

TEACHERS' KNOWLEDGE ABOUT INFORMAL LINE OF BEST FIT

STEPHANIE A. CASEY
Eastern Michigan University
scasey1@emich.edu

NICHOLAS H. WASSERMAN
Teachers College, Columbia University
wasserman@tc.columbia.edu

ABSTRACT

The purpose of this study was to investigate teachers' subject matter knowledge relevant to the teaching of informal line of best fit. Task-based interviews were conducted with nineteen pre-service and in-service mathematics teachers. The results include descriptions and categorizations of teachers' conceptions, criteria for placement, accuracy of placement, and interpretation of the informal line of best fit. Implications regarding teacher preparation for the teaching of this topic, including current status and recommendations for future preparation, are discussed.

Keywords: *Statistics education research; Statistical knowledge for teaching; Statistical association*

1. INTRODUCTION

Multiple factors have resulted in statistics' substantially increased role in school curricula worldwide, including the usefulness of statistics in daily life in today's data-driven society, the instrumental part statistics plays in a multitude of disciplines and professions, and the important role of statistics in developing critical reasoning (Batanero & Díaz, 2010). New curricula and/or curriculum standards published in countries such as Australia, Brazil, Costa Rica, South Africa, Spain, the United Arab Emirates, and the United States of America include a sizable amount of statistics, generally presented through a data-oriented, investigative approach (Batanero, Burrill, & Reading, 2011). With the proliferation of statistics into school curricula has come an increased need for teachers to have statistical knowledge for teaching (Groth, 2013). There are concerns regarding whether teachers have this knowledge (Batanero, Burrill, & Reading, 2011; Conference Board of Mathematical Sciences (CBMS), 2012a) and the CBMS (2012a) recently identified statistics as a content area for which teachers' knowledge for teaching greatly needs improvement. Research regarding statistical knowledge for teaching has not kept up with the research that has occurred regarding mathematical knowledge for teaching (Batanero, Burrill, & Reading, 2011; Groth, 2007). Shaughnessy (2007) noted this gap after conducting a review of research related to statistics learning and called for the research community to focus more explicitly on investigating statistical knowledge for teaching.

The purpose of this study is to explore the knowledge teachers have about the informal line of best fit. The informal line of best fit introduces students to the topic of

statistical association, a concept fundamental to the learning of statistics (Burrill & Biehler, 2011; Garfield & Ben-Zvi, 2004). Just as there has been a recent trend to introduce students to inference informally before learning formal procedures (see the *Statistics Education Research Journal* November 2008 special issue on Informal Inference), a predominant trend in the approach of current curricula (e.g., Australia: Australian Curriculum, Assessment, and Reporting Authority, 2012; England: Qualifications and Curriculum Authority, 2007; U.S.A.: National Governors Association Center for Best Practices, Council of Chief State School Officers (CCSSI), 2010) is to introduce students to statistical association informally by having them do linear regression informally – in other words, to fit a line to data by eye, an informal line of best fit. For example, the CCSSI (2010) states students in grade eight in the United States should:

8.SP.A.2. Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line (p. 56).

The Common Core Standards Writing Team (2011) expanded on the purpose of this standard, which included an expectation that eighth grade students be able to determine the informal line of best fit for data that has no association and come to the realization that a horizontal fitted line implies no association between the variables. The learning of informal linear regression lays the foundation for students' future study of formal linear regression (e.g., least squares regression line (LSRL), median-median line) and regression of other function forms, such as exponential and quadratic (CCSSI, 2010-see HSS.ID.B.6).

The recent inclusion of informal inference in school curricula led to the naming of research on teacher knowledge about statistical inference as a research priority (Harradine, Batanero, & Rossman, 2011). Likewise, the recent inclusion of informal regression in school curricula presents the need for research on teacher knowledge about informal regression that has not yet been met. This study aims to meet that need and inform teacher education efforts (both pre-service and in-service) to support teachers as they gain the statistical knowledge for teaching informal line of best fit.

2. BACKGROUND AND RESEARCH QUESTION

This section includes background information separated into three subsections, followed by the research question for the study. The first subsection presents the Statistical Knowledge for Teaching framework and associated work on the statistical knowledge for teaching informal line of best fit. The second subsection reviews the literature about teacher knowledge of linear regression, and the third subsection summarizes research findings regarding how people learn to reason about data that were relevant to the analysis procedures of the current study.

2.1. STATISTICAL KNOWLEDGE FOR TEACHING

Groth (2007, 2013), building on the work of Hill, Ball, and Schilling (2008), developed the Statistical Knowledge for Teaching framework as a practice-oriented theoretical structure to guide research regarding knowledge needed for teaching statistics; it was adopted for the current study for this purpose. It has two major domains: subject matter knowledge and pedagogical content knowledge. The present study attended only to teachers' subject matter knowledge; therefore, only that domain will be explicated

further. Three categories comprise subject matter knowledge: common content knowledge, specialized content knowledge, and horizon knowledge. Common content knowledge refers to knowledge developed in conventional statistics courses and is considered common because it is the knowledge one would expect a common person who knows a statistical concept to have. Specialized content knowledge, however, is specific to the work of teaching statistics. It is the subject matter knowledge that teachers need in order to do their work, like appraise student-generated strategies or utilize representations that help students learn statistics, tasks which are central to the practice of teaching. The third category of subject matter knowledge, horizon knowledge, necessitates that teachers' knowledge of statistics extends beyond the scope of the curriculum a teacher teaches, as such understanding may impact their current instructional approaches and practices. For example, Groth (2013) identified knowledge of standard deviation as important horizon knowledge for sixth grade teachers in the United States because mean absolute deviation, a measure of variability taught in sixth grade, is a precursor to the more sophisticated measure of variability, standard deviation. Taken together, these three categories specify the types of subject matter knowledge needed by teachers of statistics.

Casey and Kaplan (2014) led a study aimed to describe the knowledge needed for teaching informal line of best fit. Taking a practice-based qualitative approach, a teacher was observed as she taught her students about informal line of best fit and interviewed immediately following each observation. Records of practice were assembled to create a compilation document that essentially recreated each of the five observed class sessions along with related materials including textbook pages and student work. Analysis of the compilation documents was done by a team of five analysts, including statisticians and statistics education experts. The analysis focused on the demands on teachers' statistical knowledge involved in the practice of teaching informal line of best fit, as exemplified by the work in which the teacher actually had to engage. Elements of the knowledge description pertinent to the present study are described below.

Teachers need to be able to read scatter plots to assess whether a linear association exists between the plotted variables. They also should know the detection of an association between the variables does not imply causation. Regarding the purpose of the informal line of best fit, teachers should have a deep understanding of the informal line of best fit as an aggregate model for data as well as knowledge about different types of linear models (e.g. median-median line, least-squares regression line). They also need to know the assumptions associated with linear models and how to assess them. Teachers should be able to estimate the location of the linear trend with precision in order to assess students' placement of the informal line of best fit. They also should have knowledge of appropriate criteria for placing the informal line of best fit (as well as potential limitations of other criteria), understand the interplay between data and criteria for linear models, and be able to differentiate between algebraic linear functions and statistical linear models.

2.2. TEACHER KNOWLEDGE OF LINEAR REGRESSION

Research regarding teacher understanding of regression is scarce (Engel & Sedlmeier, 2011). Three studies that are relevant to the present study are summarized. The study most similar to the present one was carried out by Sorto, White, and Lesser (2011) with eighteen sophomore-level students in a university introductory statistics course. The purpose of the study was to determine the methods used by students for drawing an informal line of best fit and whether these methods agree with the method of least squares. Some of the participants in the study were middle school pre-service teachers. Their data, however, were not analyzed separately from the rest of the participants so one

cannot draw conclusions regarding the teachers' results alone. In their study, Sorto, White, and Lesser gave a written assessment consisting of three tasks centered around the analysis of a bivariate data set about job satisfaction scores and salary. The data set of eight points was presented in tabular and graphical (scatter plot) form. For the first two tasks, the participants were asked to draw a line on the scatter plot that best fits the data and then describe the criteria they used to do so. These assessment tasks informed the design of the tasks utilized in the present study. These types of tasks were effective at eliciting participants' thoughts regarding the informal line of best fit and so in this study the participants were given five similar tasks. The third assessment task Sorto, White, and Lesser used inspired another task in our study. It presented two scatter plots of the same data set with different lines of best fit shown, and then asked the participant which of the lines was better, how he/she made that decision, and if there was anything they could compute to support the decision. The participants' responses to these three tasks were categorized to reveal four criteria for the informal line of best fit: equal number of points on either side of the line; at the middle or average of all the points; fit the trend of the data; consider distance of the points to the line. This is the only known study to address knowledge of informal line of best fit.

