(2004). In Forbus, K., Gentner, D., & Regier, T. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 987-992). Mahwah, NJ: Erlbaum.

Numerically-Driven Inferencing in Instruction: The Relatively Broad Transfer of Estimation Skills

Edward L. Munnich (munnich@berkeley.edu)
Michael A. Ranney (ranney@cogsci.berkeley.edu)
Daniel M. Appel (dappel@berkeley.edu)
University of California, Graduate School of Education, 4533 Tolman Hall, Berkeley, CA 94720-1670

Abstract

What is the current U.S. immigration rate? Policy-makers, voters, and consumers should have a sense of quantities of this kind in order to help shape effective policies, and schools must prepare students for such roles. We examine the Numerically-Driven Inferencing paradigm (NDI), using a method in which participants: *Estimate* policy-relevant quantities, state *Preferences* for these, receive actual quantities as feedback to *Incorporate*, and offer preferences again to exhibit any policy *Changes* (EPIC). Past work has generally suggested rather poor estimation of such base rates, but there is potential for improvement as one carries out many estimates over various issues, and perhaps a benefit for taking a more analytic approach to estimation. Here we consider whether one can improve estimation skills broadly by using multiple perspectives in estimation problems, and by working out of conflicts that arise among multiple, locally coherent, numerical understandings. Using an NDI curriculum that emphasized disconfirmation, we found that estimation improved across a wide variety of questions.

What is the current annual U.S. immigration rate (including both legal and illegal immigration)? Please take a moment to estimate this quantity, and reflect on the kinds of skills you used to generate your estimate. One might assume that those who know about immigration issues are good at estimating immigration rates, while those who know about environmental issues are good at estimating per capita garbage production, but that there is no general skill for estimating across content domains. Research on estimation suggests that people can improve the accuracy of estimates in a variety of ways, including using category information (e.g., Huttenlocher, Hedges, & Prohaska, 1988), or learning relevant "seed" numbers (e.g., Brown & Siegler, 2001), but there is no indication that such benefits transfer broadly to estimation over a wide variety of quantities, to say nothing of problem solving skills more generally. However, we suggest that in domains ranging from estimation to physics problem-solving, it is important to learn to seek alternatives to initial conceptions of problems, which brings the possibility of disconfirming hypotheses. The potential value of such a strategy is illustrated Johnson-Laird and Hasson (2003), who have found that when some premises are consistent with an invalid conclusion, counterexamples are useful in rejecting the conclusion. The focus of the present paper is on the extent to which analytic estimation skills can transfer broadly, so that people might improve their estimates for quantities across a broad range of issues without specific instruction on those issues.

Theoretical Framework

This project builds on the Numerically-Driven Inferencing paradigm (NDI; Ranney, Cheng, Nelson, & Garcia de Osuna, 2001), which examines how understandings of relevant base rate information (e.g., the present U.S. immigration rate) affects people's attitudes on public policy issues (e.g., given the immigration rate, what would you *prefer* that rate to be?). With NDI's methods, people need not be asked whether they are for or against a particular issue, but rather what they would prefer the *numbers* to be. Indeed, it is not uncommon that those who consider themselves to be in favor of reducing immigration (e.g., believing the current base rate of a policy-relevant quantity to be 10%, one might prefer 5%) have more in common than they realize with those who claim to favor an increase (e.g., believing the rate to be 1%, but sharing a preference for 5%). However, if such people were only asked the extent to which they favor or oppose an issue, they would appear to be at odds. In contrast, NDI asserts that qualitative attitudes have some—albeit not necessarily direct—relationships with relevant quantities, and aims to explore the nature of the relationships. By focusing on numerical concepts, NDI can shed light on how these concepts interact with people's initial attitudes, and the extent to which learning actual values shapes subsequent attitudes: Do we maintain preferences for the same *absolute* rates, or for the same *proportions* relative to actual rates? To what extent do we shift our policy stances after surprising feedback (Munnich, Ranney, Nelson, Garcia de Osuna, & Brazil, 2003)?

