NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

American Statistical Association

The Statistics Teacher Network

www.amstat.org/education/stn/index.html

Number 66 ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability Spring 2005

Real Classroom Examples of Hypothesis Testing for Grades K-9

Angela L.E. Walmsley

Saint Louis University

The *Principles and Standards* (NCTM, 2000) challenges students to "develop and evaluate inferences and predictions that are based on data." Often, educators focus on flipping coins, rolling die, sorting categories, and creating graphs to provide experiences to meet this standard. Many students first encounter ideas about inference and predictions in high school mathematics and science classes. However, the basic ideas behind statistical hypothesis testing can be initially explored in primary grades without difficulty. Children can begin to make good decisions if they follow the pattern of traditional hypothesis testing.

What is hypothesis testing? In a traditional sense, hypothesis testing often begins with a null hypothesis that states that there is no difference between two groups, or any effect or difference is due to chance (Patten, 2004). The alternative hypothesis, then, is that a difference exists between the groups or the difference is caused by something other than chance. In statistics courses at the high school level students also encounter types of errors. At the high school level the type I error, rejecting a true null hypothesis, or in other words, a "false alarm" (Stark, 2004) and the type II error, failing to reject a false null hypothesis, are hopefully presented within the context of everyday situations. Younger children, on the other hand, need only know that errors are always possible when you are trying to verify something. For elementary children, every example of this concept can help them to see how decisions are based on scientific hypothesis without even realizing it.

Many things that adults and children experience daily are examples of hypothesis testing problems – even if they are not "statistical." Below are a few examples:

- When a child goes to see the doctor with a sore throat, the doctor is trying to determine between the null hypothesis that the child's throat is fine, and the alternative hypothesis that the child's sore throat indicates strep throat, flu, or some other illness. Because doctors are unable to always predict the best course of action, both types of errors are present. The doctor can make an error, concluding that the child has a throat infection when in fact, he or she doesn't. Or, the doctor may make a different error by stating that nothing is wrong, when in fact, the child has strep throat.

- If a child goes through the lunch line at school, and there are two types of juice cartons to pick from, the child may assume that both cartons will taste exactly the

Also In This Issue...

same (the null hypothesis). The alternative hypothesis is that they will taste different. The child may refuse to drink the juice (which are the same) because he or she believes the cartons taste different. Or, a child will drink both cartons stating they are the same juice, when in fact, one is apple and one is white grape.

Examples like these are important because educators can frame age appropriate issues and problems in the form of hypothesis testing, and then lead into a discussion of inference based on the results. The following are some examples that can be used or discussed in the classroom:

For Pre-k to 2 the NCTM *Standards* (2004) suggest that students should be able to discuss events in terms of "likely or unlikely." For grades 3-5 students are challenged to "propose and justify conclusions, and prediction based on data." In the middle grades the *Standards* want students to "use observations about differences between two or more samples… make conjectures…and use conjectures to formulate new questions."

Assume that in one class the teacher comes in with two bags of Oreo cookies, one with green filling and one with red filling. Suppose further, that when the teacher begins to pass out the cookies, some children want red instead of green because they believe the cookies do not taste the same. This kind of opportunity is perfect for hypothesis testing and inference. The teacher could propose the theory that, despite color, the cookies will all taste the same (null hypothesis). Then, each student could eat both a red and a green Oreo, and decide whether he or she believes the cookies taste the same.

Assume the teacher does this with the class, and 23 out of 25 students believe the cookies do taste the same. Thus, the teacher concludes that the cookies do taste the same despite color. The teacher then poses the question to the students, "do you think that all bags of green-filled Oreos taste the same as all bags of red-filled Oreos?" This process is causing the students to logically infer for all Oreos made. The students can probably assume yes, but they may want another taste test! A couple of students may even ask to compare to regular white filled Oreos. If the students agree that white filled Oreos also taste just like green and red ones, then they may begin to wonder (or infer) whether all Oreos taste the same despite the color of the filling. Would you assume, based on this "test," that the Oreos are "likely" to be identical in taste? An error in this case would be if the students believed the color of the filling did change the taste of the cookies, when in fact, the cookies are the same. But, assume that the teacher did check with Nabisco, and realized that the cookies do in fact differ in taste based on the color of the filling – a different kind of error has occurred. Discussion of likelihood and possible errors can be rich, even at this young age.