The other two studies were interventions that aimed to improve pre-service teachers' understanding of association, including formal linear regression. Batanero, Estepa, and Godino (1997) researched the effects of a computer-based course on pre-service primary teachers' understanding of statistical association. They determined two strategies the teachers could master to judge correctly the association between bivariate quantitative variables: while viewing the data set as a whole, realizing independence means the distribution of one of the variables does not change when conditioning by values of the other; and using the direction (increasing/decreasing) of the scatter plot to justify the description of association. They also identified key elements of a good understanding of association that they used as learning goals for the study. Those relevant to this study include the meaning of independence and distinguishing association from causation. Most of the teachers were able to meet the study's learning goals by the end of the course with the exception of the difference between association and causation: teachers showed no improvement on this learning goal. Distinguishing association from causation is a common difficulty for learners of this topic (Engel & Sedlmeier, 2011). The intervention study of Engel, Sedlmeier, and Wörn (2008) centered on a data-based modeling course for pre-service secondary teachers and had as its primary learning objective mastery of the signal and noise conception (Konold & Pollatsek, 2002) for models.

This review highlights the limited attention to research specific to teachers' knowledge about informal line of best fit. The methodology of this study, task-based interviews with teachers, provides an additional and necessary perspective on teacher knowledge about informal line of best fit. As previously mentioned, the Sorto, White, and Lesser (2011) study informed the methodology of the present study. The focus on the signal and noise conception (Konold & Pollatsek, 2002) by Engel, Sedlmeier, and Wörn (2008) as well as the findings of all three studies provide a backdrop for interpreting the present study's results.

2.3. LEARNING TO REASON ABOUT DATA

The work of researchers over the past fifty years has enabled us to know a considerable amount regarding how people learn to reason about data. The research findings relevant to our data analysis and interpretation processes are described in this subsection.

Learners new to the study of data often view data as a series of individual cases and focus their attention on specific cases in the data set, known as a case-oriented view. In contrast, experts have an aggregate view where they can see the entire data set holistically with characteristics that are invisible in any of the individual cases (Bakker, 2004). For example, when experts with an aggregate view look at a dot plot of the heights of students in a classroom, they analyze its general shape, the variability of the data, and where the data are centered to inform their analysis of the heights. In contrast, a person with a case-oriented view would focus on specific data points of interest, such as shortest or tallest person in the class. Estepa and Batanero (1996) documented the prevalence of the case-oriented view by pre-university students when analyzing scatter plots to evaluate correlations between quantitative variables. Specifically, some students used only a subset of the data when performing their analysis. Casey (2015) found that eighth grade students approach the task of finding the informal line of best fit through case-oriented approaches, such as looking at selected few data points rather than the entire data set or ignoring points they deemed to be outliers.

A robust finding that has emerged from multiple studies is that persons' prior beliefs about the relationship between two variables, based on their knowledge of the context, has a great deal of influence on their judgments of the covariation between those variables (e.g., Batanero et al., 1996; Jennings, Amabile, & Ross, 1982; Kuhn, Amsel, & O'Loughlin, 1988). If the data provided do not agree with the expectation one has about the relationship between the variables, then the data are often ignored. Persons also tend to exhibit a confirmation bias, viewing data values that confirm their expectations as more relevant than disconfirming values (McGahan, McDougal, Williamson, & Pryor, 2000) and also infer that association means a causal relationship exists between the variables (Batanero, Estepa, & Godino, 1997).

2.4. RESEARCH QUESTION

The research question for this study was: What subject matter knowledge for teaching informal line of best fit do teachers have? In particular, the study focused on exploring teachers' conceptions, placement, and interpretation of the informal line of best fit, situated in contexts relevant for teaching. Ultimately, the study explores teachers' existing statistical knowledge to better inform teacher education.

3. METHODOLOGY

3.1. PARTICIPANTS

Nineteen mathematics teachers, eleven female and eight male, participated in the study during either the Fall 2012 or Fall 2013 semester. They were recruited from enrollees in teacher education courses at three large universities in the United States. The courses from which the participants were recruited were specifically designed for teachers, with two of them focusing exclusively on probability and statistics. Although participation in the study was intentionally prior to any instruction about lines of best fit in their current courses, all participants were familiar enough with the idea of fitting lines to data from their previous coursework (fifteen of the participants had previously taken a college-level statistics course) to understand directions given to them in this study. Three of the teachers were undergraduates; the rest were working towards master's level degrees. Eleven of the teachers were pre-service secondary teachers. The remaining eight teachers were in-service: two at the elementary level, five at the secondary level, and one

at the two-year college level. Four of the teachers previously taught informal line of best fit in a secondary school setting.

3.2. METHOD

A task-based interview was administered individually to each participating teacher, led by a mathematics teacher educator at each of the universities. This method has been shown to be effective at eliciting conceptions regarding statistical concepts (e.g., Konold et al., 2002; Mokros & Russell, 1995). The interviews were videorecorded and followed a semi-structured format: all of the tasks and a majority of the questions were predetermined and established in an interview protocol (see Appendix A for a copy of the protocol), but follow-up questions created at the time of the interview by the interviewer were also allowed. Each interview took approximately 20 minutes.

The protocol for the interview consisted of a series of tasks and questions designed to prompt teachers' thinking related to aspects of line of best fit and its informal placement that are relevant for teaching the topic. The following is an excerpt from the protocol that describes how the first five tasks were to be done:

Each one involves a scatter plot graph of data. We will talk about the data on each graph and then I will ask you to determine the line of best fit for the data points. I would like you to think out loud as you determine where the line is on each graph so that I can understand how you are deciding where to place it.

I would like you to use this piece of wire [hand out piece of music wire] as the line and place it where it best fits the data. There are no right or wrong answers. I am interested in how you think about placing the piece of wire, so please tell me what you are thinking as you do this so I can follow your thoughts. Once you decide where you want the line on the scatter plot, we are going to use tape to keep the wire in place.

The participating teachers followed this procedure for a series of five scatter plots for contextual data sets with eight ordered pairs each. In general, people have a better understanding of covariation when it is presented in a meaningful context (Moritz, 2004), and therefore, all of the interview tasks referred to data sets in familiar contexts. Given the nature of the task instructions and materials (i.e., place a straight wire to represent the informal line of best fit), the datasets used in the tasks were selected so that a linear model would be appropriate. The residual plots created when a linear regression line is found for each of the tasks do not display any structure or curvature indicative of a line being an inappropriate model. In addition, the data presented were reasonable regarding both their numerical values and their variability given the contexts. Purposefully, none of the tasks had outliers or influential points so that the participants would not be distracted by the issue of dealing with such points in completing the tasks. Characteristics that changed across the tasks include the direction and strength of association. Table 1 summarizes aspects of the tasks.

A task from Coxford et al. (2003) as well as personal experience of the authors in teaching the topic provided insights regarding criteria people have for the line of best fit. These include going through the maximum number of points possible, having the same number of points above and below the line, connecting the first and last points, and forcing the line to go through the origin. The tasks were designed with these criteria in mind. For example, if a person applies the criterion of going through the maximum

number of points possible when placing the line on Task Three, then he/she will be confronted with the issue that it creates a poor best fit line. Also, Tasks One and Two present data sets where requiring the line to pass through the origin could be a relevant consideration when placing the line.

Table 1. Summary of interview tasks

Task number	Context of task	Direction of association	Strength of association
1	Golf ball bounce height versus drop height	Positive	Strong ($r = 0.96$)
2		Positive	Strong ($r = 0.97$)
3	Movie attendance versus average ticket costs	Negative	Moderate ($r = -0.76$)
4		Negative	Strong ($r = -0.92$)
5	Height versus shoe size for elementary students	None	None ($r = 0.05$)

After completing these tasks, the teacher was asked three questions relevant to his/her content knowledge for teaching students the line of best fit using an informal approach. The first question, “Could you tell me what you would say to a student that asked you ‘What is the line of best fit?’”, ascertained the teacher’s knowledge of its meaning set in a context relevant to teaching. This was followed by a similar query concerning criteria for the line of best fit: “Could you tell me what you would say to a student to help them draw the line of best fit on a scatter plot?” The third question asked “As you determined the line of best fit for each of the five scatter plots, did your thoughts about where the line of best fit gets placed change? If so, how?” to assess whether the process of completing the five tasks in the interview had been a learning experience for the teacher participant.

Another task, an adaptation of one used by Sorto, White, and Lesser (2011), followed. The setting for this task involves two fictional students, Angelo and Barbara, who were given the exact same task the participating teachers completed in Task One. Their solutions were different and are presented in this task as side-by-side scatter plots with their lines drawn in and equations of the lines given below the plots. The task asks the teacher which student’s line fits the data better and why. Angelo and Barbara’s solutions were designed to be similar in their placement and do not pass through the origin; Angelo’s line, however, goes through two of the points while Barbara’s minimizes the sum of the squared residuals (it is the least-squares regression line) and does not go through any points. One purpose of this task was to determine if a teacher has the knowledge to be able to estimate reasonably the location of the linear trend, an important component of the knowledge needed by teachers for the work of teaching informal line of best fit (Casey & Kaplan, 2014). It also forced the teacher to assess which criterion – goes through points or closest to all of the points – was more important for a line of best fit.

The remainder of the interview centered on questions asking the teachers to interpret the lines of best fit they made on previous tasks. They included requests for the teacher to describe what his/her line of best fit shows about the relationship between those variables

in context, decide if that relationship is causal, and calculate and interpret in context the slope of one of their lines of best fit. These interpretation questions address whether the teacher has the knowledge to teach students how to interpret appropriately lines of best fit, called for by current curriculum standards (e.g., CCSSI, 2010 – see 8.SP.A.1, 8.SP.A.3, HSS-ID.C.7, and HSS-ID.C.9).

