(2008). In P. A. Kirschner, F. Prins, V. Jonker, & G. Kanselaar (Eds.), *International Perspectives in the Learning Sciences: Proceedings of the Eighth International Conference for the Learning Sciences, Volume 2* (pp. 2-246 to 2-253). International Society of the Learning Sciences, Inc.

Designing and Assessing Numeracy Training for Journalists: Toward Improving Quantitative Reasoning Among Media Consumers

Michael Andrew Ranney, Luke F. Rinne,
Graduate School of Education, University of California, Berkeley, 4533 Tolman Hall, Berkeley, CA, 94720-1670

Louise Yarnall
SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025

Edward Munnich
Psychology Department, 2130 Fulton St., University of San Francisco, San Francisco, CA 94117

Luke Miratrix
Graduate Group in Science and Mathematics Education
University of California, Berkeley, 4533 Tolman Hall, Berkeley, CA, 94720-1670

Patricia Schank
SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025

Email: ranney@cogsci.berkeley.edu, lrinne@berkeley.edu, louise.yarnall@sri.com, emunnich@usfca.edu, lmiratrix@berkeley.edu, patricia.schank@sri.com

Abstract: Journalists inform multitudes of people. However, they sometimes over-focus on the narrative, failing to integrate critical quantitative information effectively. The Numerically Driven Inferencing (NDI) paradigm's research (e.g., Ranney et al., 2001; Munnich, Ranney, & Appel, 2004) suggested that a curricular module highlighting evidential/scientific thinking might enhance reporters' quantitative and analytic skills. The resulting controlled experiment involved 55 first-year journalism graduate students, our "Numbers, News, and Evidence" module, 4.5 classroom hours, 20 homework hours, and several (e.g., Pre-, Mid- and Final-test) assessments. Post-module findings indicate success: Relative to control data, the experimental group improved on the main numeracy measures: 1) estimation accuracy and 2) math competence involving simple problem solving, data analyses, and exponential growth. Students and faculty both concluded that future students should also receive numeracy modules. The module apparently influenced students' attitudes about numerical information, too. The collective results may benefit journalists, their instructors, and media consumers.

Arguably, journalists rival teachers in how much they educate the citizenry, but a bias toward narrative in journalistic writing often means less focus on using numbers to underscore key points. Quantitative data are commonly central to high-quality information, yet the skill to make full, meaningful use of statistics has long been considered a weak point among journalists (Curtin & Maier, 2001; Maier, 2003; Merritt, 2005; Meyer, 1979; Paulos, 1995) and other professionals (Levitt & Dubner, 2005). New reporters are often surprised at the importance of numbers in their work (Curtin & Maier, 2001; Maier, 2003). Such findings about journalism may not be surprising, given views about the public's collective numeracy (e.g., Paulos, 1988). However, becoming numerically adept seems ever more valued—by both society and our educational systems (e.g., Steen, 2004). Quantitative literacy practices stimulate considerable research (e.g., Gillman, 2006), which seems to be a worthy venture, given that numerical ability is central in making judgments and decisions (e.g., Peters et al., 2006).

Modest newsroom numeracy can also contribute to errors or other reporting weaknesses, such as in the *New York Times'* May 2007 front-page headline/story: "White House Said to Debate '08 Cut in Troops by 50%" (Sanger & Cloud, 2007). Although later corrected, many other news outlets repeated the conclusion, even as savvy readers noted that the possible reduction (146,000 to 100,000) was 32%, not 50%. As a more extreme example, National Public Radio reported in 2005 that a Danish company suffered $200 million-a-day losses in its Middle East branches—instead of the accurate $2 million-a-day—after a Danish newspaper published controversial cartoons. The use of a few key numerical strategies could have eliminated the error, and such thinking goes well beyond simple journalistic accuracy. For instance, reporters from *The Denver Post* received the 1986 Pulitzer Prize for public service due to such skills after a reporter noticed a discrepancy between national missing children statistics and the lack of missing children in local school districts.

Beyond the issues of errors or partial misrepresentations, statistics offered by media outlets are often rare, decontextualized, or disconnected. Indeed, some journalists are self avowedly "number-phobic"—trying to "write around the numbers," thus yielding pieces insufficient in rich, memorable, or accurate information. News consumers (e.g., readers) are often aware of the problems we note, and they sometimes view various journalists as being ill-informed or under-informing—notably on scientific issues that have quantitative bases. Journalists are occasionally even "unrepresentatively balanced" (e.g., when they try to achieve balance by "pairing competing experts" on stories about evolution or global warming, rather than by representatively portraying the relative merit of particular positions/sides with quantitative data, such as surveys of experts).

