ADDRESSING COGNITIVE AND SITUATIONAL COMPLEXITY IN THE INSTRUCTION AND ASSESSMENT OF STATISTICAL REASONING

Sharon J. Derry
University of Wisconsin-Madison, USA
Joel R. Levin
University of Arizona, USA
Rand J. Spiro
Michigan State University, USA

The cognitive-theoretic instructional and assessment approaches described here represent our efforts to develop and measure students' abilities to reason spontaneously and flexibly with statistics in the context of complex real-world activity. We report results from two instructional projects based on situated cognition, in which students were taught statistical reasoning through guided participation in simulations of authentic professional activities requiring presentation and critique of statistical arguments. Although students' statistical reasoning improved in selected ways, the approach was costly and difficult to implement and sustain. In search of more practical and powerful approaches, current experiments are investigating whether instruction based on video technologies and Cognitive Flexibility Theory can speed development of ability to think flexibly with statistics while seeing interacting themes in real-world situations.

INTRODUCTION

We live in a world in which statistical themes interact in complicated ways with sometimes emotional legal, scientific, ethical and aesthetic issues. It is also a world in which spontaneous, intuitive response can be as important as -- sometimes more important than -- deliberate, reasoned reflection. Consider that whether one is reading a newspaper, watching a political debate on television, participating in a jury trial, or merely arguing with a friend, it is "in-the-moment" reaction to argument and evidence that largely determines how such moments are experienced and later recalled when there is time to think and reflect. For active participants in argumentative discourse, such "in-the moment" reactions shape the direction of discourse itself. Thus, instruction concerned with statistical literacy must focus not only on how students think when there is time for deliberate reflection, but also on the shape of student thinking that occurs "in real time." The work discussed in this paper is about the problem of teaching students to reason statistically in a complex real-time world. Current theories of learning have encouraged us to explore different approaches to this problem, and from these explorations we are learning valuable practical and theoretical lessons.

COGNITIVE COMPLEXITY AND INSTRUCTION

Much early thinking on this matter was grounded in the symbolic processing model of cognition, which leads to the idea that statistics, as subject matter, must be encoded symbolically in memory before it can later be recalled and used in problem solving. A major objective of instruction from this perspective is to help students achieve conceptual understanding and a high degree of knowledge organization. Instructional design focuses on clarity of explanation, representation, and the structure of statistics as a discipline. There is also substantial concern with providing problem-solving practice, so that symbolically represented statistical ideas become closely associated with symbolic memory representations of problem contexts. From this theoretical viewpoint, appropriate in-the-moment responses are capabilities that result from intensive repeated problem-solving practice. Through repeated practice, students learn to activate statistical concepts rapidly and automatically, but only in those situations where the problems of real life are sufficiently similar to previously learned analogous problem cases experienced during statistics instruction.

Spiro, Feltovich and Coulson (1992) argued that helping students cope with real-world complexity involves many further considerations. One is that problems encountered in real-world contexts are often unique and more complicated than those designed for convenient practice in statistics classrooms. Thus, familiar analogies will seldom be readily available to guide students'

future problem solving. Instruction concerned with promoting more advanced forms of statistical literacy must help students learn to assemble unique situational representations flexibly from the partially analogous problem cases that were previously practiced.

Another issue is that adequate performance requires rapid, intuitive (rather than or in addition to reflective, deliberate) response, so the assembly process itself must proceed with effortless flexibility. From the symbolic viewpoint, the only way to develop a fluid, automated skill is through intensive, long-term practice. Thus, instructional designs for advanced learning must incorporate means of providing intensive practice in the process of spontaneous, real-time idea assembly. Accordingly, Cognitive Flexibility Theory (CFT) suggests that advanced instruction in complex domains should proceed by helping students interpret and reinterpret multiple real-world cases through assembly and reassembly of multiple domain concepts (Spiro et al., 1992).

A third point is that real-world cases are not mere statistical abstractions; they should be viewed from multiple perspectives -- viewpoints that interact with statistical reasoning and may even call it into question. Thus, additional lenses and considerations, such as moral, legal and aesthetic issues, are appropriate and important components of statistical reasoning instruction, and the case interpretations required of students should incorporate them (see, for example, Derry, Levin & Schauble, 1995).

Because statistical literacy assessments should be proxies for real-world performance, these also should incorporate requirements for flexible assembly and reassembly of statistical concepts in interaction with other ideas. And at least some of this assembly process should be assessed in real time. For example, given performance standards requiring flexible, in-the-moment intelligence, students might be assessed on their ability to participate in debate requiring real-time response to alternative complex arguments about real-world dilemmas.

