Exploration of Matrix Questions with Fourth Grade Students Using Eye-Tracking

Aaron Maitland
Hanyu Sun
Andrew Caporaso
Roger Tourangeau
Westat
1600 Research Blvd.
Rockville, MD 20850

Jonas Bertling
Debby Almonte
Educational Testing Service
660 Rosedale Road
Princeton, NJ 08541

Introduction

Matrix questions have been commonly used in Web surveys as well as in other types of surveys and are a standard survey questionnaire item format in all NAEP grade 8 and 12 student questionnaires as well as teacher and school questionnaires. Matrix questions are currently not used in NAEP fourth grade student questionnaires. In matrix questions, a series of items is usually presented in the rows of a table with a common set of response options usually presented in the columns of the table. The literature shows respondents took less time to answer questions when they were presented in a matrix than when they were presented individually across separate pages or screens (e.g., Couper, Traugott, and Lamias 2001; Tourangeau, Couper, and Conrad 2004). The use of matrix questions, however, may be associated with several undesirable outcomes, including higher breakoff rates (e.g., Jeavons 1998), higher missing data rates (e.g., Iglesias, Birk, and Torgerson 2001; Toepoel, Das, and van Soest 2009, found this, but Couper et al. 2001, find the opposite), as well as higher levels of straightlining (e.g. Tourangeau et al. 2004, found this, but Bell, Mangione, and Kahn 2001, Yan 2005, and Callegaro, Shand-Lubbers, and Dennis 2009, did not).

Eye-tracking has been used in psychology to study cognitive processing and visual attention (e.g., Duchowski 2007). In the survey literature, eye-tracking has been used to better understand the survey response process. For instance, eye-tracking has been used to identify questions that are difficult to comprehend (Graesser et al. 2006). It has also been used to examine how to present long lists of response options (Redline et al. 2009). Additionally, eye-tracking has been used to measure the length of fixation on definitions in Web surveys (Galesic et al. 2008). This project will use eye-tracking technology to observe how fourth grade students navigate matrix items in questionnaires in comparison to single item questions.

As part of the cognitive interview efforts to develop new survey questions for the 2017 NAEP student questionnaires, the NAEP survey questionnaire team used eye-tracking to test how fourth grade students answer matrix items. The need for this study arises from questionnaires transitioning to tablet computers and the desire to use matrix questions across all grades following recommendations of the NAEP Questionnaire Standing Committee (QSC), other survey questionnaire experts, and requests from NCES to ensure that fourth grade students can answer matrix questions if this question type is used in future operational assessments.

Study Design

Recruitment: Nine fourth grade students were recruited to participate in this eye-tracking study. In general, smaller numbers of participants can be used in eye-tracking studies since the purpose of the study is to understand eye pattern movements rather than produce survey estimates that are generalized to a population. Eye-tracking studies produce relatively rich data since multiple data points can be collected on each item per respondent. For example, most of our analyses are concerned with how respondents navigate across multiple items rather than any individual item. All study materials were cleared through Westat's Institutional Review Board. We recruited from Westat employees or acquaintances of Westat employees who have children in fourth grade. We posted advertisements on the Westat Intranet so that interested parents may contact us via email or telephone. The demographics of the students who participated in the study are shown below in Table 1.

Table 1. Demographics of eye-tracking participants

Demographics	n
Gender	
Male	4
Female	5
Race/Ethnicity	
Non-Hispanic Asian	2
Non-Hispanic Black or African American	1
Non-Hispanic White or Caucasian	5
Hispanic White or Caucasian	1
Parent's Education	
Associated degree: Occupational, technical, or vocational program	1
Bachelor's degree (example: BA, AB, BS, BBA)	4
Master's degree (example: MA, MS, MEng, MEd, MBA)	3
Doctoral degree (example: PhD, EdD)	1
Total	9

Each participating student was offered $25 to encourage participation and thank participants for their time and effort. A parent or legal guardian was required to bring their student to and from the testing site and they also received $25. The study was described as a NAEP questionnaire pretest with fourth grade students which lasted for approximately one hour.

Procedure: Interested parents were asked to sign up for a one hour time slot. At the scheduled interviewing day and time, a researcher met with the parent(s) and the student, first, to go over the consent process. The student and the parent were asked to read and sign the consent form. Next, a technician worked with the student to set up and test the eye-tracking device with the student. Mobile Eye-XG eye-tracking glasses were used as the eye-tracking device in this study. The eye-tracking device was set up by connecting the glasses and the display/transmit unit (DTU) with the eye-tracking software installed on a laptop computer. The laptop computer was used to record the eye movement during the study as well as monitor the eye-tracking data quality in real time.

Prior to beginning the survey task, the technician administered a calibration procedure to ensure that the eye-tracking device was tracking the students' eye movements accurately. Once satisfactory calibration was achieved, the student began the main task of answering the questions on the tablet computer. The technician stayed in the interviewing room to observe the session and to provide any necessary technical assistance. The technician was primarily observing the quality of the eye-tracking calibration. It was usually necessary to recalibrate the eye-tracking glasses after the student began answering the survey questions. The technician stopped the interview and assisted the student with recalibration when necessary. The student then resumed the survey once the calibration was satisfactory. The researcher also remained in the room. The researcher watched for any potential difficulty the respondent had with answering the survey questions and administered the debriefing questions in Appendix B once the student completed the eye-tracking task.