NDI builds on research in many fields, such as attitude, conceptual change, mental models, and judgment and decision-making (although NDI deals directly with base rates—not through Bayesian analyses). In particular, NDI has drawn on work in scientific conceptual change including the Theory of Explanatory Coherence (TEC; Ranney & Thagard, 1988; Thagard, 1989), which describes change as spawned by incoherence and conflicts among ideas, such that people try to revise their beliefs to increase global coherence. In an illustration of this, Ranney, Schank, Mosmann, and Montoya (1993; based on a misconception noted by Keysar, 1990) found that most participants initially believed that Berlin lay *on* the East/West German border, but revised their beliefs as they incrementally received information that could be used to disconfirm "on-border" hypotheses (e.g., they were told/reminded of the Berlin airlift, the Western Allies' agreement to halt their troops far west of Berlin, Berlin's location within united Germany, and northern and southern ends of the border). With each successive piece of evidence, participants moved toward a more accurate view of Berlin's location relative to the border, suggesting that they modified their belief networks to maintain coherence in the face of the new information.

According to TEC, evidence that is critical, germane, and credible carries considerable weight in our belief systems. Within NDI, we seek to understand when and how a particular kind of evidence that meets these criteria—numerical propositions—can catalyze knowledge-transforming effects. NDI asserts that estimates and numerical preferences are outputs of our belief systems—the tips of a "reasoning iceberg." One's understanding of an issue may be thought of as a network of ideas connected by personal experiences, media, religion, etc. When asked to estimate an immigration rate, few can simply recall it. Instead one activates various understandings about immigration that shape the estimate. Likewise, numerical preference is an output from an extensive belief network that lies below the surface of overt response. For example, one might believe the assumed immigration rate to be acceptable and simply reiterate one's estimate as one's preference (a status quo policy). However, if later surprised by the actual immigration rate, one's sense of reality is challenged, and one might come to the conclusion that prior reasoning was incorrect or incomplete.

In this conception, the iceberg's "bulk"—the belief network from which estimates and numerical preferences emerge—may be transformed by the impact of feedback. As such, NDI can offer rich, quantitative findings to cognitive scientists concerned with the dynamics of belief networks. In this paper, we consider *curricula* based on NDI, designed to facilitate the recruitment of multiple, locally coherent understandings that can mutually constrain one another. Just as feedback that conflicts with one's numerical understanding might lead to a transformation, when one spontaneously seeks to disconfirm one's own numerical hypotheses by bringing alternative numerical notions to bear, it may lead to revisions that bring one's belief network into closer alignment with facts of the world. Such a transition would be evidenced by improved estimation across a wide range of issues.

NDI Findings That Frame the Issues

To address NDI, Ranney and colleagues developed a variety of methods, including EPIC (Estimate-Prefer-Incorporate-Change), which is used in this paper: (1) Participants *estimate* a quantity that is relevant to an issue, as you did for the U.S. immigration rate at the beginning of this paper. (2) Participants indicate what they *prefer* the quantity to be; to familiarize yourself, please write down what you would prefer the U.S. immigration rate to be (including both legal and illegal immigration). (3) Participants receive correct base rate feedback to *incorporate*; now, please look at the actual immigration rate in the footnote below.[1] Finally, (4) participants indicate again what they prefer the quantity to be; has your preference *changed* now that you know the actual number? We have found that, to the extent feedback is surprising, it generally leads to nontrivial belief revision. So far, research on estimates within NDI has focused on a rather short period of time, but an obvious extension of this work is to consider (a) whether estimation skills can *improve* with targeted interventions, and (b) the extent to which there may be broad transfer.

