Statistical Journal of the IAOS 27 (2011) 173–183
DOI 10.3233/SJI-2011-0732
IOS Press

Statistical literacy: A new mission for data producers

Milo Schield
Director, W. M. Keck Statistical Literacy Project, Vice President of the National Numeracy Network, Augsburg College, Minneapolis, USA
E-mail: milo@pro-ns.net

Abstract. Statistical literacy is a new goal for statistical educators. A core element of statistical literacy for consumers is the ability to read and interpret data in the tables and graphs published by national statistical offices. A core element for producers is the ability to create tables, graphs and reports that are unambiguous and comprehensible. It appears that comprehensibility is not considered part of the mission for many national statistical offices (NSOs). Yet can the staff or users read the data generated by these agencies? The 2002 W. M. Keck Statistical Literacy survey indicates that professional data analysts, college students and school mathematics teachers and even college professors have difficulties reading such data. A common reason is confusing captions. Other reasons include user difficulties in decoding tables and in using ordinary English to describe part-whole relations. Recommendations include vetting agency tables and graphs for comprehensibility, assessing the statistical literacy of staff and users, and developing objective standard standards for using ordinary English to describe rates and percentages and for titling such tables. Establishing these standards can help teachers improve the statistical literacy of students and future leaders so they can use agency-generated data to make better decisions.

1. Statistical literacy

While the phrase "statistical literacy" has a long history, it has only recently become a goal for statistical educators. In 1979, "statistical literacy" was the title of a textbook [11]. In 1982, "statistical numeracy" was described in the Cockcroft report [8]. In 2001, "statistical literacy" was an IASE conference theme. In 2002, "Developing a statistically literate society" was the theme of the International Conference on Teaching Statistics (ICOTS-6). In 2006, statistical literacy was adopted as a goal by the American Statistical Association (ASA) in endorsing the ASA Guidelines for Assessment and Instruction in Statistics Education (GAISE) [1]. This goal is stated in the first sentence in the PreK-12 portion of the GAISE report, "*The ultimate goal: Statistical Literacy*," and in the first recommendation of the College GAISE report: "introductory courses in statistics should, as much as possible, strive to emphasize statistical literacy and develop statistical thinking . . . " A "statistically literate society" is a goal in the ASA Strategic Plan for Education [2]: "Through leadership in all levels of statistical education, the ASA can help to build a statistically literate society . . . " Assessing statistical literacy is reviewed in recent articles [5,10, 25].

The increased attention to statistical literacy does not mean there is clear agreement on its definition [25]. A lack of agreement on the definition or its relevance to the IASE mission may explain the omission of "statistical literacy" in subsequent IASE conferences. It may be that "statistical literacy" is just a buzz-word that lacks substance or staying power. Yet a 2009 MAA survey of US four-year colleges found that 17% offered a Statistical Literacy course [24].

But analyzing differences in definition and approach can be bypassed – and entanglement with a potential fad can be avoided – if there is agreement on what statistical literacy involves. Gal's statement clearly identifies a key element:

– statistical literacy involves the ability to read and interpret the data in tables and graphs published by government statistical associations [9].

This measure of statistical literacy for data consumers requires that tables and graphs produced by government statistical associations are unambiguous, clear and comprehensible. Producing such tables and graphs requires a high level of fluency in data presentation by these agencies.

2. Missions of national statistical offices

So how does statistical literacy or fluency in data presentation relate to the mission of national statistical offices? Consider these mission statements by the U.S. Census Bureau:

- Mission: "to be the preeminent collector and provider of timely, relevant and quality data about the people and economy of the United States." Goal: "to provide the best mix of timeliness, relevancy, quality and cost for the data we collect and services we provide." [32]
- Mission: "The Census Bureau serves as the leading source of quality data about the nation's people and economy. We honor privacy, protect confidentiality, share our expertise globally, and conduct our work openly. The production of high quality, relevant statistical information rests on principles that the Census Bureau holds dear. Openness to user and respondent concerns, independence and neutrality, strong statistical standards, and protection of confidentiality form the foundation for the work we do." [33]

Assessing or enhancing the comprehensibility of their tables and graphs is not listed as a high-level goal for this agency.