INSTRUCTIONAL DESIGN BASED ON SITUATED COGNITION

An alternative view of learning and instruction is situated cognition, also called the "situativity perspective" (Greeno, 1997), a family of non-dualist cognitive theories including Lave's situated social cognition theory (Lave, 1988), embodiment theory (Glenberg, 1997), activity theory (Nardi, 1996), distributed cognition (Hutchins, 1995), and sociocultural theory (Wertsch, 1998). Lave, a major spokesperson for this world view, has argued that her version of situated cognition is a critical analysis of how learning occurs in context that has been inappropriately co-opted as a theory for instructional design. Yet in recent years, this worldview has intrigued and influenced almost all instructional researchers and designers.

Designers of statistics instruction taking a true situated cognition perspective are vitally concerned with complexity, but less concerned with the structure and meaning of statistics as an abstract mathematical discipline, or with helping students encode that structure as a compact symbolic memory representation. Nor do they think of statistics concepts as lenses or schemas for seeing the world. Rather, the focus is more directly on the structure of socially relevant statistical reasoning *activity*. The objective of instruction is to immerse the student, as an apprentice, in a "culture" of statistical reasoning, such that the structure of the knowledge to be acquired becomes highly influenced by the authentic activity structures of the culture. Knowledge and learning become embodied as internalized activity structures, acquired through guided participation in socially relevant practice. In contrast with the symbolic view, knowledge need not begin with deliberate symbolic encoding of statistics as an abstract mathematical domain. In fact, because the knowledge required for intelligent activity does not reside solely within the learner's head, but is largely present in the supporting social environment and tools (statistical procedures, co-learners, computing and measuring devices) of the culture, much knowledge required for successful performance is not symbolically encoded at all. Further, because bodily activity within social and physical systems results in perceptual, intuitive attunements of which learners may not be explicitly aware, only part of what is internalized by the learner is encoded explicitly and hence measurable through traditional verbal means. Thus, all assessments, even in early phases of learning, are based on performances in social contexts, in which meaningful activity is performed in authentic social and physical settings that provide developmentally appropriate support for the work required.

With respect to the issue of when and how to introduce real-world complexity into statistics instruction, the situated cognition and symbolic views can be contrasted in important ways. The symbolic processing view is essentially dualistic: it emphasizes the need for building explicit mental representations of the mathematical domain as a family of conceptual lenses for seeing a simplified world at first, but later for helping view real-world complexity. Statistical representations are built up gradually and systematically, with complexity being introduced at advanced stages of learning. Complex domain representations and flexible habits of mind are fostered through activities in which students flexibly assemble and reassemble holistic interpretations of situations, or "cases." Statistics and other conceptual ideas are viewed as interacting lenses for helping students "see" complexity. In contrast, the situated cognition view sees students as integral parts of a complex environment. It emphasizes immersion in authentic activity within supportive cultural contexts, long before any domain knowledge that might support individual activity is acquired. Complex cognition is viewed as a process of gradual attunement and acquiring greater connection to a complex social and physical system in which the student gradually increases participation and leadership. Performance in authentic social contexts is key, and although some symbolic encoding of statistical content occurs over time, such individual encodings are only part of the community's distributed knowledge.

TWO INSTRUCTIONAL EXPLORATIONS

Although the symbolic perspective was not entirely discarded, it was the situated cognition perspective that most strongly influenced the thinking of Derry and Levin (Derry, Osana, Levin & Jones, 1998; Derry, Levin, Osana, Jones, & Peterson, 2000) when we embarked upon two explorations in innovative teaching of statistical reasoning. These were not formal courses in statistics *per se*, and they did not emphasize computation. Our goal was to develop students' abilities to reason flexibly and argue with statistical concepts in the context of complex, real-world situations. We attempted to accomplish this by basing classroom activities on authentic problem-solving tasks. With minimal direct instruction we immersed students in guided problem solving and attempted to establish norms consistent with a learning community framework. In our two projects, we taught middle school students and pre-service teachers using two types of instructional methods: (a) immersive simulations of authentic professional activities requiring presentation and critique of statistical arguments about important societal issues (which were termed "Sitsims," or situated simulations); and (b) reflective discussions with critique and interpretation of popular media presentations (e.g., movies) that dealt with complex statistical and social issues. These projects, a description of the materials and assessments we used, and the lessons we learned from them will be summarized briefly.

A MIDDLE SCHOOL UNIT IN STATISTICAL LITERACY

The following description is taken from Derry et al. (1998): "The Vitamin Wars" project was conducted with three middle-school classrooms and was a three-week instructional unit adapted to three different subject-matter contexts (science, social studies, and mathematics). In all contexts, the purpose of the unit was to improve students' abilities to think and reason statistically about real-world issues. Prior to instruction, teachers and researchers collaborated in planning and in-service training sessions. During the three weeks of instruction, students watched and discussed (with researchers and teachers as guides) a popular movie, *Lorenzo's Oil*, an emotional drama that focuses on such issues as ethics in medical research, experimental control, informed single-case observations versus randomized clinical trials, and governmental regulation of the scientific and lay community. Following the movie, students carried out lengthy mock legislative hearings dealing with government regulation of the dietary supplement and vitamin industry, an issue that was present in the news of the time and that shared several parallels with the story line in *Lorenzo's Oil*. During the three weeks of instruction, teachers and researchers made several short presentations on topics of thinking, presenting arguments, and statistical decision making. However, there were no homework assignments pertaining to statistical reasoning and a relatively small amount of time (one day) was devoted to direct instruction on statistics. Nevertheless, it was intended and expected that students' statistical reasoning would measurably improve because of participation in mentored discussions and activities.