Prior research has shown that journalists and journalism students do not adequately appreciate the potentially catalyzing effects that even a single, critical, statistic has in changing citizens' social policies (N = 112; reported by Ranney et al., 2005, etc.). Although these participants were markedly resistant to estimating rates and averages about social phenomena (e.g., killings of police vs. private citizens), ultimately their own opinions on the related social issues were powerfully swayed by the true numbers—as are the opinions of other groups. In other work, the median undergraduate was found to believe that about one in 200 U.S. fetuses are aborted; in fact, roughly one in four were aborted (e.g., Garcia de Osuna, Ranney, & Nelson, 2004). The shock of the numerical feedback causes students and journalists alike to change their policies and verbal rationales about this abortion ratio (Garcia de Osuna et al., 2004). This suggests that journalists could benefit from instruction that improves their awareness of how numbers influence their own *and their audience's* thinking .

In another research vein, 232 surveyed journalism instructors described their students' quantitative skills as notably insufficient (Yarnall, Johnson, Rinne, & Ranney, in preparation). Journalism schools mount few systematic ventures to improve students' numerical understandings, so journalists often learn numeracy skills by trial and error—and media consumers may suffer the consequences of haphazard learning.

The need for numerically sophisticated, highly analytic reporters is underlined by the many inaccurate statistical data they encounter (e.g., assertions made by politicians or other public figures). Deadline-sensitive journalists do not always challenge these claims (much as non-journalists often fail to expose hoaxes or urban legends with a bit of arithmetic or a relevant base rate). Some of the skills our project hopes to enhance are as simple as contextualizing quantities—to help yield "value-added news." Too often, journalists report that a federal budget item has increased by, say, three billion dollars without providing either a time-frame, a "whole pie" sense of the federal budget, or the base that the increase adds to from a prior period (let alone more extended historical or international norms). Similarly, both journalists and the public often ignore growth or decay rates—especially the cumulative effects of nonlinearities. Clearly, whether projecting the sea level's rise, one's retirement fund, or changes in energy use, contextualizing a short period's change with respect to a longer period represents an important part of one's numeracy. However, rarely do financial or environmental journalists invoke the helpful "Rule of 72" for estimating doubling periods. (The rule approximates a natural-logarithm function, such that dividing 72 by a percentage increase roughly yields a doubling-period; e.g., at a 3% annual population growth rate, a country's populace would double about every 24 years.)

Given such concerns, we sought to improve the public's numeracy by improving the numeracy of those who help inform them—hopefully eventually yielding enhanced individual decisions and public policies as benefits. As a first step, we infused a graduate introductory newswriting course with evidence-based scientific reasoning. We provided them with a curricular module that replaced part of their regular course, introducing relevant scientific techniques so that the students might become more quantitatively, empirically, and critically sophisticated. The module also was intended to foster metacognitive skills so students might better assess the coherence of their own—and others'—sources of information. Such students might then, upon encountering *in*coherent information, employ quantitative skills and sound reasoning to amend and enrich their news articles.

Our resulting curricular module, highlighting evidence and scientific thinking, included elements from the Numerically Driven Inferencing (NDI) paradigm (Ranney et al., 2001). NDI provides methods for revealing and confronting students' underlying assumptions about rates, magnitudes, trends, and social policies. A typical NDI method that the module sometimes used, "EPIC," involves <u>E</u>stimating a base rate (e.g., the abortion rate), offering a <u>P</u>reference for (or involving) that rate, <u>I</u>ncorporating the *true* base rate as feedback, and then checking for a <u>C</u>hange in one's preference (e.g., Ranney et al., 2001; Rinne, Ranney, & Lurie, 2006). We taught the resulting "Numbers, News, and Evidence" module to five sections (N = 55) of a first-year graduate journalism newswriting course. Each section received roughly 4.5 classroom hours during one week, 20 homework hours, and several out-of-class (e.g., Pre-, Mid- and Final-test) assessments. The brief curriculum, loosely based on one successfully used with high school students (Munnich, Ranney, & Appel, 2004), engaged students in many quantitative reasoning exercises—such as estimation, basic computation, and numerical critique.

An overarching curricular goal was to offer a journalistic numeracy that is beyond simple arithmetic—one stressing evidence and the scientific method (e.g., disconfirmation), and a synergy of quantities and verbal propositions (e.g., between evidence such as statistics and scientific hypotheses about potential social policies; cf. Ranney & Schank, 1998). The module also offered epistemological criteria (e.g., "What knowledge does the reader gain?") that good news editors use about a story's ontology (i.e., "What *is* the story?"), and the module focused on analyzing the veracity of information sources—a critical skill needed by all journalists. As described below, our most critical hypotheses regard whether the module would improve aspects of student numeracy, primarily with respect to "control" ("Late") students at the point before they received the module.

