Response Times and Misconception-like Responses to Science Questions

Andrew F. Heckler (heckler.6@osu.edu)
Department of Physics, Ohio State University
Columbus, OH 43210 USA

Thomas M. Scaife (scaife.1@osu.edu)
Department of Physics, Ohio State University
Columbus, OH 43210 USA

Eleanor C. Sayre (esayre@gmail.com)
Department of Physics, Wabash College
Crawfordsville, IN USA 47933

Abstract

Patterns of incorrect answering or "misconception-like" responses to scientific concept questions have been well documented. Here we investigate both response choices and response times to gain insight into the nature of misconception-like responses. In a series of experiments involving questions on graphs in which participants must compare the slopes of two points, we find that students answering with misconception-like responses, namely comparing heights rather than the slopes, do so consistently and more rapidly than those answering correctly. We also find in a speeded experiment, that all students are able to compare slopes and heights, but comparing heights requires less time than comparing slopes. Finally, by imposing a delay in responding that is long enough for the responder to process both slopes and heights, we find a reduction in misconception-like responses. Thus the misconception-like responses can be explained in terms of speed-accuracy trade-off models in which responders place high priority on answering quickly.

Keywords: Scientific misconceptions, graphs, response time, speed-accuracy tradeoff, physics education.

Introduction

It is well documented in science education that students often respond to scientific concept questions in regular and persistent patterns of errors (Pfundt & Duit, 2000). For example, when presented with qualitative questions about the time of flight of a projectiles with various trajectories, many students will incorrectly answer that both range and height of the trajectory influence the time of flight, when in fact only the height determines the time of flight. For convenience, we will refer to such patterns of incorrect answers as *misconception-like* responses, as we do not know whether they stem from coherent and explicit "misconceptions" of the students or some other mechanism.

While past studies of student difficulties with answering science concept questions have examined the patterns of response choices, in this study we investigated the response *times* as well as the response choices in order to address two main questions. First, are there interesting patterns of response times when comparing correct and misconception-like responses? Second, does response time data shed any light on the processes involved in choosing correct or misconception-like answers?

A number of investigators have examined response times on standardized tests. These studies use both response time and response accuracy in order, for example, to eliminate the effect of guessing and thus improve the accuracy of the tests (e.g., Schnipke & Scrams, 1997; van der Linden, 2008), or to detect cheating (van der Linden & van Krimpen-Stoop, 2003). In this study we investigate questions that evoke misconception-like responses. As we will see in Experiment 1, these incorrect responses are not guesses but rather a coherent pattern of answering.

In addition to studies on standardized tests, a long history of response time studies in a wide range of tasks has revealed the well-known phenomenon of the speed-accuracy tradeoff, namely that there often exists a monotonically increasing relation between response time and response accuracy (Wickelgren, 1977). There are two classes of models used to explain the speed-accuracy tradeoff curve. The first is the *fast-guess* model which proposes that students use a mixture of guesses, which are fast, and non-guesses, which are slow. As mentioned earlier, since there is very little guessing in the responses in this study, we will not consider this class of models.

The second class of models postulates that response choices are a result of *decision criteria* applied to evidence that accumulates over time. As time increases, the amount of information increases, thus increasing accuracy, which explains the speed-accuracy tradeoff curve (e.g., Ratcliff, 1978; Smith & Vickers, 1988).

Let us consider the decision-criteria model with respect to response times on scientific concept questions that often evoke misconception-like responses. If correct answers and misconception-like answers require different solution paths, then it is possible that the response times of the two paths will be different. For example, if the time needed for the process involved in obtaining the misconception-like answer is inherently shorter than the process for obtaining the correct answer, then one would expect the misconception-like response times to be shorter.

In addition to expecting different response times for correct and misconception-like responses, in this model the actual response time also depends on decision criteria. For example, there may be a minimum amount of information needed before a decision can be made. On the other hand, there may also be a maximum amount of time allotted for the decision. Therefore, misconception-like responses may be a result of implicit decision criteria rather than the responder's absolute ability to determine the correct answer.

