Date: 12/10/2000

To: Judy Dill, ASA/NSF/BLS/Census Research Program
American Statistical Association
1409 Duke Street
Alexandria, VA 22314-3204
judy@amstat.org

From: Milo Schield, Ph.D., Associate Professor
Department of Business and Management Information Systems
Augsburg College
2211 Riverside Drive
Minneapolis, MN 55454
E-mail: schield@augsburg.edu

Topic: Application for the ASA/NSF/BLS/Census Bureau Fellowship Program

I submit my application for the 2001-2002 Fellowship program. The following are enclosed:

A. The names and addresses of three references that may be contacted

B. Detailed research proposal

C. Curriculum vitae

D. Copies of my publications that are most closely related to this proposal
Statistical Literacy: Thinking Critically about Statistics
Statistical Literacy and Mathematical Reasoning
Statistical Literacy and Evidential Statistics
Statistical Literacy, Simpson's Paradox and Cornfield's Condition
Common Errors in Forming Arithmetic Comparisons
Statistical Literacy: Student Difficulties in Describing and Comparing Rates and Percentages

E. Appendix: Student Difficulties in Describing Percentages in Tables
Table 1352 from the 1998 U.S. Statistical Abstract

SECTION A: REFERENCES

The names and addresses of three references that may be contacted

1. Wendy Treadwell, Past President of the Association of Public Data Users (APDU)
   Phone: 612/624-4389, FAX: 612/626-9353
   Machine Readable Data Center
   E-mail: wendy@mrdc.lib.umn.edu

   University of Minnesota
   2 Wilson Library
   309 19th Avenue South
   Minneapolis, MN 55455

2. Glenn King
   US Census Bureau
   Administrative and Customer Service Division
   ACSD, 4-1109, Phone: 301/457-1171
   Washington D.C. 20233-0140
   E-mail: Glenn.W.King@census.gov

3. Frederick Conrad, Ph.D., Senior Research Psychologist
   Phone: 202/606-7513
   Bureau of Labor Statistics
   2 Massachusetts Ave. NE; Room 4915
   Washington, DC 20212.
   E-Mail: Conrad_F@bls.gov

SECTION B: RESEARCH PROPOSAL

a. Title: ***Improving the actual utility of data to accommodate a wider range of statistical literacy among users.***

b. Application for any other fellowship: NONE.

c. Abstract:

This research project addresses the need of the ASA, the Bureau of the Census (BOC) and the Bureau of Labor Statistics (BLS) to increase the utility – the usability – of data. As BOC/BLS data reach a wider audience through the web, this audience is more likely to involve statistical non-professionals. And as the audience of readers becomes less professional in their training, there is increasing opportunity for them to misunderstand the data being presented. At this point, *the primary problem is not the quality of the data but the understandability of the data*.

This project involves two related activities. The goal of the first is to "improve the readability of rates and percentages in tables." The goal of the second is to "improve the information obtainable from data by modeling." Together they will help to improve the understandability of BOC/BLS data by non-professional users.

This project has the potential for significant results. Moving forward on either of these goals would be a significant step in supporting the BOC byline: "Helping You Make Informed Decisions. Moving forward on both of these goals would signal to non-professionals the commitment of BOC/BLS to increase the actual utility of their data for all – not just for statistically trained professionals.

d. Term: 12 month split term.
June through August 2001 and mid-December 2001 though mid-August 2002.

e. Background Information on Research Topic with References.

The *utility of data* is explicitly stated as a criterion for this fellowship: "We welcome projects that could lead to improvements in the quality and *usefulness* of our data." The *utility of data* is intimately related to *statistical literacy*: the ability of users to read and interpret statistics. All too often the *utility of data* is judged in terms of its potential to explain based on the variables and statistics involved. But ultimately the *utility of data* must be judged in terms of its actual usability by non-professional users: those people with basic levels of statistical literacy. Non-professional users require a higher level of actual utility – a higher level of immediate understandability – than do professional users.

Rates and part-whole percentages require high levels of statistical literacy. 40% (483) of the 1,232 tables in the 1997 U.S. Statistical Abstract contain either percentages (383) or rates (132). Based on my experience in teaching students to read tables of rates and percentages, I predict that non-professional users will have difficulty reading a significant share of the tables involving rates and percentages. See Schield (2000b): *Statistical Literacy: Student Difficulties in Describing and Comparing Rates and Percentages.* [A copy is included in section D.]

The importance of *statistical literacy* was highlighted at the 1999 Conference of the Association of Public Data Users (APDU). Some of these conference presentations were summarized in the second issue of APDU's "*Of Significance*" journal in a paper entitled "*Making Things Add Up for the End User: Issues in Statistical Literacy.*"

- Jocelyn Tipton, Yale University Library, wrote "With numbers joining dollars as the currency of policy debates in society, media and government, public understanding of quantitative information has become as important as public access to the numeric data."
- Katherine K. Wallman, Chief Statistician, Office of Management and Budget, challenged the audience to "heed the needs of our consumers in the larger society we hope to serve … by working for *statistical literacy*." See also her article entitled, "*Enhancing Statistical Literacy: Enriching our society*" in the March, 1993 Journal of the American Statistical Association,
- Marianne W. Zawitz, Bureau of Justice Statistics, said that, "data providers have an obligation to present data clearly and accurately… with a goal of *statistical literacy*."
- Wendy Treadwell, APDU President, described ways to bring *statistical literacy* initiatives to the local level. She asserted that APDU had been "sneaking the subject of statistical literacy into its conferences." She concluded by saying, "we need to actively address how well the data we present are being interpreted…" We must see what "we … can do to address the problem of *statistical literacy* within the general population."