The Module and the Experimental Design

Five graduate news-reporting course sections (students and instructors) in an illustrious journalism program assented to participate in our experiment. For easy integration into the sections, the module was developed as a stand-alone "numeracy training." Each module was taught by the first author, but in consultation with the instructors and the research team. We streamlined the module by shifting considerable work into 20 hours of homework assignments and by conducting tests outside of the class periods. The module included in-class activities and writing homework that yielded feedback and critiques on students' prose (see Table 1).

To ensure that all students ultimately received the module, yet still test its impact with a comparison group study design, we staggered the module's timing among the groups. To better counterbalance the study, we designated two specific sections (n = 23) to take the module early in Fall, 2006, and the other three sections (n = 32) to take the module later in the semester. This schedule effectively gave us two comparison groups, the "Early" classes and the "Late" ("control") classes (see Table 1). Each participant experienced three testing periods (Pre-, Mid-, and Final-tests) and the curricular module. Many of our most critical empirical hypotheses concerned changes for the Early group from Pre-test to Mid-test, compared to the Late group (i.e., differential, controlled, changes). The Final-tests allowed us to address other interesting hypotheses, such as the longevity of the learning for the Early group and potential replications of the Early group's improvements (see Table 1).

Table 1: The experimental design, showing the module, student-writing critiques, and main testings.

Week(s)	**Early Group (n = 23; two course sections)**	**Late Group (n = 32; three course sections)**
0	**Pre**-test (**E1**)	**Pre**-test (**L1**)
1	*Curricular Module & Writing Critiques*	
2	**Mid**-test* (**E2**; an immediate post-test)	**Mid**-test* (**L2**; a redundant pre-test)
3-9		*Curricular Module & Writing Critiques*
10	**Final**-test (**E3**; a markedly delayed post-test)	**Final**-test (**L3**; a somewhat delayed post-test)

*A student's Mid-testing was an "initial post-test" (E2) for the two Early sections, yet was a "redundant, second, pre-test" (L2) for the three Late sections—allowing a rather direct module assessment.

The Curriculum

An agenda was given to students for each of three days' 1.5-hour classroom sessions. Regular section faculty participated or observed about one third of the time. Project collaborators often joined the sessions, offering logistical and other support. The 4.5 classroom hours involved about eight main kinds of activities, yielding these class time shares: 25% estimation practice and strategies such as disconfirmation, benchmarking, decomposition, coherence, "whole pie" contextualizations, and the "Rule of 72" (on nonlinear aggregation/compounding); 20% data-foraging tactics and other tips or caveats, often involving polls, misleading statistics, quantities, and the scientific method; 12% examples of superior and inferior uses of statistics in reporting; 12% on NDI-based philosophies and exercises about numerical journalism; 10% using numerical and statistical resources; 9% assignments of readings, math (arithmetic) homework, and written stories involving statistics; 6% feedback on the math and writing performances; 6% brief discussions of the readings. (Central readings included Cohen, 2001, Huff & Geis, 1954, and parts of Levitt & Dubner, 2005.)

Part of our NDI-based philosophy suggested the goal of reducing the view that "journalism is just narrative prose." Numbers alone rarely make a news story, but when aptly used, they can offer the reporter and news consumer dramatic insights that notably enhance the story. Students were not encouraged just to flood pieces with numbers; rather, they were trained to infuse their writing with the most crucial, contextualized, memorable, and veridical statistics—and to use quantitative analysis to understand story topics better.

A hallmark of the curriculum was a "Top 40" list of important quantities "that one should know (but many don't)" that were used for some estimation practice and as benchmarks to enhance number sense regarding social policies. The list (available upon request), along with the quantities' true magnitudes and their sources, were provided to the students. Its topics include populations and their influences, demographics, personal and governmental economics, housing, crime and punishment, employment, evolutionary acceptance, natural resource use/misuse, and global warming (which was a curricular content emphasis).

One curricular element was a focus on understanding that many claims have numerical bases—a focus that sometimes involved suggestions to attempt disconfirmation. For instance, students were to consider what to have asked Karl Rove (George W. Bush's adviser) after a 2006 Rove speech in which he claimed that their "tax cuts have helped make the U.S. economy the strongest in the world." Most students did not venture a guess. It was then pointed out that one could have asked, "By what measure is the U.S. economy strongest?," and "Which nation had the strongest economy before the tax cuts?," and "When did the U.S. overtake that country?" (If "strongest" means "largest," it seems that the U.S. economy became the strongest around 1910.)