To evaluate students' performance gains in their ability to reason statistically in everyday (and emotional) contexts, two assessment tasks and accompanying scoring rubrics were devised. The assessment tasks presented students with dialogue from fictional dramatic trials, inspired by popular television shows, in which witnesses and lawyers presented evidence, arguments and counterarguments to support their cases. For each trial, the student's task was to analyze the dialogue, draw a conclusion, and support the conclusion with reasons. Students' responses were judged by scorers trained to use a rubric based on current research on statistical reasoning. Taken together, the rubric and framework amount to a normative model for good thinking that both prescribes what is valued in students' reasoning and proscribes what is not. The tasks and scoring procedures were designed as general assessment tools to be shared with teachers and researchers. They are described in detail in Derry et al. (1998)

"The Vitamin Wars" implementation was much more successful in some classrooms compared to others. In one class, arguments became emotional and unfolded in unexpected ways. There were differences among classrooms in quality of mentorship, classroom discipline, and student attitudes. Moreover, it was not always possible to extract statistical issues from their interwoven contexts and foreground them as learning issues. In fact, students reported learning more about vitamins than statistics. Nevertheless, based on carefully designed assessments as described above, when student arguments were judged against a normative model of good thinking grounded in cognitive research on statistical reasoning, students in the classes that participated performed statistically better than students in comparison classes that did not.

AN INNOVATIVE COLLEGE COURSE IN STATISTICAL REASONING

Our second exploration (Derry et al., 2000) was an innovative course in statistical reasoning for a college undergraduate "fellows" program, a community-based cohort within the School of Education. The student body was very heterogeneous, both demographically and in terms of students' academic abilities and interests. Direct instruction was employed only intermittently throughout the course, with the instructional emphasis instead on guided inquiry learning in problem contexts. We tried to develop students' understanding of a few important statistical principles and concepts (rather than focus on broad content coverage or computation). Most course time was devoted to group problem solving and evidential argumentation.

The types of problems presented to the students were ill structured and tied to authentic scenarios. In one activity, students learned about scientific experimentation and measurement in the context of conducting carefully controlled experiments with fast-growing plants (Wisconsin "Fast Plants"), followed by a presentation and defense of their findings at a class "conference." In another activity, students assumed the roles of parents and teachers of grade-school students of different abilities and ethnicities. At a simulated town meeting, students argued either in support of or against a proposal to re-establish ability tracking within the town's schools.

In conjunction with the second offering of this course (the first offering had several operational problems and was deemed "uncussessful"), evidence of student growth was obtained from pre- and post-course interviews designed to assess students' ability to reason with statistical evidence from everyday sources. Both quantitative and qualitative analyses indicated that students made selected gains in their ability to reason statistically. The assessments included structured interview tasks used to assess students' ability to judge the credibility of arguments found in naturalistic website and newspaper reports of uncontrolled studies. In scoring students' responses to this task, we awarded points on the basis of criteria that reflected students' ability: (1) to reason about evidence; and (2) to recognize that there were shortcomings in the evidence presented. For example, one article claimed that of 90 troubled teens who watched an MTV show on alternative religions, 60 left home, so the show should be cancelled. The article claimed a relationship between watching the show and leaving home. Maximum points were awarded if the interviewee verbalized (in any number of ways) that: (1) there was insufficient evidence to warrant cancellation of the program; (2) the missing evidence included data pertaining to the home-leaving behavior of troubled teens who did not watch the program; and (3) even if a positive correlation existed, this would not justify a conclusion of causality, as the article asserted.

In many ways, this course was a success. Based on both empirical and qualitative analyses (though with some uncertainty due to the use of a single-group pretest-posttest research design),

the course did enhance students' ability to activate and use course-related statistical ideas to critique and reason about research in the news. In other ways, the course was less successful. Qualitative analyses of selected interview data revealed a pervasive confusion in students' thinking that may have been exacerbated by course design. Students' affective reactions to the course varied widely. Despite the community rhetoric and design of the program, extensive collaborative was viewed by some students as a burden. Both students and faculty felt overwhelmed by the amount and intensity of work involved.