The importance of statistical literacy has been highlighted elsewhere:

- David S. Moore, Past ASA President, has given many talks on *statistical literacy*. These were summarized in his 1998 Presidential Address to the ASA with the title, "*Statistics Among the Liberal Arts*." This paper was then printed in the 1998 issue of the Journal of the American Statistical Association (JASA), pp. 1253-1259.
- Cathryn S. Dippo, BLS Associate Commissioner for Survey Methods Research, published an article entitled, "*FedStats Promotes Statistical Literacy*." in the 1998 Journal of the Association for Computing Machinery,

Improving the actual utility – the immediate understandability – of data is under the direct control of data providers. Improving the understandability of their data is their primary contribution to increasing statistical literacy and must have their highest priority.

f. Statement of Relevant Work Already Accomplished by the Researcher

The following milestones have been achieved:

1. Researched the English grammar used to describe and compare rates and percentages. This research project used the Harper-Collins Cobuild 50 million-word corpus of machine-readable text at the University of Birmingham, England. The data involved totaled approximately 500 MB.

2. Developed the descriptive rules used to compare counts. These grammar rules were published in my paper entitled: *Common Errors in Forming Arithmetic Comparisons*. (1999c)

3. Developed rules to be used to describe rates and percentages. These grammar rules were published in my paper entitled: *Statistical Literacy: Student Difficulties in Describing and Comparing Rates and Percentages*. (Schield, 2000b)

4. Wrote a manuscript from which students can learn how to describe and compare rates and percentages. This manuscript has been used for over four years by a total of 700 students. This includes Describing Count-Based Data (50 pages) and Comparing Count-Based Data (80 pages).

5. Gave talks in 1999 at the Census Bureau, at Department of Labor Statistics and at the Association for Public Data Users (APDU) on student problems in reading tables prepared by the BLS and by Census. I have been invited by BLS and Census to give updated talks in early 2001.

6. Organized three conferences on Statistical Literacy at the Joint Statistical Meetings of the American Statistical Association. Speakers included Dr. Donald Rubin, Dr. John Bailar III, Dr. Jessica Utts and Victor Cohn (former science reporter for the Washington Post).

The following work is currently in progress:

1. Evaluating students' ability to read a wide variety of tables and graphs. This semester, I will have obtained about 20 student attempts on a total of 8 tables. Next semester, I will obtain data on another 8 tables. This data is the basis for forming research hypotheses on the causes of student difficulties in reading tables and graphs.

2. Testing to see if changes in the titles of tables and graphs significantly increase student comprehension. Students will be given modified versions of tables to see if their comprehension increases because of the changes.

3. Writing a booklet to teach students how to describe and compare rates and percentages contained in actual tables.

4. Researching how students describe associations both qualitatively and quantitatively. To date, students have generated some 300 examples. Half of these involve qualitative descriptions of association; half are quantitative descriptions of the slope. These 300 examples are being analyzed for underlying patterns.

g. Research Plan: *"Improving the actual utility of data to accommodate a wider range of statistical literacy among users"*

I propose a research project involving two activities:

1. Improve the readability of rates and percentages in tables.
2. Improve the information obtainable from data by modeling.

Activity 1: Improving the readability of rates and percentages in tables.
The first phase is *to identify the level of statistical literacy in the general population* in terms of their ability to properly read and describe rates and percentages contained in tables currently published by BLS and BOC. The inability to properly read these rates and percentages can be ascribed to two sources:

- Lack of statistical literacy in the general population. This is a problem in education.
- Lack of proper titles and headings in tables. This is a problem in data presentation that ultimately affects the actual utility – the understandability – of the data.

From the perspective of data providers, their primary focus must be on how to improve the utility of the data: how to better design the titles and headings in tables so non-professional readers can more easily decode and describe the rates and percentages presented in such tables. In the long term, data providers should participate in trying to raise the level of statistical literacy in the general population. This is my primary focus, but this is a long-term educational activity.

I predict that professional data providers will be surprised – if not amazed – at how much of a problem users have in reading tables which professionals consider trivial. Once the magnitude of this problem is recognized, we can move on to Phase 2 and examine different methods of reducing this problem.

The ultimate goal of this activity is to improve the readability of tables by identifying and implementing new standards for tables involving rates and percentages.

Activity 2. Improve the information obtainable from data by modeling.
Most of the data produced by BLS/Census are observational. There are two major problems involving observational data: confounding and misleading averages. Confounding is the influence of a third factor on the association between two other factors. Confounding is the major problem in using statistical associations based on observational data.

