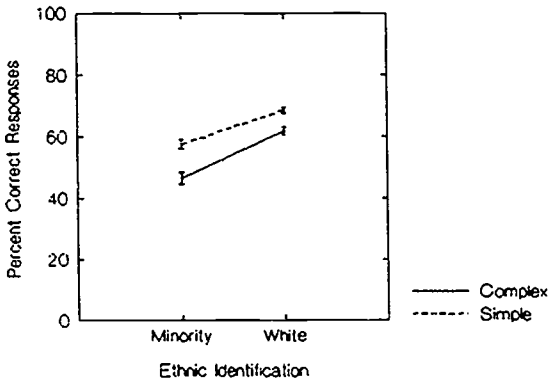
O'Neill, K., & Kubiak, A. (1992, April). *Lessons learned from examinees about computer-based tests: Attitude analyses.* Paper presented at the meeting of the National Council on Measurement in Education, Cincinnati, OH.

Figure 1. Reading Items

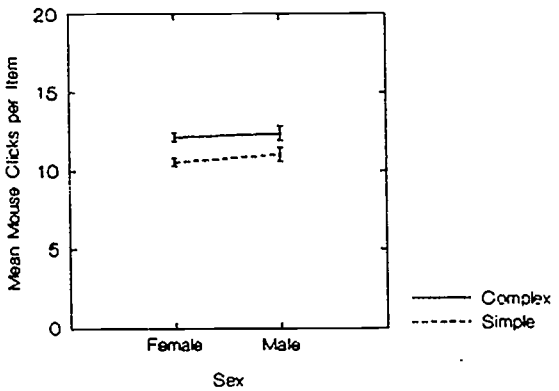
Percent Correct by Sex



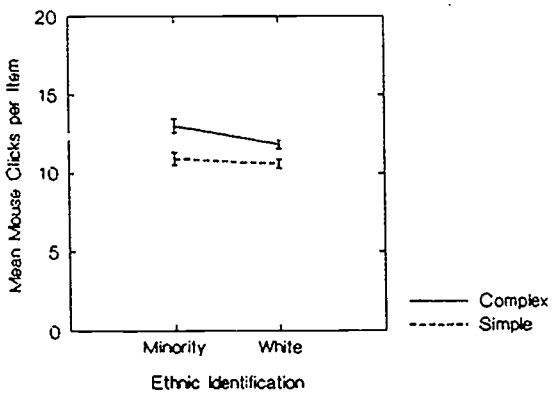
Percent Correct by Ethnic Group



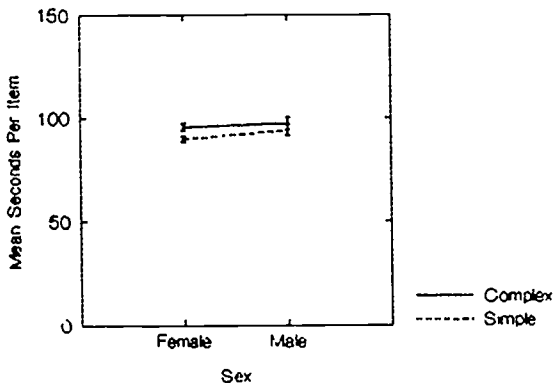
Mouse Clicks by Sex



Mouse Clicks by Ethnic Group