In sum, our two explorations in adapting situated cognition theory to classroom instruction met with both success and failure. Although students achieved desirable instructional goals, and although the approaches seemed promising in many respects and might have been perfected and sustained with additional effort, neither implementation was continued by the developers or ever fully appropriated by the programs in which they were originally implemented. In fact, these models were neither particularly feasible nor affordable in the particular institutional/educational contexts in which they were tried, and were not universally and enthusiastically embraced by the students and teachers who participated, including many who worked very hard to insure success of the projects. Thus, the search continues for effective and *practical* instructional methods that address issues of contextual complexity and adaptive flexibility in statistical reasoning (and other subject matter) instruction.

CURRENT DIRECTIONS WITH VIDEO TECHNOLOGY

Spiro (in press) has argued that instruction in complex domains must "expand perception," developing students' capabilities for complex understanding and flexible, adaptive application of knowledge and prior experience in new contexts. He believes that video technologies will enable such qualitatively different kinds of learning. At the center of his argument is the idea that processes of compacting, elaborating, sequencing, and editing digital video will enable radical acceleration in the acquisition of experience. Features of video learning environments that may be capable of producing qualitative shifts in underlying habits of mind (ways of thinking, worldviews, prefigurative lenses) include the concepts of nonlinearity and multiplicity. An example of a nonlinear sequencing "treatment" would be using video technology to criss-cross a movie or other video presentation in order to experience many varied cases of similar and/or contrasting ideas rapidly, as such ideas present themselves in action.

In a paper in preparation for AERA 2002, Derry discusses the potential of such technologies from a decidedly non-dualist perspective that fuses Eastern philosophy with Gibsonian perceptual and situated cognition theory. Her analysis emphasizes the importance of "intuitive conceptual ecological intelligence," a non-dualist form of mindful practice, and considers whether this form of learning is possible with video-based technologies such as those described by Spiro.

These are empirical questions currently being examined experimentally. For example, Spiro and Derry (working with John Stampen and Mary Leonard) are addressing the question of whether video-based instruction using nonlinear sequencing and computer-enhanced perceptual overlays that enhance thematic interactions and contrasts, can help accelerate college students' acquisition of ability to reason expertly when statistical concepts interact in complex ways with scientific, ethical, and aesthetic themes. For example, in one experiment, a group of students viewed selected segments of the movie, "Twelve Angry Men," in sequences that emphasized differences between the stochastic and non-stochastic reasoning that occurs in the movie. In a more complex experimental presentation, a second group viewed the same sequencing treatment, but in a version that employs perceptual overlays to emphasize interacting themes, such as reliability of eye-witness testimony and psychological impact of bias on thinking. Our approach to video enhancement employs color, sound, and symbolic imagery to emphasize statistical themes and complex thematic interactions in movies and instructional videos representing complex, reality-based situations.

These data have not yet been analyzed, but we predict that such instructional treatments will substantially enhance students' subsequent "transfer" experiences, measured in terms of their "in-the-moment" awareness of complexity, ability to discriminate particular instances of statistical concepts rapidly, and ability to respond to the presence of interacting themes. The

ICOTS conference presentation will include a demonstration of this approach and will summarize available experimental data. It will also discuss assessment techniques we have devised to measure transfer of statistical reasoning conceptualized as both situated cognitive flexibility and intuitive conceptual ecological intelligence.

ACKNOWLEDGEMENT

This research was made possible by grants from the Joyce Foundation and the National Science Foundation, Division of Undergraduate Education.

REFERENCES

Derry, S.J., Levin, J.R., Osana, H., Jones, M.S, & Peterson, M. (2000). Fostering students' statistical and scientific thinking: Lessons learned from an innovative college course. *American Educational Research Journal, 37,* 747-773.

Derry, S., Levin, J.R., & Schauble, L. (1995). Stimulating statistical thinking through situated simulations. *Teaching of Psychology*, *22*, 51-57.

Derry, S.J., Osana, H., Levin, J.R., & Jones, M.S. (1998). Developing middle school students' statistical reasoning through simulation gaming. In S.P. Lajoie (Ed.), *Reflections on statistics: Agendas for learning, teaching and assessment in K-12* (pp. 175-198). Mahwah, NJ: Erlbaum.

Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1-55.

Greeno, J.G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5-17.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge MA: MIT Press.

Lave, J. (1988). *Cognition in practice*: Mind mathematics and culture. New York: Cambridge University Press.

Nardi, B.A. (Ed.). (1996). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, MA: MIT Press.

Spiro, R.J. (in press). Experience acceleration with digital video cases: New directions in Cognitive Flexibility Theory. *Educational Technology*.

Spiro, R.J., Feltovich, P.J., & Coulson, R.L. (1992). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains In T.M. Duffy and D.H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 57-75). Mahwah, NJ: Erlbaum.

Wertsch, J.V. (1998). *Mind as action*. New York: Oxford University Press.