What Gaze Data Reveal About Coordinating Multiple Mathematical Representations

Theodore Wills (twills@temple.edu)
College of Education, 1301 W. Cecil B. Moore Avenue
Philadelphia, PA 19122 USA

Thomas F. Shipley (thomas.shipley@temple.edu)
Department of Psychology, 1701 North 13th Street
Philadelphia, PA 19122 USA

Briana L. Chang (briana.chang@temple.edu)
College of Education, 1301 W. Cecil B. Moore Avenue
Philadelphia, PA 19122 USA

Jennifer G. Cromley (jcromley@temple.edu)
College of Education, 1301 W. Cecil B. Moore Avenue
Philadelphia, PA 19122 USA

Julie L. Booth (julie.booth@temple.edu)
College of Education, 1301 W. Cecil B. Moore Avenue
Philadelphia, PA 19122 USA

Abstract

40 High school students were given a battery of paper and pencil tests, which collectively assessed a variety of spatial abilities, graph and table competencies, conceptual mastery of calculus, and achievement in common topics from typical precalculus and calculus courses. In addition, students completed a computer-presented measure of Coordinating Multiple Representations (CMR), in which they had to assess whether two mathematical representations (e.g. an equation and a graph) depicted the same underlying mathematical function. Gaze data were captured during this measure, using a Tobii T60 eye tracker. Findings suggest that good or poor performance on several paper measures is associated with distinct and specific gaze behaviors. Better achievement scores are associated with fewer fixations near the centerline of the graph, and with fewer point-plotting and function scanning behaviors. These findings are discussed in terms of differing approaches or strategies for engaging in CMR.

Keywords: Eye tracking; coordinating multiple representations, graphs.

Introduction

Students of mathematics are often required to engage with a variety of representations. They may be asked to understand information presented in the form of equations, graphs, tables, and text. The ability to pass flexibly between two or more such representations, a skill called Coordinating Multiple Representations (CMR), is seen as a hallmark of well developed competence, and is associated with better math achievement (Gagatsis & Shiakalli, 2004). However, while there is more than one way to successfully execute CMR, there is little existing research that delineates the sorts of granular strategies that are associated with superior math competencies.

This study focuses on CMR strategies associated with graphical representations. Researchers in graph comprehension have delineated a distinction between locally and globally represented information, and the different behaviors or strategies that are necessary to successfully extract these types of information (e.g., Shah, Freedman, & Vekiri, 2005.). Leinhardt et al. (1990) draw a distinction between lower-order techniques in graph interpretation, like point reading, and those associated with higher-order information, like evaluating intervals of increase or decrease. More recently, theorists have argued for meaningful, hierarchical distinctions between *reading the data* (e.g., reading the value of a single point), reading between the data (e.g., find differences between points, or compute sums or aggregations of points), and *reading beyond the data* (e.g., infer information that is not explicitly presented; Friel et al., 2001).

Differentiation of graphical coordination skills is supported by other research that outlines the relative challenges of different CMR tasks. For example, it is common practice for students to manually graph data provided in tabular format (Kozma, 2000; Roth & Bowen, 2003), but this type of coordination is more challenging with non-linear functions (Demana, Schoen, & Waits, 1993). Likewise, younger students often focus on lower-order information, such as specific points, rather than understanding the function as a continuous entity with a global structure (Elia, Panaoura, Eracleous, & Gagatsis, 2007; Elia et al., 2008; Even, 1998; Monoyiou & Gagatsis,

2008). Importantly, such lower-order approaches may be inadequate when engaging with more complex functions (Carlson et al., 2010).

Therefore, depending on the task and the specific form of the graph, effective information extraction may require preferentially attending to the most relevant local regions, or conversely, efficiently ignoring the least relevant regions. In other cases, coordinating and integrating multiple regions may allow access to more global kinds of information, such as trends or patterns, which may be a more efficient path to resolving the task. In this paper we examine gaze behaviors of high school students who are coordinating a graph with either an equation, or a table.

Given the lack of previous research linking specific gaze behaviors with CMR strategies, we used an exploratory approach to the eye tracking data. We examine both lower-order visual behavior corresponding to specific local areas of graphs, as well as higher-order behaviors that require sequential coordination of multiple areas. Finally, we explore how these behaviors are related to measures of math achievement.