3.3. ANALYSIS

Analysis of the data began with the creation of written transcripts for the interviews. Next we read the transcript, viewed the videotaped interview, and studied the placement of lines on the first five tasks for each teacher. This led to an iterative process of analysis by the authors: moving back and forth between viewing of the data sources and identifying, discussing, and classifying the teacher responses. Major points of interest, relevant to answering our research questions or emerging from initial examination of the data, were the focus of the first round of analysis. These major points included conception of the line of best fit, criteria used to place the line of best fit, response to Task Six (Angelo or Barbara) and justification, responses to interpretation questions at the end of the interview, use of case or aggregate reasoning (Bakker, 2004), and comparison of the lines placed for tasks one through five amongst the teachers and with the least-squares regression line. Data relevant to each of these points of interest were assembled and organized, along with corresponding transcript lines if appropriate, for each teacher. Initial description for each teacher's conception of the line of best fit was predominately provided by his or her response to the question, "Could you tell me what you would say to a student that asked you 'What is the line of best fit?'" but was also informed by relevant statements made by the teacher throughout the entire interview. Similarly, the teacher's criteria for fitting the line emerged from his/her work to complete all of the tasks as well as the explanation given by the teacher of how he/she would help a student draw the line of best fit. The interview protocol allowed us to pull data from multiple points in the interview that were relevant to a major point or research question.

Following this first round of analysis, we discussed the findings thus far, including themes and ideas for further analysis. For example, it was noted that the teachers varied much more in their response to Task 5 (in both their placement of the line and criteria for doing so) than the first four tasks, which led to a separate analysis of Task 5 in the next round of analysis. Other points also attended to in the next round of analysis were calculation and interpretation of the slope of the line of best fit (Casey & Kaplan, 2014), use of specific methods when fitting the line of best fit (to a subset, disregarding points teachers considered to be outliers, through the origin, context knowledge) (Bakker, 2004; Casey, 2015), and the role of context in informing responses (Casey & Kaplan, 2014). To synthesize the data relevant to teachers' conceptions and criteria for the line of best fit, the constant comparative method (Glaser & Strauss, 1967) was used collaboratively to categorize the responses. A tag was assigned to each teacher's dominant conception as well as dominant criteria (Task 5 was tagged uniquely), identifying the categorical designation of that conception or criteria. The categories emerged from the data; they were not pre-determined. Any cases of disagreement regarding the categorization given to a teacher's response were discussed until agreement was reached on the use of the coding framework. This synthesis revealed trends regarding teachers' conceptions and criteria.

Analysis concluded with a synthesis of the data relevant to our major research points as well as examination of relationships amongst the data variables collected. These included demographic information (gender, level of degree sought, pre-service or in-service, grade level certification, taken college statistics, taught line of best fit) as well as

categorizations previously established in the analysis. In particular, the results were studied to determine potential associations between all relevant variables; for example, whether the slope of the line placed on Task 5 was associated with teachers' statements regarding the context of the variables when completing the task, or with teachers' educational status (pre-service or in-service). Fisher's exact test (two-sided) was utilized to determine whether noted associations between two categorical variables were significant. The test is particularly suited for this study's small sample size, as it is based on exact p -values, not ones estimated from approximate distribution curves that require larger sample sizes and expected values to use methods like the chi-squared test. Additionally, the test is relatively conservative in that its actual rejection rate is frequently slightly below the nominal significance level; those variables determined to be associated by Fisher's exact test often have a stronger association than that required of other tests to be deemed statistically significant. In particular, given the small sample size of this study, we elected to err on the side of understating significance rather than overstating it. Thus, we performed Fisher's exact test (two-sided) using Preacher and Briggs' (2001) software to determine if associations were significant.

4. RESULTS

First, we describe the different conceptions and criteria that came out of our qualitative analysis, since these categories are then used and referenced further in association with other data. Next, we present the participants' overall responses to the tasks, highlighting particularly important aspects of their line placements, as well as detailing additional findings from subsequent analyses on Task 5. Lastly, we discuss some findings pertaining to their interpretations of various features of the line of best fit.

4.1. CONCEPTIONS AND CRITERIA

Analyzing the teachers' responses throughout the entire recorded interview – including their discussion of each of the specific tasks and in responses to interview questions – allowed us to capture the predominant personal conceptions these teachers held regarding the informal line of best fit. The following categories are our response to addressing the question: what meaning about the line of best fit best articulates the way the teacher predominantly understood the line of best fit? Table 2 describes each of the conceptions, an illustrative excerpt from the interviews with the teacher participant noted by identification number in brackets, and the number of participants that predominantly defined the informal line of best fit in this way.

Similarly, analyzing the teachers' responses yielded the following criteria for placing the informal line of best fit. These criteria categories were our response to addressing the question: what specific criteria were used as the participant decided precisely where to place the informal line of best fit? We coded each of the interview transcripts according to evidence of these criteria in their responses and in their placement of a line of best fit. Due to the variety of tasks, however, a large number of different criteria were used by the participants throughout the different tasks. For example, at times teachers used the following criteria: “through as many points as possible,” or “if there's no visible association, the line is horizontal,” or “find a linear subset of the data.” However, such instances were the exception, not the rule. (In fact, much of the variety existed in response to Task 5.) So we identified a dominant criterion for each participant, which was the criterion most frequently used and evident throughout all the tasks as well as responses to the interview questions (Table 3). For the ‘closest’ and ‘sum deviation’

Table 2. Conception categories

Conception	Description (The line of best fit is...)	Example excerpt	Number of participants
Model	The model to show the general relationship between the variables in the population.	“it is just a line that, um, as closely approximates the actual true relationship as possible.” [T7]	5
Signal	The meaningful signal for data that eliminates the inherent noise.	“the line of best fit is...something that’s uncomplicated...something simple and easy to visualize, [trying] to find a relationship...” [T5]	2
Typical	The bivariate equivalent for determining middle or typical values; sometimes named as average, other times as median.	“The line that averages all the – the average of the points.” [T6]	3
Representer	The best representation of all the sample data, where the line accounts for the data at hand rather than a more general relationship.	“...a line that represents the set as a whole as opposed to any particular data point.” [T1]	7
Predictor	The line that enables predictions for values not in the data set.	“so we can predict where it might be in the future... gives you an idea where the answer is going.” [T17]	2

Table 3. Categories of criteria

Criterion	Description (To place a line of best fit, look for...)	Example excerpt	Number of participants
Equal Number	Equal number of points above and below the line (while most of these instances had exactly equal numbers, at times, they were not exactly but roughly equal numbers above and below, as the teacher adjusted accordingly)	“...looking for half above, half below...” [T11]	5
Pairs	Each point above the line has a symmetric pair below the line so that the distance to the line is the same for each member of the pair; pairing off individual data points to determine line of best fit	“...I inspect them in pairs I suppose...mirrored pairs ... this pair [has a] similar distance from the line, ...[then] these two ... [then], these two...” [T2]	2
Closest	The line that is closest to all points (implies total distance from points to line is minimized)	“...just eyeballing the distance from the line to the points... finding where I am closest to the most data points” [T3]	6
Sum deviation	Deviations for points above the line to sum to the same as the sum of deviations for points below the line (they did not necessarily make measurements, but referenced this as they placed their line)	“...there needs to be the same distance total away from the line with the points that fell below the line as there is with the points that fall above the line...” [T14]	6

categories, eight of the teachers indicated that distance and deviation were measured vertically while the responses of the remaining four teachers were imprecise.

4.2. PLACEMENT OF INFORMAL LINE OF BEST FIT

While the teachers in this study at times drew on differing criteria and held different conceptions about the line of best fit, the placements of their lines on the tasks were notably similar overall and close to the least-squares regression line – with one major exception discussed in Section 4.3.

As a representative example, Figure 1 shows all 19 teachers' lines for Task 2 along with the least-squares regression line, providing a visual for how similar the placement of the lines was across all the participants as well as how accurate the lines were in their proximity to the least-squares regression line. Indeed, while there are minor variations, the placements of their lines are notably more similar than they are different.

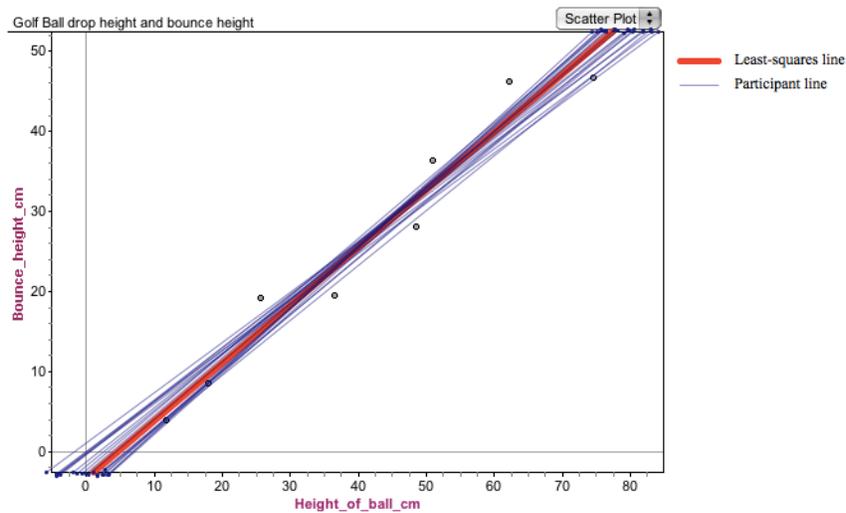


Figure 1. Placement of lines on Task 2 for all participants ($n=19$).

