Title
Can Middle-School Students Learn to Reason Statistically Through Simulation Activities?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 17(0)

Authors
Osana, Helena P.
Derry, Sharon J.
Levin, Joel R.

Publication Date
1995

Peer reviewed

Can Middle-School Students Learn to Reason Statistically Through Simulation Activities?

Helena P. Osana
osana@macc.wisc.edu

Sharon J. Derry
sderry@macc.wisc.edu

Joel R. Levin
levin@macc.wisc.edu

Department of Educational Psychology
University of Wisconsin-Madison
1025 W Johnson Street
Madison, Wisconsin 53706

Abstract

This paper describes the implementation and quasi-experimental evaluation of a three-week instructional project designed in accordance with theories and assumptions of constructivism and socially situated cognition. Our goal was to develop students' ability to reason about real-life problems, where "good reasoning" was conceptualized in terms of a normative thinking model derived from cognitive research in decision making, probabilistic reasoning, and argumentation. In the spring of 1994, students in two middle school classrooms worked in teams that collected evidence, constructed arguments, and prepared presentations while engaged in activities that culminated in a mock legislative hearing. Through instruction and mentoring, students were encouraged to use statistics and probability as tools for reasoning. The effectiveness of the program was evaluated by comparing the written arguments of students from the two treatment classrooms with those of students from eight comparison classrooms. Students' arguments were scored in terms of how well they captured essential features of model reasoning and avoided particular thinking fallacies. That the reasoning abilities of students developed through social negotiation and shared problem solving was suggested by findings showing significant differences in performance between treatment and comparison classrooms.

This study examined the effects of a three-week instructional unit implemented in two middle-school classrooms during the spring of 1994, for the purpose of improving students' individual and social thinking skills within real-world problem contexts. Our instructional approach assumed that theories of probability and statistics provide reasonable normative models upon which to build knowledge about "good thinking," an assumption supported by scholarly argument, theory, and empirical research (Kahneman & Tversky, 1973; Kuhn, 1991; Nickerson, 1991). Our approach also assumed that processes of social negotiation and shared problem solving within meaningful, activity-oriented environments are integral components of "real-world" apprenticeship that can and should be modeled in the classroom (Brown, Collins, & Duguid, 1989; The Cognition and Technology Group at Vanderbilt, 1990; Dewey, 1938; Greeno, 1989; Lave, 1991; Resnick, 1987; Vygotsky, 1978). The idea that all learning is situated, in that it emerges from the physical, cognitive, and social contexts in which it occurs, permeated our research efforts. We draw from researchers who have pointed to the importance of context and culture in the attempt to understand children's thinking, and from classroom research that appropriates the socially situated cognition perspective (see Brown & Campione, 1990; Lave, 1992; Resnick, Salmon, Zeitz, Haley Wathen, & Holowchak, 1993).

Consistent with these views, students in our study were introduced to concepts about argument, evidence, statistics, and research, as these concepts were needed to help them make sense of issues that arose during viewing and discussion of a popular film (Lorenzo's Oil). Then, students were expected to employ and extend this knowledge in the context of small-group research activity that required them to take a position on a particular issue that was present in the news of the time -- government regulation of the vitamin and dietary supplement industry. Finally, students were required to present and defend their arguments and counterarguments within the context of a four-day mock legislative hearing wherein the fate of a proposed regulatory bill was decided. During activities, researchers and teachers were present in the classrooms as mentors who attempted to coach and scaffold students' transitions from novices to more skilled reasoners (Collins, Brown, & Newman, 1989).

Our study can appropriately be viewed as a "design experiment" (Brown, 1992) in that it tested the efficacy of instructional design based on the theoretical assumptions outlined above. As the activities and simulation format obliged students to work collaboratively in small groups within the context of particular problems, and as relatively little direct instruction was provided, we hoped that students would be able to construct useful knowledge through social negotiation and shared problem solving, and that this knowledge would enable them to reason more effectively about real-world problems. We hypothesized that the students would acquire greater understanding of the inherent uncertainty in the world, and would acquire knowledge of probability and statistics as tools for thinking about problems.

To obtain evidence regarding the effects of instruction, we compared both the individual and social reasoning of students who received the intervention to that of students from comparison classrooms that did not participate in the program. The basis for this comparison was students' written responses to questions about fictional court scenes. Based on a detailed coding system derived from our theoretical analysis of what constitutes good and fallacious thinking in these contexts, student responses were coded by trained raters for evidence of statistical and nonstatistical thinking. Mean comparisons between treated and non-treated classrooms indicated that a nondirect, simulation-based instructional approach can foster maturation of students' reasoning skills. Details follow.

Method

The investigation was conducted in a racially integrated and socioeconomically diverse middle school in the Midwest. Ten intact, home-base classrooms in the eighth grade participated in the study. These classes had not been constituted with respect to ability, race, or socioeconomic status.