To extend this example for middle school students, the teacher could use a variety of bags of different colored Oreo cookies, and ask the students to create a scientific study to test the assumption that all cookies will the taste the same. The teacher could also expand this lesson to include research on the Oreo cookie to see if their conjectures are correct.

One compelling example for classroom discussion that leads into error types for middle school students is based on the following scenario: a man is on trial for murder. The assumption is that the man is innocent (the null hypothesis). The prosecutor presents evidence in the hope that the jury rejects this hypothesis. If convicted, the man will receive the death penalty or be sentenced to life imprisonment. What is worse: finding an innocent man guilty (putting someone in jail for life or putting someone to death), or finding a guilty man not guilty (letting a guilty murderer go free)? This is a great example that ties into the Constitution and legal system; however, the importance of hypothesis testing clearly depicts the situation, as well as the possible errors, that can be made in this very important decision. This example has impact with high school and college students, and it brings out great discussions of the importance of strong hypothesis testing; and if one can't be certain in the testing, the importance of the possible errors is crucial. Many of my students state that finding an innocent man guilty is worse; however, there are always a few who believe that letting a guilty murderer "back on the streets" is the worse error. What is evident in this case is the logical structure of hypothesis testing and the gravity of the possible errors that still exist.

From the examples above, great discussions can take place for students as they work through the logic of hypothesis testing and logic reasoning in a fun and practical way. An educator can use examples for students to discuss options that help them make good decisions and realize to what extent they can infer for others. An educator can use "experiments" such as the Oreo one described above to allow children to discover answers and infer those answers to how it affects other populations.

References

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*, Reston, VA: NCTM.

Patten, M. (2004). *Understanding Research Methods*. Glendale, CA: Pyrczak Publishing.

Stark, P.B. (2004). *Hypothesis Testing: Does Chance explain the Results. Statistics Tools for Internet and Classroom Instruction with a Graphical User Interface*, Chapter 19. Retrieved from the World Wide Web, Dec. 25, 2004: *http://www.stat.berkeley.edu/users/stark/SticiGui/Text/ch19.htm.*

Illumination through Representation: An Exploration Across the Grades

Lawrence M. Lesser, Associate Professor
University of Texas at El Paso

I enjoy sharing the 19th-century John Godfrey Saxe poem "The Blind Men and the Elephant" with students of all ages because they readily see how seeing things from only one point of view can fail to capture the entire phenomenon. In fact, two different partial views can appear to suggest completely opposite or incompatible conclusions! Likewise with data.

Let's start with a situation that can be explored by students starting in 3rd or 4th grades, if the students can compare fractions with unequal denominators. In fact, this could add much interest and variety to a lesson drilling those skills. (Teachers could construct this with manipulatives and decorated boxes, or maybe use "hats" such as Mitchem (1989).) Suppose the question of interest is: "Is it more likely to draw a white ball from a striped box or from a solid box?"

"red solid box":	1 white ball and 2 black balls.
"red striped box":	3 white balls and 5 black balls.
"green solid box":	3 white balls and 2 black balls.
"green striped box":	2 white balls and 1 black ball.
Striped boxes(combined):	5 white balls and 6 black balls.
Solid boxes (combined):	4 white balls and 4 black balls.

The probability of drawing a white ball from the red solid box is 1/3, which is less than the probability (3/8) of drawing a white ball from the red striped box. Now the probability of drawing a white ball from the green solid box is 3/5, which is less than the probability (2/3) of drawing a white ball from the green striped box. So for the red boxes and for the green boxes, the striped boxes have a better chance of drawing a white ball than the solid boxes do. However, when we combine the striped boxes, the probability of drawing a white ball is 5/11, which is LESS than the probability (4/8) of drawing a white ball from the combined solid boxes.

With only 19 balls and mere fraction arithmetic, we have encountered a probability representation of Simpson's Paradox (no relation to Homer or O.J.!), an intriguing situation in which (to give a succinct "verbal representation") a comparison can be reversed when data are grouped. (Students have seen mathematical "reversals" before this, by encountering that 2 < 4 and yet 1/2 > 1/4, not to mention -2 > -4.) This phenomenon is not contrived: Lesser (2001) lists cited instances of occurrences in many natural situations, including university admission rates (male versus female), fertility rates (rural versus urban), death rates (young versus old), death penalty cases (black versus white), categories of federal tax rates, and various baseball statistics.