This study proceeds as follows. In the first experiment, we established that our example science concept question evokes misconception-like responses, and we characterized the difference in the response times of the correct and misconception-like responses. In Experiment 2, we measured and characterized the response times needed to process the main underlying tasks required for obtaining the correct and misconception-like responses. In Experiment 3, we impose a minimum time to respond in order to determine whether this will affect the response choices.

Experiment 1

The first experiment investigates a well-known student difficulty with interpreting graphs commonly used in math and physics courses at the high school and introductory university level. Specifically, when a variable of interest corresponds to the slope of a line at a given point, students instead often attend to the value (i.e. the height) of the point on the line rather than the slope. For example, When students are presented with a position versus time graphs for an object (see Figure 1) and asked "at which point does the object have a higher speed?", many incorrectly answer according to the higher point rather than the greater slope (McDermott, Rosenquist, & van Zee, 1987).

Experiment 1 was designed to achieve three goals. The first goal was to replicate the misconception-like response pattern indicating a tendency to attend to values (heights) of points on a line rather than slopes at those points. The second was to determine whether this pattern was due to students' fundamental inability to compare the slopes of points on a line, or if it was instead a function of familiarity with the question context. Finally the third goal was to compare the response times of students answering correctly vs. those answering incorrectly to determine if there was a pattern in response times corresponding to answer choice.

Experiment 1 used a between-subjects design employing three conditions: math graphs, kinematic graphs, and electric potential graphs (see Figure 1). Each condition presented a series of graphs and participants were asked to compare two points on a curved (or straight) line on the graph. Figure 1 presents examples of the graphs in the three conditions, including the question posed for each graph.

In addition to the fact that the series of graphs in the three conditions were identical (except for the labels on the axes), the questions posed for each graph are also conceptually analogous. In particular, the math graph condition asked for a comparison of the slopes at two points (magnitude of slope = $|dx/dt|$), and the other two conditions also effectively asked for a comparison of slopes since *speed* is the slope for the position-time (kinematic) graph (speed = $|dx/dt|$), and the magnitude of *electric field* is the slope of the electric potential (V)-position (x) graph (magnitude of electric field = $|dV/dx|$).

The three graph conditions were also at differing levels of familiarity for the participants. The math graphs were the most familiar, as they are introduced in standard curricula before and throughout high school. The kinematic graphs were the next most familiar. They are typically introduced in high school physics or physical science courses, and used frequently in the university level physics course that was a prerequisite to the physics course in which the participants were enrolled at the time of the study. Finally, the electric potential graphs were the least familiar, as they are not part of standard pre-university curriculum and most participants saw them for the first time in physics course in which they were enrolled at the time of the study.