Experimental Design: The questionnaire administered during the eye-tracking study highlighted five content areas: grit, self-efficacy, desire for learning, school climate, and technology use. All items are part of the larger item pool that was developed for pre-testing in cognitive labs, reviewed by NCES and various expert groups. Each participant received 60 items drawn from all five content areas. These items were presented as either discrete (i.e., one item per page) or in a matrix (i.e., an item containing multiple sub-items) format. Specifically, each participant received 15 discrete items, and 9 matrix items. The matrix items varied in size. Three matrix items contained two sub-items, three matrix items contained five sub-items, and three matrix items contained eight sub-items.

Two versions of the questionnaire were created (Booklet A and B) and shown in Appendix A. In Booklet A the participant will first be administered nine matrix items regarding technology, desire for learning, grit, school climate, and self-efficacy. The participant will then be administered 15 discrete items regarding grit, self-efficacy, and school climate (see Table 2). In Booklet B the participant will first be administered 15 discrete items regarding school climate, desire for learning, and grit. The participant will then be administered matrix items regarding technology, grit, school climate, and self-efficacy (see Table 2).

This study design was a mixed within-subjects and between-subjects design that varies the format of survey questions—discrete versus matrix. The overall design is shown in Table 2. Each discrete question was presented on a separate screen followed by its response options. With the matrix condition, a series of items was presented in the rows of a table, sharing a common set of response options.

Table 2. Experimental Design

Item	Number	Form A		Form B	
Set	of items	Construct	Format	Construct	Format
1	2	School Climate	Matrix	School Climate	Discrete
2	5	Desire for Learning	Matrix	Desire for Learning	Discrete
3	8	Grit	Matrix	Grit	Discrete
4	2	School Climate	Matrix	Technology	Matrix
5	5	Grit	Matrix	School Climate	Matrix
6	8	Self-Efficacy	Matrix	Technology	Matrix
7	2	Technology	Matrix	School Climate	Matrix
8	5	School Climate	Matrix	Grit	Matrix
9	8	Technology	Matrix	Self-Efficacy	Matrix
10	2	Grit	Discrete	Grit	Matrix
11	5	Self-Efficacy	Discrete	Self-Efficacy	Matrix
12	8	School Climate	Discrete	School Climate	Matrix

Respondents were randomly assigned to receive one of the two questionnaire formats described in Table 2: (A) first answer questions in the matrix format and then switch to the discrete format; or (B) first answer questions in the discrete format and then switch to the matrix format. Questions in item groupings 1-3 and 10-12 were analogous questions across versions A and B in both the discrete and matrix formats. The matrix questions in item groupings 4-9 were about the same construct, but were shown in a different order across versions A and B.

Following the survey, a set of debriefing questions were given to the student (see Appendix B). The debriefing questions will assess their satisfaction as well as cognitive difficulties with the matrix questions. The length of the self-administered questionnaire was approximately 30 minutes, followed by a debriefing session of approximately 5-10 minutes.

<u>Data Preparation</u>: Once the eye-tracking videos were recorded they were prepared for analysis using the ASL Results Plus Monitor Tracking software package. There are several steps involved in preparing the eye-tracking data for analysis.

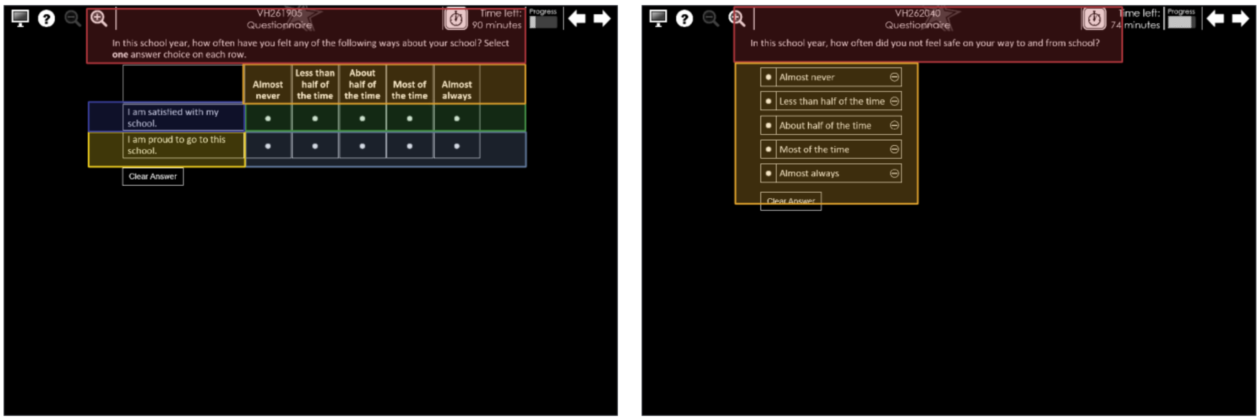
1. Each video is uploaded and parsed in to 24 separate events corresponding to the 24 unique screens on which each discrete or matrix item is presented in the survey.

