LEARNING, IDENTITY, AND STATISTICAL DATA ANALYSIS

Paul Cobb and Lynn Hodge
Vanderbilt University
USA

The sociocultural perspective draws attention to students' development of a sense of who they are in relation to statistics as an integral aspect of their learning. This focus on students' construction of identities as doers of statistics relates directly to a number of issues that are of immediate concern to most teachers including students' interest in and motivations for studying statistics. We present the results of a classroom design experiment in which a group of 12-year-old students developed identities as people who chose to engage in, saw value in, and viewed themselves as competent at developing data-based arguments. We also discuss aspects of the design experiment that appeared to play an important role in supporting the students' development of these positive orientations towards statistics.

INTRODUCTION

Analyses of students' learning of statistics typically focus on their development of particular content-specific concepts and cognitive skills. However, recent theorizing and research within the sociocultural tradition draws attention to a second aspect of learning that concerns students' development of a sense of who they are in relation to statistics. From this perspective, students' development of identities is an integral aspect of their engagement in the activities of a community such as that constituted by the teacher and students in a classroom (Wenger 1998). Further, students' emerging identities as doers of mathematics are viewed as forms of individuality that are defined with respect to the types of competence that membership in the classroom community entails. Nasir (in press) clarifies that students' developing statistical understandings and their emerging identities are closely related and co-evolve. In addition, she notes that a focus on students' construction of identities as doers of statistics is directly related to several issues that are of immediate concern to most teachers. These include students' interest in statistics, their persistence and motivation while studying statistics, and the choices they make about whether to continue to study statistics and to embark on careers that involve the use of statistics (cf. Boaler & Greeno, 2000; Wenger, 1998).

THE CLASSROOM DESIGN EXPERIMENT

The analysis that we report in this paper is based on classroom design experiment that we conducted over a 14-week period with a group of 11 American eighth-grade students (thirteen years old). Seven of the students were African American, one was Asian American, and three were Caucasian. During the previous school year, we had conducted a prior design experiment with the students that had focused on the analysis of univariate data. The overall goal of the follow-up eighth-grade experiment was to support the students' development of increasingly sophisticated ways of analyzing bivariate data sets as part of the process of developing effective data-based arguments. To this end, we attempted to ensure that the students analyzed data sets that they viewed as realistic for purposes that they considered legitimate when we developed the instructional activities. As an illustration, in one instructional activity the students analyzed data on carbon dioxide levels in the atmosphere for the period 1957 – 1979 as part of a series of activities that focused on global warming. As second example, the students investigated possible inequities in men's and women's salaries in the United States by analyzing data sets that showed salary against years of education for men and for women. In these and the other instructional activities we used, the students were required to write a report in which they presented their findings to a particular audience. This requirement oriented the students to take account of the purposes for which the data had been generated.

In the first phase an instructional activity involved the teacher and students talking through how they could generate data that would enable them to address a particular problem or issue. In these data creation discussions, the teacher and students first clarified why the problem or issue at hand was potentially significant. Next, they discussed aspects of the phenomenon under consideration that were relevant to the issue under investigation and considered how they would measure these attributes in order to generate the required data. Against this background, the teacher introduced the data the students were to analyze. In the second phase of instructional activities, the students analyzed the data individually or in small groups by using a computer-based analysis tool that we developed for the experiment. This tool provided the students with a range of options for organizing bivariate data sets that were inscribed as scatter plots. The final phase of the instructional activities consisted of a whole class discussion in which a computer projection system was used to support the students' presentation of their analyses.