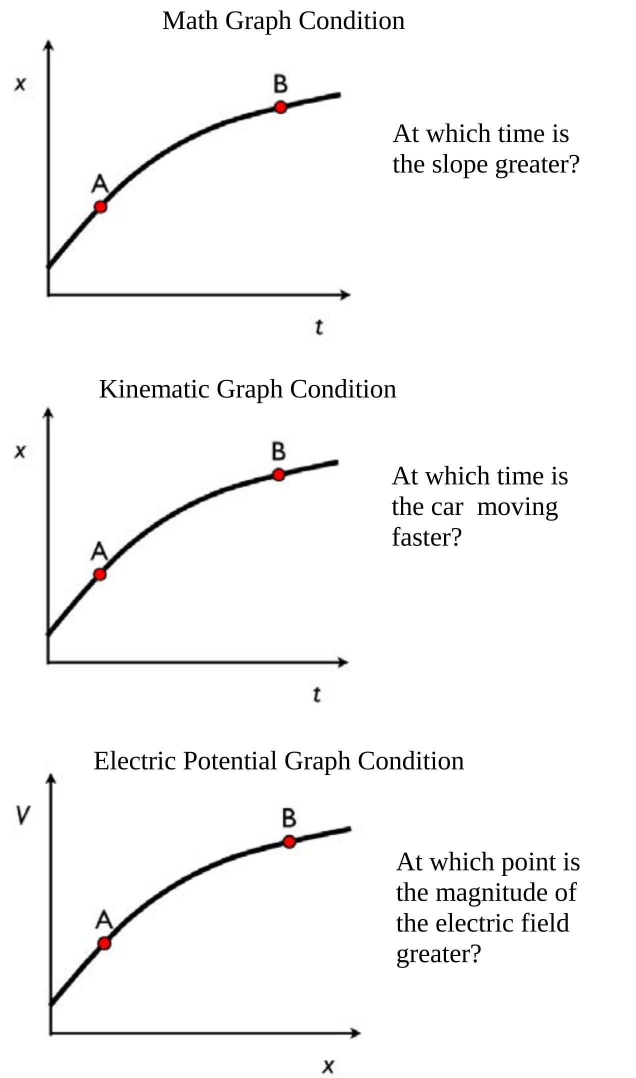

Figure 1. Examples of the graphs and questions used in the three conditions in Experiment 1. The answer choices for all three were: "A", "B", or "the same at A and B".

Method

Participants Participants were enrolled in one of two undergraduate calculus-based introductory physics courses. The first course covered the topic of classical mechanics and the topic of the second course covered electromagnetism. The courses are part of a three-course introductory physics series, and are typically populated with engineering majors. Participants received partial course credit for participation, and the participation rate for both courses was > 95% of all students enrolled in course.

Participants were randomly chosen to be placed in each condition. For the math graphs condition, 28 participants were chosen from the mechanics course and 49 were chosen from the electromagnetism course, for a total of 77 participants. For the kinematics graphs condition, 94 participants were chosen from the mechanics class. For the electric potential graphs condition, 38 students were chosen from the electromagnetism course.

Procedure, materials and design All testing was presented to individual participants on a computer screen in a quiet room. They proceeded through testing at their own pace, and their response choices and response times were electronically recorded.

In each condition students were presented with a series of graphs and asked to compare relevant values at two points on each graph. Participants were given no feedback as to the correctness of their answers. See Figure 1 for examples of graphs and specific questions asked.

Testing consisted of a comparison of two points on 14 graphs (presently serially) with various curve shapes: 8 graphs in which the higher point had a lower slope (these are the difficult "target" questions), 2 graphs in which the higher point had a higher slope, 2 graphs in which both points had the same slope, and 2 graphs in which the two points had the same height but different slopes. The graphs types were placed in a fixed random order, and this sequence was presented to all participants in all conditions. Thus the graphs were mixed such that the correct response was not always "A", and not always the lower or higher point. Our previous pilots studies did not reveal any significant effects of order of graph type on answering. Furthermore, Experiment 3 uses a design to counterbalance for order, with similar results to Experiment 1. Therefore we are confident that the results here are not an artifact of question order.

Results

Analysis of response choices We first report on the performance on the "target" questions, namely those graphs in which the higher point has a lower slope (see Figure 1 for examples). These type of questions are important for investigating graph difficulties, since the correct answer choice (the point with the greater slope, but with a lower height on the graph) is opposite of the common "misconception" that, for example, "the higher point has greater speed".

Figure 2 presents the average scores for the target questions for each condition. The averages depended strongly on the graph type, with scores of 94% for the Math graphs, 72% for the Kinematic graphs an 47% for the Electric Potential graphs (One-way ANOVA with Bonferroni adjusted post-hoc comparisons, ps < 0.0001). Thus the less familiar the graph context, the lower the score.

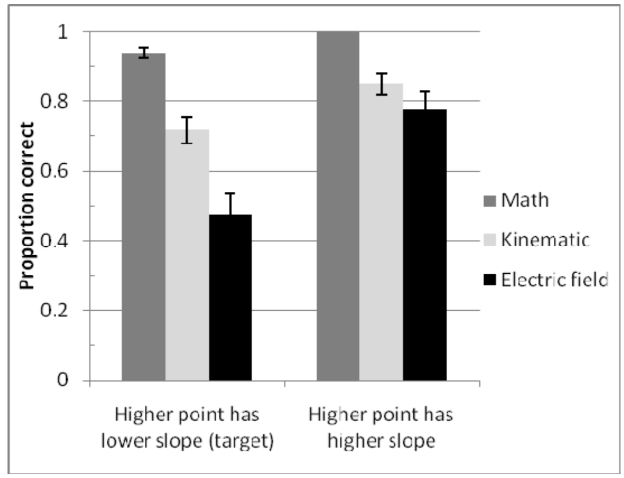

Figure 2. Experiment 1, mean scores for the Math graphs, the Kinematic graphs, and the Electric potential graphs conditions. Scores are shown for target questions in which one of the points has a higher slope but lower value, for "aligned" questions in which one of the points has a higher slope and higher value. Error bars are 1 S.E.M.