Now consider the aims of the UK Office of National Statistics [27].

- to provide authoritative, timely and accessible statistics and analysis that enable decision making across UK society, anticipate needs and support public accountability
- to be a trusted and leading supplier of national government statistical expertise and surveys
- to maintain a dynamic portfolio of statistical sources, which reflects changing data needs
- to deliver the sources portfolio in a way that meets user expectations of quality within the available resources
- to minimise the burden on respondents for all survey collections
- our people, systems and processes are able to develop the current business and to respond rapidly to changing demands
- to identify social and technological changes that will impact on what we do and how we do it

These aims mention providing statistics that "enable decision making" but make no mention of measuring the comprehensibility of their statistics. Yet banks and insurers are required to ensure that their forms and publications meet certain objective standards for comprehensibility.

Consider the 2008 IAOS conference: Reshaping Official Statistics [13]. Areas covered included:

> "Use of administrative data in the statistical system, Use of administrative data in official statistics, Challenges of building register based or other administrative based statistics, More efficient use of statistical data, Questionnaire design and testing, User demands for official statistics, Electronic reporting, and Process orientated statistical production."

Aside from possible "user demands," it appears that the comprehensibility of data was not a significant agenda item at this conference. Questionnaires may undergo extensive testing, but is there any testing of whether the tables and graphs produced are comprehensible by the general public?

Yet there are signs that comprehensibility may be emerging as a goal for some NSOs.

- In 2008, the International Statistical Literacy Project (ISLP) presented "programs of some National Statistical Offices (NSOs) whose only purpose is to increase the level of statistical literacy of the public." "By a successful program, we mean a program that has reached the front page of the National Statistical Office web site, that is, a program that constitutes an intrinsic part of the general public." [17]
- The focus on projects, such as "the Census in Schools" project, indicates some support by NSOs for education and statistical literacy for consumers.
- The Statistics Education Unit of the Australian Bureau of Statistics identified criteria for statistical literacy and presented statistical literacy competencies by grade in school [3].

In summary, statistical literacy does not appear as a high priority with many National Statistical Offices although this may be changing. Once NSOs view statistical literacy as central to their mission they can extend their production-styles missions (to generate data that is accurate, timely, and relevant to their user's needs) to include market-driven missions: to generate accurate

No. 149. Death Rates for Injury by Firearms, Sex, Race, and Age: 1995

[**Death rate per 100,000 population.** Deaths classified according to the ninth revision of the *International Classification of Diseases*]

ITEM	5-14 yrs. old	15-24 yrs. old	25-34 yrs. old	35-44 yrs. old	45-54 yrs. old	55-64 yrs. old	65-74 yrs. old	75-84 yrs. old
MALE								
Firearms: White	2.5	31.4	26.1	21.2	19.6	19.9	26.1	39.8
Black	5.5	140.2	94.4	46.6	32.1	24.3	22.0	20.9
Accidents: White.	0.7	1.8	0.8	0.6	0.5	0.4	0.6	0.7
Black.	0.8	4.3	1.5	(B)	(B)	(B)	(B)	(B)
Suicide: White	0.8	15.4	15.1	14.2	14.9	16.6	23.9	38.2
Black.	(B)	13.2	11.9	7.6	6.9	7.5	10.2	13.9
Homicide: White	0.9	13.6	9.8	6.3	4.0	2.8	1.5	0.8
Black.	4.1	121.0	80.7	38.3	24.6	15.9	10.8	(B)

Fig. 1. U.S. Death Rates for Injury by Firearms, Sex, Race and Age.

and timely data that is comprehensible by and useful to decision makers.

But are the tables and charts published by National Statistical Offices comprehensible? This is a critical question for NSOs that justify their existence by producing data that is supposedly useful in making decisions. Consider two groups of data consumers: (1) journalists, the staff of politicians, politicians who vote on legislation and on the NSOs' budgets, leaders who make business and social decisions, and the general public, and (2) professional data analysts at NSOs, college professors, college students and school mathematics teachers. This paper presents data on the statistical literacy of the second group: professional data analysts, college professors, college students and school mathematics teachers.