TABLE 1: Baseball Season Batting Averages

	1989		1990		both years combined	
	Justice	Van Slyke	Justice	Van Slyke	Justice	Van Slyke
Hits	12	113	124	140	136	253
Outs	39	363	315	353	354	716
At-bats	51	476	439	493	490	969
Bat.Avg	.235	.237	.282	.284	.278	.261

For example, Friedlander (1992) shows that there have actually been occasions when one baseball player had a higher batting average than another player in two consecutive years, but a lower average for the combined two-year period, such as the example in Table 1. David Justice and Andy Van Slyke each enjoyed moderately long (13+ seasons) careers, each gathering 1500+ hits, 3 All-Star designations, and numerous playoff appearances. Students may happen to know that Justice was married for four years to celebrity actress Halle Berry!

A classroom example is provided by Movshovitz-Hadar and Webb (1998, p.113) for two students, Pátria(P) and Bruce(B):

TABLE 2: Data on Classroom Tests

	fall semester		spring semester		full school year	
	P	B	P	B	P	B
Tests passed	5	8	6	4	11	12
Tests failed	1	2	8	6	9	8
Tests taken	6	10	14	10	20	20

Table 3 is a 2x2x2 table or numerical representation of a university hiring scenario (with simplified numbers):

TABLE 3: Hiring Data (by gender and academic area)

	Social Sciences		Physical Sciences		OVERALL	
	male	**female**	**male**	**female**	**male**	**female**
Hired	5	30	50	15	55	45
Denied	15	50	30	5	45	55
Total Applied	20	80	80	20	100	100

It is straightforward to verify that women are hired at a higher rate than men within each of two departments, but at a lower rate than men when the data from both departments are pooled together. Being aware that hiring rate comparisons between men and women might be reversed when department categories are combined is an example on the list of essentials for citizenship by NCED (2001). If an "overall" 2x2 table of combined data had not been provided, some students might not have the habit or instinct to make it themselves to check for reversal.

This numerical table representation, however, does little to explain WHY or WHEN this happens, and a textbook (or teacher) is unlikely to offer an additional representation. We will share examples of additional representation (and the additional understanding they reveal) from the author's chapter (Lesser 2001) in the NCTM Yearbook on representations, the newest of the NCTM Standards (2000).

One of the simplest geometric representations merely requires working with the areas of rectangles within a unit square. Each row of Figure 1 shows the step-by-step assembling of hiring rate representations for the data in TABLE 3. For the last square in each row of Figure 1, we merely compare shaded proportions of the side-by-side rectangles.

By middle school, students have learned how to compute the arithmetic mean of a set of data, and this can be interpreted as the balance point or fulcrum for the data. This leads to the platform scale model for comparing hiring rates, as shown in Figure 2. Students can actually build a physical model for this with simple materials.

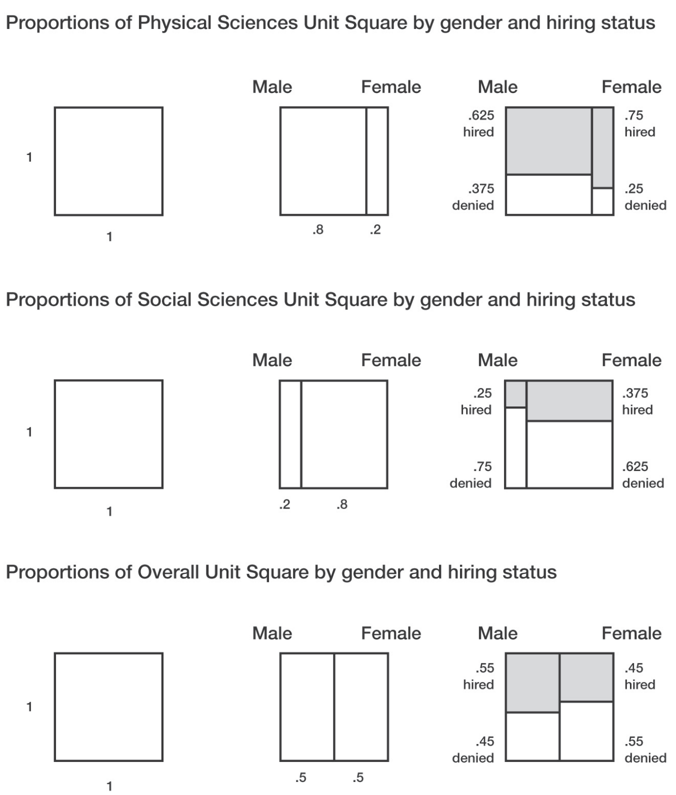


Figure 1: Unit Square Representation.
(Reprinted with permission from The Roles of Representation in School Mathematics, *copyright 2001 by the National Council of Teachers of Mathematics. All rights reserved.)*