Illustrations of the kinds of alternative conceptions that people can have comes from Munnich et al. (2003), who reported differential patterns of estimation for the same underlying question: One group was asked to estimate the number of abortions in the U.S. *per million live births,* while a second group drawn from the same undergraduate class estimated the number of abortions in the U.S. *per million fertile women* each year. The results showed a striking contrast in numerical understanding, depending on how the question was framed: For the *per-women* question the median response (10,000) was *half* the correct answer at that time, but for the *per births* question, the median estimate (10,000 as well, coincidentally) was *33.5 times* too low at the time the study was conducted.[2] Could people perhaps improve their estimates of abortions per live births by considering how many abortions there are per fertile women? More broadly, what might happen when people bring together alternative conceptions and resolve conflicts on their own, without external feedback? To address this issue, McGlothlen (2003) interviewed high school students as they produced estimates and numerical preferences for a variety of issues, and reported on their online reasoning processes. She coded responses as analytic—containing relevant numerical information and constraints—or holistic—based on a feeling or general sense of the issue. McGlothlen found that estimates reached through an analytic process were significantly more accurate than those reached through a holistic process. This leads us to the following hypothesis:

> An analytic approach invokes multiple locally coherent numerical representations that provide mutual constraints among themselves, leading to more refined, more globally coherent, and hence more accurate estimates than would be observed if only one representation were invoked.

To discover whether there is a causal relationship between invoking multiple representations and accuracy, we might manipulate the degree to which people take an analytic approach. Below, we discuss an experiment in which the analytic process is explicitly emphasized in the instruction given to one group of students, and the accuracy of this

[1] The U.S. Census Bureau reports that the annual U.S. immigration rate, including legal and illegal immigrants, is 0.4%.

[2] Garcia de Osuna, Ranney, and Nelson (2004) observed a median of 5,000, sixty-seven times too low.

group's estimates, pre- and post-instruction, is compared to a parallel group who received no such instruction. If those taught analytic strategies show greater estimation accuracy, it would provide causal evidence for the benefit of an analytic approach.

Previous Curricular Interventions

In several recent studies, our group has observed estimation accuracy benefits, arising from practice with forming estimates and generating preferences. These activities are unusual for math or science classes, for which problems are generally solvable in straightforward ways by applying formulas and principles. Our curricula illustrate the utility of mathematical and scientific reasoning through our use of problems about issues that students find interesting. We ask students for societally-relevant opinions, which is virtually unheard of in math classes. These factors motivate students in ways that standard curricula may not, and shows benefits for estimation ability with relatively little practice.

In one such intervention, Curley (2003) and Howard (2003) gave fifth-grade science camp students standard physics labs about the stopping distances of vehicles. An Experimental class received NDI problems to frame the labs, while a Control class did not. Both classes took a pretest with estimation and preference problems, then a posttest with a different set of items three days later.[3] In this between-subjects design, Curley and Howard observed improvement in estimation accuracy for *both* classes on items about U.S. household income and the number of alcohol-related automobile crashes. Notably, the only NDI experience that the Control class received was during the pretest, suggesting that exposure to such items alone might be sufficient to improve estimation abilities.

In a later study, Juan (2003) found similar effects among eighth-grade Algebra students. Her Experimental class received one NDI problem per day for three days, followed by graphing activities and class discussions on estimates and preferences. In contrast, the Control class received standard algebra instruction. All students took a pretest and a posttest, in which they estimated twelve quantities (six per test). Questions dealt with issues such as California's population and teachers' salaries. On each test, students estimated and offered preferences before and after feedback for two items, and simply estimated for the remaining four items. Experimental students showed a significant overall gain between pre- and posttest estimates, while the Control class showed only a marginally significant improvement. However, an additional sign test between groups showed no advantage for the Experimental class.

The studies discussed up to this point showed minimal benefit for the curricula themselves—while both Experimental classes improved, the Control classes may have also benefited just from their pretest experience with NDI. This raised the possibility that one may improve estimation merely by working on NDI problems. Before drawing this conclusion, we carried out the following experiment, which lasted much longer than past interventions, and focused not on content areas like physics or college costs, but on analytic techniques aimed at improving students' general estimation abilities.