Method

Participants
Participants included 40 high-school students from pre-calculus and calculus classes at two high-achieving public suburban schools, one in New Jersey the other in Pennsylvania. Their mean age was 16.6; 45% were male; 77% were White, 18% Asian, 3% Black, and 5% other races. As a proxy for SES, median parental education was a Bachelor's degree.

Procedure
Parental consent and student assent were acquired, after which students were tested individually in a session lasting approximately 70 minutes. Participants completed a series of paper and pencil measures. Participants then completed the computer-presented CMR measure described in more detail below.

Paper and Pencil Measures
To assess math achievement, one of our measures was comprised of 11 released items from AP Calculus exams. A second measures was comprised of 11 released graph and table items from the National Assessment of Educational Progress (NAEP), and was intended to assess basic skills associated with graph and table reading (in the absence of coordination with a second representation. We used the "Understand function representations" subscale of the Pre-Calculus Concept Assessment (PCA) (Carlson et al., 2010), which feature tables and graphs, to assess CMR skill. Finally, a researcher-designed Calculus Conceptual Measure (CCM) was included. This measure was designed to assess conceptual mastery as opposed to procedural knowledge.

CMR Eye Tracking Measure (ET CMR): stimuli and task
12 pairs of mathematical representations (graphs, tables and equations), were presented. Across trials, half of the pairs represented the same underlying function (match) and half did not (mismatch). Each possible pair of representations was presented four times, with left-right position and match-mismatch fully counterbalanced. The participants' task was to decide for each trial whether the two representations (e,g, an equation and a graph) matched or did not, while verbalizing their thoughts. No tools such as paper, pencil or calculator were provided, and all work and answers had to be provided verbally to assure continuous tracking of gaze. Four of the CMR items, which required coordination of a table and an equation, are not analyzed here. A sample item is shown in Figure 1. For the current report, our questions of interest focused on coordinations involving graphs. For this reason, the 4 items that consisted of equation-table coordinations were excluded from our analyses.

Do the equation and graph represent the same function?

$$f(x) = -2x^2 + 25$$
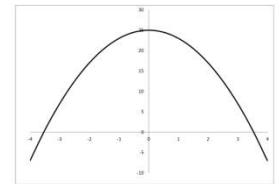
Remember to say what you are thinking out loud, and to say your final answer out loud as well. When you are finished, and ready to proceed, **PRESS ENTER**.

Figure 1: A sample item from the eye tracking CMR task. This example pairs an equation with a graph (other relevant items paired a graph with a table).

Eye Tracking Apparatus
We used a Tobii T-60 remote eye tracking system, which tracks eye gaze with a sampling rate of 60 Hz. Stimuli were presented using a Lenovo T430 laptop running Tobii Studio 3.1. Screen resolution was 1280 x 1024 pixels and participants were seated 65 cm from the screen. With this arrangement the display subtended 32.8° of visual angle, while individual representations subtended between 3.6° and 6.5°.

Analysis
Each participant's gaze data was assessed for problems of calibration that might affect coding. In particular, calibration issues were expected to impact first-order coding, where sensitive placement of a fixation within a

specific Area of Interest (AOI) might lead to error. Ten individuals with substantial calibration error were excluded from the first-order analysis. The specific second-order codes analyzed here proved to be more robust with respect to calibration problems, primarily because they are based on large AOI's, and/or are measures of movement as opposed to locale. For this reason, all participants were retained for analyses of second-order codes (patterns of results did not change when they were excluded).

Coding Gaze Data Each graph was partitioned into a set of AOIs. Some AOIs were associated with areas of the graph, independent of the specific function. For example, every graph featured seven vertical "stripes," centered on the integer X values ranging from -3 to 3. In addition, there were customized AOIs corresponding to potentially meaningful features of graphed functions, such as intercepts, the origin, points of local minima and maxima, and one encompassing the entire contour of the plotted function. In this way, a single point (e.g. the origin), might lie within multiple overlapping AOIs (see Figure 2).
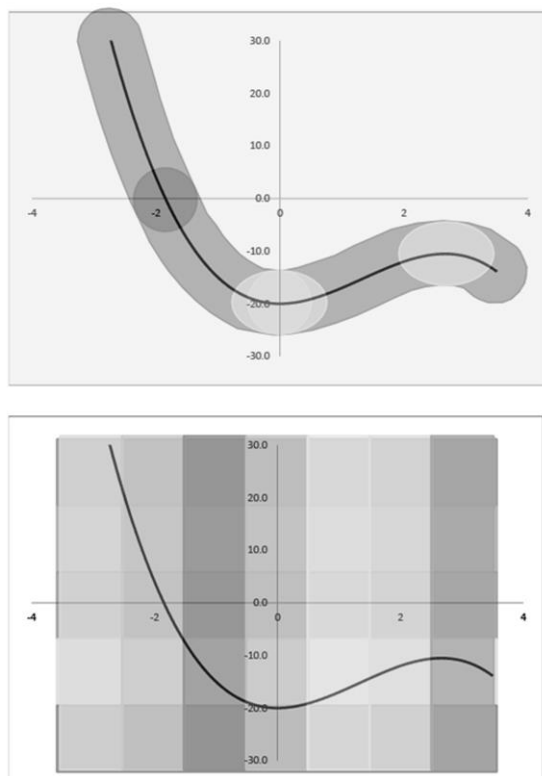