3. Statistical illiteracy

From 1998 to 2002, students taking Statistical Literacy at Augsburg College studied tables of rates and percentages presented in the U. S. Statistical Abstract. This exercise indicated that students had difficulties reading these tables. The Director at that time, Glenn King, provided copies of the U. S. Statistical Abstract for use by these students and participated in a preliminary survey to identify the level of statistical literacy in reading summary statistics presented in tables, graphs and statements. As an example, here are two tables from the U.S. Statistical Abstract that college students in non-quantitative majors found difficult to read.

In Fig. 1, students found the title confusing [31]. They could see data classified by age, sex and race, but not by firearm. They did not realize that the "by" in "by firearms" was short for "caused by." "Classified by" and "Caused by" may have been conflated to save space in the title.

In Fig. 2, "by" means three things: "classified by", "caused by" and "distributed by." [31] Since rate tables seldom involve "by" as "distributed by", students failed to see that the rows were parts (numerators). They mistakenly presumed the rows were wholes: pre-existing conditions which led to the different death rates. So they mistakenly said, "Among those with 'diseases of the heart', the death rate was 152 per 100,000 in 1990." A correct statement would be, "The rate of death due to diseases of the heart was 152 per 100,000 population in 1990."

4. 2002 statistical literacy survey

In 2002, an international survey of statistical literacy was conducted by the W. M. Keck Statistical Literacy Project [21]. This survey focused entirely on informal statistics – the ability to describe and compare rates and percentages as presented in table, graphs and statements. Here are the average error rates for each of the four groups surveyed: college teachers (29%), professional data analysts (45%), college students (49%), and school mathematics teachers (55%). The actual survey is also available [20].

The Appendix contains the results for professional data analysts. Here are some highlights:

- 30% had difficulty reading a simple 100% row table [Q23]
- 43% were unable to identify an invalid comparison using data in a 100% row table. [Q28]
- About half were unable to correctly classify descriptions of percentages in a two-way half table [Q30-35] or in one-way half tables [Q44-48, Q49-50 and Q54-56]
- About half were unable to correctly classify descriptions of rates in a rate table [Q60-63]

No. 139. Age-Adjusted Death Rates, by Selected Causes: 1990 to 1996

[Rates per 100,000 population. For explanation of age-adjustment, see text, Section 2. The standard population f
is the total population of the United States enumerated in 1940. See also headnote, Table 138]

CAUSE OF DEATH	1990	1991	1992	1993	1994	1995
All causes	**520.2**	**513.7**	**504.5**	**513.3**	**507.4**	**503.9**
Major cardiovascular diseases	189.8	185.0	180.4	181.8	176.8	174.9
Diseases of heart	152.0	148.2	144.3	145.3	140.4	138.3
Rheumatic fever and rheumatic heart disease	1.5	1.4	1.3	1.3	1.2	1.1
Hypertensive heart disease [1]	4.8	4.7	4.8	4.9	5.0	5.1
Hypertensive heart and renal disease . .	0.5	0.5	0.5	0.5	0.5	0.4
Ischemic heart disease	102.6	99.1	95.7	94.9	91.4	89.5
Other diseases of endocardium	2.5	2.5	2.6	2.6	2.6	2.6
Acute mycocardial infraction	53.7	51.5	49.1	47.5	45.6	43.8
Old mycocardial infraction and other . . .	47.8	46.6	45.7	46.5	45.0	44.9
Hypertension [1]	1.9	1.9	2.0	2.2	2.2	2.3

Fig. 2. U.S. age-adjusted death rates by selected causes.

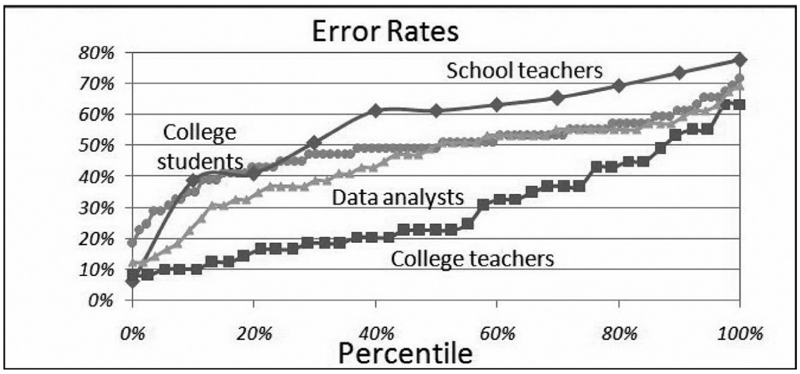

Fig. 3. Error rates by occupation and percentile.