Likewise, on Task Six, nearly all teachers (17 out of 19) correctly selected Barbara's line (the least-squares regression line) as a better line of best fit than Angelo's, which forced the line to go through specific points in a way that was unrepresentative of the more standard approach of minimizing the sum of the squared vertical deviations. However, this brought up one additional point of analysis: while most of the teachers demonstrated an aggregate view of the data, a few were much more prone to relying on specific cases for determining the line; we classified these teachers as having a "case-to-aggregate" view of data. Our description is representative of an emerging aggregate view, where specific cases still played a notable role regarding what the teachers attended to. This often meant that teachers seemed particularly attentive to specific points on the graph or points in the data set, such as: requiring the line go through (0,0) (e.g., Figure 2a: T18's response to Task 1 where she forces the line to go through the origin); fitting the line to a "linear subset" of the data points, two or more points that closely follow a linear pattern (e.g., Figure 2b: on Task 4, T1 fit the line to the first, second, third, seventh, and eighth points only); or removing points deemed to be outliers from consideration (e.g.,

Figure 2c: on Task Three, T12 ignored the eighth point, stating “...I feel like this one [point farthest to the right] is kind of abnormal. It’s like an outlier so I tried to kind of ignore that one.” [T12]).

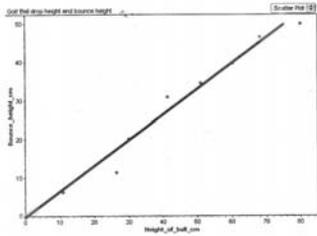


Figure 2a. Attending to origin on Task 1.

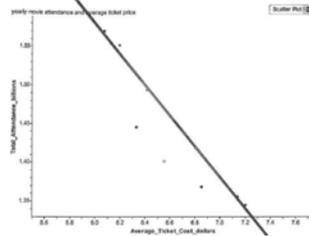


Figure 2b. Attending to linear subset on Task 4.

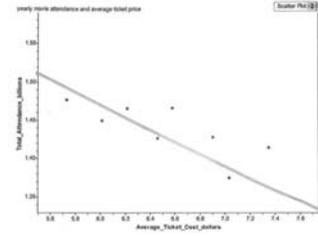


Figure 2c. Ignoring outlier on Task 3.

Not unexpectedly, when we coded, for example, for evidence that a participant ever removed “outlier” data from consideration while placing the line of best fit, Fisher’s exact test found these were significantly associated ($p=0.038$) with having had a case-aggregate view (Table 4).

Table 4. Association between treatment of “outliers” and aggregate approach

Predominant view	Never removed points deemed as outliers	Removed points deemed as outliers at least once	Total
Aggregate	11	1	12
Case-aggregate	3	4	7
Totals	14	5	19

As mentioned, teachers’ placements of informal lines of best fit were strikingly similar, despite the varying criteria or conceptions present in their responses. In fact, the demographic data of the teachers were not associated with any particular dominant conceptions or criteria, nor any other variables introduced during the analysis for that matter. The demographics of the various teachers in this study were not explanatory with regard to knowledge of informal lines of best fit. (This may be somewhat surprising in fact; for example, the grade-level teaching variable (elementary, middle, secondary) potentially could be indicative of participants’ responses, but this was not the case from the teachers in this study.) Similarly, while the categories for dominant conceptions and criteria were useful, for most of the tasks they seemed to have little relationship to the placement of teachers’ lines. While many of the additional analyses between the different variables of interest did not add further insight into the teachers’ understanding of line of best fit – the variables were understandably associated – we discuss below one task that merits further inspection and illuminates specific findings from the study: participant’s work on Task 5.

4.3. PLACING A LINE FOR DATA WITH NO ASSOCIATION

In contrast to the other interview tasks, teachers’ placement of the informal line of best fit on Task 55 (in which the data displayed no association) exhibited much wider variation. Figure 3 shows all 18 teachers’ placed lines for the task (one refused to place a line) along with the least-squares regression line, demonstrating a much more varied collection of responses. The lines are differentiated according to their subsequent coding

in relation to slope. The teacher who refused to place a line on this task stated there isn't a line of best fit because "there is really not a relationship to these points" [T16].

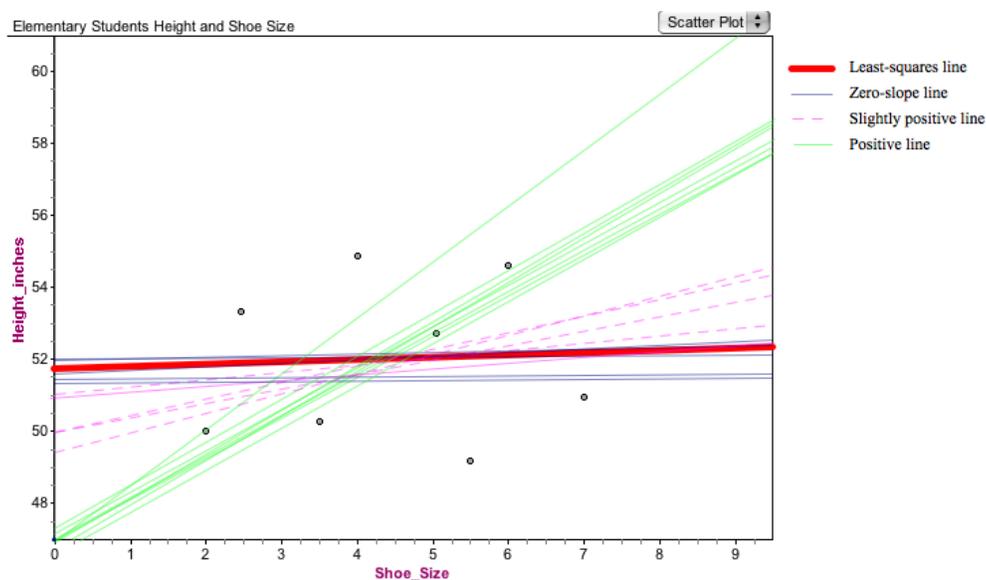


Figure 3. Placement of lines on Task 5 for ($n=18$) participants.

Task 5 was different from the other tasks in that it asked teachers to place a line of best fit for data that had no visible linear association. What is evident from the data is that the teachers' interactions with this task in terms of placing an informal line of best fit were substantively different from the others. In an attempt to better understand this variation, teachers' line placements on Task 5 were coded according to their slope: approximately zero-slope (Z), a slightly positive slope (S), and a positive slope (P). In addition, one teacher refused to place a line altogether on this task, based on the sense that it was incorrect to fit a line to data that clearly had no linear trend. We describe some results from further analyses trying to understand the teachers' responses to this task.

The line placements on Task 5 were neither associated with any particular criteria or meaning, nor linked to any demographic information. Teachers that utilized different criteria and meanings throughout the other tasks were found within each of the three slope-groups in the Task 5 analysis. However, an interesting connection in terms of the criteria the teachers mentioned in Task 5 emerged. In terms of consistency with criteria, all eight of those in the "P" group switched the criterion they used to answer Task 5 from what had been their dominant criterion on other tasks. For example, one participant, whose dominant criterion throughout the tasks had been finding the line closest to all points, altered his criterion on Task 5 to consider going through as many points as possible: "For this one it might actually make sense to do a lot of rotation and see what I get out of it... It goes through three data points, which minimizes their impact on the line of fit. Is there a way to get a fourth?... I think that's about as good as I can do" [T7]. Similarly, 3 out of 5 of those in the "S" group also switched the criterion they used for task 5. However, for those who correctly placed a zero-slope line, they either kept the same dominant criterion evident in previous tasks (2 of the 5), or switched to the criterion (3 of the 5), "no association means horizontal," directly demonstrating an understanding of the meaning of a zero-slope line in this scenario. Thus, we find evidence that those

who were unable to place a zero-slope line successfully in this context (those in the “P” and “S” groups) were unable to take their dominant criterion from other tasks and apply it in a meaningful way to this task.