Seconds per Item by Sex



Seconds per Item by Ethnic Group

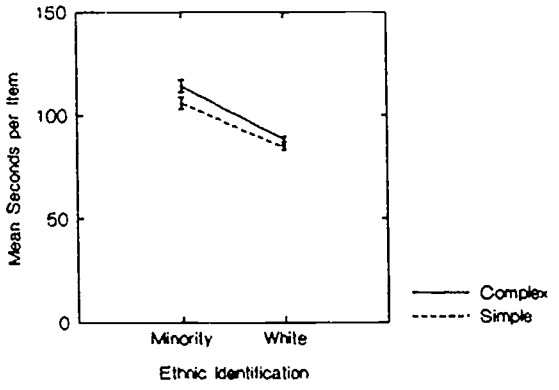
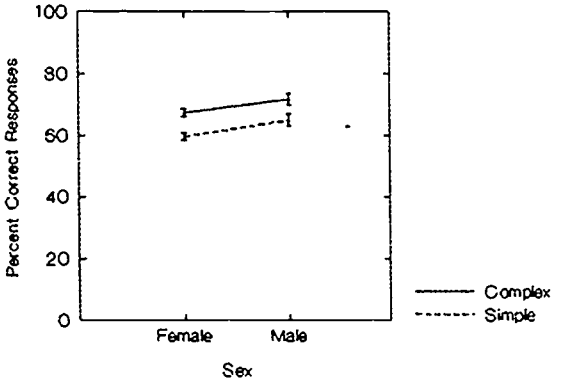
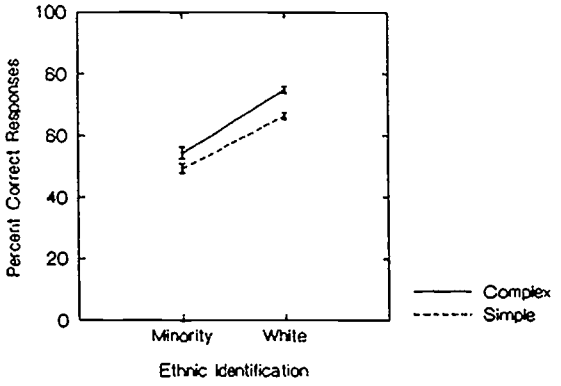


Figure 2. Mathematics Items

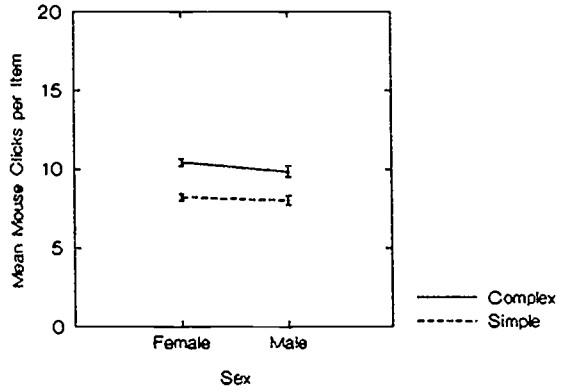
Percent Correct by Sex



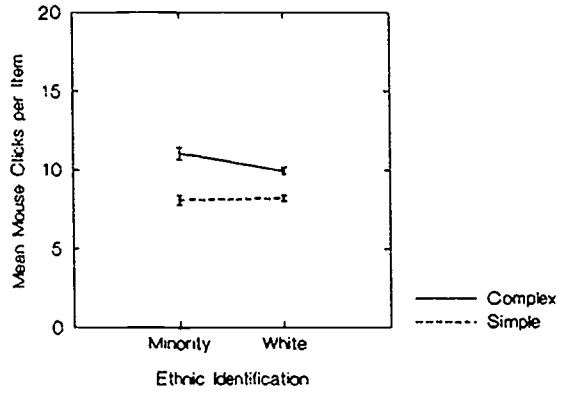
Percent Correct by Ethnic Group



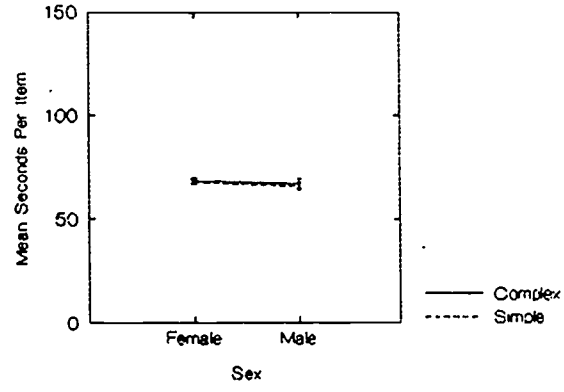
Mouse Clicks by Sex



Mouse Clicks by Ethnic Group



Seconds per Item by Sex



Seconds per Item by Ethnic Group