Note: these are error rates for professional analysts working for national statistical offices.

Figure 3 presents the distribution of error rates by percentile within each of the four groups.

Focusing on averages may draw attention away from the range of scores within each group. Specifically, the best scorers had error rates of 10% to 20%; the worst had error rates of 60% to 80%. These high error rates within each group indicate that statistical illiteracy is widespread.

5. Explaining statistical illiteracy

One might expect that professional data analysts would do better than school teachers, and that school teachers would do better than college students. So why did school teachers average lower – and professional data analysts didn't do much better – than college students?

In Table 1, the lowest error rates are for college faculty (29%, top row) and for native English speakers (43%, left column). The highest error rates (shaded cells) for each row typically involve those who learned English after childhood (52%, right column). It appears that being a non-native English speaker may be a risk factor for statistical illiteracy.

One reason that school teachers had lower scores than the college students is language: most school teachers were non-native speakers while most college students were native speakers. Language helps explain why professional data analysts had lower scores than college teachers: most data analysts were non-native speakers; most college teachers were native speakers.

This distribution of respondents by English background may seem unexpected. In fact, all those labeled school teachers are school mathematics teachers in South Africa. Over half of those labeled data analysts are staff at the South African Statistical Society while the others are staff at the US Census Bureau and at STATS. While these are unusual groups or mixtures for those in first-world agencies, they may reflect some of the challenges for NSOs whose English-language statistics are being read by an increasing number of non-native speakers.

Table 1
Error rate (count) by occupation and english speaking background

Error Rate (Count)	English Speaking Background			
Occupation	Native speaker	Learned as child	Adult Learner	ALL
College Teacher	29% (27)	27% (10)	35% (2)	29% (39)
Professional Analyst	39% (24)	48% (23)	54% (7)	45% (54)
College Student	49% (80)	47% (5)	63% (2)	49% (87)
School Teacher	47% (4)	63% (6)	41% (1)	55% (11)
ALL	43% (135)	45% (44)	52% (12)	56% (191)

6. The Jenkinson project

As can be seen in the preceding examples and the results in Appendix A, the titles for tables and graphs can be ambiguous when describing rates and percentages. Consider these two sets of titles. Given the small changes in syntax within each set, do all the members of the following sets indicate the same part-whole relationship?

- The suicide rate of males, the males' rate of suicide and the male rate of suicide
- The percentage of smokers who are men, the percentage of male smokers, the percentage of smokers among men

In each case at least two members indicate different part-whole relations [18]. The moral: small changes in syntax can create big differences in semantics. Understanding these conditional relationships is crucial to understanding the various ways we have of dealing with context – of taking into account the influence of related factors.

Of course these are subtleties compared to a simple reversal of part and whole as appeared in a 2009 AP story: *Study says too much candy could lead to prison* [26]. The AP stated that "Of the children who ate candies or chocolates daily at age 10, 69 percent were later arrested for a violent offense by the age of 34." But the AP got it backward. The true statistic was the inverse: 69% of violent criminals ate candy daily at age 10.

In 1949, Jenkinson [14] noted the difficulty of titling tables of percentages. He proposed "a search for a systematic presentation which focuses attention on basic problems in percentage description." That was more than 60 years ago! This isn't rocket science; it won't cost billions or even millions. This Jenkinson "project" should be carried out if data producers are to avoid the charge of being statistically illiterate. Schield suggested some rules [18] but the producers of tables need to set their own standards. Some may think the use of the word "project" is rather grandiose. But based on my experience with NSOs, getting agreement on when to use *percent* and when to use *percentage* may prove to be a project all by itself.

One might even argue the some of the statistical illiteracy of users is caused by the lack of fluency in data presentation by data producers: either the data producers can't produce tables and charts that are comprehensible, or they don't have objective standards that would guide users in reading their tables and charts. If NSOs would provide such standards, they could guide teachers in teaching students how to be statistically literate consumers.