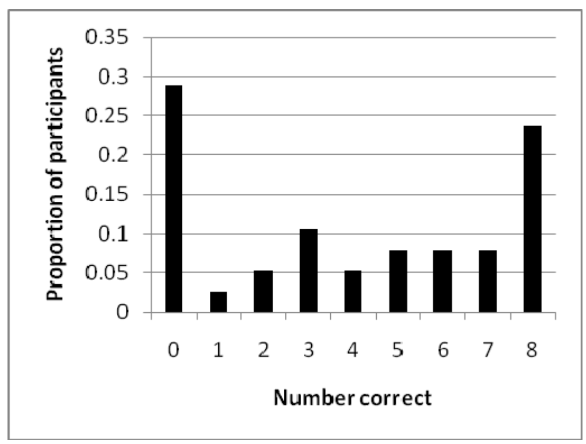

Figure 3. Distribution of scores on target questions for the Electric Potential graph questions. Rather than a random binomial distribution, this distribution is bimodal, indicating that most participants are not guessing.

The patterns of specific answer choices also revealed that answering was not random, and those choosing incorrect answers consistently choose the main misconception-like answer. There are two kinds of evidence to support this. First, Figure 3 presents the score distributions for students answering the electric potential graph questions. If the answering were random, one would expect a binomial distribution of scores. Instead Figure 3 shows a strong bimodal distribution, with most students either answering all questions correctly or answering all questions incorrectly. Note that over 95% of the incorrect answers were the main misconception-like distracter, namely the point with the higher value; few incorrect answerers chose that the points had the same electric field.