2. For each event, a monitor tracking procedure is performed to specify to the software package where the monitor is located within the video.

3. The background for each survey item is uploaded and the areas of interest (AOI) are defined for each item – these are the specific pieces of the questionnaire items on which fixation data is desired. The

AOI are ‘painted’ on to a screenshot of a survey item which serves as the background; two examples of this are shown in Figure 1. Figure 1 illustrates the background and AOI for a two subitem matrix question and one discrete question. The two subitem matrix question on the left has six AOI defined including the question introduction, response categories, two subitems, and two answer spaces. The discrete question has AOI defined for only the question and the response categories.

Figure 1. Illustration of AOIs for a two subitem matrix and one discrete question



4. Once the background and AOI are specified for an item, each event corresponding to that item is configured. This involves specifying to the software where, within the event, the background can be seen and only has to be done on one frame of the video where the background can be seen clearly (for this project the background was the entire screen).

5. Finally, a rectangular fixation level raw data file is output for analysis. A fixation occurs when the respondent focuses on a single location for at least 100 milliseconds. In order to illustrate the structure of the data, an example of the raw data file including the key variables for fixation sequence and fixation duration is shown in Table 3. The example in Table 3 demonstrates a respondent who has five initial fixations on the question. The next five fixations are on the response categories. Then the respondent has two more fixations on the question followed by one final fixation on the response categories. The duration of each fixation varies between 100 and 160 milliseconds. The final column shows the start time for each fixation and can be used to calculate the time to first fixation from each AOI.

Table 3. Example of key variables in raw (fixation level) data file

Respondent	Item	AOI	AOI Type	Fixation Sequence	Fixation Duration	Start Time
1	1	1	Question	1	150	150
1	1	1	Question	2	100	250
1	1	1	Question	3	120	370
1	1	1	Question	4	110	480
1	1	1	Question	5	100	580
1	1	2	RC	6	100	680
1	1	2	RC	7	100	780
1	1	2	RC	8	150	930
1	1	2	RC	9	110	1040
1	1	2	RC	10	100	1140
1	1	1	Question	11	150	1290
1	1	1	Question	12	160	1450
1	1	2	RC	13	100	1550

Quality control is performed throughout the process. For example, a research analyst watches the videos using the analysis software and observes how accurately the software is mapping the students' eye movements to AOI. Each event was graded based on its eye tracking quality as well as the quality of the AOI identification within the event. Poor data quality was most common in the beginning of the survey which resulted in more missing data for the initial questions. This was because respondents tended to shift their position following calibration and prior to beginning the survey.

One item presented significant data quality issues and was excluded from the analysis. This was matrix item two in Version A of the survey. This item was different from the others because it spanned beyond the length of one screen so respondents had to scroll in order to see the entire item. In order for the analysis software to capture AOI data accurately for such a stimulus (which can be presented in an infinite number of ways depending on how the user scrolls through it), a special method of AOI specification is required. This method involves specifying the AOI in a video that serves as the item's background as opposed to a single screen shot like every other item. Upon reviewing the AOI fixation data for item two using this method it was found that the software could not accurately differentiate between different components of the question once the screen moved.

<u>Analysis Plan</u>: We have four research questions to guide our analysis. First, we examine whether matrix items require more effort to answer than discrete items for fourth grade students. Second, we investigate how the processing of sub items change within a matrix. Third, we examine how the processing of questions change over time. In order to address these research questions, we examine difference in the mean number of fixations per word and the mean duration per word for matrix and discrete questions. The number of fixations is related to the amount of information that a respondent is processing (Ares et al., 2014). The duration of fixations is related to the amount of difficulty that the respondent is having extract information from an AOI (Ares et al., 2014).

Fourth, we investigate how fourth grade students navigate through matrices. In order to address this research question, we examine data on time to first fixation for each AOI in the matrix. Time to first fixation measures attentional capture of an AOI (e.g. subitem) and the order in which the AOIs are viewed (Ares et al., 2014).

Finally, we examine respondents' preferences for matrix and discrete questions. Data from self-reported difficulty about the different types of questions will be used to address this research question.

Findings

<u>Differences in Level of Effort for Matrix and Discrete Questions:</u> We first present results that show the level of effort that respondents undertake to answer questions in matrix and discrete questions. The table compares five sets of analogous questions at the beginning and end of each version of the questionnaire that were asked in both matrix and discrete formats.[1] With matrix items, introductions were only presented once at the beginning of the entire matrix. With discrete items, introductions were given at the beginning of each question. Table 4 shows more mean fixations and longer mean durations for discrete items compared to matrix items. Respondents took an average of 64.3 fixations to process the introduction and subitems matrix items and an average of 81.9 fixations to process discrete items. In addition, it takes 57.2 fixations to process the response categories and answer space with matrix items and 66.8 fixations to process response categories with discrete items.

[1] Matrix question Q2 required scrolling that makes it difficult for the eye-tracking software to accurately capture the area of interest where a respondent is looking. Therefore, we excluded matrix question Q2 and discrete questions Q3-Q7 from this analysis.