Measures Employed in This Experiment

Each Pre-, Mid-, and Final-test generally involved seven parts. For five parts, multiple parallel tests were created so that no student saw the same item twice; versions were counterbalanced across groups, test times, and sections. These five parts were: (1) a mathematics assessment with ten basic arithmetic items (modeled on a journalistic math web-test from IRE, Doig, 2006) and seven items requiring the analysis of data presented in tables, charts or graphs; (2) three exponential growth problems (to which the "Rule of 72," if known, could be applied); (3) a 22-item estimation skills test (e.g., "What was the U.S. population in 1900?" or "What percentage of U.S. homes contain a firearm?"); (4) a three-item measure of students' abilities to generate issue-relevant quantities; and (5) a three-item measure of students' abilities to generate disconfirming arguments. The other two parts were (6) Viswanathan's (1993) Preference for Numerical Information (PNI) attitude survey and (7) a Need For Cognition (NFC) survey. (The NFC data were inconclusive, partly due to a ceiling effect, so are not discussed herein.) In addition, a final assessment instrument that allowed students to evaluate the module itself was administered during Week 15, just prior to debriefing students on the experiment.

Results and Discussion

We hoped to build quantitative competence on several levels. In addition to (a) improving basic computational skills important to journalism, we were also interested in (b) increasing students' abilities to generate reasonable estimates of unknown quantities, (c) deepening students' capacities to think about how (and which) numbers might convey information about particular issues, and (d) helping the students develop more critical eyes regarding the numbers and statistics they encounter. Perhaps the more salient of these goals were the first two, improving students' mathematical and estimative abilities. Basic math skill is often emphasized as a fundamental indicator of numeracy. Additionally, we believe that the ability to estimate real-world quantities (e.g., the federal U.S. budget) is equally important—as this kind of estimation ability reflects an individual's knowledge-based competence in accessing and effectively using quantitative knowledge in a variety of contexts. One could be a superb mathematician (e.g., in computation or logic), yet still lack the contextualized real-world knowledge, number sense, and/or strategic skills to estimate a population or the median price of a house.

Both Basic and "Rule of 72" Mathematics Skills Improved

The small amount of instruction, practice, and feedback yielded a marked improvement in students' basic math skills (e.g., involving percentages and word problems, and interpreting tables and graphs). Students' Pre-test accuracy was rather modest, given the fairly simple items. For the "Early" (experimental) group, the mean accuracy before the module (E1 in Table 1) was 68%. However, their mean score grew to 81% correct on the Mid-test following the module (E2). A linear mixed-model regression analysis in which students were treated as a random factor indicated that this increase was significantly greater than that observed for the "Late" (control) group ($z = 3.70$; $p < .001$; all tests are *two-tailed* herein), who had not yet received the curriculum and showed no change between the Pre-test (L1) and the Mid-test (L2; 76% vs. 75%). When the Late group received the module, their scores increased as well; their Final-test scores (L3 = 84%) were significantly higher than their Mid-test scores (L2) ($z = 3.84$; $p < .001$; and higher than their L1 scores, as well). Further, the Early group held its math gains, showing no significant decline in skills even nine weeks later (at E3).

At first, students scored only a mean of 39% on exponential growth problems (with partial credit response scoring). However, after learning the Rule of 72 during the curriculum, Early students posted a mean score of 70% correct on the Mid-test (E2), indicating that many could learn and use the Rule of 72. Further, students retained their knowledge of the rule; even after their nine-week post-curricular delay (at E3), they still scored 65% correct on average, which was not significantly lower than their scores at E2. Interestingly, scores on these items were uncorrelated with those for the other items in the math assessment. This suggests that the Rule of 72 can be learned and used effectively by students with a variety of levels of mathematics ability.

Estimate Accuracy Improved

The 66 estimation items were diverse, often challenging (e.g., "How many U.S. domestic commercial passenger flights occur per month?"), and greatly varied in difficulty. If one sorts the items by how far their respective median estimates were from their respective true values, the least extreme median estimate error was only 9%, while the most extreme median estimate error was 1011%. To analyze the estimation error data, we first transformed the raw error scores (|estimate – actual|) by taking log(|estimate – actual| +1). This reins in extreme values, which represent a common problem in examining estimation data. (1 is added to the error score in order to prevent it from being negative or undefined under the transformation.) A linear mixed-model regression analysis was conducted on the transformed error scores with items and students treated as crossed random factors. Results show that the students in the Early group reduced their estimate error scores between the Pre- and Mid-tests significantly more than did the Late students, who were assessed twice but had not yet received the module ($z = 3.90$; $p < .001$). For given values of the other explanatory variables (item and subject), the decrease in error roughly represents a relative reduction in raw (untransformed) error of 36%.