Two other findings shed additional light on this task. First, the context of the data (shoe size and height of elementary students) proved to be a barrier. For those who did not place a zero-slope line, many discussed the context as the reason for why they chose to make the slope positive (P) or slightly positive (S) – they expected the two variables to be related in this way (none of those with zero-slope lines did so) due to their contextual knowledge of these variables. Fisher’s exact test found that those who referenced the specific context in Task 5 in their discussion of placing the line of best fit was significantly associated ($p = 0.045$) with a non-zero slope (Table 5). This behavior by persons assessing association has been previously noted by several researchers (e.g., Batanero, Estepa, Godino, & Green, 1996; Jennings, Amabile, & Ross, 1982; Kuhn, Amsel, & O’Loughlin, 1988). Although it is the case that shoe size and height are often positively associated, simply ignoring the data at hand in preference of conforming to preconceived expectations shows the teachers were unable to objectively assess whether the data displayed were associated or not. Similarly, those who looked at a subset of the data to fit their line at some point throughout all the tasks (either removing points considered to be outliers or finding a subset of points that closely follows a linear pattern) were also significantly more likely ($p = 0.033$) to have struggled with placing a non-zero sloped line on Task 5. (Table 6 presents the data) In fact, 40% (2/5) of the slightly positive, whereas 75% (6/8) of the positive, and 100% (1/1) of the no line, groups had, at some point, used only a subset of the data to place a line of best fit during the interview. Many of those teachers who placed a positive or slightly positively sloped line on Task 5 did so by fitting the line to a linear subset, thereby exhibiting confirmation bias (McGahan, McDougal, Williamson, & Pryor, 2000).

Table 5. Association on Task 5 between slope of line and use of context

Slope of fitted line	Referenced context	Did not reference context	Total
Positive or slightly positive slope, or no line	8	6	14
Zero-slope line	0	5	5
Total	8	11	19

Table 6. Association on Task 5 between slope of line and relying on a subset

Slope of fitted line	Relied on a subset at least once to fit a line	Never relied on a subset to fit a line	Total
Positive or slightly positive slope, or no line	9	5	14
Zero-slope line	0	5	5
Total	9	10	19

Results, particularly related to the teachers’ approaches and responses to fitting an informal line of best fit in the context of data with no association, along with the subsequent analyses presented, provide some important considerations for thinking about teachers’ knowledge regarding lines of best fit. We examine some of these implications further in the discussion section.

4.4. INTERPRETATIONS OF INFORMAL LINE OF BEST FIT

This section presents two analyses related to the teachers' interpretation of informal lines of best fit. In particular, it has already been noted that the distinction between association and causation is particularly difficult for people, including teachers, to understand (Batanero, Estapa, & Godino, 1997; Engel & Sedlmeier, 2011), with people often incorrectly assuming that an association between two variables implies causation. This study is no different. When asked to respond to whether the linear association in Task 3 suggested that movie ticket cost caused attendance trends (namely that higher ticket prices caused lower attendance), most of the teachers in this study (14/18) said that it had. (One teacher did not respond to this question.) While this affirms others' findings about this confusion, one further analysis with this group of teachers was illustrative in that one demographic variable was significantly associated with this difference: whether or not teachers had taught about line of best fit previously. Those who had taught the topic previously were less likely ($p = 0.019$) to claim that there was a causal relationship from the data. (None of the other demographic variables, however, such as having taken a statistics course or grade-level teaching, was associated with this difference.) Table 7 presents the data and Fisher's exact test for significance. Indeed, this sheds some additional clarity on this common mistake in the context of teachers, and is somewhat reassuring – that those who have taught the topic are more likely to understand this distinction.

Table 7. Association between causation and having taught lines of best fit

Teaching experience	Yes to causation	No to causation	Total
Had taught lines of best fit	1	3	4
Had never taught lines of best fit	13	1	14
Total	14	4	18

Lastly, participants were asked to calculate the value of the slope of their best fit line for Task 3 and interpret its meaning in context. The calculation of slope proved quite difficult for these teachers. We focus, however, less on the difficulty the teachers had calculating a correct value for the slope, and more on their interpretations of that value which is content secondary teachers are expected to teach (e.g., CCSSI, 2010 – see HSS-ID.C.7). First, we note that all but two of the teachers successfully set their interpretation of slope in the context of the problem. However, we also note that nearly all of the teachers expressed a deterministic view of the negative slope, that an increase of x would cause a decrease of y . For example, a typical response was: “Basically every dollar and seventy cents you raise the ticket price, point one billion people less would go to movies” [T2]. Their descriptions implied that the line of best fit determined a “precise” and constant rate of change, similar to interpreting the slope of a line, whereas in fact, statisticians would interpret the slope of a line of best fit as an approximation or an average value – the typical change one would expect to see on average, but never “always.” Only two of the teachers gave a less deterministic interpretation; for example, one of them stated “For each dollar that you would add to the cost of a ticket, you’d expect 0.06 billion less people to go to the movies” [T10]. While the two sample interpretations about slope presented above are very similar, the slight change in emphasis is important. There was such a deterministic emphasis within the teachers’ responses, with very little language providing a more statistical emphasis on expectation

or long run trends, that it felt important to present. Their descriptions of slope appeared much more appropriate for thinking about the slope of a line rather than the slope of a line of best fit. Indeed, we provide further thoughts in the following discussion about the teachers' interpretations, connecting to larger subject-matter differences between mathematics and statistics that were pertinent to informal line of best fit.

5. DISCUSSION

We expand on the results outlined above, discussing some of the primary findings from the study and their implications, which we organize into three sections. This is followed by a section regarding future directions to build on the present study.

5.1. CONCEPTIONS, CRITERIA, AND PLACEMENT OF INFORMAL LINE OF BEST FIT

The line of best fit had five different meanings for the teachers in this study: model, signal, typical, representer, and predictor. Signal was the meaning Engel, Sedlmeier, and Wörn (2008) advocated for teachers to have. Thus, it was encouraging to see that some of the teachers from this study held that meaning. Looking across the meanings, we saw distinctions between teachers who focused on the data at hand (signal and representer meanings) and those who focused on the general relationship between the variables being graphed (model, typical, and predictor meanings). There was also a natural distinction regarding how the teachers dealt with and considered inherent variability of the data when modeling with a line, either primarily incorporating (representer and typical meanings) or reducing (predictor, signal, and model meanings) it. These distinctions provide further insight into what teachers conceive the purpose of a line of best fit to be and have implications for how they would teach this topic to students.

The teachers in this study used numerous criteria, such as “through the most points” and “fit to a linear subset of points”, as they determined where to place the line of best fit on the various tasks. Although particular criteria may have been more dominant during a teacher's interview, we found that the teachers were often interweaving and intermingling a variety of ideas as they informally placed their lines of best fit. Interestingly, the dominant criteria that emerged were markedly similar to those found by Sorto, White, and Lesser (2011) as three of the four criteria (equal number, closest, and sum deviation) were also found in their work with university students. The use of the pairs criterion (see Table 3) as a dominant strategy by two of this study's participants, however, has not been previously noted. Generally, while these teachers may have had different primary ways of thinking about the criteria used to place the informal line of best fit, these were not necessarily “competing” ways of approaching the tasks, but complementary ones, resulting in very similar line placements. This differs from the results of identical work with students, whose criteria and ideas were competing and resulted in widely different line placements (Casey, 2015).

5.2. HORIZONTAL LINE OF BEST FIT

Overall, the teachers in this study appeared to have a reasonable knowledge base regarding informal line of best fit. Their placements, despite utilizing different criteria and expressing different conceptions of best fit lines, were relatively consistent and closely resembled the least-squares regression line in all but one case: bivariate data that displayed no association. One of the primary findings from the study is specifying this

particular area as needing to be addressed more with prospective and practicing teachers. Bivariate data with no association, and lines of best fit with zero slopes, are particularly important ideas in statistics that have implications for more advanced inference tests. In addition, they provide a great context to address, understand, and interpret slopes of lines. We discuss both of these aspects.

In terms of statistical subject matter knowledge, the teachers displayed a relatively knowledgeable understanding of lines of best fit, likely common and comparable to what other professionals would know. This is evidence of common content knowledge in the Statistical Knowledge for Teaching framework (Groth, 2007, 2013). Their variation and struggle in response to Task 5, however, may be indicative of another facet of their knowledge – horizon content knowledge (HCK). The teachers’ work on Task 5 reveals not so much an inability to place a line to data – as they were able to do so effectively in other cases – but rather that the situation with non-associated data is qualitatively different from the previous tasks in terms of fitting a line. The findings on Task 5 from this study support such a distinction. We explain this in terms of understanding descriptive versus inferential aspects of linear models, for which we argue inference should be considered part of the mathematical horizon.