Figure 2: Two images of the same graph showing different sets of AOIs. The top image shows AOIs associated with meaningful features of the graph, including the function, all intercepts, local minima and maxima, and the origin. The bottom image shows vertical and horizontal regions common to all graphs in the stimulus set.

To facilitate flexible analysis of this very large data set, Microsoft Excel was used to create a system of automated coding of fixation data (hereafter termed AC for ease of reference). Tabulation of first-order codes was straightforward, and consisted of simply recognizing gazes that fell within defined AOIs. Higher level coding was more complex. For example, one common strategy associated with graph comprehension is plotting, in which the reader scans either vertically or horizontally, most frequently to associate a value on the X or Y axis with a specific point on the function. We expected this strategy to be expressed in the gaze data by consecutive hits within the graph that were displaced either horizontally (for a Y value) or vertically (for an X value). However, we did not have a priori values for the extent of separation between the two fixations, nor did we have estimates of reasonable tolerances for deviation from strictly horizontal or vertical displacement.

For this reason, the third author coded a sample of the data (>30%), and these codes were compared with those of the AC. Disagreements were examined to determine whether there were systematic discrepancies, and whether these were attributable to inappropriate criteria being applied by the researcher or by the AC. Where appropriate, the researcher adjusted her criteria, and coded additional trials, or the parameters of the AC were adjusted. This process was iterated until the researcher and AC reached high levels of agreement, with at least 90% of eligible fixations receiving the same code (or affirmative absence of a code) from both sources. In this paper we discuss the plotting codes described above (PlotY = vertical, PlotX = horizontal), as well as a Scan code. This latter code reflects the strategy of examining the structure of the drawn function, and is operationalized by consecutive hits on the function AOI, that also satisfy a minimum displacement threshold.

Results

Measures
Correlations among the paper measures are shown in Table 1. It is worth noting that as a whole, the ET CMR measure did not correlate with any of the other measures, perhaps because performance on this untimed test was generally very high. However, in the analyses that follow we do find associations between other measures and specific gaze patterns during the CMR task.

First-Order Codes
We first examined gaze behavior for all items that featured a graph, regardless of whether they were paired with a table or an equation. A consistent pattern emerged in which participants who emphasized the left and right edges of the graph showed better performance on the paper measures than those who emphasized the vertical midline. For example, the total number of fixations on the central vertical stripe was negatively associated with the PCA [$r(30)$= -.37, p<.05], This pattern also applied to total fixations on specific features that are necessarily located in the central vertical AOI. For example, fixations on the origin were negatively associated with scores on the PCA [$r(30)$=-.49,

$p<.01$], and marginal for the AP scores [$r(30)=-.32$, $p=.08$]. Likewise, fixation on the Y-intercept were negatively associated with the PCA [$r(30)=-.46$, $p<.05$].

Table 1
Intercorrelations among All Measures (N = 40)

	1	2	3	4
1. APC				
2. NAEP	-.04			
3. PCA	.412**	.208		
4. CCM	.671**	.111	.447**	
5. ET CMR	.194	.228	.093	.110

*Note. * p < .05, ** p < .01.*

In contrast, total number of fixations in the vertical AOI associated with the X=-3 region of the graph was positively correlated with the subset of CCM items that featured graphs [$r(28)=.38$, $p<.05$]. Similarly, the X=3 region was positively correlated with scores on the NAEP [r(30)=.50, $p<01$] and PCA [$r(30)=.38$, $p<.05$].