An alternate explanation is given by Lehohla [15], who noted the dramatic increase in interest in official statistics by non-professionals. Tables and charts that are readily comprehended by data professionals may appear ambiguous or incomprehensible to non-professional data consumers.

7. Training government employees

When the Director of Statistics in South Africa, Pali Lehohla, realized that some of his staff had difficulty reading simple tables and graphs, he recognized that other government employees charged with making appropriate decisions using data were even more likely to have such difficulties. He wondered about the possibility of setting up training programs for all such government employees. This seemed like an impossible task in 2002 – given the ongoing lack of resources. But web-based solutions are emerging.

Programs for improving statistical literacy include on-line delivery of multiple choice exercises [23], a web-program that analyzes a user's ability to use ordinary English to describe and compare rates and percentages as presented in graphs, tables and statements [6,7] and an entire web-based, statistical literacy course for consumers [12] now offered at Augsburg College.

8. Supporting policy goals

National Statistical Offices may become involved in supporting policy goals. The United Nations Development Group has identified a number of Millennium Development Goals (MDGs). Achieving these goals requires enhancing the "statistical capacity and literacy across country partners in order to increase data availability and enhance data use and support evidence-based policy-making [28–30]."

If NSOs become responsible for supporting policy goals as an integral part of their mission, they may also take on responsibility for ensuring that their data is useful to and understandable by policy makers. This may entail taking on responsibility for ensuring that policy makers and their staff are statistically literate consumers – that they have the necessary skills to read and interpret the data provided by the NSOs.

9. Other ways of supporting statistical literacy

There are other ways to support or improve the statistical literacy of users besides improving the comprehensibility of tables and graphs. These include

- Flagging averages where the distribution is bimodal. This may occur when some of the members have none of the characteristic being measured. E.g., the average US family spends more on pets than on alcohol; the average American is drinking less and working less.
- Improving data accessibility: Some PDF tables are not designed to take advantage of the new Acrobat 9 commands ("Copy as a Table", "Save as a Table" and "Open Table in Spreadsheet") so that numeric appearing data is not actually available as numeric data.
- Focusing attention on the importance of how the definition of a group can influence the size of a number. As Joel Best [4] noted, all statistics are socially constructed. Consider these examples: (a) OPEC countries supply 50% of US oil imports, but only 30% of US oil usage. (b) The average US farm is 440 acres; the average US family farm is 326 acres and (c) Annual income is $43K for households, $53K for families and $62K for married couples [22].
- Promoting multivariate thinking [16]: All too often data is classified by factors that are secondary – such as by geographic region – when other factors are more highly correlated with the outcome in question. In some cases, taking these other factors into account can increase, decrease or even reverse the observed association between the secondary factor and the outcome of interest. Helping users be aware of this possibility might enhance their statistical literacy and increase their trust in the veracity of the data when it appears to say different things depending on what is taken into account.

But these all enhance usability and interpretation which seems to be secondary to enhancing the comprehensibility of the data presentation.

10. Conclusion

Data comprehensibility is a growing challenge for governmental statistical agencies as their users are less likely to be data analysts. But if professional users – the staff of administrative officials – cannot accurately interpret the data provided by the agency, they may make bad decisions. Providing useful data to support better decisions is a primary justification for such agencies. Thus agencies may need to promote statistical literacy for consumers externally and fluency in data presentation internally in order to avoid jeopardizing their core mission. This may require agency leadership in establishing rules for using ordinary English to describe percentages and rates and for titling tables. With objective rules, teachers can increase statistical literacy among school and college students which in turn may improve the statistical literacy of future managers and leaders.

11. Recommendations

If national statistical offices are to be more effective in promoting evidence-based decision making that uses the data they provide, they should consider extending their missions to include "comprehensibility": to provide accurate and timely data that is useful to and comprehensible by users.