Users with typical levels of statistical literacy are not aware of the problem of confounding so they treat statistical associations as though they were absolute – just like in arithmetic. In the US in 1996, 13.1% of black babies were low birth weight and 6.5% of white babies were low birth weight. These readers presume that this 2 to 1 ratio is attributable entirely to the difference in race. They don't see that it might possibly be explained by other related factors. They don't see that this 2 to 1 association might be changed after taking into account a related factor. And there is nothing in the table to help them understand the relation between race and age of the mother.

Tables are the simplest way of indicating the influence of confounding factors. For example, in understanding the percentage of babies who have low birth-weight, relevant confounders are the age of the mother, prenatal care, along with the health, education, income and marital status of the mother. The 1999 US Statistical Abstract, Table 121 shows these percentages by the smoking status, age and race of the mother. Table 122 shows the percentage of mothers who drank during pregnancy. There is no table that relates the presence of low birth weight babies to the alcohol

use by the mother. Ultimately the problem is one of space: there simply isn't room to show tables with all these relevant confounders.

A related problem is the presence of misleading averages. This can occur when the averages reflect quantitative differences. In the Consumer Expenditure Survey, there are many relevant explanatory variables in explaining consumer unit expenditures (e.g., the number of people and the number of children) and consumer unit incomes (e.g., the number of wage earners). Yet all the data shown are averages over units having greatly differing internal characteristics. Even if all 2x2 cross-tabulations could be presented, this does not address the need for more complex cross-tabulations based on the number of people, the number of adults and the number of wage earners. One issue is the lack of space. But even if space were available, we want to summarize our knowledge as the amount of these quantitative determiners increases. (E.g., for each additional person in the family what happens to the expenditure for meat?)

Averages can also be misleading when they reflect qualitative differences. In the Consumer Expenditure Survey, the 1997-98 cross-tabulated table for those under age 25, shows an average expenditure for alcohol of $330 per year. But in many families, this is zero. Depending on the proportion of such families, the average amount in families with non-zero expenditures can be much higher. There is no way to obtain an estimate of the average expenditure for alcohol per consumer unit – *among just those units that spend money for alcohol.* The same is true for education. Showing the average amount spent on education is almost meaningless when perhaps half of those in the group are not spending anything on education.

SUMMARY: Observational data involves two major problems: confounding and misleading averages. Both problems could be "solved" by generating more tables but then space becomes a relevant constraint.

SOLUTION: A possible solution to all these problems is to summarize these relationships using regression equations – either linear or logistic. The advantage is two fold. First, many more variables can be included than would be possible in a table. Second, the influence of confounding between the variables analyzed can be taken into account in a systematic manner. Given different models, users can see the impact of a given variable on the association between other predictor variables and the outcome variable of interest.

The purpose of this activity is to investigate the feasibility – the costs and the benefits – of generating regression models.

DETAILED PLAN BY ACTIVITY AND PHASE:

Activity 1: Improving the readability of rates and percentages in tables.

Phase 1: Identify the problem

1. Meet with BOC/BLS managers to discuss the importance of the problem and the level of commitment needed to effect a substantial change.
2. Educate BOC/BLS personnel on the difficulties users have in describing rates and percentages and on the difficulties user have in decoding rates and percentages in tables.
3. Assemble a task force to oversee the pilot study of statistical literacy and readability of rates and percentages in tables currently published by BLS and Census. This pilot study is similar to the process I have used in teaching several hundred college students. It will involve the use of training materials that I have prepared and tested.
4. Identify 50 subjects who can serve as sample non-professional users. Identify 10 professionals who will serve as a sample of professional users. These 60 people will need 20 hours for training, 10 hours for study, 5 hours for testing and 5 hours for debriefing as follows:
   a. Test their native ability to accurately decode and describe rates and percentages in actual tables of rates and percentages currently produced by BLS/Census.
   b. Train them in using proper grammar to describe rates and percentages in 'toy tables:' tables with one word for the part and one word for the whole.
   These words are used to head specific columns or rows. These tables need no title.
   Now test these readers again on the same actual tables previously presented.
   c. Train them on decoding *part* and *whole* in actual tables with multiple variables in which the title gives partial clues as to the part or whole status of each.
   Now test them again on the same actual tables previously presented.
   d. Analyze the results and summarize for presentation and/or publication.
5. Identify data providers interested in measuring user understanding of ratio data in tables.
6. Identify data providers interested in working on improving the utility of tabular data.
7. Establish operational goals to improve the readability of tabular rates and percentages.

Phase 2: Test proposed solutions

1. Work with the task force to identify potential solutions to improve data utility.
2. Design a pilot test of these potential solutions.
3. Analyze the results and summarize for presentation and/or publication.

Phase 3: Summarize and institutionalize the results

1. Identify BOC/BLS staff interested in developing a short book.
2. Develop a booklet on reading tables of rates and percentages.
3. Investigate the importance of this focus to the mission of the bureau.
4. Establish internal standards on how rate and percentage tables should be presented to improve communication with non-professional readers.
5. Investigate the feasibility of establishing new operational sub-units to carry on these activities in a systematic fashion.

Activity 2: Improve the information obtainable from data by modeling associations.