Table 4. Processing of information in matrix and discrete questions

Question components	Matrix		Discrete	
	Mean fixation	Mean fixation duration	Mean fixation	Mean fixation duration
Introduction	17.5 (12.6)	249 (152)	81.9 (66.7)	248 (170)
Sub item	46.8 (42.1)	259 (174)		
Response categories	14.9 (10.4)	270 (208)	66.8 (48.4)	291 (217)
Answer space	42.3 (34.6)	315 (265)		
Overall	121.5 (77.1)	278 (214)	148.7 (112)	267 (194)

N=9.

The findings suggest that respondents read introductions and response categories once or only a few times with matrix items but almost every time with discrete items. The format of matrix items seems to facilitate the cognitive processing of respondents with matrix items.

Next we looked at variation in fixations and duration within respondents across items in the questionnaire. Figure 2-1 illustrates this within respondent variability in mean fixations for respondents answering questionnaire version A. Each respondent is represented by a line in the figures. Questions Q1-Q9 are matrix questions and the data points represent mean fixations per item within each matrix. The rest of the questions are discrete questions and the data points represent the number of fixations on the question as a whole since there is only one question per screen. Figure 2-2 illustrates the within respondent variability in mean fixation duration for respondents answering questionnaire version A. The figures show that some respondents required more fixations answering the first matrix item (Q1) compared to the other items. However, it appears that mean fixations decrease once the respondents learned how to answer the matrix items.

The mean fixation durations vary across respondents but the overall pattern appears to be similar between matrix items and discrete items. There are also some other spikes in the data throughout the questionnaire that may be related to difficulty with individual items.

Figure 2-1. Mean Fixation with Version A

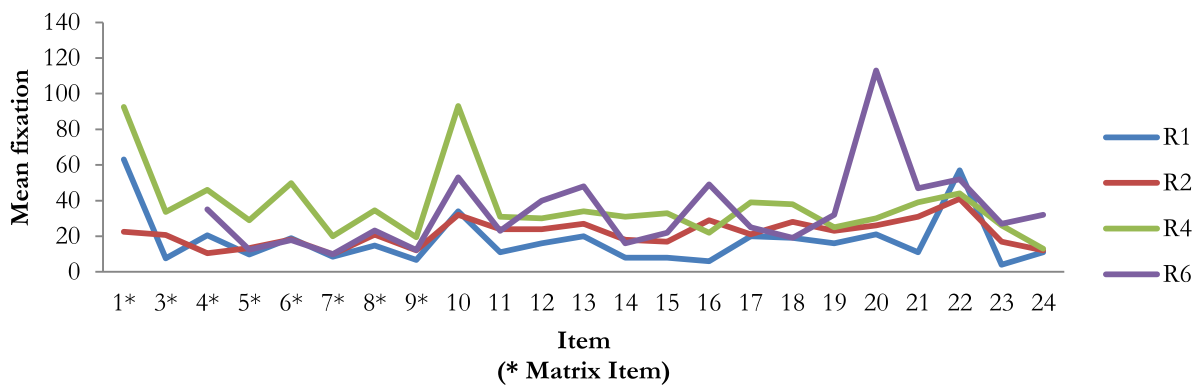
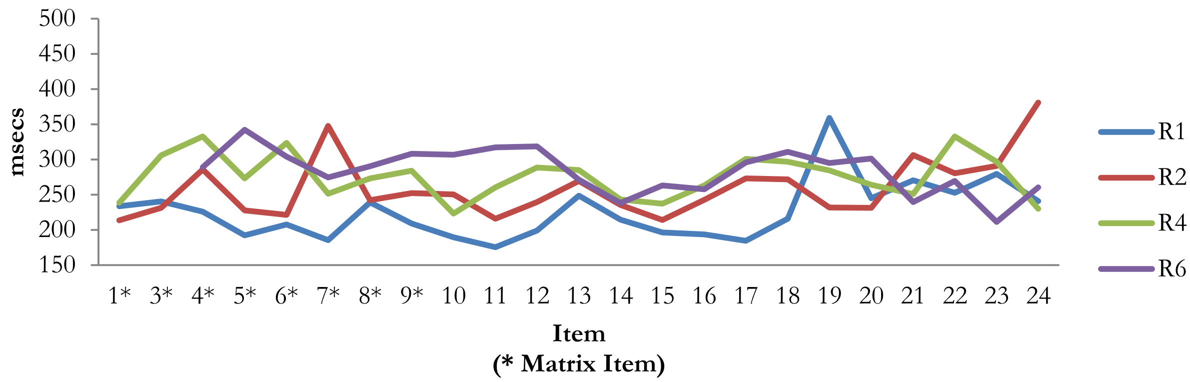


Figure 2-2. Mean Fixation Duration with Version A



Figures 3-1 and 3-2 show a similar story for questionnaire version B. Questions Q16-Q24 are matrix items in these two figures. The fixation data in Figure 3-1 are once again presented in terms of the number of fixations per item within a matrix.