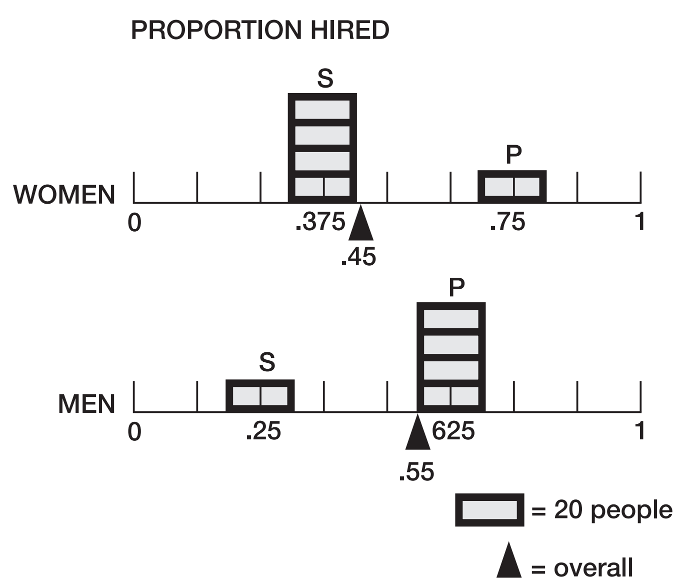


Figure 2: Platform Scale Representation.
(Reprinted with permission from The Roles of Representation in School Mathematics, *copyright 2001 by the National Council of Teachers of Mathematics. All rights reserved.)*

An interpretation of the weighted mean other than a fulcrum point employs a trapezoid. Specifically, the fact that a segment parallel to the nonparallel sides is the weighted mean of the two parallel sides(bases). Each base is weighted by the proportion of the trapezoid's height travelled towards that base to reach the weighted mean segment. In this figure (Figure 3), the parallel sides are vertical, of course.

Due to space constraints, teachers are invited to discuss and explore these representations (and how they show when Simpson's Paradox can occur) with their colleagues or students and/or to consult Lesser (2001). Lesser (2001) also offers examples of still more representations (mostly for high school teachers) such as: a circle graph representation, a vector(slope) representation, and a representation using determinants of 2x2 matrices. Also, Lord (1990) shows a representation using linear transformations.

We hope we have shown the added value of non-numerical representations, but there is one thing more that we can get from the original numerical information. A lesser-known but easy-to-check necessary condition for Simpson's Paradox is the existence of a plausible confounding factor for which there is a greater difference of percentage points than for the original factor of interest (Schield 1999). In our example, the male and female hiring rates are 55% and 45%, respectively – a difference of 10 percentage points. When we investigate the possible confounding factor of department, we see that the physical science and social science departments' hiring rates are 65% and 35%, respectively, a difference of THIRTY percentage points. Since 30 > 10, we would know Simpson's Paradox is possible here.

Let's close with the following simplified salary data. Ask students to explore and discuss which gender appears to be paid better at this company, supporting their answers with appropriate mathematical reasoning and representations!

Support Staff Employees at the Widget Company
70 males, each earning $20,000
90 females, each earning $30,000

Executive-Level Employees at the Widget Company
30 males, each earning $90,000
10 females, each earning $100,000