1. Using data from the Current Population Survey, the Consumer Expenditure Survey, the Crime Victimization Survey and the Health survey, select variables of interest.
2. Select related variables; investigate the feasibility of modeling these variables.
3. Build relevant cross-sectional regression models using OLS or logistic regression.
4. Examine the sensitivity of these models to the classification and grouping of data.
5. Investigate the costs and benefits of generating such summary equations.
6. Investigate the costs and benefits of including these equations along with the tables.
7. Present these preliminary findings to the agencies involved.
8. Repeat the process with a new group of outcome variables to further investigate the sensitivity of the models to changes in explanatory factors.
9. Summarize the project for presentation and/or publication.
10. If feasible, establish operational goals to generate multivariate models for selected variables.

h. Significance of expected results.
I believe this project will have a major impact on the focus within the Bureau of Labor Statistics and The Census Bureau. Historically, quality has received more emphasis than utility from these data producers. This project aims to increase the focus on the utility of the data. The byline for the Census Bureau is "Helping You Make More Informed Decisions." To achieve this goal, users must be able to read and interpret the data. Improving the readability of tables (Activity 1) and identifying the relations between relevant confounders (Activity 2) will be a major leap forward toward the achievement of this goal with a wider audience.

This change in focus will have significance both externally and internally.

Externally, it will result in a substantial increase in the usability of BLS/Census data by non-professionals in high school and college classes. It will elevate the teaching in the Social Sciences by highlighting the inter-relationships between variables.

Internally, this fellowship could lead to some new directions in statistical research. I toured the psychology "labs" at the Bureau of Labor Statistics. The primary goal of these labs is to improve the accuracy of data entry. I asked how much effort was spent on seeing how users read the data being produced. My guide said he was unaware of any such efforts at this time. Such research is sorely needed and this fellowship will help direct efforts in that direction.

I believe that Activity 1 of this fellowship will lead to an increased focus on how to make data more readable. This will require an inter-disciplinary effort by statisticians, table designers, editors and teachers as well as by psychologists, linguists and those involved in the communication of data. Currently far more is spent on insuring the quality of the data than on insuring the readability of the data. At this point the biggest problem in the communication of data to the reader is not the quality of the data, but the reader's inability to understand the data.

I believe that Activity 2 of this fellowship will lead to an increased focus on the multivariate modeling of observational data. For a given table, various models could be presented as footnotes in the CD-ROM versions. Such models could indicate the influence of relevant confounders that exist within the data being presented. At the Census Bureau for example, the US Statistical Abstract presents data on low birth-weight babies. Relevant confounders might be the mother's age, race, education and marital status as well as her smoker/drinker status. There simply isn't space to present all possible tables relating these factors. And many of these separate tables might be very misleading even though each would be true. The influence of marital status could easily be a reflection of age. The presentation of the partial contributions of related factors in various multivariate models could be a major step forward in educating readers on the importance of context in interpreting the meaning of these statistics. At the Bureau of Labor Statistics, the Consumer Expenditure Survey could better communicate the complex inter-relationships between demographic variables (age, sex, race, family size and family composition) and non-demographic variables (education).

i. Advantages of conducting this research at BLS/BOC.
The key activity requiring direct access is running a pilot study to determine what fraction of the general population is able to accurately read the tables produced by BLS/BOC. It is one thing to hear an outsider say that only 50% of his students could accurately read a table (See Appendix E). It is another to obtain this result from a study organized, conducted and evaluated by BLS/BOC personnel. And even after the extent of the problem is identified, there is a second reason to have on-going access to BLS/BOC personnel. The people interested in improving the utility of data may be located in a variety of different units. It will take time to locate these people, develop relationships and identify and prioritize tasks that fit in with the ongoing activities of these two large institutions. It will take time to learn the constraints that BLS/BOC editors have in writing

the titles and headings for tables. It will take time for both of us to be able to communicate on a subject that is so familiar it is almost overlooked.

Updating of BLS/Census guidebooks on how to design and title tables will definitely require access to BLS/Census personnel. While BLS and Census certainly have guidelines that are more than adequate on the layout of complex tables of counts, there is definitely room for improvement in the titles of tables involving rates and percentages.

Multivariate modeling of specific data (e.g., the amount borrowed and the amount spent on interest as a function of assets and income or the percentage of low birth weight births) will require access to BLS/Census data and to the staff familiar with that data.

j. Requirement for research support and work facilities.

Direct support: Desk, phone and PC computer with MS Word, Excel, web access and appropriate statistical software: Minitab, SPSS, JMP and/or SAS.
Indirect support: A pilot study involving 50 people for a total of 40 hours each is an equivalent of around four person-years.

k. Budget required for appointment (Approximately $100,000).

- Salary costs ($90k): Current salary: About $70K gross per 10 month contract teaching 9 courses per year. For a 12 month appointment (e.g., June-August 2001, January-August, 2002) I request about an annual salary of about $90,000 to cover living expenses in the DC area.
- Benefits (zero): Benefits would be paid by Augsburg College as long as I teach at least 4 courses in the regular academic year. Otherwise, they must be covered by the fellowship.
- Research assistance costs ($3,000: $1,500 for Aug, 2001 and $1,500 for Aug 2002). I expect to present preliminary and final results of this research at the Joint Statistical Meeting of the American Statistical Association. Costs of registration, travel, lodging and food for six nights at the ASA Joint Statistical Meeting are about $1,200 - $1,500.