A Focal Experiment

Method

Two high school geometry classes participated, each with 27 students. Both classes received a normal geometry curriculum, but the Experimental class spent 12% of their time over a ten-week period on activities centered on six NDI questions. Activities included discussions and written reflection aimed at promoting the analytic responses that McGlothlen (2003) found to correlate with successful estimation. In addition, explicit connections were made between logical argumentation about issue-relevant quantities and the argumentation required in geometric proof. Due to limited space, we omit discussions of possible benefits regarding student motivation and the transfer of argumentation skills from NDI to geometric problems.

Table 1: Pretest, Intervention, and Posttest Questions (Group B received the pre- and posttest in reversed order)

Pretest (Group A)	**Intervention**	**Posttest** (Group A)
Average US age	CA population	Cars per driver
Athlete salary	College vs. H.S. grad earnings	College degrees (%)
College cost	H.S. dropout rate	Homes-with-computers
Miles driven/Year	Athlete salaries by gender	% Female teachers
Commute time	Poverty line	Garbage per person
Incarceration rate	US oil imports	Hours of sleep
Soda calories		Inflation
Homes-with-TVs		Voting percentage
US population		Teacher salary
		Car price

On Thursday of each week, students generated estimates and preferences for a given quantity as homework (see Munnich et al., 2003, for examples of how such items are worded). In class on Friday, they discussed their estimates in groups and generated a group estimate, in which they: (a) provided a consensus estimate, (b) explained their rationale for that number, (c) provided rationales for a number considerably higher than their estimate, and for a number considerably lower than their estimate. This was followed by a short class discussion to acquaint students with alternative approaches that their classmates had taken. Between hearing classmates' arguments and generating rationales for estimates and preferences other than their own, students were encouraged to engage in problems analytically, considering the strengths and weaknesses of various constraints that might be placed on the estimate.

The following Monday, students' original estimates and

[3] When we ask for preferences, the objective is to assess how students' numerical understandings affect their preferences, not to make a normative assessment.

preferences were returned, along with the actual number as feedback. From Monday to Tuesday, they generated final preferences, based on the feedback and any insights they gleaned from group discussions. Finally, on Tuesday, students discussed their preferences in groups and generated arguments that might be used by one who preferred (a) decreasing the quantity, (b) maintaining the status quo, and (c) increasing the quantity. As with estimation discussions, this was followed by a whole class discussion on preferences.

To measure the intervention's effects, both Experimental and Control classes were given ten NDI items as a pretest, and then, ten weeks later, ten different NDI questions as a posttest (an immigration rate item was excluded when it became clear that responses were bizarrely high in many cases; students also reported numerous misinterpretations). Questions were counterbalanced so that the items that half of each class saw on the pretest (Group A in Table 1) appeared on the posttest for the other half of each class, and vice versa. Students were asked to generate estimates and preferences, and were then handed a separate sheet of paper with the actual quantity, which also elicited a Likert surprise rating and their final preferences. Students received two items per day for five days (as past studies indicated that fatigue sets in when students receive many items on a single day). Each of the ten problems was presented in the same order to all students, minimizing any benefit for discussing items with classmates in unintended ways.

Results and Discussion

All estimates (for both classes, pre- and posttest) were ranked by proximity to the actual value for each item, in order to put data from questions with different scales onto one common scale (i.e., accuracy rankings). Between-group analyses on the rankings assessed whether there were differences between the classes, and whether each class improved from pre- to posttest. Mann-Whitney tests showed no reliable pretest difference between Experimental and Control classes (z=1.04, n.s.). On the posttest, however, the Experimental class estimated reliably more accurately than Controls (z=3.29, p<.001). Further, while there was no difference among Controls on pre- and posttests (z=0.41, n.s.), Experimentals showed a significant improvement (z=2.74, p=.003). These effects indicate that the intervention led to improved estimation of novel quantities (i.e., transfer).