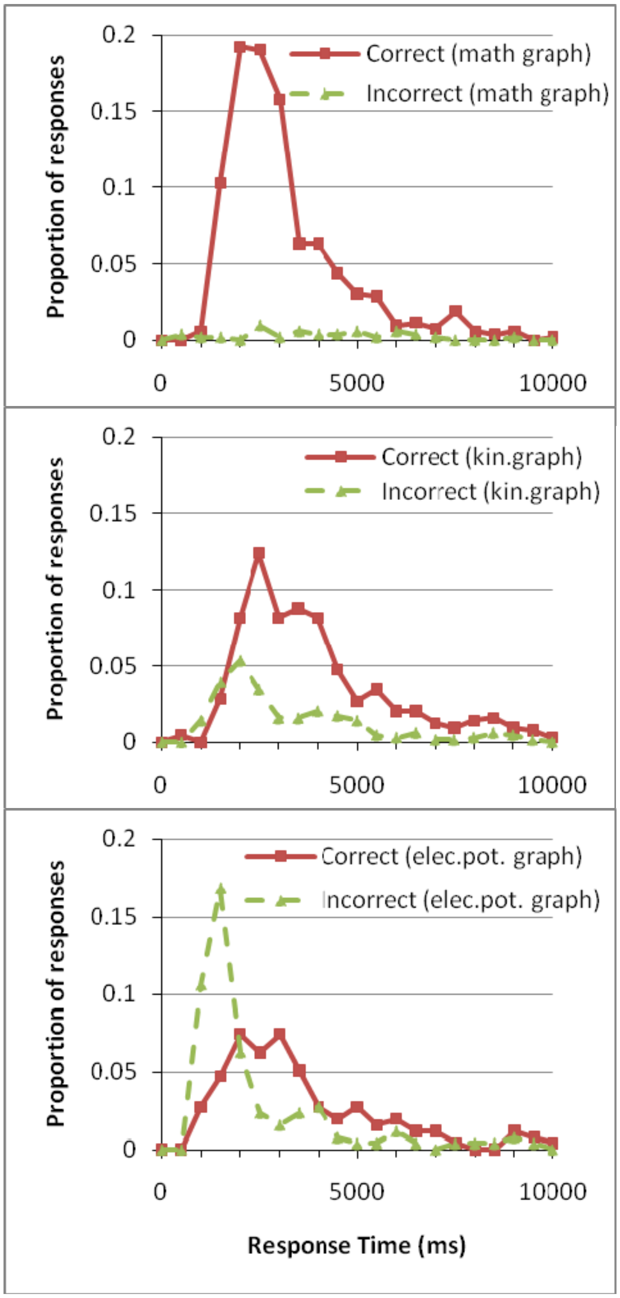

Figure 4: Experiment 1 distribution of response times on target questions in the math (top), kinematic (middle), and electric potential (bottom) graphs conditions. The area under the curves represents the total proportion of correct or incorrect (misconception-like) responses for each condition.

The second kind of evidence to support the fact that students answering incorrectly consistently chose the misconception-like distracter comes from the non-target type of questions. For example, Figure 2 shows the score for the questions for which the point with the highest values was also the point with the highest slope. In all three conditions, participants scored higher on these "aligned" questions in which the slope and values were both greater compared to the target questions in which the point with the higher slope had the lower value (paired t–tests, $ps < 0.003$).

Analysis of response times Figure 4 presents the distribution of response times for each question in each condition, separated out by all questions answered correctly and those answered incorrectly. Note that the response times for all students were pooled together, so this graph represents both between student and within student data mixed together.

There are two main points about the data presented in Figure 4. First, for the kinematic and electric potential graphs conditions, the response times are shorter for the incorrect answers than the correct answers (Mann-Whitney U test used because of long tails in distribution, $ps < 0.0001$). The peaks of the distribution for the incorrect answers are about 500 ms earlier than for the correct answers. There are so few incorrect responses for the math graphs that no reliable comparisons can be made for that case. Second, the peaks of correct answers for all three conditions are at the same place (about 2000 ms) for all three conditions.

At first glance, the fact that the response times for the incorrect responses are shorter than the correct responses may not be a surprise: the speed-accuracy tradeoff is a well known phenomenon. However, as discussed earlier, the incorrect answers are not random guesses, so one cannot conclude that the shorter response times are due to fast guessing. Rather, there is a pattern to the guessing.

This leads us to the question of whether there is an inherent difference in time required to perform the two different response modes, which in Experiment 1 translate to systematically correct vs. "incorrect" (misconception-like) responses. The underlying task to determine the correct answer is to compare the slopes at the two points and the underlying task to determine the misconception-like answer is to compare the heights of the two points. Therefore, in Experiment 2 we determine the time required to perform these two basic tasks.

Experiment 2

The goal of Experiment 2 is to compare the response times for the tasks of comparing the heights of two points vs. comparing the slopes at two points.

Method

Participants Eighteen undergraduate students participated, receiving partial credit for a calculus-based introductory physics course.

Procedure, materials and design The procedure was similar to Experiment 1. Participants were presented with examples depicting various position time curves for a car (see Figure 1 for an example). For each graph, two points on the curve were marked, indicating the position and time of the car at two different times. In a within-subject design, participants were asked to determine as quickly as they can without making a mistake either which point was higher, or at which point the slope was greater. The test was administered in blocks of 9 questions of the same type (compare height or compare slope). Question type blocks were presented in an alternating sequence, with 2 blocks for each question type, for a total of 4 blocks (36 questions).

Results

The mean score for both the compare height questions and the compare slope questions was >97%. Because the response times in the first two blocks were initially relatively high and decayed to an asymptote within 3-4 questions, and the times were near a steady asymptote in the second two blocks, we only compared the response times in the second two blocks (third and forth block). The response times in the first two blocks showed the same trend. Figure 5 presents the distributions of response times for the height and slope comparison tasks. The mean response time was significantly lower for the comparison of height questions (788 ms) verses the comparison of slope questions (1216 ms), (paired-sample $t(17) = 7.04$, $p < 0.001$, $d = 1.28$).