  In August 2001, I have applied to give a half-day seminar on describing and comparing rates and percentages. Part of that presentation would be on the difficulty of describing rates and percentages in tables and graphs. I also plan to give a contributed paper: "Statistical Literacy: Describing Associations Qualitatively and Quantitatively." Part of this paper would be to introduce the difficulties of reading tables and graphs whose titles are, by necessity, often shorter than would be most desired.

  In August 2002, I plan to present the results of this research jointly with representatives of BLS and Census at the Joint Statistical Meeting of the ASA.
- Hardware/Software Costs (around $1,000): I need access to JPM or SAS on my home laptop.
- Travel Costs ($3,000): Two round trips from Minneapolis to Washington are required for a split appointment. As MIS Coordinator at Augsburg College in Minneapolis, MN, I need to make another trip during the summer and three trips during the spring for a total of six round trips. At $500 per round trip, this totals $3,000.

SECTION C: RESUME AND EMPLOYMENT HISTORY

Summary: Dr. Milo Schield has presented papers on statistical literacy at the American Statistical Association, at the Bureau of the Census (BOC) and at the Bureau of Labor Statistics (BLS) as well as at the Making Statistics Effective in Schools of Business, the International Conference on Mathematics Education (ICME-9, Tokyo), the International Conference on Teaching Statistics (ICOTS-5) and the Association for Public Data Users (APDU). He has organized three sessions on statistical literacy for the American Statistical Association and he has done original research on users' understanding of rates and percentages as presented in tables and graphs. His research has helped to codify some of the unstated rules of English grammar governing descriptions and comparisons of rates and percentages. Dr. Schield has been invited to give a keynote talk on Statistical Literacy at the next International Conference on Teaching Statistics (ICOTS-7) in South Africa in 2002.

EDUCATION AND EMPLOYMENT HISTORY

Employment History at Augsburg College

1985 – 1990	Assistant Professor of Business & MIS, tenure track
1990 – 1991	Associate Professor of Business & MIS, tenure-track
1991 – 2000	Associate Professor of Business & MIS, tenured

Other Teaching Employment:

1968 – 1971	Instructor, Department of Physics, University of Iowa
1978 – 1983	Adjunct Instructor, National College
1985 – 1991	Adjunct Instructor, Quantitative Methods Dept., University of St. Thomas

Other related employment:

1976 – 1978	Fox & Co., CPA	Senior Management Consultant
1978 – 1984	St. Paul Insurance Co.	Senior Operations Research Analyst

Professional Background and Post-Secondary Education

1958 – 1962	Iowa State University	B.S.	Physics and Economics
1963 – 1965	University of Illinois	M.S.	Physics and Mathematics
1965 – 1968	Rice University	Ph.D.	Space Physics
1968 – 1970	University of Iowa	4 classes	Economics & Statistics
1978	National Exam	CMA	Certificate in Management Accounting
1978	National Exam	Actuarial	Passed Exam #1
1985 – 1988	University of Minnesota		Business, Government and Society

PUBLICATIONS IN STATISTICS:

Copies of the following are contained in this application:

Schield, Milo (2000b). *Statistical Literacy: Student Difficulties in Describing and Comparing Rates and Percentages*, 2000 ASA Proceedings of the Section on Statistical Education (to be published next June).

Schield, Milo (2000a). *Statistical Literacy and Mathematical Reasoning*. University Working Group, International Conference on Mathematics Education (ICME-9), Tokyo.

Schield, Milo (1999c). *Statistical Literacy: Thinking Critically about Statistics. Of Significance* journal. The Association of Public Data Users, Volume 1.

Schield, Milo (1999b). *Common Errors in Forming Arithmetic Comparisons. Of Significance* journal. The Association of Public Data Users, Volume 1.

Schield, Milo (1999a). *Statistical Literacy, Simpson's Paradox and Cornfield's Condition.* 1999 ASA Proceedings of the Section on Statistical Education, p. 106.

Schield, Milo (1998b). *Statistical Literacy and Evidential Statistics.* 1998 ASA Proceedings of the Section on Statistical Education, p. 187.

Copies of the following are available on my web site: www.augsburg.edu/ppages/schield or at www.augsburg.edu/ppages/~schield

Schield, Milo (1998a). *Teaching Bayesian and Classical Statistics.* International Conference on Teaching Statistics: ICOTS-5. Invited paper.

Schield, Milo (1997). *Interpreting Statistical Confidence.* 1997 ASA Proceedings of the Section on Statistical Education, p. 137.

Schield, Milo (1996). *Using Bayesian Reasoning in Classical Hypothesis Testing.* 1996 ASA Proceedings of the Section on Statistical Education, p. 274.

MEMBERSHIPS

American Statistical Association (ASA), Section on Statistical Education
International Association of Statistical Education (IASE)
American Mathematical Association of Two Year Colleges (AMATYC)

ORGANIZATION OF SPECIAL CONTRIBUTED SESSIONS AT THE ASA:
I have organized three special contributed sessions at the national Joint Statistical Meetings of the American Statistical Association.