Figure 3-1. Mean Fixation with Version B

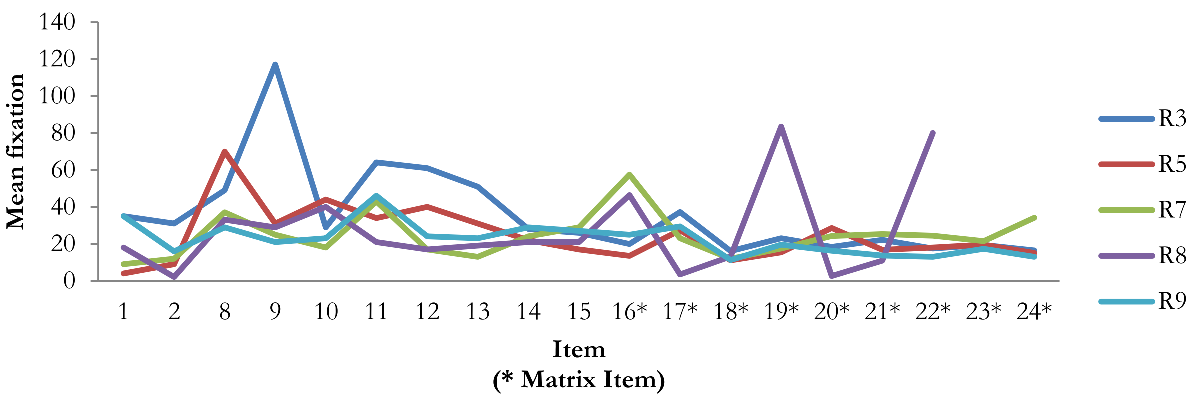
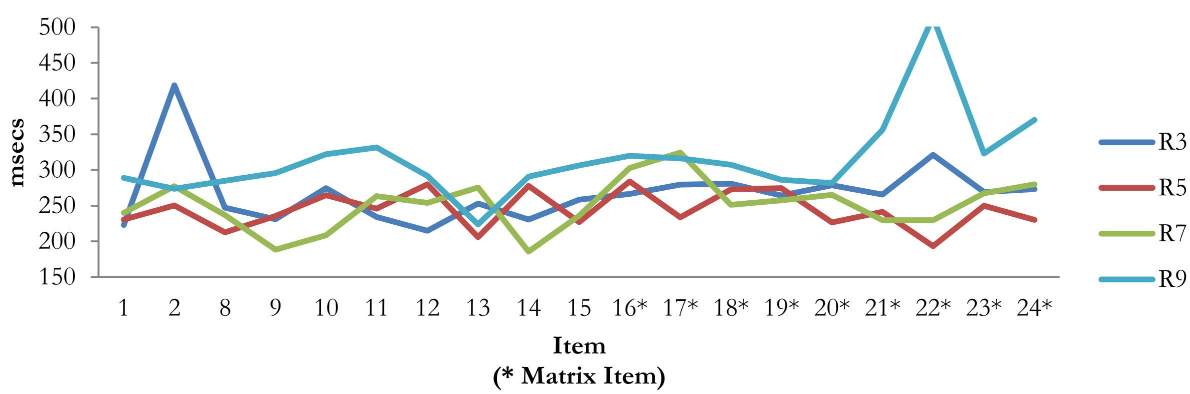


Figure 3-2. Mean Fixation Duration with Version B



Differences in Processing of Items Within a Matrix: Next we examine respondents' fixations and duration throughout the course of completing matrix items. The mean number of fixations per word throughout matrices of different sizes is shown in the top panel of Table 5 and the mean duration is shown in the bottom panel of the table. The table shows very little variation in fixation until the answer space for the last sub item in the table for matrices with five and eight sub items. It appears that the respondents had fewer fixations on these answer spaces compared to the answer spaces for other sub items. Overall there are only minor differences in mean duration across sub items and answer spaces.

Table 5. Processing of information within matrices

Sub item	Two sub items		Five sub items		Eight sub items	
	Sub item wording	Answer space	Sub item wording	Answer space	Sub item wording	Answer space
Mean number of fixations per word						
1	0.8 (0.7)	2.4 (3.1)	0.9 (0.7)	2.0 (1.1)	0.9 (0.6)	1.7 (1.4)
2	1.4 (1.5)	1.8 (2.4)	1.1 (0.6)	1.2 (0.7)	1.4 (0.8)	1.5 (1.2)
3			1.1 (0.6)	1.0 (0.6)	1.2 (1.4)	2.0 (1.4)
4			1.3 (0.9)	1.4 (0.8)	1.1 (0.9)	1.6 (1.1)
5			0.8 (0.5)	0.9 (0.4)	1.2 (0.9)	1.6 (1.3)
6					1.0 (0.5)	1.6 (1.2)
7					0.7 (0.7)	1.1 (0.7)
8					1.1 (1.3)	0.9 (0.7)
Mean duration per word in milliseconds						
1	204 (212)	771 (1105)	238 (196)	607 (409)	233 (188)	546 (506)
2	381 (410)	675 (955)	259 (151)	394 (236)	371 (243)	446 (336)
3			253 (152)	278 (207)	355 (453)	605 (445)
4			357 (271)	448 (392)	285 (283)	533 (464)
5			178 (185)	281 (157)	305 (256)	516 (505)
6					249 (153)	595 (688)
7					189 (210)	408 (338)
8					259 (349)	321 (255)
Number of questions	6		5		6	
Number of respondents	9		9		9	