To the casual observer, placing a line for Task 5 does not make intuitive sense – the data in fact do not appear to be linear, and, as such, the line does not describe a relationship. In this sense, teachers’ struggles with this task are somewhat expected and the one teacher’s refusal to place a line is understandable. However, fitting a line to non-associated data has important implications, not necessarily for describing the present data, but for making statistical inferences about their association. Teachers need to consider not just whether a line may model or describe a scatter plot of bivariate data, but also recognize the inference about the two variables that can be made by doing so – namely, is the slope (statistically) significantly different from zero? (While there are exceptions such as data with quadratic relationship, zero slopes frequently indicate no association between the variables in question.) It is only in the knowledge of statistical inference about the slope of a best-fit line – statistical knowledge beyond the immediate task of modeling with a line the relationship exhibited by data – that teachers could make sense of the task, both for themselves and for their students. In this way, recognizing that fitting a line to data that do not appear to be linearly related makes sense not descriptively but in terms of the implications for making inferences based on slope is important for teachers. Since teachers’ knowledge of statistical inference in this case could meaningfully impact their instruction about informally fitting lines to data (i.e., that fitting lines is not just about identifying descriptive “trends” in the data but also about inferring association or non-association), we see their struggles as evidence of a lack of HCK related to lines of best fit. Middle and secondary mathematics teachers need to be prepared to help students grapple with the purpose and implications of fitting such lines for all scenarios (Common Core Standards Writing Team, 2011), and be prepared to help students understand the significance of fitting lines to data, both descriptively and inferentially. Without knowledge of the mathematical horizon related to statistical inference, which was evident from participants’ responses to Task 5, teachers may be unaware of or unprepared for situations and questions that could arise in their teaching about informal lines of best fit, for which they need to be knowledgeable and able to point students toward particularly important and future statistical ideas. Indeed, based on this study, we would argue for including inference tests as a component of HCK for teaching informal line of best fit, in addition to different types of linear and non-linear models and affordances/limitations of predictions as described in Casey and Kaplan (2014).

For teachers, this additional understanding about non-linearly associated variables provides some further thoughts about teaching not just lines of best fit but lines in general. The ability to point students toward important concepts in the discipline is essential as a mathematics educator. Understanding both that data with no linear association would have a zero-slope best fit line, and that a slope of zero for a line of best fit often indicates no association between variables, is important. Fundamentally, it requires that teachers understand and interpret very deeply the implications from zero-slope lines. Indeed, many teachers teach students about slope and help students understand slope as a fundamental aspect of linear functions. Zero-slope lines are frequently discussed less, as both their equations (e.g., $y=5$) and their graphs seem to help less with understanding the more general notion of linear equations and of slope. Indeed, zero-slope lines and vertical lines (e.g., $x=5$) are usually only superficially treated and discussed as two “extra” cases. In fact, because of the statistical importance, we would argue that the teaching of zero-slope lines should be given more attention during the discussion of lines (not just lines of best fit), both in the classroom as well as in teacher education and professional development. Teachers and students alike should have a depth of understanding about association that moves beyond simply positive and negative slopes, but also to the inferences about zero-slopes: namely that the variables are not associated. Meaningful discourse around lines and linear equations that have a zero-slope, particularly moving away from interpretations of slope in terms of “rise over run” and more toward conceptions about covariation between variables (in this case having no impact on each other), both in secondary classrooms and teacher education, could help foster deeper understanding about and interpretations of lines of best fit. Thus, the teaching of slope connected to lines and linear equations provides one place to further bolster and support students’ (and teachers’) development of statistical understanding. This is an important application of linear models and lines of best fit, and as such needs to be developed further.

5.3. STATISTICS AND MATHEMATICS

In addition to informing the knowledge base for teaching informal line of best fit (specifically connected to their struggles on Task 5 and their lack of HCK), we found the work on the tasks evidence of broader differences between mathematics and statistics that merits further inspection. These disciplinary differences have been discussed before (e.g., Moore, 1998; Rossman, Chance, & Medina, 2006; Scheaffer, 2006), but this study informs the ongoing discussion in at least one specific way: a confusion between when and how to use deterministic versus flexible approaches. (By deterministic we mean known, algorithmic, and more rigid approaches, whereas by flexible we mean non-deterministic, aware of limitations, and context-specific approaches.) In fact, while it is the case that mathematics tends to be more deterministic whereas statistics requires greater flexibility and attention to context (Rossman, Chance, & Medina, 2006), both approaches are used and necessary in statistics. The teachers from this study, however, regularly confused when each approach was appropriate.

Indeed, throughout the study, participants seemed unclear about flexibility and the role of context in statistics. On one hand, the teachers considered the context as essential when, in fact, they should not have. For example, sometimes the teachers used the context to provide specific data points that must be on a best-fit line – such as the line having to go through (0,0) “because if the ball isn’t dropped from that height, it’s not going to bounce at a height. So, I know that has to be one of the points” (T12) – or used their knowledge of the variables to infer some particular trend or association that was not

present in the data (e.g., on Task 5, "...generally the taller you are, the bigger your shoe size is. So that kind of factors in..." (T2)). At other times, they discarded points they deemed to be outliers or only focused on a subset of the data when informally fitting their line. In fact, statisticians would argue that these contextual and somewhat flexible approaches with data are inappropriate when fitting a best fit line – the data exist, so there is little room for such thinking. A best fit line, at least a least-squares regression line, has a deterministic formula – the goal, in some form or another, is to minimize the vertical deviation of the data points from the linear model. However, the teachers incorrectly approached fitting a line to the data flexibly, making use of context as an important component in this process or removing undesirable data points when, in fact, their approaches should have been more deterministic. On the other hand, precisely when more flexible and contextual approaches would be appropriate, even requisite, the teachers showcased less flexibility. In statistics, context is important in a way it is not always in mathematics: interpreting the slope of the best-fit line is one such instance. Although context is integral for understanding and describing the slope of a best-fit line, a few teachers did not use the specific context when describing it (e.g., rather than talking about a change with the two variables, the slope was described as "negative two over three"). More prominent, however, the teachers frequently held a deterministic view of the slope from the data, where a more flexible and less rigid interpretation is necessary. Best fit linear models, in fact, do not determine that for every increase in x there is going to be or must be some specific impact on y . Rather, the slope of the line simply suggests a typical, average, or expected change and association between the two variables.

The finding from this study related specifically to teachers' confusion between when more flexible and more deterministic approaches are useful in statistics further informs disciplinary differences between mathematics and statistics about which teachers need to be knowledgeable. Additional clarity regarding when each of these is appropriate in statistics is necessary. Teachers need to develop knowledge not only of general differences between mathematics and statistics, which statistics educators have been noting for years (e.g., Moore, 1998; Rossman, Chance, & Medina, 2006; Scheaffer, 2006), but also specific differences such as this one. This becomes an important component of acquiring statistical knowledge for teaching, and suggests the need to more explicitly address this disciplinary difference with teachers.

5.4. FUTURE DIRECTIONS

This study utilized a relatively small convenience sample of nineteen mathematics teachers from the United States. Inevitably, the sample has limitations in that the teachers were not randomly selected from the population of all mathematics teachers, and thus the results are not necessarily generalizable. However, given that there were no significant knowledge or performance differences between the teachers from the three different universities in the United States, it seems likely that this study's results are indicative of a broader population of United States' teachers. Despite differences based on the preparation the teachers received across the country, their knowledge regarding the informal line of best fit was similar in the results of this study. In addition, the five tasks themselves have inherent limitations. Given the teachers' work on task 5, it would have been interesting to explore other task situations with non-linearly associated data, including some where the context and variables might be considered less strongly related (e.g., shoe size and course grade) or perhaps related but non-linearly (e.g., quadratic data). Also, since greater variability makes fitting lines more difficult, which may have added to the teachers' difficulties with task 5, it would have been interesting to look at

other tasks with data that appeared linearly associated but had smaller r values. Additional tasks could provide further insights into teachers' knowledge about informal lines of best fit. Although the evidence in this study indicates some general themes about and implications for teachers' statistical knowledge, further study could help clarify the generalizability of these findings. Similarly, the questions about causality were limited to only one of the task settings, and did not include scenarios (such as the first two tasks, drop height versus bounce height) that came from more deterministic, causal relationships. Additional contexts in the set of tasks would allow further generalization for our findings.

There are many directions that future work in this area could take. Regarding further research, as previously stated, this study was carried out with teachers in the United States only. Hence, similar studies could be done in other countries that include informal line of best fit in their curricula standards to describe teacher knowledge of informal line of best fit from a more global perspective. It would also be helpful to the field for an expert-level study to be added to the current teacher-level study and Casey's (2015) student-level study. Fitting a line of best fit informally, by eye, in many ways feels more like an art than a science. It would be informative to know how expert statisticians approach the task of informally fitting a line and to use the findings to improve the teaching of this topic to novice learners.

Another direction future work could take concerns teacher education. Development of curricula for use with teachers to address the needs raised previously in Sections 5.1 and 5.2 is a fruitful and desired endeavor as such curriculum does not currently exist. In particular, a curriculum that helps teachers deeply understand differences between mathematics and statistics as disciplines as well as in specific instances like linear mathematical functions versus linear statistical models is sorely needed for it is mathematics teachers that are called to teach statistics in school (pre-university) settings. Additionally from this study, exploring ways to help teachers grasp the importance of linear models for determining association, and the role that the value of slope (and statistical tests on that slope) plays in statistical inference, is also necessary. This is particularly so given the present situation in the United States where future middle and secondary teachers generally take at most one statistics course (Bargagliotti, 2014), and the content of the statistics course(s) often falls well short of what teachers need to teach current curriculum standards (CBMS, 2012b).

