INTRODUCTION: Milo Schield, Professor of Business Administration at Augsburg College, an elected member of the International Statistical Institute, and the editor and webmaster of www.StatLit.org. I have reviewed papers for *The American Statistician* (ASA) and for *Numeracy* (NNN). I have been chair of the Department of Business Administration at Augsburg and am currently the coordinator of the Management Information Systems (MIS) area. I teach traditional statistics at the undergraduate and graduate level (MBA); I teach statistical literacy for students in non-quantitative majors. I have written nine invited chapters or articles in journals, books or conference proceedings. I have presented two invited webinars for E-COTS. I have presented 56 contributed papers on statistics and statistical literacy. The Schild family home is in Bellenberg, Lippe, Germany ☺

SRTL9 Theme #4: What are rudimentary ideas of models and modelling and how are they expressed among young students? (e.g., what is a model? what does it mean to model?)

Title: Comparisons, Named Ratios and Weighted Averages as informal Doorways to Modelling

Introduction and Overview: Models typically involve mathematical relationships between variables that are expressed algebraically. For example, Galileo noted that the period of a pendulum (T) varied as the square root of the length (L): $T \propto Sqrt(L)$. This paper argues that arithmetic comparisons are a doorway to models. Schield (1999) examined four kinds of arithmetic comparisons: simple differences, simple ratios, percent change and times change. The simple ratio includes the part-whole ratio: arguably one of the first ratios that children learn. Schield (2011) studied the use of informal grammar (prepositions and relative clauses) to express part-whole relationships. E.g., "Three out of four Doctors like Crest." Schield (2000) and Schield and Burnham (2007) analyzed the use of named ratios such as Rate (Prevalence and Incidence), Percentage (Proportion) and Chance (Odds, Risk, Likelihood and Probability) to express part-whole relationships or numerator-denominator relationships; Likely and Prevalent are the named ratios used to compare ratios. All of these named ratios are used to express simple arithmetic relationships using ordinary English. Finally, Schield (2006) argued that weighted averages are the keystone to more abstract algebraic models where something is being 'taken into account' or 'controlled for.' These uses of informal grammar are argued to be doorways to modelling. Schield (2008) argued that those teaching arithmetic at the school level are encouraged to rely on these informal uses as a doorway to the abstract modelling.

Literature Review: These are contained in the papers mentioned in the introduction. The primary source for analyzing ordinary English is the Cobuild WordBank: the world's largest database of written and spoken English. Cobuild was accessed in two stages. In 1999 we downloaded and analyzed 380,000 lines of text (190 mb@512 bytes per line) involving 99 categories of named ratio descriptions and comparisons. In 2010-11, we made a second pass and downloaded 3,219 files (2,034 mb) containing more than 4 million lines of text in more than 500 categories. This research spanned several man-years.

Theoretical Framework and Methodology: We tried to avoid imposing any kind of theoretical structure on the sentences being analyzed. We tried to identify the most common grammatical patterns used to describe descriptions and comparisons of ratios. We tested our results on more than a thousand students to see which forms were most problematic.

We found the comparisons of ratios to allow users to take into account (control for) at least two different types of confounders. The test-base comparison specified the two groups being compared while the underlying named ratios took into account (controlled for) the size of the different groups involved.

We found that phrases such as "after controlling for" and "after taking into account" to be fairly common in the Cobuild database. Students quickly realized that controlling for the size of different size groups could change their ranking. For example, unemployment (the number unemployed) may be higher in California than in Wyoming, but the unemployment rate (the percentage of workers who are unemployed) may be higher in Wyoming than in California. But they had no idea that rates or percentage could change after controlling for a third factor.

Our article in Stats magazine (Schield, 2004) featured a graphical device that allowed students to take into account the influence of a relevant confounder. After using this graphical technique, students had a much better understanding of how controlling for a factor could influence both the size of a ratio but also the comparison of two ratios.

Finally Schield (2008) summarized these findings and argued that bright high school students who planned to select non-quantitative majors (majors that did not require a quantitative course) should take statistical literacy as an alternative to Algebra II. According to the American Freshman Survey (2012), 29% of college freshmen took a statistics course in high school; 15% took AP statistics. This means that 14% took a non-AP statistics course. Schield (2008) noted that 60% of college graduates are in majors that are required to take calculus or statistics. The other 40% are in non-quantitative majors: majors that don't require a specific quantitative course.

Schield (2008) argued that secondary schools should offer a course in statistical literacy as an alternative to Algebra II for the 40% of college-bound students who plan to graduate in non-quantitative majors.