Processing of Questions Over Time: We also examined the processing of all questions across versions of the questionnaire. The first three and last three sets of items in the questionnaire contain similar question content across forms of the questionnaire. Hence, the level of processing can be directly compared for these sets of items across versions of the questionnaire. Figure 4 shows the mean number of fixations across each series of questions in each version of the questionnaire. The data show that there are relatively large differences between the matrix and discrete items in terms of mean fixation for the first set of items. This difference is much smaller for the remaining sets of items. Unfortunately, we were unable to use the data for the second set of items, because the matrix screen required scrolling making it difficult for the eye-tracking software to accurately capture eye-movements for that set of items. It is also worth noting that that there are fewer mean fixations for matrix than discrete format in the last three sets of items. It seems to suggest that respondents spent more time to process the first matrix question when it was given at the beginning of the survey. After that first matrix question, however, the amount of information respondents processed was lower in matrix than discrete questions.

Figure 4. Mean fixations across each series of questions in each version of the questionnaire

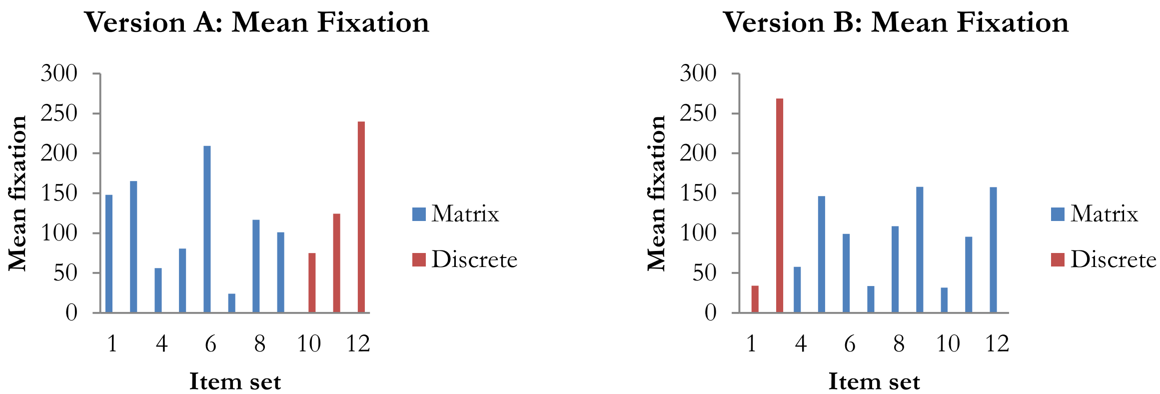


Figure 5 shows the mean fixations per item within each series of questions in each version of the questionnaire. The data tells the same story— there are more mean fixations with matrix than discrete items for the first set of items, but by the end of the questionnaire there are fewer mean fixations with matrix than discrete items. This supports the notion that there may be some initial learning about matrix items that requires deeper processing, but there seems to be less processing on matrix items compared to discrete items once respondents learned how to answer matrix items.

Figure 5. Mean fixations per item within each item set

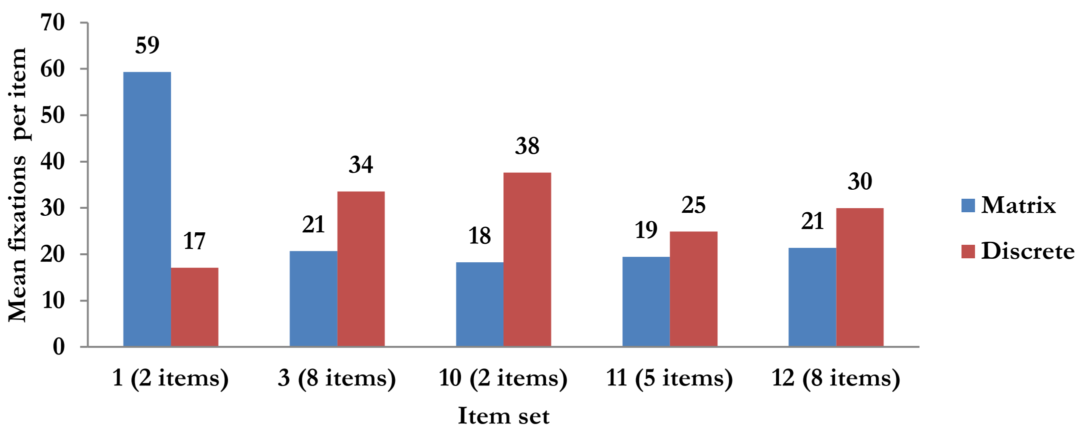


Figure 6 shows the mean fixation durations across each series of questions in each version of the questionnaire. The differences are small between matrix and discrete items across all series of questions. The data suggests that matrix and discrete questions are similar on comprehension difficulty.