National statistical offices should consider the following actions:

1. Survey users on their statistical literacy as consumers: their ability to accurately decode the data presented in agency-generated reports.
2. Tables and graphs should be ranked on their level of incomprehensibility and then analyzed to see if user errors can be decreased by improving the titles

3. Survey staff on their fluency in data presentation: their ability to create tables and graphs that are unambiguous, clear and comprehensible. Establish appropriate training programs for staff as needed.
4. Initiate the Jenkinson project: Generate rules (a) for the use of grammar (such as "percent" vs. "percentage", the use of "by" and the use of adjective-noun phrases to indicate part-whole ratios) and (b) for the titling of tables of ratios. Identify all tables, graphs and statements that might be subject to these rules and see which ones are in violation of these rules. Determine if the general rules must apply to all cases or if there are relevant exceptions.
5. Establish responsibilities for increasing the fluency in data presentation among staff. Establish a center (a) to oversee the process of establishing rules and standards, (b) to survey the level of statistical literacy among newly-hired staff, (c) to establish appropriate training programs to improve their level of fluency in data presentation, (d) to monitor compliance to these rules and standards in all agency-generated data – statements, graphs and tables, (e) to oversee the process of modifying these rules and standards as needed, and (f) to generate periodic reports on compliance and progress.
6. Establish responsibilities for increasing statistical literacy among data consumers. Establish a center (a) to promulgate the rules for using ordinary English to describe rates and percentages and for titling tables, (b) to identify programs to increase statistical literacy for consumers, (c) to establish relations with educators at the school and college level who can enhance statistical literacy among their students, (d) to decide how to work with and provide support for these groups, (e) to decide what types of education to provide independently of these groups, and (f) to monitor progress on these strategic goals and modify these goals as needed.
7. Determine whether statistical literacy should be included in their mission statement and if so, determine what goals are appropriate to achieve this mission.

In so doing, they would be following the advice of the 1949 director of the U.S. Census Bureau, J. C. Capt, who said "The primary purpose of [Census] publications is to provide reliable and needed information to the users of statistics – a purpose that can be best served if the tabular materials are made easy to read and to understand." [10].

In this way government service agencies can promote the use of evidence-based decisions using agency-generated data.

Acknowledgments

Glenn King, a former editor of the U. S. Statistical Abstract, introduced me to Jenkinson's book, allowed me to include Census Bureau staff in my survey of statistical literacy and provided copies of US Statistical Abstract to the Augsburg students who used them in their Statistical Literacy class. Tom Burnham helped design the 2002 survey. Brian Phillips (IASE) authorized administering this survey at ICOTS-6 in South Africa. Jacqueline Galpin (Wits University), Nirmala Naidoo (Statistics-South Africa) and Pali Lehohla (Statistician-General, South Africa) assisted in conducting the survey. This work was performed as part of the W. M. Keck Statistical Literacy Project "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts."

Appendix: 2002 Statistical Literacy Survey Results

This paper presents the results for most of the graphs and tables. The full results are presented elsewhere [19]. After presenting the question and proposed answer, the correct answer is shown in brackets [] followed by the error rates for four groups: college students, school teachers, college mathematics teachers and professional data analysts. The score for data analysts is bolded.

Do the following statements *accurately describe* the data shown in Fig. 4.

Q9. *20% of smokers are Catholic.* [Yes] Error rates: 20%, 0%, 5%, **0%**
School mathematics teachers and data analysts had the lowest error rate possible: 0%.

Q10. *Protestants (40%) are twice as likely to be smokers as are Catholics (20%).*
[No; Smoker is part in the statement; whole in the graph.] Errors: 63%, 82%, 18%, **59%**.

Do these statements accurately describe the data shown in Table 2?

Q23. *25% of females are blacks.* [No] Error rates: 44%, 55%, 10%, **30%**.

Table 2
100% row table

	SEX		
RACE	Male	Female	TOTAL
Black	75%	25%	100%
White	50%	**50%**	100%
Other	40%	60%	100%
TOTAL	50%	50%	100%

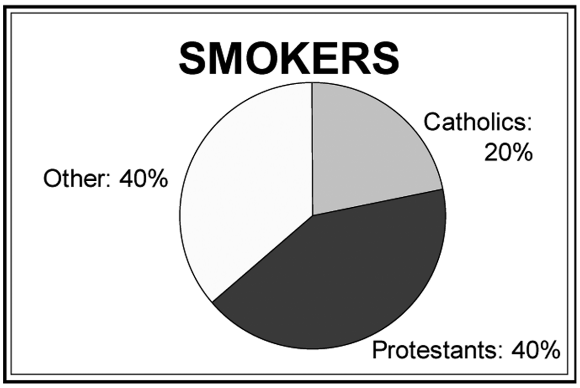

Fig. 4. Pie Chart of Smokers by Religion.