We generated two sets of data during the design experiment. The first data set was generated to document the students' developing statistical understandings whereas the second set of data was generated in order to investigate the identities that the students were developing as doers of statistics. The first data set consisted of video-recordings of all 41 classroom sessions conducted during the design experiment, copies of all the students' written work, video-recorded post-interviews conducted with all the students, and one set of field notes of the classroom sessions. The analyses that we have completed of this data indicate that the students' learning of significant statistical ideas was substantial. In particular, they came to view bivariate data sets as distributions located within a two-dimensional (i.e., multiplicative) space of values and to develop relatively sophisticated ways of analyzing data of this type as part of the process of developing data-based arguments (Cobb, McClain, & Gravemeijer, in press). They also developed a relatively deep understanding of the process of generating data and came to anticipate that the legitimacy of the conclusions that they would be able to draw from data depended crucially on the soundness of the process by which those data were generated (Cobb & Tzou, 2001). The specific issues that they raised when they considered how data might be generated to address particular questions indicated a concern for both the representatives of samples and the control of extraneous variables. A further analysis documents that the students came to view statistical data analysis as an activity that was worthy of their engagement (Hodge, 2001). Taken together, these findings indicate both that the students were afforded access to significant statistical ideas as they fulfilled their obligations as members of the classroom community and that they had come to view these obligations positively. The design experiment classroom would therefore appear to constitute an appropriate case in which to investigate issues relating to students' development of identities as doers of statistics.

The second data set that was generated to investigate the students' developing identities consists of a second set of field notes of the classroom sessions and total of 41 audio-recorded interviews conducted with the 11 students while the design experiment was in progress. These semi-structured interviews were usually scheduled shortly after classroom sessions and focused on the students' interpretations of classroom events with a particular emphasis on both their understandings and valuations of their obligations, and their assessments of their own and other students' competence. Five target students were interviewed on a weekly basis throughout the design experiment and the remaining six students were interviewed at least twice. The interviews were typically conducted with pairs or groups of three students because pilot work indicated that group interviews elicited richer responses than did interviews conducted with individual students.

These data were analyzed by using a variant of Glaser and Strauss's (1967) constant comparison method in which conjectures formulated while analyzing one interview are tested and, if necessary, revised while analyzing subsequent interviews. Six themes or sets of issues were identified in the course of this analysis: The students' understandings of both their general and specifically statistical obligations, and of the nature of statistical activity, their assessments of their own competence, their perspectives on the resources that supported their learning, and their views on the extent to which the class was a collective community. We discuss each of these themes in turn as we present the results of the analysis.

THE STUDENTS' PERSPECTIVES ON THE DESIGN EXPERIMENT CLASS

The observations the students made about their obligations and, indeed, most other aspects of the design experiment classroom were remarkably consistent. With regard to their general classroom obligations, all 11 students noted that their participation in classroom discussions involved explaining their own analyses and asking questions in order to understand other students' analyses. Furthermore, the way in which the majority of the students' framed their statements (e.g. "your job", "you have to", etc.) implied that these were, for them, obligations. For example, one student explained: "You talk about your way, or you add something to someone else's way. You can't just say that you agree or you disagree. Ms. M [the statistics teacher] makes you explain it. You have to ask questions about things that you don't understand" (Megan, 9-1-98).

Five of the students also went on to explain that it was not sufficient to state a conclusion or to demonstrate *how* they had analyzed the data. They also had to explain *why* they had analyzed the data a particular way. These five students all justified this obligation in terms of the importance of the listening students understanding the explanation. For example, one student commented, "I knew what they [the other students] did [in their analyses of the data] so I didn't want them to tell me what they were doing, but what were they thinking, yeah, what was your intention" (Valerie 10-6-98). Similarly, a second student argued, "If you don't talk about what you were thinking about then we don't know if it all is okay…we can't figure out if it is a good point" (Sally 8-27-98). Taken together, the students' comments indicate that they viewed the obligations of explaining their analyses and asking questions in order to understand others' explanations as reasonable and valued them positively. Furthermore, it appeared that an acceptable explanation involved not merely demonstrating how the data had been analyzed but explaining why this way of organizing the data was relevant to the question or issue being addressed (e.g., determining whether there are inequities in men's and women's salaries).

The students all indicated that they valued the fact that an analysis resulted in a number of different viewpoints rather than one correct answer. The following comment is representative in this regard: "It [analyzing data] makes you think of different possible answers and everyone in the classroom could have gotten lots of different answers. So it's real fun to compare them" (Mark, 12-16-98). The majority of the students also said that they sometimes changed their views as they listened to others' explanations and two students clarified explicitly that, for them, the purpose of discussions was to identify the best rather than right answer. All the students also indicated that that the statistics class involved "thinking" and "ideas." It appears, from some of their other comments that "thinking" involved making judgments about the data.