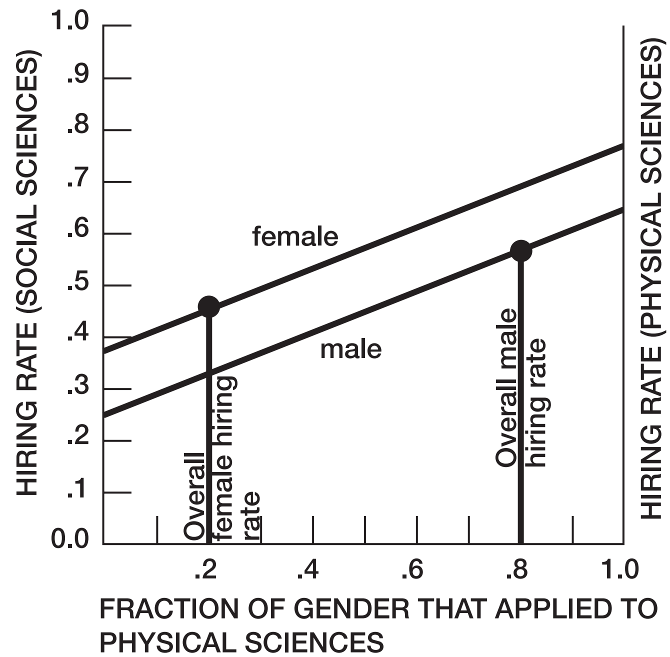


Figure 3: Trapezoidal Representation.
(Reprinted with permission from The Roles of Representation in School Mathematics*, copyright 2001 by the National Council of Teachers of Mathematics. All rights reserved.)*

Acknowledgement

This work is partially supported by the El Paso Mathematics and Science Partnership funded by the National Science Foundation, EHR 0227124.

References

Friedlander, R.J. (1992). Ol' Abner Has Done it Again, *American Mathematical Monthly*, 99, 84.

Lesser, L. M. (2001). Representations of Reversal: An Exploration of Simpson's Paradox. In Albert A. Cuoco and Frances R. Curcio (Eds.), *The Roles of Representation in School Mathematics*, pp. 129-145. Reston, VA: NCTM.

Lord, N. (March 1990). From Vectors to Reversal Paradoxes. *Mathematical Gazette*, 74, No. 467, 55-58.

Mitchem, J. (April 1989). Paradoxes in Averages. *Mathematics Teacher*, 82, 250-253.

Movshovitz-Hadar, N. and J. Webb. (1998). *One Equals Zero and Other Mathematical Surprises*. Berkeley, CA: Key Curriculum Press.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.

National Council on Education and the Disciplines. (2001). *Mathematics and Democracy: The Case for Quantitative Literacy*. Princeton, NJ: Woodrow Wilson National Fellowship Foundation.

Schield, M. (1999). Simpson's paradox and Cornfield's conditions. *Proceedings of the ASA Section of Statistical Education*, 106-111. Alexandria, VA: American Statistical Association.

Book Reviews

Bock, David; Velleman, Paul; De Veaux, Richard. 2004. *Stats, Modeling the World.* Pearson, Addison Wesley, New York.

David Bock and his co-authors have written an AP stat book that takes a different look at course structure and students' learning styles. They have included the standard material in shorter chapters and have added many "learning devices" that may appeal to a variety of teachers and students. Each chapter has an abundance of exercises, graphics, and sections that are aimed at improving students' retention of the material. Ample ancillary material is available.

Yates, Daniel; Starnes, Daren; Moore, David. 2005. *Statistics Through Applications.* W. H. Freeman and Company, New York.

Statistics Through Applications is a text designed for a one- or two-semester non-AP statistics course. The volume covers producing data, organizing data, chance, and inference with greater emphasis on using the calculator than actually "crunching" the numbers. A variety of exercises, web explorations, and cartoons are included in each chapter. Ample ancillary material accompanies the text including tests from a CD-rom that may be modified using a word processor.

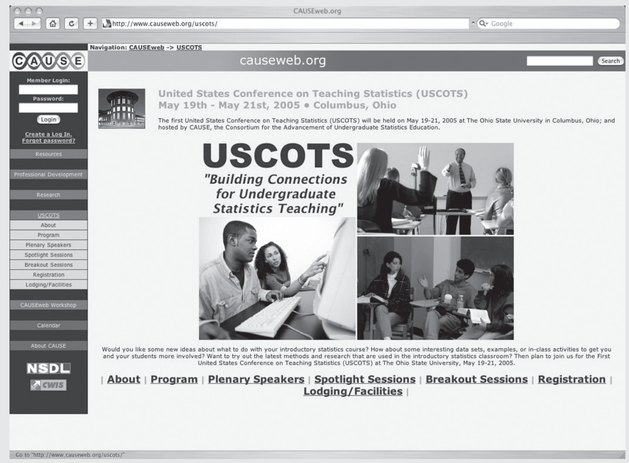
Millard, Ron and Turner, John. *Activities and Projects for High School Statistics Courses*, 2004. W. H. Freeman and Company, New York.