Q24. *25% of blacks are females.* [Yes] Error rates: 13%, 27%, 15% and **15%**.

A majority of school teachers, almost half the students and almost a third of the data analysts failed to spot an incorrect statement as being wrong. Over a fourth of the school teachers and about 15% of the other three groups failed to identify a correct statement as being correct.

Q25. *25% is the percentage of blacks among females.* [No] Error rates: 39%, 18%, 15%, **30%**.

Q26. *25% is the percentage who are females among blacks.* [Yes] Errors: 22%, 9%, 13%, **17%**.

Do you think these statements below *accurately compare* the 25% circled with the 50% immediately below it in Table 2? {Correct analysis: Female is common part; the races are wholes.}

Q28. *Females are two times as likely to be white as to be black.* [No; 'white' is part in question but 'female' is part in the graph]. Error rates: 45%, 45%, 23%, **43%**.

Do you think the following statements *accurately describe* the 20% circled in Table 3?

Q30. *20% of runners are female smokers.* [No] Error rates: 56%, 55%, 28%, **57%**.

Q31. *20% of females are runners who smoke.* [No] Error rates: 53%, 73%, 33%, **57%**.

Q32. *20% of female smokers are runners.* [Yes] Error rates: 61%, 64%, 56%, **56%**.

Q33. *20% of smokers are females who run.* [No] Error rates: 43%, 73%, 26%, **39%**.

Q34. *Among female smokers, 20% is the percentage of runners.* [Yes] 67%, 82%, 54%, **57%**.

Q35. *Among females, 20% is the percentage of smokers who are runners.*
[Yes; Q35 is the semantically the same as Q34.] Error rates: 49%, **73%**, 69%, **48%**.

Do you think these statements accurately compare the circled 20% with the 10% below it? {Correct analysis: Runner = common part, smoker and sex = whole.}

Q36. *The percentage of runners is twice as much among female smokers as among male smokers.* [Yes.] Error rates: 43%, 64%, 49%, **48%**.

Q37. *The percentage of smokers who run is twice as much among females as among males.* [Yes.] Error rates: 41%, 73%, 51%, **56%**.

Notice that when a percentage table lacks any 100% totals, the error rates are high among all groups for both descriptions (Q30-Q35) and comparisons (Q36-Q37).

Do the following statements accurately describe the 26.2% bolded in Table 4?

{Correct analysis: The ALL margin values are averages so Sex and Race are wholes. Years delimit wholes. So the part must be in the title. For several reasons the part is "smoking."}

Q44. *In 1990, 26.2% is the percentage of smokers who were black.* [No] 72%, 82%, 28%, **59%**

Q45. *In 1990, 26.2% of blacks were smokers.* [Yes] Error rates: 57%, 64%, 31%, **37%**.

Q46. *In 1990, 26.2% of smokers were black.* [No] Error rates: 72%, 82%, 33%, **59%**.

Q47. *In 1990 26.2% was the percentage of black smokers.* [No] Errors: 70%, 64%, 23%, **59%**

Table 3
Two-way half table

PERCENTAGE WHO ARE RUNNERS			
	Non-smoker	Smoker	Total
Female	50%	20%	40%
Male	25%	**10%**	20%
Total	37%	15%	30%

Table 4
Percentage of smoking prevalence among U.S. adults 18 and Older

Year	All	Males	Females	Whites	Blacks
1955		56.9	28.4		
1965	42.4	51.9	33.9	42.1	45.8
1970	37.4	44.1	31.5	37.0	41.4
1980	33.2	37.6	29.3	32.9	36.9
1990	25.5	28.4	22.8	25.6	**26.2**

Table 5
Low birth weights

Percent of Births with Low Birth Weight			
State	1990	1995	1996
U.S.	7.0	7.3	7.4
AL	8.4	9.0	9.3
AK	4.8	5.3	5.5
AZ	6.4	6.8	6.6
AR	8.2	8.2	8.5
CA	5.8	6.1	6.0

Q48. *In 1990, 26.2% is the percentage of blacks who are smokers.* [Yes] 51%, 36%, 23%, **44%**.