To explore the effect's loci for the Experimental class, planned Mann-Whitney comparisons were performed separately on each item. Participants improved significantly on three items (U.S. population, z=2.49; cars per driver, z=1.97; hours of sleep, z=1.96; ps<.05), and marginally on three other items (college cost, z=1.48; teacher salary, z=1.36; miles driven/year, z=1.19; ps<.10). Among these items, we see patterns of both near and relatively far transfer from the intervention. The only one of these items that was directly related to one of the intervention questions was that on U.S. population (i.e., related to an question on California's population in the intervention). The other items range from those that seem to have only an indirect relationship with intervention items (e.g., "teacher's salary" may be related to "H.S. vs. college grad earnings," although teachers' incomes are closer to the incomes of high school graduates than to those of other college graduates), to items that have no obvious relationship with the intervention problems (e.g., hours of sleep the average person gets).

The results point to benefits from an intervention focused on analytic approaches to estimation. Looking more closely, we found both transfer among highly similar questions, as well as the relatively far transfer of general estimation skills to seemingly unrelated quantities. These findings are in line with the hypothesis that multiple numerical representations provide constraints on one another and can lead to more globally coherent estimates. The difference between the classes was that, although both had experience with estimation and giving preferences on the pretest, the Experimental class received a curriculum that engaged them in discussions of multiple perspectives in estimation and numerical preference. These results are rather surprising if one believes that estimation ability is not a broadly transferable skill. Given the variety of topics covered by items on the pre- and posttests, it is unlikely that the Experimental class could have learned the vast array of new facts about the world necessary to drive observed improvements. Rather, they appeared to use their extant numerical knowledge about the world more constructively than before.

Why did students improve broadly in estimation? McGlothlen (2003) found that those who invoked a richer repertoire of analytic tools estimated better than those who used a more holistic/feeling approach. With this in mind, one explanation for the Experimental students' improvement is that the intervention moved them towards a more comprehensive approach to estimation. Our lab is conducting ongoing research to examine other possible causes for students' improved performance. One such possibility is that Experimental students enjoyed the curriculum and were simply more motivated than Controls to complete the posttest exercises. If this were the source of improvement, we would expect Experimental students to spend more time on solutions, and report more interest in the task, but we would not expect to see greater richness in the strategies they employ. Another possibility is that Experimental students benefited from the recency of their practice with estimation during the intervention. If this caused the difference between the groups, then, again, although Experimentals gave more accurate estimates, we would not expect subsequent analyses to show that they used richer strategies than Controls. We cannot reject this possibility at present, but we find it highly unlikely, as estimation curricula are generally quite taxing for students: Our prior results indicate that without a particularly engaging curriculum, more recent practice leads to a performance decrement, presumably due to fatigue.

General Discussion

Many propositions inform our social preferences (e.g., Ranney & Schank, 1998), but to illustrate the role played by numbers, consider whether your immigration preference would change if you made an estimate that was highly inaccurate. What sort of numerical feedback might call your assumptions into question, leading you to a different preference? Preferences are central to human cognition, and *numerical* preferences provide useful sources of evidence regarding conceptual change. Numerical preference represents a concrete way in which mathematics is relevant to our lives, and contributes to discussions of quantitative literacy in math education. By ignoring base rates, voters or political candidates may take stands that conflict with what they would otherwise prefer. Of course, some people take *absolute stances* on particular issues, such as completely eliminating abortion; as such, they imply that the numbers are irrelevant to their beliefs on the issue, and we would not expect them to change their preferences after feedback very often (Ranney et al., 2001). For those who indicated nonzero preferences, Munnich et al. (2003) found two main patterns: First, those who were less surprised by base rates generally *proportionately rescaled* their preferences—those who preferred halving the abortion rate initially, still preferred halving the actual rate when it was revealed. This suggests the base rate was belief-relevant, but that it did not inspire dramatic revisions of belief networks. Second, those who were more surprised by feedback showed *policy shifs*—accommodative belief revisions—for instance, those who preferred halving the abortion rate initially, but were surprised by the actual rate, indicated final preferences notably more or less than half of that rate (see Garcia de Osuna, Ranney, & Nelson, 2004, for more discussion of the qualitative nature of such shifts).