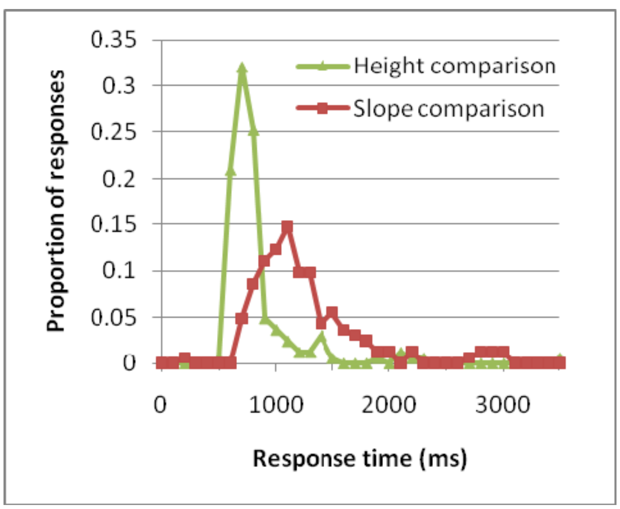

Figure 5: Distribution of response times for the height comparison and slope comparison tasks in Experiment 2.

Figure 5 is similar to the results from the electric potential graphs in Figure 4, with the participants choosing the point with the greater height answering significantly faster than those choosing the point with the greater slope. The main difference is that the peaks in Experiment 2 are earlier and the widths are narrower. One possibility for the difference is that in Experiment 2, the participants were asked to answer as quickly as possible. Therefore the time to peak represents a typical minimum time needed to perform the task.

These results suggest that the difference in response times between the correct answer (comparing slopes) and misconception-like answer (comparing heights) is due to these answers employing different procedures to complete, and these two procedures require different amounts of time.

Experiment 3

The results of Experiments 1 and 2 demonstrate that response times of misconception-like responses are shorter than those of correct responses, and the underlying task necessary for determining the misconception-like response (comparing heights) takes less time than the task necessary for determining the correct response (comparing slopes). Considering the decision-criteria model discussed earlier, one way to help explain misconception-like responses on these questions is to propose that students self-imposed a decision criterion that gave high priority to answering quickly. In this case, then students may have tended to choose the information that was processed first, namely information about the relative heights of the points, and this lead to an incorrect response. The information about the relative slopes would lead to the correct answer but took longer to process, so it was excluded from the decision.

Experiment 3 aims to test this idea by imposing a minimum time delay before responding. That is, participants are shown the question and may answer only after a short delay. If the delay is long enough to allow for the processing of both faster solution (comparing heights) and slower solution (comparing slopes), then they would have both kinds of information available. This could then result in participants with the delay answering more frequently with the response consistent with the slower process compared to participants who had no delay imposed. The delay was set to 3 seconds, since the majority of participants who answered correctly in Experiment 1 did so within this time.

Method

Participants A total of 72 undergraduates enrolled in a calculus-based introductory physics courses in electromagnetism participated, receiving partial course credit for participation. Participants were randomly assigned to one of two conditions: 37 in the delay condition and 35 in the control condition.

Procedure, materials and design The procedure was similar to Experiment 1. Participants in the control condition were presented with the same graphs as in the electric potential graph condition in Experiment 1. Participants in the delay condition were presented with the same graphs. However, before the questions began they were presented a screen with the following message: “On each slide, you will see the question with a message at the bottom of the screen. At first the message will read: ‘Take a moment to carefully consider your answer.’ While this message is displayed, you will not be able to answer the question. After a couple of seconds, the message will change and prompt you for an answer. Please press the key that corresponds to your answer at that time." They were then given a simple math-fractions problem as an example of the delay, then they proceeded to the graph questions.

Therefore the students in the delay condition were required to wait 3 seconds before responding. The only other difference in Experiment 3 was to randomly assign students within each condition into one of two question-order conditions, to counterbalance for any question order effects. Note that there were no significant differences in performance between the control in Experiment 3 compared to the electric potential graphs condition in Experiment 1.