Figure 6. Mean fixation durations across each series of questions in each version of the questionnaire

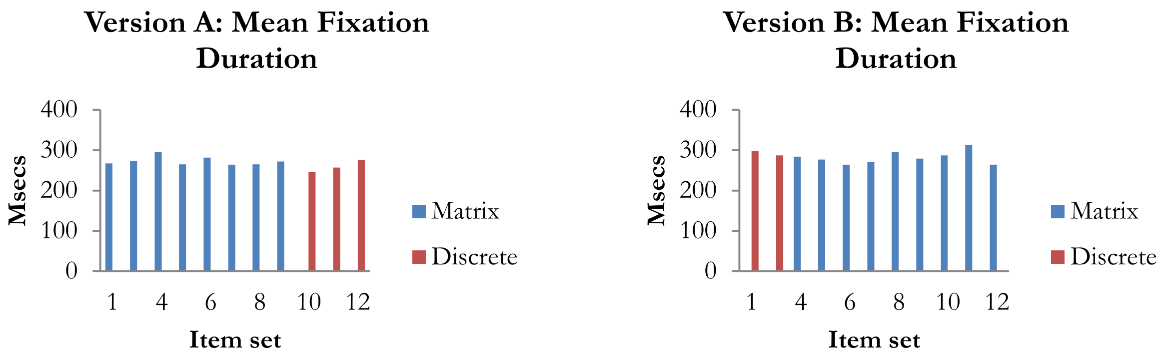
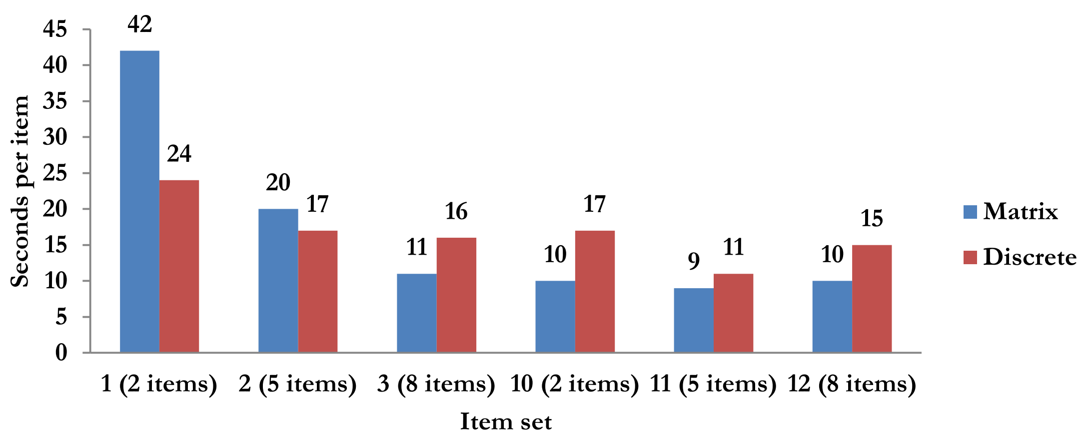


Figure 7 shows the average amount of time per item within each series of questions in each version of the questionnaire. The data suggests that there are relatively large differences between matrix and discrete items in terms of mean amount of time with the first set of the items, but by the end of the questionnaire there are small differences between matrix and discrete items. Hence, respondents spent more time at the beginning of an interview to learn how to answer a matrix question; however, once they learned how to answer the matrix items the amount of time that they spent on matrix items was either similar to or slightly less than the time spent on discrete questions.

Figure 7. Mean fixation durations per item within each item set



<u>Navigation Through Matrices</u>: We also looked at how respondents navigate through matrices. Our primary goal was to determine how linearly respondents navigate through a matrix. A linear sequence would involve the respondent reading the introduction followed by the response categories and then proceeding to the sub items and answer spaces. Table 6 shows the sequence in which respondents looked at different areas of interest. We use the mean time to first fixation to understand when, on average, respondents are focusing their attention on an AOI for the first time. Respondents appear to be more likely to follow a linear pattern on matrices with two or five sub items. It appears that the respondents are drawn to the sub items more quickly for an eight item matrix compare to two and five sub item matrices. Our experimental design does not allow us identify precisely why this pattern might occur. It is possible that the additional amount of information in a longer matrix draws the respondents' eyes to the sub items more quickly; however, it is also possible that respondents had learned the response categories by the time they saw an eight item matrix since every respondent sees both a two and five sub item matrix before an eight sub item matrix. In any event, a linear pattern clearly emerges on the latter half of eight sub item matrices.

Table 6. Navigation through matrices

Question part	Two sub items		Five sub items		Eight sub items	
	Sequence	Mean first fixation	Sequence	Mean first fixation	Sequence	Mean first fixation
Introduction	1	1.4 (1.6)	1	0.9 (1.1)	1	4.2 (10.0)
Response Categories	2	6.8 (7.2)	2	3.6 (4.0)	5	13.6 (19.9)
Sub item 1	3	7.4 (7.1)	3	6.5 (5.2)	2	5.6 (4.9)
Answer Space 1	5	11.1 (8.4)	4	10.3 (9.6)	4	13.4 (8.9)
Sub item 2	4	9.2 (8.0)	5	15.0 (9.6)	3	9.4 (8.6)
Answer Space 2	6	13.3 (9.1)	7	21.7 (9.1)	8	20.6 (12.7)
Sub item 3			6	20.0 (9.5)	6	14.0 (8.0)
Answer Space 3			8	24.9 (10.0)	7	17.9 (10.3)
Sub item 4			9	26.8 (11.8)	9	20.8 (12.4)
Answer Space 4			10	27.6 (16.0)	10	25.3 (14.7)
Sub item 5			11	30.2 (14.4)	11	31.0 (20.6)
Answer Space 5			12	33.8 (14.4)	12	36.6 (24.6)
Sub item 6					13	38.2 (27.3)
Answer Space 6					14	47.3 (26.6)
Sub item 7					15	48.7 (26.1)
Answer Space 7					16	51.0 (30.9)
Sub item 8					17	52.9 (33.5)
Answer Space 8					18	55.3 (35.2)
Number of questions	6		5*		6	
Number of respondents	9		9		9	

*Matrix (Q2) is excluded from this analysis due to poor data quality.