Presentation: Our presentation will survey the results of our research on the use of ordinary grammar to describe and compare ratios. The use of ordinary English will be connected to the level of mathematics used in various grades at the primary and secondary level. It will also review the use of the graphical approach (along with a proportional reasoning and an algebraic approach) to controlling for a binary confounder – and student understanding and appreciation of these three approaches. With this background, we will argue that the school mathematics should be infused with the use of ordinary English to express descriptions and comparisons of ratios – and to identify those factors taken into account by models. In so doing, students will be better prepared to handle the use of various models in future math courses.

Contribution to the Theme: Statistical modelling should connect to informal techniques using ordinary English. By focusing on the use of ordinary English, this paper may offer a unique prospective on students' rudimentary ideas concerning statistical models. And by focusing on ordinary English, it allows K-6 teachers to introduce important statistical concepts without the need for formal training in mathematics or statistics.

BIBLIOGRAPHY:

Pryor, J., et al (2012). American Freshman National Norms 2012. Cooperative Institutional Research Program (CIRP) at the Higher Education Research Institute (HERI) at the University of California Los Angeles (UCLA). Copy at www.heri.ucla.edu/monographs/theamericanfreshman2012.pdf

Schield, Milo (1999). Common Errors in Forming Arithmetic Comparisons, 1999 APDU *Of Significance*, pp. 47-48. See www.StatLit.org/pdf/1999SchieldAPDU2.pdf.

Schield, Milo (2000). Statistical Literacy: Difficulties in Describing and Comparing Rates and Percentages. 2000 American Statistical Association Proceedings of Section on Statistical Education, pp. 76-81. See www.StatLit.org/pdf/2000SchieldASA.pdf.

Schield, Milo (2006). Understanding Confounding from Lurking Variables using Graphs. *STATS Magazine*, American Statistical Association. Fall 2006. pp. 14-18. Copy at www.StatLit.org/pdf/2006SchieldSTATS.pdf.

Schield, Milo and Thomas Burnham (2007). Grammar of Statements Involving "Chance", 2007 American Statistical Association Proceedings of the Section on Statistical Education. [CD-ROM] pp. 2235-2242. See www.StatLit.org/pdf/2007SchieldBurnhamASA.pdf.

Schield, Milo (2008). Quantitative Literacy and School Mathematics: Percentages and Fractions in *Calculation vs. Context: Quantitative Literacy And Its Implications for Teacher Education*. Edited by Bernard L. Madison and Lynn Arthur Steen. 2008 Mathematical Association of America. P. 87-107. See www.statlit.org/pdf/2008SchieldMAA.pdf or www.maa.org/Ql/cvc/cvc-087-107.pdf.

Schield, Milo (2011) Describing Arithmetic Relations Using Informal Grammar. 2011 American Statistical Association Proceedings of the Section on Statistical Education. [CD-ROM] P. 911-925. See www.StatLit.org/pdf/2011SchieldASA.pdf

BACKGROUND OF THE CONFERENCE:
http://www.isi-web.org/8717
http://blogs.uni-paderborn.de/srtl/files/2014/05/SRTL9-Preliminary-Announcement_final.pdf

The Ninth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-9) will be held from 26 July 2015 to 1 August 2015 at the Waldhotel Nachtigall in Paderborn - Schloss Neuhaus, Germany. The conference theme will focus on "**Reasoning about Models and Modelling in the Context of Informal Statistical Inference**."

Our discussions suggest that further ideas and pedagogical approaches related to **models and modelling** are relevant and important to study to allow us to better integrate and/or extend previous work in the context of reasoning about informal statistical inference [ISI].

We invite focused contributions that study one or more of the following themes in relation to statistics education:

1. Why bring models and modelling into the research and practical arenas at all? (e.g., what are philosophical, historical, epistemological and/or practical reasons for introducing models and modelling in statistics education research? How is mathematical modelling the same and/or different than statistical modelling?)
2. According to G. Box some models are useful: What are the utilities and purposes of model and modelling? What is a model / modelling for?
3. How does reasoning about models and modelling develop in the context of learning to make ISIs from data?
4. **What are rudimentary ideas of models and modelling and how are they expressed among young students? (e.g., what is a model? what does it mean to model?)**
5. How are ideas related to models and modelling understood and used by students in making ISIs? (e.g., what ideas are needed to understand and use models? what does it mean "to understand a model?")
6. What are innovative tasks, tools, or sequences of instructional activities that may be used to help these ideas emerge?
7. How can technology help to develop students' reasoning about models and modelling in the context of making ISI?
8. What are ways to assess reasoning about models and modelling?

What new approaches can be used to help teachers develop students reasoning about models and modelling?