Two of the 10 classrooms received the instructional treatment. One social studies teacher and one science teacher offered their home-base classes for the administration of the intervention. The remaining eight classrooms served as comparison classrooms. Pretest and posttest measures were administered to all classes to measure the development of statistical reasoning in the treatment groups. In addition, videotape data were collected during the instructional simulation in the two treatment classrooms, although the current paper will focus primarily on the quantitative analysis.

Pretest and Posttest Instruments

All students in the 10 classrooms received a paper-and-pencil pretest one week before the start of the instructional treatment. There were two forms of the test; half the students received one form and the other half received the other. Each form consisted of a transcript of a fictional courtroom drama (inspired by the popular television show, "Law and Order"), followed by two questions requiring students to take a side or position on the drama and then to provide written justifications for their position. Of the two questions, one required statistical thinking for an appropriate justification; the other did not. We were interested in testing whether or not statistical knowledge, if learned, would be inappropriately generalized; that is, whether or not the students would apply statistical reasoning to a situation where it was not appropriate.

The first form of the test (the Andrevil Trial) involved a boy named Kenny who suffers from a severe form of cancer. His mother voluntarily agreed to allow her son to participate in Dr. Birch's experiment in which an experimental drug, Andrevil, is tested. In the scene, Kenny and his mother are suing the doctor because Kenny's condition had only deteriorated throughout the course of the experiment. The doctor wishes to complete his experiment (with Kenny as a participant) and arrive at a valid conclusion about the effects of the drug. In the second version (John's Trial), a young boy is tried in adult court for violently harming a friend. The lawyer defending him argues that John is not responsible for his actions because he possesses an extra y-chromosome, a condition shown in the 1960s and 1970s to be linked to violent behavior. The prosecuting attorney claims that the results of those studies are inconclusive and thus cannot be used as evidence in the case.

The posttest was administered approximately 10 days after the instructional intervention had been implemented. Students who received the Andrevil Trial on the pretest received John's Trial on the posttest, and vice versa.

Testing Procedure

To enable us to assess the effects of treatment on both individual and social reasoning, students were either tested individually or in groups. With each classroom, a random sample of 4 to 5 students was chosen to take the pretest individually in the library. The remaining students in the two treatment classrooms had been assigned to instructional groups by their teacher, and these were maintained during testing. In the comparison classrooms, the students were randomly assigned to groups of 4 to 5 students.

In the library, the students were told to work without communicating with their neighbors. They were given 12 minutes to read the text, and 5 minutes to respond to the questions. The students working in small groups in the classrooms were given 12 minutes to discuss the story amongst themselves, but each student was required to provide individual responses to the questions, and was given a separate response sheet with which to do so. Students were given 5 minutes to respond to the questions, but were permitted to discuss their answers with other group members.

Instructional Intervention

In each of the two treatment classrooms, the intervention lasted approximately three weeks. Each teacher offered one 70-minute class period each day during this 3-week period. These 70 minutes consisted of the students' regular 50-minute class time with the addition of a 20-minute home-base period.

The first four days of the instruction concentrated on viewing and discussing important issues raised in the film Lorenzo's Oil. This film served as a vehicle for introducing topics related to reasoning with statistics and probability, evidence and argumentation, and medical research. Small-group discussion guides were designed to encourage appropriate thinking about important statistical and scientific issues. Class discussions following small-group activities were orchestrated by the teacher and project researchers. The next three days consisted of three instructional units that focused on statistical concepts, evidence and argumentation, and effective communication of data and arguments, respectively. These units were taught by the project researchers in collaboration with the teachers.

During the next eight days, students participated in the simulation activity. They were placed in small groups that represented interest groups that actually existed within the community: those who opposed regulation, those who were in favor of regulation, and those who held balanced positions on the issue, such as the legislature and the press. Within each group, each student played the part of a particular member of the community, and their duties and responsibilities were outlined. In general, students remained in the same groups for both instruction and testing.

The legislative group was asked to draft a bill proposing some form of regulation of the vitamin and dietary supplement industry. The ultimate responsibility of the legislature was to decide the fate of this bill after considering arguments presented during a four-day mock legislative hearing by the pro- and anti-regulation groups. Before the hearing, the students in pro- and anti-regulation groups worked within their groups studying data pertaining to the issue (made available to them by the researchers), developing arguments and counterarguments, and preparing to present their positions at the hearing.

The teacher and the project researchers mentored the activities of small groups. The aim of mentoring was to help keep the students motivated, structure their activities, remind them of pertinent statistical concepts, and monitor their metacognitive strategies.