A second finding was that participants who did not continue to fixate on the orienting question at the top of the screen after the first item tended to perform better on the paper measures compared with those who did. After the first item, fixations on the question were negatively correlated with scores on the NAEP, PCA, CCM and ET CMR measures (all p's < .05), and marginally with APC.

We were also interested in gaze behavior at the beginning of each trial. We speculated that by restricting analysis to the first few visits of a trial, we might capture relatively automatic or default strategies that might later be obscured during subsequent processes like rechecking. Consecutive fixations within a representation were aggregated into visits. The first visit to a representation is often (though not always) characterized as an orienting event. Further, if it precedes all visits to the alternate presentation it does not afford sufficient opportunity for coordination between representations for meaningful analysis. For this reason we chose as our time point the conclusion of the second visit to the alternate representation (the equation or table), as this is the earliest time point that allows for both orientation and coordination. Since we believed that early behavior may depend strongly on the identity of the alternate representation, we further separated analyses between equation-graph (EG) items, and table-graph (TG) items.

For the EG items, one interesting pattern emerged that was related to the left-right positions of the graph and equation. Students who answered more than one item correct on the AP measure preferentially distributed gaze duration (log transformed) to the left-hand representation compared with those who scored one or zero items, $F(1, 115) = 4.38$, $p<.05$, $\eta^2=.04$. In short, higher AP scores were associated with a stronger left-right bias. These students were more likely to work left-to-right, regardless of which representation occupied the left-hand position.

Second-Order Codes
We selected the PlotX, PlotY and Scan codes for analysis because they were expected to correspond to common graph reading behaviors. As expected, these codes were prevalent, with every participant enacting each code at least once, and with mean counts of 22.7 (SD=11.2), 5.7 (SD=3.5), and 25.8 (SD=13.0), respectively.

Overall vertical visual plotting (PlotY), was associated with lower graph competency, as it was negatively associated with both PCA scores [$r(40) = -.41$, $p<.01$) , and scores on the subset of AP items that featured graphs [$r(40)=-.34$, $p<.05$]. In contrast, overall horizontal visual plotting was more common, but was not differentiated among individuals with different competencies.

The Scan code (consecutive hits along the function) was also associated with poor performance on the PCA, but only when the data was restricted to two items that were presented on early trials of the CMR measure [$r(40)=-.37$, $p<.05$].

Discussion
Our findings indicate that individuals with different levels of competency in math enact different gaze behaviors when asked to coordinate multiple representations that include a graph. Stronger skills, as operationalized by our paper measures, are associated with less attention to the center of the graph, more to the outer edges, and a tendency to engage the left-hand representation first irrespective of its form. In addition, under some circumstances stronger students are less likely to engage in the specific strategies of plotting individual values (as operationalized by the PlotY code), or to systematically scan along the function (Scan code).

Taken together, these findings may reflect a weaker student who is visually drawn to the graph, and then to features (like the origin and Y-intercept) that are most visually salient. In contrast, the stronger student works left-to-right, regardless of the graph's position, and is more likely to fixate on non-central areas of the graph.

Alternatively, weaker students might be emphasizing the center of the graph for reasons that are not strictly perceptual. For example, a more accomplished practitioner of CMR may be more willing to engage with areas of the graph that are not close to zero. First, these values are more computationally difficult for "plug and chug" calculations. Further, examining these regions of the

graph might afford assessment of the overall structure of the function in a way that is more efficient than scanning any two points. For instance, one could rapidly evaluate whether the tails of a cubic function match the expected directions with fixations to the left and right areas of the graph (even if the fixations did not fall precisely on the function).

Another possible interpretation is that while all students enact the Scan and Plot codes, stronger students do so more efficiently and reliably. In contrast, the weaker students are more likely to repeat these steps, either because they realize they have not executed them correctly, or simply because they are not confident that they have. This would be consistent with findings that greater expertise is associated with more efficient gaze behaviors (e.g., Gegenfurtner, Siewiorek, Lehtinen, & Saljo, 2013)

There is also a question about the extent to which the participants are intentionally and knowingly enacting specific strategies, and the extent to which they are prone to enacting (or avoiding) strategies as a function of habitual practice. For example, our stimuli were constructed in such a way that the Y-intercepts in both representations always matched. This design choice was made because we felt that evaluation of the Y-intercept is trivially easy for students at this level of mastery, and that allowing a mismatch determination based on this comparison alone would lead to ceiling effects. This design feature does raise the possibility, however, that an observant participant might conclude during the course of the test that the Y-intercept is not useful in our task, and might adjust their approach in response. We do not have the power to definitively test this possibility. However, examining the strengths of the relevant associations as a function of trial does not suggest that this sort of within-task strategic shift accounts for the overall results.