These themes of interpretation and sense making were also evident when the students were asked what they had learned. For example, one student explained that although she already knew terms such as mean, median, and range from previous classes, she had learned when to use them to describe data. She then went on to clarify: "Most of the time I don't use the average. I like using the range. I use the range when the points are spread out. If the points are around in a really small area you probably want to use the median since that would be a better way to let someone know about the points" (Megan 12-16-98).

As a second representative illustration, another student said that he had learned to use the median to compare data sets and justified the use of this statistic by saying, "Because that is the highest concentration of data" (Mike, 12-16-98). However, when the interviewer asked if it is good to always use the median he elaborated, "No, but on some it is…like the speed reading graph. That wouldn't really make sense…the scores were scattered all over". It is evident from these and other comments that the students were developing an understanding of when it was appropriate to use particular statistics to compare data sets. Their responses also indicate that their goal when they analyzed data was to identify trends and patterns that were relevant in enabling them to understand the situation or phenomenon from which the data had been generated (e.g., changes in the level of carbon dioxide in the atmosphere). In this regard, it appears that the graphs the students created by using the computer tool were, for them, texts about these situations. All the students in fact commented that they enjoyed learning about these situations and a majority spoke of it being necessary to "look at the graph and see what is says" and "look at the graphs and decide what they tell you." These comments indicate that the students were acting in a world that they experienced as being very real when they analyzed data.

Turning now to consider they students' views of their own competence, all 11 students said that they viewed themselves as succeeding in the statistics class. In doing so, they spoke of knowing what they were doing, of coming up with good ideas, and of other people in the class talking about their analyses. These responses indicate that they assessed their competence to a large extent in terms of their ability to contribute in substantial ways to class discussions. In addition, we infer that the students had the experience of knowing what they were doing because the graphs they created by using the computer tool were, for them, texts about situations from which the data were generated. It was because they could use the computer tool in this way that they could rely on their own resources to make decisions rather than, say, the judgments of the teacher.

The students all commented on the resources that supported their learning and enabled them to achieve competence even though they were not questioned about this issue explicitly. Six of the 11 students described either the discussions or the computer tool as a resource. For example, when asked what advice she would give to a new student in the class, one student recommended, “You look for like what you think stands out in the graphs. Then you sort of use that. It's kinda easier to use the tool to explain what you saw, you know, when you tell everyone your conclusion” (Sally 10-24-98). Similarly, another student explained, ”You have to look for what's consistent and then talk about it. You can use the tool to explain it. That helps me a lot” (Brad 9-24-98). Other students explained why they found the discussions valuable. For example, one student reported, “I like the discussions because I can share ideas and get some back. You really have to think about what you did and what other people did” (Stacey 8-26-98). Another student volunteered: “It helped me as far as listening to other people and their opinion. It gave me a way to look at things differently. Normally I would just skim over things and come up with a conclusion…that would be it. Now I see what everyone else came up with…sometimes I change my way if it sounds better” (Janet 9-23). In addition, six of the students reported that they developed insights about the data they were analyzing as a consequence of working with others in a small group. Finally, the majority of the students indicated that they viewed the teacher as a resource and spoke of her both explaining the purposes of analyses clearly and valuing their contributions to discussions.