Analysis and Results

A few introductory comments must be made before reporting the results of this study. First, it was not possible to assign the 10 participating eighth-grade classrooms to treatment and control conditions. This, in the face of positive effects, of course limits our ability to argue that the intervention directly caused the improvement in statistical reasoning abilities in the two treatment classrooms. To the greatest extent possible, this limitation was addressed in our statistical analyses.

Second, the individual students in the 10 classrooms were not regarded as the "units of analysis" (see Levin, 1992) because within-classroom independence among students could not be reasonably assumed. Indeed, one objective of the study was to help the students build cooperative reasoning skills. The treatment required that the students work together in small groups within each classroom; the performance of the individuals engaged in these joint problem-solving activities affected the thinking and learning of the other students in the class, both during the simulation and while being tested. Therefore, classrooms served as the units of analysis for this study.

The first results reported are based on students' responses to the statistical question on both the pretest and posttest measures (as opposed to the nonstatistical question). The analyses for both pretest and posttest performance were conducted with testing condition (individual versus group testing) and story version (the Andrevil Trial and John's Trial) as two within-classroom factors. For further details regarding the analyses conducted, see Derry, Osana, Levin, & Jones (in press).

Scoring

The written responses on both the pretest and posttest were coded using a scheme developed by the researchers. This assessment rubric was based on a theoretical model for good reasoning representing a synthesis of research and theory on normative thinking models, statistical reasoning, and evidential argument (e.g., Kahneman & Tversky, 1973; Konold, 1989; Kuhn, 1991; Shaughnessy, 1992). The scheme comprised 13 categories, each one representing one way in which an argument can be (correctly or incorrectly) supported. More specifically, categories were assigned for using correlational reasoning, for considering the degree of strength of relationships, and for referring to specific counterpositive evidence. Categories were also developed for recognizing the distinction between real and chance relationships, for noting that correlation does not imply causation, for taking into account methodological quality of the research, for realizing that further research might be needed before conclusions can be drawn, and for commenting on any risk/benefit tradeoffs. Categories were additionally developed for responses that included elements of deterministic reasoning (Konold, 1989) and unsubstantiated opinion. For a complete description of this scoring scheme, see Derry et al. (in press).

Point values were assigned to each of the categories in the scoring scheme. Higher values were attached to the categories that were central to model responses developed for each story (Andrevil Trial and John's Trial). Negative point values were assigned to undesirable responses. An overall score was computed for each student and each question by summing together the values representing points awarded or subtracted for a particular response. Student scores were in turn averaged to obtain a small-group mean, and these were averaged to obtain a mean for the class.

Findings

Statistical question. There was no significant difference between the mean of the experimental classrooms on the pretest and the mean of the comparison classrooms, $\underline{F} < 1$. (Missing data for one of the repeated measures cells required that one comparison classroom be dropped from this analysis).

The analysis of posttest performance was conducted after adjusting students' posttest scores by their corresponding pretest scores (Levin & Serlin, 1993). All subsequent results reported here are based on these adjusted posttest scores. The mean on the posttest for the experimental classrooms (3.61) was statistically higher than the mean for the comparison classrooms (2.27), $\underline{t}(7) = 2.21$, $\underline{p} = .032$, one-tailed -- a result confirmed by an exact permutation test.

In addition, the main effect of story version was found to be statistically significant, with the mean of the posttest scores on the Andrevil Trial (3.88) higher than the corresponding mean on John's Trial (2.00), $\underline{t}(7) = 3.19$, $\underline{p} = .015$. The main effect of testing condition was found not to be significant, and no interactions were detected.

Nonstatistical question. We were interested in seeing whether or not the students who received the instructional intervention would overgeneralize their knowledge and apply statistical reasoning to the nonstatistical question on the posttest. In other words, did these students transfer their knowledge of statistics to a situation in which statistical reasoning was not appropriate?

The answer to this question, it seems, is no. The students in the treatment classrooms were compared to the performance of the students in the comparison classrooms on the posttest. Again, the classrooms were used as the units of analysis and testing format was the within classroom factor. Virtually no students in any of the 10 classrooms who received John's Trial on the posttest used statistical reasoning on the nonstatistical question. For the Andrevil Trial, no statistical difference was found between the nonstatistical question mean of the posttest scores for the two experimental classrooms (1.37) and the corresponding mean for the eight comparison classrooms (1.24), $\underline{t} < 1$ in absolute value.

Discussion

Possessing well-developed higher-order thinking skills and reasoning abilities is becoming increasingly important in the face of the complex problems of our constantly changing world. Thinking about, and considering solutions to, social and scientific problems requires not only knowledge and information from various domains, but also process-oriented skills such as thinking statistically and making reasoned decisions. Furthermore, problems that are complex and uncertain in nature cannot be solved by one person working alone; they require the planning, evaluating, and decision making of people working in teams. Dealing with problems of global proportions, such as AIDS and overpopulation, or even problems of local interest, such as building a new school in the community, demands the joint knowledge and problem-solving efforts of various people.