Results

As shown in Figure 6, participants in the Delay condition score significantly higher than those in the control condition (70% vs. 49%, $t(70) = 2.07$, $p = .04$, $d = .5$).

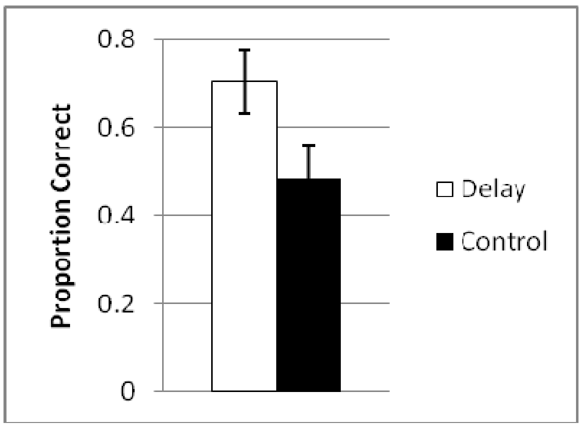

Figure 6: Results of Experiment 3. Error bars are 1 S.E.M.

Discussion and Conclusion

There are three main results of this paper. First, in Experiment 1 we found a clear difference in the pattern of response times for correct vs. misconception-like responses. This cannot be explained by correct vs. guessing responses because the misconception-like responses are not guesses, rather a consistent pattern of answering. For the particular example used in this study, we found that students will often compare heights of points on a graph, even in cases when they are supposed to compare the slopes. The participants answering with the misconception-like response tend to respond more quickly than those answering correctly.

Second we found in Experiment 2 that participants were able to compare heights and slopes with near-perfect accuracy, and it takes longer to compare slopes than heights. This response-time pattern is consistent with Experiment 1.

Third, when a delay for responding is imposed on the participants, they tend to answer correctly more frequently. This suggests that participants are able to arrive at the correct answers for these kinds of questions, but there is another factor influencing their responses.

The basic structure of the decision-criterion model may at least qualitatively provide an explanation for these results. The key feature of the model is that there exists a set of criteria for responding. Let us hypothesize two criteria that can explain the results. The first criterion is the need for information about the comparison of the two points that is *plausibly relevant*. The second criterion is the need for rapid responding. If the information on the comparison of heights is plausible enough, the responder who is free to respond at any time may tend to use *only* the height information since it is obtained quickly, and thus respond consistently and incorrectly. If, on the other hand, a time delay were imposed that was long enough to allow the responder to process both height and slope comparison information, then the response choice will be based on *both* height and slope information (and an additional decision is made on which is more relevant). This could naturally result in an increase in respondents choosing the correct answer.

Therefore, these results suggest that for the graphs questions studied here, an implicit tendency to answer rapidly coupled with the fact that an incorrect answer with sufficient plausibility is arrived at rapidly may be at least partially responsible for the misconception-like answers. The respondents are capable of answering correctly, but instead they tend to answer quickly. This prevents them from processing additional relevant information and considering alternative possibilities that may be more valid.

Acknowledgments

This research is supported by a grant from the Institute of Educational Sciences of the U.S. Department of Education (#R305H050125).

References

McDermott, L. C., Rosenquist, M. L. & van Zee, E. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics*, *55*, 503.

Pfundt, H., Duit, R. (2000). *Bibliogaphy: Students' Alternative Frameworks and Science Education* (5th Ed.). Kiel, Germany: Institute for Education Science.

Ratcliff, R., 1978. A theory of memory retrieval. *Psychol. Rev.* 85, 59–108.

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item-response times with a two-state mixture model. *Journal of Educational Measurement*, 34, 213–232.

Smith, P.L., Vickers, D., 1988. The accumulator model of two-choice discrimination. J. Math. Psychol. 32, 135–168.

van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. Journal of Educational and Behavioral Statistics, 33, 5-20.

van der Linden, W. J., & van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant response patterns in computerized adaptive testing. *Psychometrika*, 68, 251–265.

Wickelgren, W.A., 1977. Speed-accuracy tradeoff and information processing dynamics. *Acta Psycholog.* 41, 67–85.