6. CONCLUSION

The landscape of school mathematics is increasingly shifting to include more and more statistical concepts, especially due to the broad application of statistics in today's data-driven world. This shift has implications for mathematics teachers – who are frequently responsible for teaching this content – and particularly their statistical content knowledge for teaching. Based on this study, which explored teachers' knowledge of informal lines of best fit, we found that teachers had a relatively strong ability to place lines of best fit accurately. Indeed, their lines were similar to the least-squares regression line on most tasks. However, their varying conceptions and criteria, which at times were inaccurate despite producing relatively good approximations for placing a line of best fit, point to some significant gaps in their knowledge. Such holes or misconceptions in their content knowledge may become increasingly evident in consideration of the complexity of teaching; slightly inaccurate descriptions from a teacher, or using widely different criteria in various situations, can lead to increasingly problematic situations for students' statistical development. In particular, we found that the teachers lacked knowledge that

lines with a horizontal slope are frequently indicative of a lack of linear association between two variables, as well as broader purposes for best fit lines (i.e., that fitting lines serves more than just descriptive purposes). While it is perhaps not sensible to fit a line to non-linear data, the inferences based on slope that can be made are powerful. Teachers need this horizon knowledge to provide a thorough and thoughtful treatment of linear association with students. In addition, we found that the teachers demonstrated a lack of statistical sense, confusing mathematical and statistical approaches. Indeed, their confusion between when statistics calls for more flexible approaches (e.g., interpreting the slope of a regression line) versus when it calls for more deterministic approaches (e.g., not forcing regression lines to conform to mathematical functions, say going through (0,0)) was evident from this study. These disciplinary distinctions between mathematics and statistics need to be clear to teachers. Teacher education and professional development with teachers needs to address such disciplinary distinctions more explicitly. This study adds to extant literature some specific ideas related to knowledge for teaching informal lines of best fit, with conclusions that we find have concrete implications for the statistical preparation of mathematics teachers.

ACKNOWLEDGEMENTS

We acknowledge the assistance of David Wilson with conducting interviews as well as Mike Shaughnessy for his remarks on earlier versions of the manuscript.

REFERENCES

- Australian Curriculum, Assessment, and Reporting Authority (2012). *The Australian Curriculum: Mathematics*. Sydney, Australia: Author.
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.
[Online: [http://iase-web.org/documents/SERJ/SERJ3\(2\)_Bakker.pdf](http://iase-web.org/documents/SERJ/SERJ3(2)_Bakker.pdf)]
- Bargagliotti, A. (2014). Statistics: The new ‘it’ common-core subject. *Education Week*, 33(19), 20–21.
- Batanero, C., Burrill, G., & Reading, C. (Eds.) (2011). *Teaching statistics in school mathematics: Challenges for teaching and teacher education. A Joint ICMI/IASE study: The 18th ICMI study*. New York: Springer.
- Batanero, C., & Díaz, C. (2010). Training teachers to teach statistics: What can we learn from research? *Statistique et enseignement*, 1(1), 5–20.
- Batanero, C., Estepa, A., & Godino, J. (1997). Evolution of students’ understanding of statistical association in a computer-based teaching environment. In J. Garfield & G. Burrill (Eds.), *Research on teaching statistics and new technologies* (pp. 191–206). Voorburg, The Netherlands: International Statistical Institute.
- Batanero, C., Estepa, A., Godino, J., & Green, D. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151–169.
- Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education. A Joint ICMI/IASE study: The 18th ICMI study* (pp. 57–69). New York: Springer.
- Casey, S. (2015). Examining student conceptions of covariation: A focus on the line of best fit. *Journal of Statistics Education*, 23(1), 1–33.

- Casey, S., & Kaplan, J. (2014). Statistical knowledge for teaching: The case of informal line of best fit. Manuscript submitted for publication.
- Common Core Standards Writing Team (2011). *Progressions for the common core state standards in mathematics (Draft)*.
 [Online: http://commoncoretools.files.wordpress.com/2011/12/ccss_progression_sp_68_2011_12_26_bis.pdf]
- Conference Board of the Mathematical Sciences (CBMS) (2012a). *The Mathematical education of teachers II*. Providence, RI and Washington, DC: American Mathematical Society and Mathematical Association of America.
 [Online: <http://cbmsweb.org/MET2/>]
- Conference Board of the Mathematical Sciences (CBMS) (2012b). *Report: Statistics and probability, high school*. Providence, RI and Washington, DC: American Mathematical Society and Mathematical Association of America.
- Coxford, A. F., Fey, J. T., Hirsch, C. R., Schoen, H. L., Burrill, G., Hart, E. W., & Watkins, A. E. with Messenger, M. J., Ritsema, B. E., & Walker, R. K. (2003). *Contemporary mathematics in context: A unified approach Course 1*. New York: Glencoe McGraw-Hill.
- Engel, J., & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. In C. Batanero, G. Burrill, and C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A Joint ICMI/IASE study: The 18th ICMI study*. (pp. 247–258). New York: Springer.
- Engel, J., Sedlmeier, P., & Wörn, C. (2008). Modeling scatter plot data and the signal-noise metaphor: Towards statistical literacy for pre-service teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *A Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Roundtable Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education.
 [Online: http://iase-web.org/documents/papers/rt2008/T4P8_Engel.pdf]
- Estepa, A., & Batanero, C. (1996). Judgments of correlation in scatter plots: An empirical study of students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 25–41.
- Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397–409). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago, IL: Aldine.
- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38(5), 427–437.
- Groth, R. E. (2013). Characterizing key developmental understandings and pedagogically powerful ideas within a statistical knowledge for teaching framework. *Mathematical Thinking and Learning*, 15(2), 121–145.
- Harradine, A., Batanero, C., & Rossman, A. (2011). Students' and teachers' knowledge of sampling and inference. In Batanero, C., Burrill, G., & Reading, C. (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A Joint ICMI/IASE study: The 18th ICMI study* (pp. 235–246). New York: Springer.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372–400.

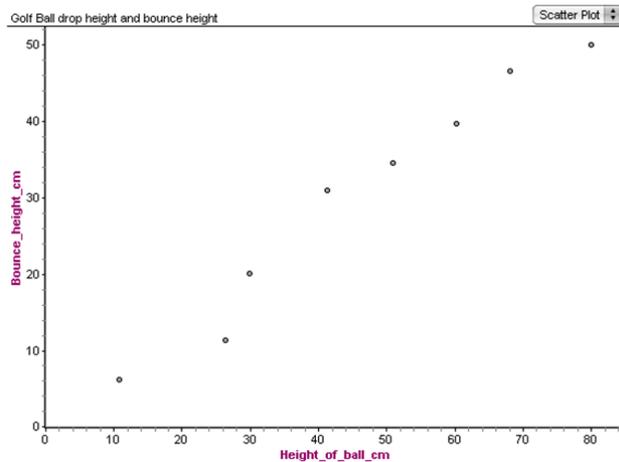
- Jennings, D., Amabile, T., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211–230). Cambridge, England: Cambridge University Press.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259–289.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In *Proceedings of the International Conference On Teaching Statistics*, Durbin, South Africa, June 2002. [Online: http://iase-web.org/documents/papers/icots6/8b2_kono.pdf]
- Kuhn, D., Amsel, E., & O'Loughlin, M., (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- McGahan, J. R., McDougal, B., Williamson, J. D., & Pryor, P. L. (2000). The equivalence of contingency structure for intuitive covariation judgments about height, weight, and body fat. *Journal of Psychology*, 134(3), 325–335.
- Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39.
- Moore, D. S. (1988). Should mathematicians teach statistics? *The College Mathematics Journal*, 19(1), 3–7.
- Moritz, J. B. (2004). Reasoning about covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 227–256). Dordrecht, Netherlands: Kluwer.
- National Governors Association Center for Best Practices and Council of Chief State School Officers (CCSSI) (2010). *Common Core State Standards (Mathematics)*. Washington DC: Author.
- Preacher, K. J., & Briggs, N. E. (2001). Calculation for Fisher's exact test: An interactive calculation tool for Fisher's exact probability test for 2×2 tables [Computer software]. Available from <http://quantpsy.org>.
- Qualifications and Curriculum Authority (2007). *The National Curriculum 2007*. Earlsdon Park, Coventry: Author.
- Rossmann, A., Chance, B., & Medina, E. (2006). On important comparisons between statistics and mathematics, and why teachers should care. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance* (pp. 323–333). Reston, VA: National Council of Teachers of Mathematics.
- Scheaffer, R. L. (2006). Statistics and mathematics: On making a happy marriage. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance* (pp. 309–321). Reston, VA: National Council of Teachers of Mathematics.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Charlotte, NC: Information Age Publishing.
- Sorto, M. A., White, A., & Lesser, L. (2011). Understanding student attempts to find a line of fit. *Teaching Statistics*, 33(2), 49–52.

STEPHANIE A. CASEY
 Department of Mathematics
 Eastern Michigan University
 Ypsilanti, MI 48197 USA

APPENDIX A: INTERVIEW PROTOCOL

Introduction: In just a few minutes I am going to give you six tasks to complete. Each one involves a scatter plot graph of data. We will talk about the data on each graph and then I will ask you to determine the line of best fit for the data points. I would like you to think out loud as you determine where the line is on each graph so that I can understand how you are deciding where to place it.