The remaining theme that came to the fore as we analyzed the interviews concerned the extent to which the students viewed the class as collective community. All but three of the students indicated that they viewed all the other students as competent. For example, one student said that she felt regular or equal when the asked if like a good student in the class. When the interviewer followed up by asking if anyone stood out, she clarified, “Yeah, I think we are all equal” (Kate, 10-23-98). The other student who participated in this group interview then added, “Yeah, everybody felt smart. You're like, ‘Hey, look what we did. We talked about global warming, we analyzed these graphs’” (Valerie, 10-23-98). Similarly, a third student responded as follows when asked who was smart in the statistics class: “I can't think of one person. We all do a good job because she [the teacher] explains what we're suppose to do. We can talk to our partner about it” (Janet, 10-6-98). For his part, a fourth student commented, “I think we're all smart. We talk about things so everyone knows what we're talking about. We kind of go at the same speed” (Sean, 12-15-98). In light of the students’ responses, it is worth noting that the analyses they developed differed significantly in terms of their sophistication throughout the design experiment (Cobb et al., in press). We therefore interpret Sean’s observation that the students all “went at the same speed” to mean that, in his view, they were all able to make substantial contributions to class discussions. The students’ assessment of both their own and others’ competence are encouraging from the perspective of equity in that they indicate that none of the students perceived either themselves or others to be marginalized.

DISCUSSION

The analysis we have presented indicates that that the identities the students were developing as doers of statistics involved a strong sense of personal agency in identifying trends and patterns in data that they judged to be relevant to the question or issue under investigation. In this regard, it is noteworthy that their accounts of classroom activities and events focused primarily on their own activity rather than that of the teacher. The findings of an investigation reported by Boaler and Greeno (2000) enable us to clarify the broader significance of this result. In the investigation that they describe, Boaler and her colleagues interviewed students who were in Advanced Placement Calculus classes at six US public high schools. The mathematics classes in two of the schools involved discussions and collaboration among students and were, in Boaler and Greeno's (2000) characterization, generally consistent with current US mathematics reform recommendations. In contrast, the students in the remaining four schools typically worked individually and were expected to complete tasks by applying methods and strategies introduced by the teacher. Interviews that Boaler and colleagues conducted with the students reveal that most of the students in the discussion-based classes valued both the creativity involved in completing the relatively open-ended tasks and the opportunities to collaborate with other students. In contrast, many of the students in the more traditional classes indicated both that they found their experiences of doing mathematics in these classes distasteful and they had come to dislike mathematics and would choose not to study it further. As Boaler and Greeno demonstrate, these students' negative valuations stemmed from a conflict that they experienced between their current identities and the type of person that they believed they would have to become in order to be "a mathematical person." For example, a number of these students noted that in their traditional mathematics classes they had to give up agency and follow methods prescribed by the teacher. Other students described themselves as creative or verbal people and indicated that "math just wasn't that way." For these students, it appeared that the identities they would have to develop in order to become mathematical persons were in conflict with who the students viewed themselves to be and who they wanted to become. In contrast, the identities that the students were invited to develop in both Boaler and Greeno's discussion-based classes and in the design experiment class appear to be compatible with the types of people that students wanted to become. In our view, this compatibility accounts in large measure for the design experiment students' persistence, valuing of, and developing confidence in analyzing data.

We can clarify an important feature of the design experiment classroom that supported the students' development of identities as "statistical people" by drawing on Wenger's (1998) notion of an *economy of meaning*. Wenger (1998) uses the metaphor of an economy to emphasize that the various meanings developed by the members of a community such as that constituted by a teacher and students in a classroom compete to define events and are necessarily valued differentially. This notion therefore brings relations of power to the fore in that power is concerned with whose definition of a situation wins out. In the case of the traditional classes studies by Boaler and Greeno, the teacher's interpretations were highly valued and those of the students were treated as invalid if they differed from those of the teacher. The teachers in these classrooms were the sole validators of whose ideas were legitimate and whose were invalid. In contrast, it is apparent from the interviews we conducted that the design experiment students viewed themselves and the teacher as together constituting a community of validators. In this classroom, the teacher did not express her institutionalized authority in action by attempting to regulate or control the students' statistical reasoning directly. Instead, she initiated and guided the establishment of normative criteria for what counted as a legitimate data-based argument. As the analysis we have presented documents, most of the students were explicitly aware of these criteria and were therefore able to contribute to classroom discussions by validating or challenging others' analysis in ways that were consistent with the teacher's instructional agenda. This contrast between traditional classrooms and the design experiment classroom is particularly significant given Wenger's observation that changes in the realm over which students perceive themselves to have some control has profound implications for both their learning and their developing identities. The relatively wide realm of control that the design experiment students experienced is reflected in their sense of knowing what they were doing and of being able to rely on their own resources as they analyzed data.