2000 Session Theme: **Statistical Literacy**
Organizer and Chair: Milo Schield, Augsburg College

1. John Bailar, Professor of Epidemiology, Univ. of Chicago. *Thinking Big About Statistics*.
2. Phillip Shively, author of *Cross-Level Inference*, Professor of Political Science, University of Minnesota. *Cross-Level Inference as an Identification Problem.*
3. Chamont (Wei-hong) Wang, author of *Statistical Hypothesis Tests, Sense and Nonsense*, Professor of Statistics, The College of New Jersey. *A Case Story in the Teaching of Observational Studies*.
4. David Jabon, PhD, Director of the Quantitative Reasoning Program and Carolyn Narasimhan, Dean of Sciences, DePaul University. *A First Year Interdisciplinary Quantitative Reasoning Program.*
5. Joseph H. Abramson, author of *Making Sense of Data*, Professor of Epidemiology, Hebrew University. *Teaching Statistics for Use in Epidemiology*.

1999 Session Theme: **Critical Thinking on Observational Studies**
Organizer: Milo Schield, Augsburg College
Chair: John Bailar, Professor of Epidemiology, University of Chicago

1. Victor Cohn, author of *News and Numbers*, former Washington Post Science Reporter, Visiting Professor at the Harvard School of Health. *How to Help Reporters Tell the Truth* [Vic was diagnosed with cancer. He was unable to attend, so, with his permission, I gave his talk and wrote his final report for publication. He died shortly after this conference.]
2. Milo Schield, Augsburg College. *Simpson's Paradox and Cornfield's Conditions.*
3. Thomas Wonnacott, University of Ontario, Author of several statistics textbooks. *Population Growth and Prosperity. Lessons from Complex Observational Studies*.
4. Donald Rubin, Professor and Chair of Statistics Department at Harvard University, author of several books and many articles. *Teaching Causal Inference in Experiments and Observational Studies*.
5. Reviewer: Gudmund Iverson, Professor of Statistics, Swarthmore College.

1998 Session Theme: **New Directions in Introductory Statistics**
Organizer and Chair: Milo Schield, Augsburg College

1. Jessica Utts, author of *Seeing Through Numbers*, Professor of Statistics, University of California, Davis: *Educating Everyone: Statistical Methods and Statistical Literacy.*
2. Gary Smith, author of *Reasoning with Statistics*, Professor, Department of Mathematics at Pomona College. *Statistics for Liberal Arts Students*.
3. Gudmund Iverson, author of several books on statistics, Professor at Swarthmore College. *Teaching Statistics Without Formulas.*
4. Donald Macnaughton, President of MatStat Consulting. *Eight Features of the Ideal Introductory Statistics Course.*
5. Milo Schield, Augsburg College. *Evidential Statistics.*

Milo Schield: Talks On Statistics and Statistical Literacy

Date	Place	Invited by	Topic(s)
Aug 94	JSM Am. Stat. Assoc, Toronto, Canada		Sampling
Aug 95	JSM Am. Stat. Assoc., Florida		Correlation/Causes
June 96	MSMESB: Making Statistics Effective in Schools of Business. Anchorage Alaska		Statistical Literacy
July 96	Statistical Education Workshop, Sydney	Pamela Shaw	Statistical Literacy
July 96	SISC-96 Sydney International Statistical Conference, Australia		Bayes & Hyp. Test
Aug 96	JSM Am. Stat. Assoc., Chicago, Illinois		Bayes & Hypothesis Test
Sep 96	RSS Centre for Statistical Education & the University of Nottingham, England	Tony O'Hagan and Anne Hawkins	Bayes & Classical: Hypothesis Tests
Oct 96	De Montfort University, England	Nick Longford	Bayes & Hypothesis Test
Oct 96	University of York, York, England	Peter Lee	Bayes & Hypothesis Test
Nov 96	University of Edinburgh, Scotland	Tom Leonard	Bayes & Hypothesis Test
Dec 96	European Business Management School University of Wales, Swansea	Assad Jalali-Naini & Alan Watkins	Bayes & Hypothesis Test
Dec 96	University of Plymouth, England	Chris Ricketts	Resampling
Mar 97	MCOTS Oshkosh, Wisconsin	K.L.D. Gunawardena	Bayes & Hypothesis Test
Aug 97	JSM, ASA Anaheim, California		Confidence Interval
March 98	WestCoTS: Colorado Springs, Colorado	Jim Rutledge	Evidential Statistics
June 98	MSMESB, Making Statistics Effective in Schools of Business. Univ. Iowa, Iowa City	John Cryer, Organizer	Evidential Statistics
July 98	Xi'an Statistical Institute, Xi'an China		Statistical Literacy
July 98	ICOTS-5: Singapore Malaysia	Jeff Witmer	Teaching Inference: Bayes vs. Classical
Aug 98	JSM Am. Stat. Assoc. Dallas Texas	Organized session	Evidential Statistics
Nov 98	University of Northern Iowa, Cedar Falls, Iowa	Joel Haack and Kirmani	Teaching Bayesian & Freq. Stats.
16 Feb 99 17 Feb 99	University of Ballarat, Ballarat, Australia, Victoria	Lyn Roberts	Statistical Literacy; Reading Tables
16 Mar 99	University of Technology Sydney UTS Sydney, Australia, NSW	Peter Petocz and Beverly Moore	Evidential Statistics
17 Mar 99	University of Newcastle, Newcastle, Australia NSW	Keith Dear, Gita Mishra & Bob Gibberd	Simpson's Paradox & Minimum Effect Size
18 Mar 99	Macquarie University Sydney Australia NSW	Pamela Shaw	Evidential Statistics
23 Mar 99	Statistical Society of Australia, New South Wales Branch, Univ. of Sydney.	Jennifer Kelley, Ed Bosworth & Eric Sowey	Simpson's Paradox & Minimum Effect Size
24 Mar 99	University of Wollongong, Wollongong, Australia, NSW	Catherine Milne, Anne Porter & David Griffiths	Evidential Statistics
9 Aug 99	JSM Amer. Stat. Assoc. Baltimore Md.		Simpson's Paradox
10 Aug 99	JSM Amer. Stat. Assoc. Baltimore Md.	Organized session	Statistical Literacy
24 Oct 99	US Bureau of the Census	Glenn King	Reading Tables
25 Oct 99	US Bureau of Labor Statistics	Frederick Conrad	Reading Tables
25 Oct 99	APDU: Assoc. of Public Data Users	Wendy Treadwell	Reading Tables
15 Mar 00	Western Conference on Teaching Statistics		Large Datasets
			Grammar of Rates
26 July 00	International Conference on Mathematics Education (ICME-9, Tokyo)	By invitation only. Invited by Lynn Steen	Statistical Literacy and Mathematics
9 Aug 00	JSM Am Stat. Assoc. Indianapolis, IN	Organized Session	Statistical Literacy
8 Aug 00	JSM Am Stat. Assoc. Indianapolis, IN		Describing & Comparing Rates and Percentages