Task 1:

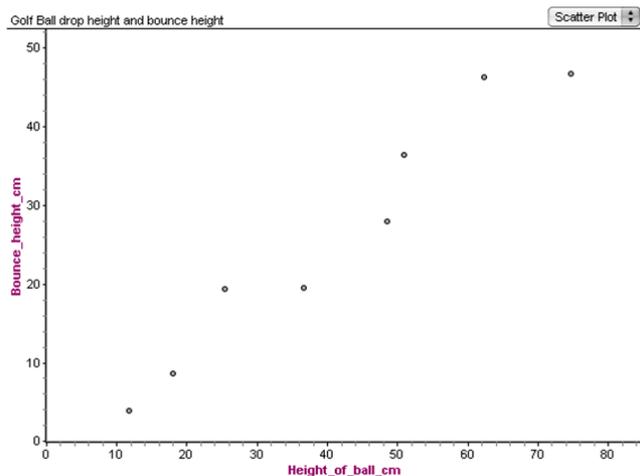


Here is the first scatter plot which shows data about how high a ball bounces when it is dropped from different heights. For example, this point shows the golf ball was dropped from 10.8 centimeters and bounced back up to 6.1 centimeters. I would like you to use this piece of wire as the line and place it where it best fits the data. There are no right or wrong answers. I am interested in how you think about placing the piece of wire, so please tell me what you are thinking as you do this so I can follow your thoughts. Once you decide where you want the line on the scatter plot, we are going to use scotch tape to keep the wire in place [Allow teacher to work through the task].

After each line of best fit is completed (if teacher did not mention when talking aloud), ask follow-up questions where applicable, including:

- Why did you choose to draw the line there?
- What criteria did you use in deciding where to draw the line?
- Did your criteria change for this graph? (for tasks after the first)
- How did you decide on that criteria?
- Could you say more about how you decided at the end that it should be angled a bit more?
- For scatter plot #1 & #2 (if applicable): I noticed you moved the line so that it went through here (point to the intersection of the axes). Can you talk about why you thought that was important?

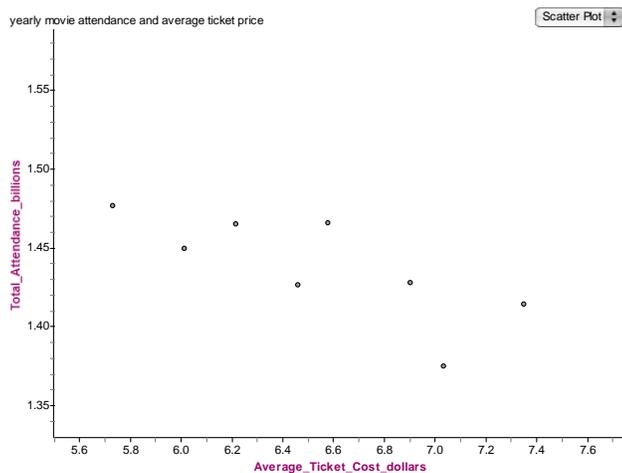
Task 2:



[Same introduction as in Task 1]

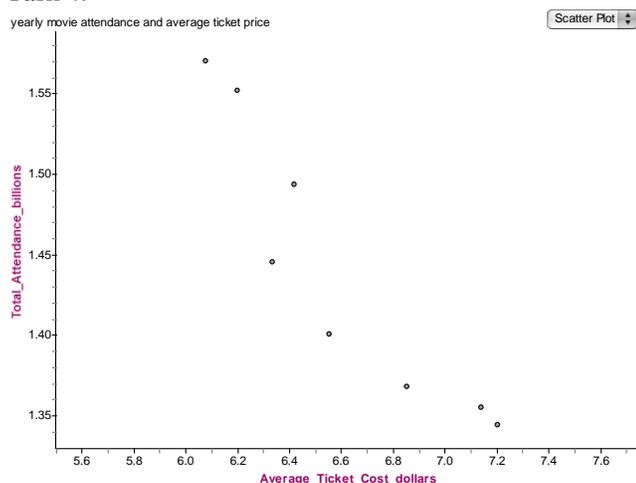
Task 3:

The next two scatter plots display data from a different setting.



The data on this graph [Task 3] estimate the total number of movie tickets sold over an entire year and also estimate the average price of a ticket for that year based on sales in a sample of cities. For example, this point shows that during one year when the average price of a ticket was about \$6 the total number of people attending movies was about 1.45 billion people. Where do you think the line of best fit would be for this data?

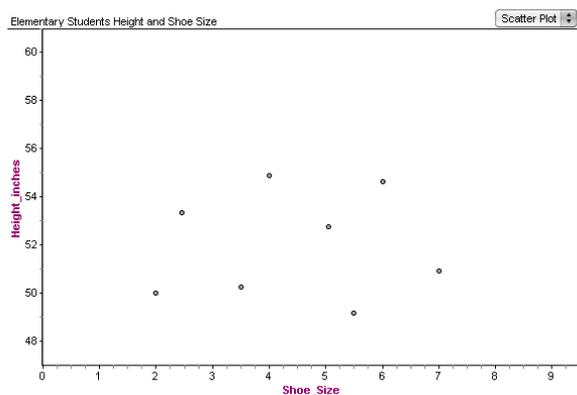
Task 4:



The next graph [Task 4] shows similar data [to Task 3] that is calculated using movie sales data from a different set of cities during the same 8 years. Where do you think the line of best fit would be for this data?

Task 5:

The next scatter plot displays data from a different setting.



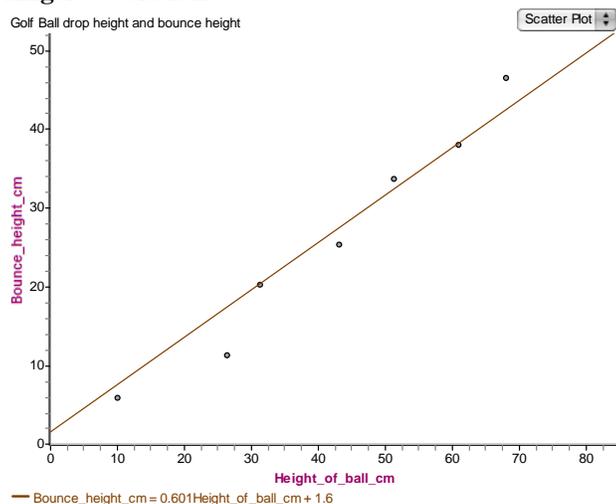
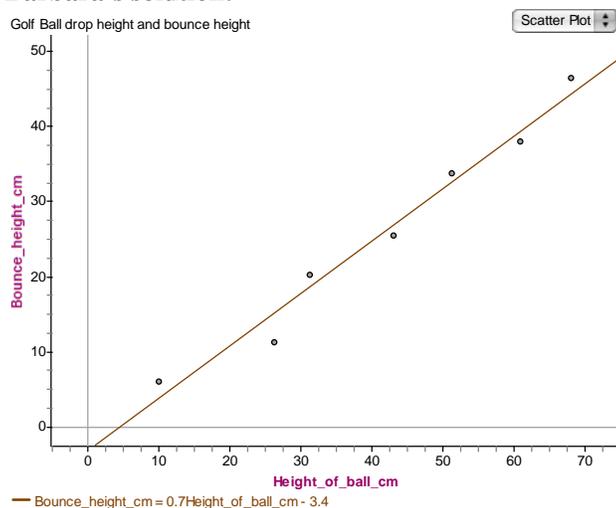
An elementary teacher wrote down the height in inches and shoe size for each of 8 students in his class. He plotted the data in this scatter plot [Task 5]. Where do you think the line of best fit would be for this data?

After Tasks 1-5, ask the following questions regarding the line of best fit:

- 1) Could you tell me what you would say to a student that asked you “What is the line of best fit?”
- 2) Could you tell me what you would say to a student to help them draw the line of best fit on a scatter plot?
- 3) As you determined the line of best fit for each of the 5 scatter plots, did your thoughts about the where the line of best fit gets placed change? If so, how?

Task 6:

Two students, Angelo and Barbara, were given the same task you were given in task 1: to find the line of best fit for the data on a golf ball's drop height and bounce height. They had two different solutions. Which student's line fits the data better and why?

Angelo's solution:**Barbara's solution:**

After Task 6, ask the following questions related to their responses on Tasks 1-3:

- 4) (Show teacher task #1 again) Can you talk about what the line shows about the relationship between bounce height of a ball and the height it was dropped from?
- 5) (Show teacher tasks #1 & #2 again side by side) The line you placed for task #2 is different in some ways than the line that you placed in task #1. Please talk about some of the ways they are different.

- 6) (Show teacher task #3 again)
- a. Can you talk about what the line shows about the relationship between attendance totals for movies and the price of a movie ticket?
 - b. Would you take a moment and calculate an estimate of the slope of the best fit line you determined? You may write at the bottom of the paper.
 - c. Can you talk about the meaning of (say the slope value here) in the context of movie attendance totals and ticket prices?
 - d. Would you say that the price of ticket caused an increase or decrease in movie attendance?
- 7) [*For in-service teachers only*] Have you ever taught students about how to find the line of best fit? (*If yes*) Can you talk about that experience? [In particular, how they explained its meaning and criteria. May need to ask additional questions to bring that out.]