CONCLUSION

Given the generally positive student comments that we reported, it is important to clarify that the students did make two critical observations. First, several students said that the format of the class was very repetitive and indicated that they would have enjoyed more variety. Second, the majority of the students were also critical about the number of class sessions spend conducting analyses that focused on a single theme such as global warming. However, the students' valuations of both their obligations as members of the classroom community and the nature of what they were learning were consistently positive. We have focused on these latter issues as they relate directly to the identities that they were developing as doers of statistics.

In the first part of this paper, we briefly summarized previously completed analyses that document that the students' statistical learning was relatively substantial. The interviews indicate that none of the students perceived themselves to have been silenced as the teacher achieved her instructional agenda but instead viewed both themselves and the other students to be substantial contributors to class discussions. These two developments bring to the fore the aspects of the instructional design that made it possible for the teacher to achieve her agenda by building on the students' explanations of their analyses. We have already noted that we attempted to ensure that the students analyzed data sets that they viewed as realistic for purposes that they considered legitimate when we developed the instructional activities. In addition, the computer tool was designed to fit with the students' current ways of reasoning about data while simultaneously serving as a primary means of supporting the reorganization of that reasoning (see Cobb et al., in press).

Beyond these features of the design, it was also essential that the instructional activities constitute a coherent sequence in the sense that, at any point in the design experiment, they gave rise to a range of different analyses from which the teacher could guide the emergence of significant statistical issues as topics of conversation. The teacher was therefore able to transcend what Dewey (1980) termed the dichotomy between process and content by systematically supporting the emergence of key statistical ideas while simultaneously ensuring that the students' activity was imbue with the investigative spirit of data analysis from the outset. This was crucial given that an investigative or exploratory orientation is not merely a means of supporting learning but is instead central to data analysis and constitutes an important instructional goal in its own right (Biehler & Steinbring, 1991). In the process, the teacher afforded her students access to identities and competencies that have what Bruner (1986) termed clout within both the school and wider society.

ACKNOWLEDGEMENTS

The members of the research team who conducted the design experiment on whish this paper is based were Paul Cobb, Kay McClain, Koeno Gravemeijer, Cliff Konold, Erna Yackel, Jose Cortina, Lynn Hodge, Maggie McGatha, Nora Shuart, and Carrie Tzou. The computer tool used in the experiment was designed by Koeno Gravemeijer, Paul Cobb, and Kay McClain.

REFERENCES

Biehler, R., & Steinbring, H. (1991). Entdeckende Statistik, Stenget-und-Blatter, Boxplots: Konzepte, Begrundungen and Enfahrungen eines Unterrichtsversuches [Explorations in statistics, stem-and-leaf, boxplots: Concepts, justifications, and experience in a teaching experiment]. *Der Mathematikunterrich,t, 37*(6), 5-32.

Boaler, J., & Greeno, J.G. (2000). *Identity, agency, and knowing in mathematical worlds.* In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 171-200). Stamford, CT, Ablex.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA, Harvard University Press.

Cobb, P., McClain, K., & Gravemeijer, K. (in press). Learning about statistical covariation. *Cognition and Instruction*.

Cobb, P., & Tzou, C (2001). *Supporting Students' Understanding of Data Creation.* Manuscript submitted for publication.

Dewey, J. (1980). Democracy and education. In J.A. Boydston, Baysinger, P.R., and Levine, B. (Eds), *The middle works of John Dewey, 1899 -1924,* Vol.9. Carbondale: Southern Illinois University Press**.**

Glaser, B.G., & A.L. Strauss (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.

Hodge, L.L. (2001). *Students' emerging identities as doers of mathematics in two contrasting classroom microcultures.* Unpublished dissertation, Vanderbilt University, Department of Teaching and Learning.

Nasir, N.S. (2002) (in press). Identity, goals, and learning: Mathematics in cultural practice. *Mathematical Thinking and Learning.*

Wenger, E. (1998). *Communities of practice*. New York: Cambridge University Press.