Standardized residuals (predicted values – observed values) indicated that one entire item's estimates and 15 other individual estimates were predicted poorly by the model (i.e., residuals with z-scores > 5 or < -5). An outlier analysis was therefore conducted by re-running the model with the item's responses after the extra 15 responses were removed. The effect described above is again significant ($z = -2.48$, $p = .01$), but its size is smaller in this analysis; the module led to a relative reduction in raw (untransformed) error of 20% (cf. the 36% from above) for given values of the other explanatory variables. (In addition, with the outliers removed, a student's gender is a significant covariate of estimation error, $z = 2.74$; $p < .01$—a male-favoring effect that is not significant when all the data are included because the outliers increase the regression coefficients' standard errors, such that coefficients would have had to be markedly larger for statistical significance.) In sum, with or without outlier data excluded, the module positively affected estimation accuracy in the journalism students.

It seems worth noting that estimation accuracy (as we noted for "Rule of 72" accuracy) was uncorrelated to basic mathematical accuracy. This suggests that people who are the best calculators are not necessarily the best estimators (e.g., at estimating the state of California's annual governmental budget).

Preference for Numerical Information (PNI) Scores Changed

One hypothesis was that a numeracy curriculum might change students' attitudes toward quantitative material, and it seems to have been partly supported. The change, however, was clearly not uniformly in the positive direction. The gain for the Early and Late groups combined, from Pre-Test to Final-test (E1+L1 vs. E3+L3), was marginally significant ($p = .06$)—as was the more narrow comparison between E1+L2 versus E2+L3, which contrasts participants' PNI scores just before and just after they received the module (considering their group membership; also $p = .06$). Although Early students' PNI scores did not increase significantly between the Pre-test and Mid-test (perhaps partly due to a ceiling effect), the Late group's gain between the Pre-test and Final-test (L1 vs. L3) was statistically significant ($p < .05$—although the majority of this effect, which was between L2's Mid-test and L3's Final-test, was not itself significant). A numerical majority of students increased their PNI scores after the module, but a numerical minority of students decreased their PNI scores; classroom and extra-classroom observations suggested that some decreases may be due to some students initially thinking that they were more numerically adept than their later modular experiences indicated (e.g., one student realized that she could no longer divide 72 by six on paper; in contrast, another who had just finished multivariate calculus was trying to determine what her next math class should be). That is, the module (and perhaps the ceiling effect) may have caused an attitude *polarization*, increasing most students' PNI scores and decreasing some of the others' scores. A post-hoc analysis of the magnitude of the *change* in PNI scores (ignoring the direction of this change) is consistent with this hypothesis; the Early group's PNI scores changed significantly more than did the Late group's at the Mid-test ($t(40.3) = 2.22$; $p = .025$). Students were given the same PNI survey (i.e., that from Viswanathan, 1993) at each test-time, though, so repetition effects (potentially in either direction) were possible; of course, it may also be that changes in PNI scores reflect students' reactions to the module (either positively or negatively), rather than more central changes in underlying attitudes toward numerical information. Future research can likely disambiguate the causes of the PNI changes we observed.

Post-hoc inter-instrument analyses suggest that PNI correlates with a number of other dependent variables, but the multiple testings with the same PNI form remains a concern for some of the effects. (Space constrains this discussion, but several inter-instrument results seem solid—e.g., that initial math skills correlate with math homework scores, of course; $F(1,48) = 18.6$; $p < .0001$. The inter-covariate relationships observed may yield insights regarding further development of NDI as a theory, e.g., about the nature of numeracy.)

Quantity-Listing and Disconfirmation Items

We wondered whether the module changed students' abilities or propensities to generate and articulate quantities that may be relevant to a particular issue (e.g., nuclear proliferation). To try to measure this, students were given three minutes to write down as many quantities as possible (e.g., "the number of American warheads") that are relevant to a particular societal issue. The module's effects on such responses were not striking—partly because students often simply seemed to write quantities almost as fast as possible throughout the time provided. Therefore, students seemed limited more by their writing speeds than by their abilities to generate issue-relevant quantities. Despite this, the module may have had some qualitative effects on the responses students gave. For instance, the module increased the average number of words written per listed quantity; the Pre-test to Mid-test gain for the Early group (E2 – E1) was significantly greater than that for the Late group (L2 – L1; 20% more words per quantity for the Early group vs. 5% more words per quantity for the pre-module Late group: $z = 2.02$, $p < .05$). Thus, the curriculum may have caused students to generate more complex quantities or to describe more richly the quantities they thought were related to a particular issue.