This is a soft-cover volume designed to accompany the statistics texts written by Daniel Yates (*Statistics Through Applications* and *The Practice of Statistics*). It contains activities and articles to create black-line masters that can be of use in a variety of courses from non-AP stat to college level statistics courses. The book contains 19 activities, seven complete articles, and instructions for a research project. Teachers who are looking for new ideas for their courses would certainly find a few to interest their classes.

Larsen, Michael. 2004. *Internet Companion for Statistics.* Thomson/Books/Cole, Australia. (*www.duxbury.com*)

Michael Larsen has assembled a short book that contains a multitude of Internet suggestions and accompanying questions and challenges. The topics include everything from probability to confidence intervals and hypothesis testing. Teachers of a wide variety of statistics courses from high school through college may find use for the suggestions of sources and activities in this book.

The First United States Conference on Teaching Statistics Aims to Build Connections for Statistics Teachers from All Disciplines



Would you like some new ideas about what to do with your introductory statistics course? How about some interesting data sets, examples, or in-class activities to get you and your students more involved? Want to try out the latest methods and research that are used in the introductory statistics classroom? Then plan to join us for the First United States Conference on Teaching Statistics (USCOTS) at Ohio State University, May 19-21, 2005.

The goals of USCOTS are to hold a national conference that focuses on undergraduate level statistics education (including AP Statistics); to share ideas, methods, and research results regarding what teachers want to know about teaching statistics; to facilitate teachers incorporating new ideas, methods, and resources into their existing courses and programs; and to promote connections between all teachers of undergraduate level statistics.

To register for USCOTS, visit the USCOTS website at http://www.causeweb.org/uscots. On the registration form you can sign up for a spotlight session. For more information, contact Deb Rumsey, USCOTS Organizer and Program Chair: rumsey@stat.ohio-state.edu.

Input Sought From Statistical Education Community On Pre K-12 Statistics Guidelines

Christine Franklin

The writing team of the report "A Curriculum Framework for Pre K-12 Statistics Education" requests comments and suggestions on the report from the statistical education community. The American Statistical Association (ASA) supported the development of this Framework though funding of a Strategic Initiative Grant submitted by the ASA Advisory Committee on Teacher Enhancement in March 2003. A draft of the report is available for public viewing at *http://it.stlawu.edu/~rlock/gaise/*.

Two primary objectives of the Framework are to provide a conceptual framework for Pre K-12 statistics education and to help educators work toward developing statistically literate citizens who can use statistics to make reasoned judgments, evaluate quantitative information, and value the role of statistics in everyday life. The report is also designed to provide stakeholders such as writers of state standards, writers of assessment items, educators at teacher preparation programs, curriculum directors, and Pre K-12 teachers with guidance in developing standards in statistics and data analysis as part of the Pre K-12 mathematics curriculum. The foundation for the Framework rests on the National Council of Teachers of Mathematics' (NCTM) Principles and Standards for School Mathematics (2000). The Framework is intended to support and complement the objectives of the NCTM Principles and Standards, not to supplant them.

The complete report is in two parts. The first part provides an overview of the goals and need for statistics education, a thorough discussion of the differences between statistics and mathematics, and a detailed presentation of the Framework Model. The second part is an appendix that presents descriptions of learning activities at each of three developmental levels; these provide illustrations for teachers of the types of activities that should be used in the classroom to effectively promote statistical literacy.

The ASA/NCTM Joint Committee and the ASA Executive Committee of the Section on Statistical Education have endorsed these guidelines "in spirit," and the Board of Directors of the American Statistical Association is expected to consider recommendations on the report later this year. **Please send comments or suggestions regarding the Framework to Christine Franklin (*chris@stat.uga.edu*).**

References

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: The Council.

Dear Readers,

From the Editors:

We hope that you find this issue of interest. Please pass it or a copy on to a fellow teacher as we hope that everyone can discover something of interest.

The fall issue will be filled with ideas for activities for many grade levels. Do you have any ideas that you would like to share with others? STN is soliciting articles on integrating probability and data analysis into upper elementary and middle school subjects OTHER than mathematics. What can you suggest? Send us a summary of any of your ideas. Thanks.

Sincerely,
Beth Lazerick, *stneditor@aol.com*
Murray Siegel, *siegel@gssm.k12.sc.us*

Send mailing address corrections and additions to: Madge Haven: *madge@amstat.org*

Statistics Teacher Network
C/o American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415
USA

Let us know your Zip +4!