Even when considering the same issue, people can arrive at markedly different estimates and policies, depending on how the issue is framed (cf. Schwarz, 1999). As noted earlier, when Munnich et al. (2003) asked for the number of abortions per live births, the median response was 33.5 times too high. With their estimates so far off, what happened with these people's preferences? After feedback, they showed a policy shift—a 64% more reductive policy than they had initially indicated. By contrast, when participants estimated the number of abortions per fertile women, the median estimate was much closer—half the actual number. Rather than shift policies, for the fertile-women variant, participants merely rescaled their preferences to adjust to their new understanding of the number. In other words, when a quantity (e.g., the number of abortions performed each year) is framed in different ways, people show vastly different abilities in estimating the quantity, and this strongly affects their preferences after they learn the actual numbers.

Our hypothesis in this paper focused on the estimation side of NDI, but there are also implications for preference. When an intervention successfully fosters estimation ability, what might we predict, regarding people's preferences? One possibility is that as estimates improve, feedback-driven surprise will abate, and policies will stabilize, producing *less* subsequent policy shift. However, it is also possible that when estimates improve, people might become more sensitive to numbers, and attach more importance to small errors, yielding *more* policy shift**.** Note that while some of our past studies showed framing effects, they did not focus on people who recruited relevant facts to frame issues in different ways for *themselves*. When an individual integrates multiple constraints without prompting, the effects may be quite different than what we see with more passive participants. In analyses of the preference data gathered along with the estimation data reported above, we find support for both possibilities—while some participants appear to shift policies less after intervention, others seem to be more sensitive to small changes in numbers, and thus shift less. In aggregate these effects largely cancel each other out. A more in-depth analysis of individuals' changes in estimation ability, surprise levels, and preferences is being conducted to determine how each phenomenon contributes to the overall pattern of results. One possible benefit of this research may be in teaching people to construct policies that are less susceptible to rhetoric. That is, as people adopt more analytic strategies (assuming this is why estimates improve in our curricula), when they hear a quantity in advertisements or on the news, they might think of the issue several different ways and generate a preference that is constrained by other numbers they have considered.

Beyond transfer to tasks involving numerical understanding, what other forms of transfer might exist? NDI problems can be considered examples of "Fermi Problems," after the physicist who famously posed queries such as "How many piano tuners are there in Chicago?" Few, if any, can simply *recall* answers to Fermi questions, but through successive approximations and drawing on other known quantities, one can approach the correct answer. When Fermi questions are posed—often by potential employers or as classroom exercises—the implicit assumption is that one's answers are indicative of general analytic ability and creativity in problem solving. It is not difficult to imagine that NDI-type interventions might benefit reasoning about the location of Berlin relative to the former East-West border: With analytic techniques, one could do for oneself what Ranney et al. (1993) did for their participants—foster the integration of multiple, mutually constraining, perspectives into a solution.

More broadly, was Fermi's physics problem-solving ability related to his ability to estimate the number of piano tuners in Chicago? Much of the problem solving literature indicates little *general* transfer of problem solving skill across divergent domains (Singley & Anderson, 1989), so this may initially seem unlikely. However, we note that one of Ranney and Thagard's (1988) participants ("Pat") reached a more sophisticated understanding of projectile motion through the same kinds of processes that we have argued to underlie strong numerical reasoning. Pat initially believed that a ball dropped by a walking person would fall straight to the ground. Later on in her verbal protocol, she contemplated the motion of a ball thrown obliquely upwards, and decided that it would follow an arc-shaped trajectory. Upon realizing this, it occurred to her that, from the zenith of its trajectory to the ground, the ball would descend analogously to a ball dropped while walking. Accordingly, she concluded that the two trajectories must have a similar arc-shape. Pat thus revised her view of the path of the dropped ball to a (more accurate) curved trajectory. This example illustrates the potential generality of the analytic skills that are useful in numerical reasoning: In both physics and estimation, we seem to benefit from using alternative representations, and then resolving conflicts among them. The degree to which one skill transfers to another is a worthy topic for future research.