OTHER

Educational Testing Services (ETS) Reviewer: I was invited to be a reviewer of an ETS computer-based assessment of "Core Skills." The goal is to assess the workplace skills of college graduates or of college students at the end of a 2-year or 4-year academic program.

Book Reviewer: I have reviewed two statistics books (Statistics By Example and the Student Edition of Minitab, Version 12) at the invitation of the publisher.

Software Beta Tester: I have been a beta tester for Minitab statistical software for the last three versions. I am also listed on their accredited author program list.

CONFERENCE ATTENDANCE

Attendance at ongoing conferences on statistical education:

1994 – present: Attended the annual national meeting of the American Statistical Association. During this five-day meeting, I normally attended at least 3 sessions (1.5 hours each) on statistical education. I usually attended a half-day workshop as part of the pre-conference activities.

1994 – present: Attended the annual national meeting of the Making Statistics Effective in Schools of Business (MSMESB). During this two-day meeting, I normally attended almost all of the sessions on education.

Attendance at specific workshops or conferences on teaching statistics:

1993: Received an NSF scholarship to attend a weeklong national statistics workshop (invitation only) on teaching of statistics to majors in the social sciences and humanities.

1995 Received an ASA scholarship to attend a weeklong national statistical education workshop (invitation only) on the teaching of statistics to majors in mathematics and science.

1996 Attended the Australian National Statistics Education Workshop (AuSEW) at Sydney.

1996 Attended the Sydney International Statistics Conference, Sydney, Australia.

1996 Spent a semester at the Royal Statistical Society Centre for Statistical Education studying the teaching of statistics in Great Britain. [Augsburg Sabbatical]

1997 Attended the Midwest Conference on Teaching Statistics (MCOTS)

1998 Attended the Western Conference on Teaching Statistics (WesCOTS)

1998 Attended the International Conference on Teaching Statistics, Singapore (ICOTS-5)

1999 Attended the National meeting of the American Mathematical Association at Two Year Colleges in Pittsburgh, PA.

2000 Attended the Western Conference on Teaching Statistics (WesCOTS)

2000 Attended the International Conference on the Psychology of Mathematics Education in Hiroshima, Japan. (PME-24)

2000 Attended the International Conference on Mathematics Education in Tokyo. (ICME-9)

SECTION D: RELEVANT PUBLICATIONS

The following are copies of my publications that are most relevant to this fellowship:

STATISTICAL LITERACY:

Schield, Milo (1999c). *Statistical Literacy: Thinking Critically about Statistics.* *Of Significance* journal. The Association of Public Data Users, Volume 1.

Schield, Milo (2000a). *Statistical Literacy and Mathematical Reasoning*. University Working Group, International Conference on Mathematics Education (ICME-9), Tokyo.

Schield, Milo (1998b). *Statistical Literacy and Evidential Statistics.* 1998 ASA Proceedings of the Section on Statistical Education, p. 187.

CONFOUNDING IN OBSERVATIONAL STUDIES

Schield, Milo (1999a). *Statistical Literacy, Simpson's Paradox and Cornfield's Condition.* 1999 ASA Proceedings of the Section on Statistical Education, p. 106.

DESCRIBING RATES AND PERCENTAGES IN TABLES

Schield, Milo (2000b). *Student Difficulties in Describing and Comparing Rates and Percentages* 2000 ASA Proceedings of the Section on Statistical Education (to be published next June).