Respondent Preferences for Matrix or Discrete Questions: At the end of the interview we conducted a short debriefing interview. We were primarily interested in whether the respondents had any preference for matrix or discrete questions. Hence, in the debriefing interview we explained the difference between a matrix and a discrete question and showed them example of each. We then asked them to rate each type of question on a five point scale from very easy to very difficult. Table 7 illustrates that the students thought that both types of questions were easy to answer. Six respondents thought that matrix questions were very easy and three respondents thought that matrix questions were somewhat easy. In contrast, four respondents thought that discrete questions were very easy and five thought that these questions were somewhat easy.

Table 7. Respondents' rating of the difficulty of different types of questions

	Matrix	Discrete
Very easy	6	4
Somewhat easy	3	5
Neither easy nor difficult	0	0
Somewhat difficult	0	0
Very difficult	0	0
Total	9	9

We also asked respondents whether it was easier to answer the discrete questions or the matrix questions. As shown in Table 8, six respondents thought the matrix questions were easier, one thought that discrete questions were easier and two were not sure which one was easier.

Table 8. Respondents' preferences for different type of questions

Question type	Preference
Discrete questions	1
Matrix questions	6
Not Sure	2
Total	9

Conclusion

The NAEP eye-tracking study investigated the use of matrix type questions for fourth grade students on NAEP questionnaires. Our findings show that there are some differences between how respondents process matrix and discrete questions. In general, the largest amount of processing occurs when respondents answer the first matrix question that they encounter in the questionnaire. The first matrix question produces the largest difference between matrix and discrete type questions. However, the differences in the amount of processing required between the two question types are relatively minor after the first matrix question. In other words, the respondents learn how to answer these types of questions relatively quickly. Furthermore, we see little evidence from the eye-tracking data or from observation of the students in the lab suggesting that fourth grade student have difficulty navigating matrix questions. With a few exceptions, they seem to proceed more or less linearly as one would expect through these types of questions. In addition, respondents do not seem to have problems with scrolling to see the entire matrix when necessary. Moreover, respondents expressed preference for answering matrix questions compared to discrete items.

In order to prevent any difficulty with some of the first matrix items that a student encounters, the NAEP questionnaire could include some screens that train the respondents on the different types of questions in the instrument. For example, a brief screen could show the different parts of a matrix and how to answer the sub items.

References

Ares, G., Etchemendy, E., Antunez, L., Vidal, L., Gimenez, A & Jaeger, S. (2014). Visual Attention by consumers to check-all-that-apply questions: Insights to support methodological development. *Food Quality and Preference*, 32, 210-220.

Bell, D. S., Mangione, C. M., & Kahn, C. E. (2001). Randomized testing of alternative survey formats using anonymous volunteers on the World Wide Web. *Journal of the American Medical Informatics Association*, *8*(6), 616-620.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer.

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-Tracking data new insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, *72*(5), 892-913.

Callegaro, M., Shand-Lubbers, J., & Dennis, J. M. (2009). *Presentation of a single item versus a grid: Effects on the vitality and mental health subscales of the sf-36v2 health survey.* Paper presented at the annual meeting of the American Association for Public Opinion Research, Hollywood, FL, May.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The design of grids in Web surveys. *Social Science Computer Review*, *31*(3), 322-345.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public opinion quarterly*, *65*(2), 230-253.

Graesser, A. C., Cai, Z., Louwerse, M. M., & Daniel, F. (2006). Question understanding aid (quaid) a web facility that tests question comprehensibility. *Public Opinion Quarterly*, *70*(1), 3-22.

Iglesias, C. P., Birks, Y. F., & Torgerson, D. J. (2001). Improving the measurement of quality of life in older people: the York SF-12. *Qjm*, *94*(12), 695-698.

Jeavons, A. (1999, February). Ethology and the Web. Observing respondent behaviour in Web surveys. In *Proceedings of the ESOMAR Worldwide Internet Conference Net Effects* (Vol. 2).

Redline, C., Tourangeau, R., Couper, M., Conrad, F., & Ye, C. (2009). The effects of grouping response options in factual questions with many options. In *Annual Conference of the Federal Committee on Statistical Methodology. Available at: http://www.fcsm.gov/09papers/Redline_IX-B.pdf*.

Toepoel, V., Das, M., & Van Soest, A. (2009). Design of web questionnaires: The effects of the number of items per screen. *Field Methods*.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*(3), 368-393.

Yan, T. (2005). *Gricean effects in self-administered surveys*. College Park, MD: University of Maryland, unpublished doctoral dissertation.