Summary

It is critical that citizens and consumers be able to make decisions on numerically laden issues. We found that people can improve their numerical understandings through activities emphasizing the consideration of multiple perspectives and the integration of mutual constraints, and we discussed possible implications of such findings for individuals' policy stances. We propose that improvements in estimation abilities arose from an analytic approach that this intervention cultivated, leading students to seek evidence that might disconfirm their initial hunches. Such an approach might have value beyond the numerical and policy realms, with respect to more general reasoning and problem solving skills. In these ways, classroom interventions that test aspects of the emerging theory around the Numerically-Driven Inferencing paradigm have the potential to answer questions of fundamental interest to both cognitive science and society.

Acknowledgements

We thank the students and teachers who participated in this study, as well as Mandy Bachman, Morgan Curley, Christine Diehl, Barbara Ditman, Karen Draney, Sujata Ganpule, Cirila Howard, Josette Juan, Florian Kaiser, Lilian McGlothlen, Michelle Million, Janek Nelson, Luke Rinne, Mirian Song, Mark Wilson, and the UCB Reasoning Group for their helpful comments. This work was funded by a UCB faculty research grant and an AERA/IES Postdoctoral Fellowship.

References

Brown, N. & Siegler, R. (2001). Seeds aren't anchors. *Memory & Cognition, 49,* 405-412.

Curley, M. (2003). *An EPIC curriculum: An examination of a curriculum to promote reasoning for conceptual change.* Unpublished Master's Project, University of California, Berkeley.

Garcia de Osuna, J., Ranney, M., & Nelson, J. (2004). Qualitative & quantitative effects of surprise: (Mis)estimates, rationales, & feedback-Induced preference changes while considering abortion. *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Howard, C. (2003). *An EPIC Quest for Justification: The Effects of a Numerically-Based Intervention on Students' Estimates and their Justifications.* Unpublished Master's Project, University of California, Berkeley.

Huttenlocher, J., Hedges, L., & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review, 95*, 471-488.

Juan, J. (2003). *An EPIC curriculum with attitude: The extension of a novel curriculum involving estimations and attitudes about higher education.* Unpublished Master's Project, University of California, Berkeley.

Johnson-Laird, P. & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory & Cognition, 31(7),* 1105-1113.

Keysar, B. (1990). *East meets west at the Berlin wall: Mental maps and the changing world order*. Unpublished data.

McGlothlen, L. (2003). *High school students reasoning with numbers: Interviews using the estimate, predict, incorporate, and change (EPIC) method*. Unpublished Master's Project, University of California, Berkeley.

Munnich, E.., Ranney, M., Nelson, J., Garcia de Osuna, J., and Brazil, N. (2003). Policy shift through Numerically-Driven Inferencing: An EPIC experiment about when base rates matter. *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society.* (pp. 834-839). Mahwah, NJ: Erlbaum.

Ranney, M., Cheng, F., Nelson, J., and Garcia de Osuna, J. (2001). *Numerically driven inferencing: A new paradigm for examining judgments, decisions, and policies involving base rates.* Paper presented at the Annual Meeting of the Society for Judgment & Decision Making.

Ranney, M. and Schank, P. (1998). Toward an integration of the social and the scientific: Observing, modeling, and promoting the explanatory coherence of reasoning. In S. Read & L. Miller (Eds.), *Connectionist models of social reasoning and social behavior*. Mahwah, NJ: Erlbaum.

Ranney, M., Schank, P., Mosmann, A., & Montoya, G. (1993). Dynamic explanatory coherence with competing beliefs: Locally coherent reasoning and a proposed treatment. In T.-W. Chan (Ed.), *Proceedings of the International Conference on Computers in Education: Applications of Intelligent Computer Technologies* (pp. 101-106).

Ranney, M., & Thagard, P. (1988). Explanatory coherence and belief revision in naive physics. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 426-432). Hillsdale, NJ: Erlbaum;

Schwarz, N. (1999). How the questions shape the answers. *American Psychologist*, *54*, 93-105.

Singley, M. & Anderson, J. (1989) *Transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435-502.