An important curricular theme and motivation was that people are often insufficiently critical or skeptical of the numbers they receive. We therefore encouraged students to disconfirm the pieces of numerical information they receive as if they were tenuous hypotheses. To measure students' abilities to do this, we developed what we call "Pat" items, in which a fictional colleague (Pat) indicates that certain quantities have particular values (e.g., that the U.S. generates 20% of its energy from nuclear power). Unbeknownst to students, one third of the numbers they received were correct (e.g., the 20% number just given), while the rest were actually high or low. Students were asked to generate reasons why each value might be too high or too low (for four minutes; two apiece). Although students generally (across both groups) increased the number of disconfirming reasons they gave during the semester (e.g., from Pre-Test to Final-Test; $z = 3.07$; $p < .005$, when controlling for covariates), no statistically significant effects on the number of disconfirming reasons offered can be directly attributed to the module. However, in parallel with the quantity listing items, an effect (albeit marginally significant here) for the average number of words written per Pat-reason is noted (i.e., the gain in average words from Pre-Test to Mid-test for the Early group, relative to the Late group; $z = 1.63$; $p = .10$). This suggests a possible effect of the module on how richly students wrote about their disconfirming reasons.

Preliminary analyses suggest that the module may have changed the distribution of the *kinds* of disconfirmation reasons students offered for the Pat items (using six codes; overall, over 65% of the reasons received the code indicating causal relationships about semantic elements connected to the focal societal issue).

Students' Assessments of the Module and Conceptual Changes About Numeracy

An anonymous survey yielded helpful feedback and comments that ought improve the module's future use. One item asked whether future students in their program should take a larger module as an intensive, one-unit (2-3 week) course: 80% responded "yes," and 11% responded "maybe," or "it depends," or "if you change [something]." Only 9% responded "no." Other feedback suggested that the module's future incarnations might be "tracked" into three levels to reflect students' strikingly divergent incoming math skills. Many students commented—via the module evaluation or directly/anecdotally—that they were positively affected by the module (intellectually and/or emotionally; e.g., in contrast to prior math curricula). The regular journalism instructors also commented that students brought their new numeracy skills into their post-module coursework (e.g., a student telling a faculty member, "I'm going to put *this* new fact into my personal Top 40 list!").

We observed shifts in students' views of the utility of numeracy in journalism. New journalists are not always prepared for present newsroom realities, as in the following example of an apparent conceptual change. A student expressed some initial skepticism about the module early in the first session, naively asking, "Why should I know these numbers? My research staff will handle that for me." The module's instructor responded by noting that newsroom research staff is being cut dramatically, and that without knowing important quantities such as those in the "Top 40" list, the student's staff (if one existed) would likely tell him the questions he *should* have asked during his interviews upon his return to the newsroom. The first of these responses reflects a concern that dramatically grew during the experiment's semester (Fall, 2006). Media and advertising changes have yielded newspaper profitability pressures, as well as employee layoffs and buyouts (e.g., Poynter Institute, 2007; Wilkinson, 2000; newspapers sold per U.S. household have dropped from 1.3 to .5 since the 1920s).

Future Directions

We are exploring several future research vectors. Some may involve a web site we developed to inform people about this inquiry realm: morenumerate.org. Besides a project description, it offers a numeracy quiz (a taste of the Top-40 numbers and the "other 66" estimation assessment items) and links to publications and resources (e.g., math tutorials for journalists and readings about understanding statistics). In addition, we hope to offer workshops to help journalism educators develop their own numeracy modules, and may provide our module directly (e.g., to journalism classes or newspapers, as in-services). Furthermore, while our module incorporated an emphasis on global warming, world events may suggest altering or sharpening that focus.

As noted earlier, some students thought they were more numerically adept than their later module experiences showed. To the degree that such curricular "reality checks" occur, they may represent desirable motivational effects: part of how we motivated students to engage the module was to show that they lacked (like most people) strong statistical grounds for many of their social policies and assumptions.

The participants' journalism faculty wish us to teach their students again—probably as an independent multi-week course, which would yield even more skill-appropriate practice and instructional attention. Other assessment data (and those noted above) suggest that, rather than "adopting" extant classroom units, assorting students into numerately appropriate curricula (e.g., using a pre-test) may be a wise future approach. Although "tracking" is often aptly criticized (e.g., as stigmatizing children), subgrouping adults who have highly variable pre-curricular functional math abilities may be acceptable. One might weight students' initial math skills, math training, estimation accuracy, and (perhaps) PNI—possibly in descending order. One could then offer modules matched to subgroups' sophistication levels. Alternatively, one could take a mastery perspective and instruct students only until they reach a criterion. This has its own risks, though, as the training may seem remedial—even though the skills trained hardly represent commonplace newsroom competence (and some journalism programs do have remedial courses—even in writing). Students may wish to quickly "test out" to "get on with 'real' journalism." (Testing out would likely be rare, given NDI findings to date.) When it exists, journalistic numeracy training is often isolated, and cognitive research (e.g., in problem solving or decision making) shows that such skills usually result in isolated application. Ideally, journalism education will eventually view numeric and scientific analytic skills as elements that ought to be integrated into an entire program of study.