It has become imperative, therefore, to examine whether and how such complex reasoning skills can best be developed by schools. This was the purpose of our study, which constituted a "design experiment" that tested an approach to thinking-skills education based on socially situated cognition. We hypothesized that students would develop statistical reasoning abilities, knowledge about effective argumentation, and collaboration skills, by engaging in an activity that simulated real-world practice.

When their arguments were judged against models of good thinking grounded in cognitive research on statistical reasoning and evidential argument, students in two classrooms that participated in our three-week instructional program performed statistically better than students in eight comparison classrooms that did not participate. Also found was that the students who participated in the instructional program did not inappropriately transfer what they had learned. They apparently were able to discriminate between contexts in which statistical reasoning was appropriate and contexts in which it was not.

As mentioned earlier, we cannot draw definite causal conclusions from the results. We maintain, however, that the results suggest that our instructional intervention was instrumental in helping students build statistical reasoning abilities, as well as develop knowledge about argumentation and collaboration. To strengthen the claim that our intervention contributed to student learning, this study needs to be replicated with various improvements. Other than ensuring the equivalence of treatment and control groups through randomization, refinements to the instructional units, the simulation activities, and the materials provided to the students are needed.

To the extent that our study represented a quasi-experimental test of instructional design based on a "situated social practice model" (Derry & Lesgold, in press), our results should encourage educators who believe that effective thinking-skills instruction should emulate the mentored, situated, social interactions that foster natural evolution of individual and social thinking abilities.

Acknowledgments

Funding for this project was provided by the National Center for Research in Mathematical Sciences Education. The authors acknowledge significant contributions by Velma Hamilton Middle School Teachers Tom Bauer, Kim

Vergeront, Doris Dubielzig, and Joan Unmacht. We also thank Leona Schauble for her assistance with instruction and for many valuable suggestions throughout the project. Jasmina Milinkovic is thanked for help in planning and data collection. Ian Duncan, of the University of Wisconsin-Madison Department of Veterinary Medicine, is thanked for his presentation on ALD and the film "Lorenzo's Oil".

References

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences, 2, 141-178.

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking, or a context by any other name. Human Development, 21, 108-125.

Brown, J. S., & Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18(1), 32-42.

The Cognition and Technology Group at Vanderbilt (1990). Anchored instruction and its relationship to situated cognition. Educational Researcher, 19 (6), 2-10.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction. Essays in honor of Robert Glaser (pp. 453-494). Hillsdale, NJ: Erlbaum.

Derry, S. J., & Lesgold, A. (in press). Instructional design: Toward a situated social practice model. In D. C. Berliner & R. C. Calfee (Eds.), Handbook of educational psychology. New York: Macmillan.

Derry, S. J., Osana, H. P., Levin, J. R., & Jones, M. (in press). Statistical gaming. To appear in S. Lajoie (Ed.), Reflections on statistics: Agendas for learning, teaching and assessment in K-12. Hillsdale, NJ: Erlbaum.

Dewey, J. (1938). Experience and education. New York: Collier Books.

Greeno, J. G. (1989). A perspective on thinking. American Psychologist, 44 (2), 134-141.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, pp. 237-51.

Konold, C. (1989). Informal conceptions of probability. Cognition and Instruction, 6, 59-98.

Kuhn, D. (1991). The skills of argument. New York: Cambridge University Press.

Lave, J. (1991). Situated learning in communities of practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), Perspectives on socially shared cognition (pp. 63-84). Washington, D. C.: American Psychological Association.

Lave, J. (1992). Word problems: A microcosm of theories of learning. In P. Light & G. Butterworth (Eds.), Context and cognition: Ways of learning and knowing (pp. 74-92). Hillsdale, NJ: Lawrence Erlbaum.

Levin, J. R. (1992). On research in classrooms. Mid-Western Educational Researcher, 5, 2-6, 16.

Levin, J. R., & Serlin, R. C. (1993, Apr.). No way to treat a classroom: Alternative units-appropriate statistical strategies. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Nickerson, R. S. (1991). Modes and models of informal reasoning: A commentary. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), Informal reasoning and education (pp. 291 - 309). Hillsdale, NJ: Lawrence Erlbaum.

Resnick, L. B. (1987). Learning in school and out. Educational Researcher, 16, 13-20.

Resnick, L. B., Salmon, M., Zeitz, C. M., Haley Wathen, S., & Holowchak, M. (1993). Reasoning in conversation. Cognition and Instruction, 11 (3 & 4), 347-364.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.) Handbook of research on mathematics teaching and learning (pp. 465-494). Reston, VA: National Council of Teachers of Mathematics.

Vygotsky, L. S. (Trans., 1978). Mind in society: The development of higher-psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and Trans.) Cambridge, MA: Harvard University Press.