Whether it is possible to differentiate among these possible interpretations is an open question. It may be that there are no gaze behaviors that would serve to disambiguate different strategies. Perhaps it is the case that either the strategies (or the gaze patterns that enact them) are too dependent on the detailed contextual factors of individual graphs to be detected given the statistical power in the present study. For example, gazes of individual students may be sensitively dependent on features of the stimuli like left-right position, or order of the function, or the specific pairing of representations.

On the other hand, we may be able to develop stronger evidence for some interpretations through any of several future activities. First, we plan to enhance the gaze data with concurrent analysis of the think aloud protocols which were collected simultaneously with the eye tracking task. This approach has the promise of revealing which behaviors are being intentionally and consciously enacted, and which may reflect more implicit processes.

Second, additional second-order codes may afford clarification. For example, we plan to develop visual sequences to operationalize the coordination of the Y-intercept across both representations. This code may reveal different competencies associated with each representation type, as well as competencies in coordinating a specific value across different representation pairs. This will also help to determine whether participants are making conscious adjustments to the relatively low informational value that the Y-intercept has for our specific stimulus set.

Finally, we plan a follow up study with data collection at two time points from each participant, to measure changes over time. Importantly, this future data will expand our available data. Additionally, it will provide developmental data which may inform the detailed relationships among acquired math competencies and the emergence of different gaze behaviors.

References

Carlson, M., Oehrtman, M., & Engelke, N. (2010). The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction, 28*(2), 113-145. doi:10.1080/07370001003676587

Demana, F., Schoen, H. L., & Waits, B. (1993). Graphing in the K-12 curriculum: The impact of the graphing calculators. In T. A. Romberg, T. P. Carpenter & E. Fennema (Eds.), *Integrating research on the graphical representation of functions* (pp. 11-40). Hillsdale, NJ: Erlbaum.

Elia, I., Panaoura, A., Eracleous, A., & Gagatsis, A. (2007). Relations between secondary pupils' conceptions about functions and problem solving in different representations. *International Journal of Science and Mathematics Education, 5*(3), 533-556.

Elia, I., Panaoura, A., Gagatsis, A., Gravvani, K., & Spyrou, P. (2008). Exploring different aspects of the understanding of function: Toward a four-facet model. *Canadian Journal of Science, Mathematics & Technology Education, 8*(1), 49-69.

Even, R. (1998). Factors involved in linking representations of functions. *The Journal of Mathematical Behavior, 17*(1), 105-121. doi:10.1016/S0732-3123(99)80063-7

Friel, S. N., Curcio, F. R., Bright, G. W. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. Journal for Research in Mathematics Education, 32, 124-148.

Gagatsis, A., & Shiakalli, M. (2004). Ability to translate from one representation of the concept of function to another and mathematical problem solving. *Educational Psychology, 24*(5), 645-657. doi:10.1080/0144341042000262953

Gegenfurtner, A., Siewiorek, A., Lehtinen, E., & Säljö, R. (2013). Assessing the quality of expertise differences in the comprehension of medical visualizations. Vocations and Learning, 1-18.

Kozma, R. (2000). The use of multiple representations and the social construction of understanding in chemistry. In M. J. Jacobson & R. Kozma (Eds.), *Innovations in science and mathematics education: Advanced designs, for technologies of learning.* (pp. 11-46). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Leinhardt, G., Zaslavsky, O.,& Stein, M. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60*, 1-64.

Monoyiou, A., & Gagatsis, A. (2008). A coordination of different representations in function problem solving. *Proceedings of the 11th International Congress of Mathematics Education*. Monterrey, Mexico.

Roth, W., & Bowen, G. M. (2003). When are graphs worth ten thousand words? An expert-expert study. *Cognition and Instruction, 21*(4), 429-473. doi:10.1207/s1532690xci2104_3

Shah, P., Freeman, E. G., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In A, Miyake, & P. Shah (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp 426-476). New York, NY: Cambridge University Press.