General Discussion and Conclusions

We believe that our research findings are noteworthy and encouraging—especially (a) students' marked gains in mathematics ability, (b) students' gains in estimation accuracy, (c) most of the results regarding students' changing attitudes about numerical information, and (d) the overall positive view by students and faculty of the module's efficacy and future utility. Furthermore, we have a clearer vision of what the module's next iteration should be (to better produce rich, useful, creative, and memorable journalistic writing)—and of the socio-cultural contexts that may yield even greater success in subsequent manifestations of the curriculum.

Our module subsumed standard and non-standard approaches to journalistic numeracy. Standard ones included, in part, "refresher" elements—basic arithmetic and how to interpret statistics and common economic indicators, etc. Non-standard ones, such as employing the "Top 40" statistics and the Rule of 72, were designed to develop better, dynamic, journalism-relevant, quantitative reasoning skills; we used them due to data showing journalists' lack of deep quantitative knowledge, which leads to widely acknowledged problems regarding overinterpreting or not questioning numerical and statistical data (e.g., Cohn & Cope, 2001; Stamm, Williams, Noel, & Rubin, 2003). Our study replicated past findings with journalists (Maier, 2003), with our participants initially scoring 68% correct on basic computation skills; however, our module improved students' scoring to 81%, which was not diminished nine weeks later. Regarding the module's non-standard numeracy teaching tools, we found some evidence that the journalism students came to elaborate more about numerical information and (albeit a more tenuous effect) to more elaborately question numerical information. Increased elaboration is an early indicator of developing deeper knowledge (e.g., Chi, DeLeeuw, Chiu, & LaVancher, 1994).

Overall, our findings indicate that even a rather short and broadly targeted conceptual numeracy curriculum can yield results that would seem desirable to journalists, their instructors, and the readers that they serve. The data, as well as feedback received from the participants and their regular instructors, suggest that a longer curriculum would be required to reinforce many of the module's elements—particularly regarding how to deploy estimation strategies better (including how to disconfirm, critique, and analyze initial estimation hypotheses better)—and how to integrate one's improving numeracy skills into one's news writing. We hope to implement a multi-week module soon, to help further enhance journalism's use of high-quality information.

Available space allows only the briefest of explications about how the experiment's results help develop NDI theory, but traces of this effort are marbled into this report. We plan to continue to explore the relationships among many of the measures discussed above (e.g., the possible importance of PNI), as well as the lack of such relationships (e.g., among abilities in estimation, exponential growth, and more basic mathematics).

This research venture was borne from the simple yet grand notion that numerical professional development for budding journalists might yield a societal impact that is comparable to professional development for budding teachers. We have not abandoned this hypothesis, and hope that future work can yield significant impact, both directly and indirectly, on journalists and the people they wish to inform. A web site (morenumerate.org) represents a further effort to inform journalists, their educators, and the general public.

In summary, the results of our experiment indicate that our module exhibited general success in meeting the overarching goal of offering improved numerical lenses to postsecondary journalism students. First, students' mathematical skills (both basic and exponential) increased and remained increased. Second, students came to estimate more accurately novel quantities after experiencing the module, compared to control participants. Third, even students' attitudes towards numbers seemed affected by the module (i.e., it apparently influenced their PNI scores). Finally, the students and their journalism faculty saw utility in the module, such that they believed that an incarnation of it should be provided in subsequent years.

References

Chi, M. T. H., DeLeeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Cohen, S. (2001). *Numbers in the newsroom: Using math and statistics in the news*. Columbia, Mo: Investigative Reporters and Editors, Inc.

Cohn, V., & Cope, L. (2001). *News & numbers: A guide to reporting statistical claims and controversies in health and related fields* (2nd ed.). Ames, IA: Iowa State Press.

Curtin, P. A., & Maier, S. R. (2001). Numbers in newsrooms: A qualitative examination of a quantitative change. *Journalism & Mass Communication Quarterly, 78* (4), 720-738.

Doig, S. (2006). *Math test for journalists*. Investigative Reporters and Editors, Inc. (Accessed 8/30/2006 from http://www.ire.org/education/math_test.html).

Garcia de Osuna, J., Ranney, M., & Nelson, J. (2004). Qualitative and quantitative effects of surprise: (Mis)estimates, rationales, and feedback-induced preference changes while considering abortion. In K.

Forbus, D. Gentner, & T. Regier (Eds.). *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society* (pp. 422-427). Mahwah, NJ: Erlbaum.