Appendix: Student Difficulties in Describing Percentages in Tables

BACKGROUND: Students in my statistical literacy classes spend half a semester learning how to describe rates and percentages as presented in tables and graphs. Initially these tables are "toy" tables – tables having minimal complexity. Typically, these tables have one word for each column and each row. The rows and columns each have a single word as the heading. Typically there is no title. These toy tables are column tables, row tables or total tables. Once students have demonstrated their proficiency with these toy tables, then they are given actual tables from the publications of the BLS and Census. The students are asked to describe a specified rate or percentage using a complete sentence that accurately describes the part and the whole(s) involved. Photocopies are made of the students' written sentences. These sentences are transcribed and analyzed to identify common errors in reading these tables. Once a certain type of error is recognized, it is often easy to explain how the titling or layout of a particular table might engender this kind of mistake.

Consider the following study based on Table 1352 of the 1998 U.S. Statistical Abstract.

No. 1352. Participation in Job-Related Continuing Education and Training, by Country

[**Percentage of the employed population 25 to 64 years old.** Data refer to all job-related education and training organized, financed, or sponsored by authorities, provided by employers, or self-financed. Job-related continuing education and training refers to all organized, systematic education and training activities in which people take part in order to obtain knowledge and/or learn new skills for a current or a future job, to increase earnings, improve job and/or career opportunities in current or other fields, and generally to improve their opportunities for advancement and promotion]

COUNTRY	Year	Total	Male	Female	COUNTRY	Year	Total	Male	Female
United States [1] [2]	1995	34	31	36	Germany [1] [7]	1994	33	35	31
Australia [1] [3]	1995	38	38	38	Greece [4]	1994	1	1	1
Austria [4]	1995	8	8	7	Ireland [4]	1994	4	3	6
Belgium [4] [5]	1994	3	3	2	Italy [4] [8]	1995	1	1	1
Canada [1]	1993	28	27	30	Spain [4]	1995	3	2	5
Denmark [4]	1995	15	13	17	Sweden [9]	1996	42	39	44
Finland [1] [6]	1995	45	45	45	Switzerland [1]	1996	35	38	31
France [1]	1994	40	38	43	United Kingdom [4]	1995	12	11	13

[1] During the 12-month period preceding the survey. [2] Excludes full-time students. [3] Excludes persons enrolled only in full-time programs at any time during the survey period and people pursuing only on-the-job training. [4] During the 4-week period preceding the survey. [5] Covers all forms of continuing education and training. [6] Includes only the training of employees which is sponsored by employers. [7] Excludes initial training of students over 25 years old in vocational schools and in the dual system and forms of continuing vocational training other than formal courses. [8] Excludes persons enrolled in ordinary school courses. [9] During the 6-month period preceding the survey. Includes only training supplied or sponsored by employers.

Source: Organization for Economic Cooperation and Development, Paris, France, *Education at a Glance*, annual (copyright).

APPROACH: Using the Excel form of this table, I deleted all those countries where the time period of the survey was less than 12 months. I also removed a great deal of material from the title along with all of the footnotes. This was done to help students focus on correctly identifying the whole(s) and the part in describing a particular percentage.

Participation in Job-Related Continuing Education
and Training During the Past 12 Months (1994-1996)
Percentage of the employed population 25-64 years old.

COUNTRY	Total	Male	Female
United States	34	31	36
Australia	38	38	38
Canada	28	27	30
Finland	45	45	45
France	40	38	43
Germany	33	35	31
Sweden	42	39	44
Switzerland	35	38	31

1998 US Statistical Abstract, Table 1352.

TASK. For females, describe the 45% for Finland. Use "percentage."

CORRECT ANSWER: In this table, *participation in Continuing Education* is the *part* in each *part-whole* percentage. A correct answer would be, "45% is the percentage *of* the female employed population, age 25-64 in Finland *who* had participated in job-related continuing education and training during the past 12 months (1994-1996)."

RESULT: 10 of these 20 students (50%) failed to identify "participation" as the *part* in this table. 6 of these 20 students (30%) identified "employed" as the part.

FINDINGS: These 20 students selected the following terms as *part:* "participation" (10), "employed" (6), female (1), Finland (1) and one student had no clue about part versus whole. Of these 20 students, 6 omitted the word "participation", 3 omitted the word "employed", 3 omitted the word "Finland", and one omitted the word "female". Some students made multiple mistakes.

ANALYSIS: In the title of this table, the *part* (participation) was contained in a self-standing phrase that had no part-related keyword such as "of participation" or a relative clause "who participated." The only part-whole keyword in the title was *of*. These students had been taught that *percentage of* can introduce either the *part* or the *whole*.

HYPOTHESIS ABOUT CAUSE OF ERROR: The hypothesis depends on the error being made:

- "Employed" as *part* error (6 in 20): Consider the table heading: "Percentage of the employed population…" Hypothesis: These students treated the adjective (employed) in this phrase (employed percentage) as the *part* and treated the noun (population) as the *whole*. Normally such an adjective can be restated as a relative clause (percentage of the population who are employed). And in this transformed case, "employed" is properly read as being the *part*.
- "Female" as *part* error (1 in 20): Hypothesis: This student treated "*Population*" as including both male and female. So when asked to identify a percentage in the female column, this student had to treat female as the *part* – even though the female and male percentages did not add up to the total percentage for that country.

NEXT ACTION: These hypotheses need to be validated. This test will be given again to different group of students after restating the title to see if the mistakes decrease significantly.