Gillman, R. (Ed.) (2006). *Current practices in quantitative literacy*. Washington, D.C.: Mathematical Association of America.

Huff, D., & Geis, I. (1954). *How to lie with statistics*. New York: Norton.

Levitt, S. D., & Dubner, S. J. (2005). *Freakonomics*. New York: HarperCollins.

Maier, S. R. (2003). Numeracy in the newsroom: A case study of mathematical competence and confidence. *Journalism & Mass Communication Quarterly, 80* (4), 921-936.

Merritt, D. (2005). *Knightfall: Knight Ridder and how the erosion of newspaper journalism is putting democracy at risk*. New York: American Management Association.

Meyer, P. (1979). *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Bloomington, IN: Indiana University Press.

Munnich, E. L., Ranney, M. A., & Appel, D. M. (2004). Numerically-Driven Inferencing in instruction: The relatively broad transfer of estimation skills. In K. Forbus, D. Gentner, & T. Regier (Eds.). *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society* (pp. 987-992). Mahwah, NJ: Erlbaum.

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Hill and Wang.

Paulos, J. A. (1995). *A mathematician reads the newspaper*. New York: Anchor Books.

Peters, E., Vastfjall, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17 (5), 407-413.

Poynter Institute. (2007). *The State of the News Media 2007: An Annual Report on American Journalism*. (Accessed 6/27/2007 from http://stateofthemedia.org/2007/narrative_newspapers_intro.asp?media=3)

Ranney, M., Cheng, F., Garcia de Osuna, J., & Nelson, J. (2001, November). *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 and Decision Making, Orlando, FL.

Ranney, M., Munnich, E., Lurie, N., & Rinne, L. (2005, May). *Talk is often cheap, but self-explanations can aid learning: Discourse and dialogue in Numerically Driven Inferencing.* Poster presented at the Talk and Dialogue: How Discourse Patterns Support Leaning conference, Pittsburgh, PA.

Ranney, M., & 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* (pp. 245-274). Mahwah, NJ: Erlbaum.

Rinne, L., Ranney, M., & Lurie, N. (2006). Estimation as a catalyst for numeracy: Micro-interventions that increase the use of numerical information in decision-making. In S.A.Barab, K.E. Hay, & D.T. Hickey (Eds.) *Proceedings of the Seventh International Conference of the Learning Sciences* (pp. 571-577). Mahwah, NJ: Lawrence Erlbaum.

Sanger, D., & Cloud, D. (2007, May 26). White house is said to debate '08 cut in Iraq combat forces by 50% [with correction and corrected title]. *New York Times*. (Accessed 10/14/07 from http://www.nytimes.com/2007/05/26/washington/26strategy.html).

Stamm, K., Williams, J. W., Jr., Noel, P. H., & Rubin, R. (2003). Helping journalists get it right: A physician's guide to improving health care reporting. *Journal of General Internal Medicine, 18* (2), 138-145.

Steen, L. A. (2004). *Achieving quantitative literacy: An urgent challenge for higher education*. Washington, D.C.: Mathematical Association of America.

Viswanathan, M. (1993). Measurement of individual differences in preference for numerical information. *Journal of Applied Psychology, 78* (5), 741-752.

Wilkinson, E. J. (2000, May). *Leveraging the Newspaper's Value in a Marketing Society*. ASTECH Intermedia Executive Roundtable on Strategic Marketing, New York (Accessed 6/27/2007 from http://www.inma.org/subscribers/papers.cfm?which=22).

Yarnall, L., Johnson, T., Rinne, L., & Ranney, M. A. (in preparation). *How Post-Secondary Journalism Educators Teach Reporting in the Digital Age: Are Appropriate Analytic Skills Conveyed?* SRI International, Institute for Analytic Journalism, and University of California, Berkeley.

Acknowledgements

We especially thank Tom Johnson for his wisdom and great effort throughout this project. Andrew Galpern and Myles Crain were also mainstays of this research, and their help was much appreciated. We also heartily thank Edgar Busovaca, Mirian Song, Sophia Rabe-Hesketh, Michelle Million, Janek Nelson, Natasha Arora, Rachel Ranney, Lauren Barth-Cohen, Mark Wilson, Charlie Litchfield, Chris Gade, Jose Gutierrez, Nicole Migliarese, Tom Wickens, Caslyn Cole, Naoko Kuriyama, Nancy Zhang, Theresa Duong, Min Su Chung, Dav Clark, and Barbara Ditman, and for their help—as well as the journalism instructors and the students involved in this study. Finally, we thank the William and Flora Hewlett Foundation, and several other sources (both extramural and intramural, and for faculty, graduate students, and undergraduates) for their support of this project.