The error rates for college students and school math teachers are – with one exception – more than 50%. This indicates these groups cannot distinguish correct and incorrect statements.

The high error rates may be due to combining "percentage" and "prevalence" in the same phrase. A non-ambiguous title would be "Percentage of U.S. Adults 18 and Older who Smoke." But subsequent surveys with unambiguous titles indicate the primary problem is with this kind of table: percentage tables where there are no 100% margins and where – in some cases – the groups involved are exclusive and exhaustive (e.g., Male and Female).

Do you think the following statements accurately describe the 6.0 circled in Table 5?

Analysis: States and years are whole delimiters. Countable whole is 'births'; part is 'low weight.'

Q49. *In the US in 1996, 6% of low-weight births were in Calif.* [No] 60%, 73%, 21%, **50%**.

Q50. *In the US in 1996, the percentage of California births among low-weight births was 6%.* [No] Error rates: 67%, 82%, 28%, **72%**.

Q51. *In the US in 1996, 6% of Calif. births were low-weight.* [Yes] 38%, 55%, 10%, **37%**.

Does this statement accurately compare 6.0% in California (CA) with 9.3% in Alabama (AL) in Table 5?

Q52. *In the US in 1996, there were more low weight births in Alabama (AL) than in California (CA).* [No] {Statement compares counts; Table presents rates} 67%, 64%, 31%, **70%**.

The title should have used "percentage" in place of "percent." Many respondents though that everything following "percent of" was the whole, "Births with low birth-weight," and expected this whole to be distributed among the various parts – the states. The only problem was the lack of the 100% sum. A better title would be "Percentage of Births that have Low Birth-Weights."

Do you think the following statements accurately describe the 10% circled in Table 6?

Q53. *10% of the women who received an HIV test were 40–44.* [No] 78%, 64%, 21%, **57%**.

Q54. *10% of these women 40 to 44 received an HIV test.* [Yes] Errors: 34%, 55%, 26%, **35%**.

Q55. *Among those women who received an HIV test, the percentage of those 40–44 was 10%.* [No] Error rates: 61%, 73%, 31%, **65%**.

Q56. *Among these women 40–44, the percentage who received an HIV test was 10%.* [Yes]. Error rates: 38%, 45%, 15%, **31%**.

Table 6 is poorly captioned. The title should have used "percentage" in place of "percent," so that the rel-

Table 6
Percent of women, 15 to 44, who received selected medical services

CHARACTERISTIC	**HIV Test**	**Pregnancy test**	**Pap smear**
Total [Table 198]	**17.3**	**16.0**	**61.9**
AGE AT INTERVIEW			
15-19 years old	14.6	16.1	33.5
20-24 years old	23.7	27.4	68.7
25-29 years old	23.6	25.3	70.9
30-34 years old	18.5	17.4	69.5
35-39 years old	14.2	8.1	62.9
40-44 years old	10.0	4.3	62.7

Table 7
Death rates, by leading cause

No. 143. Death Rates, by Leading Cause--States (Top 5 and Bottom 5). Source: 1998 US Statistical Abstract
[Deaths per 100,000 resident population estimated as of July 1.]

STATE 1995	Total	Heart disease	Cancer	Cerebro-vascular dis-eases	Accidents and adverse effects	Motor vehicle acci-dents	Chronic obstructive pulmonary diseases	Dia-betes mellitus	HIV	Suicide	Homicide
United States	**880.0**	**280.7**	**204.9**	**60.1**	**35.5**	**16.5**	**39.2**	**22.6**	**(NA)**	**11.9**	**8.7**
D.C.	1,244.2	302.4	267.2	66.8	34.8	12.3	24.2	39.5	117.8	7.0	56.8
West Virginia	1,107.0	378.9	259.4	67.9	40.4	21.2	60.0	32.8	0.0	15.1	5.5
Florida	1,081.3	351.6	263.5	69.9	38.1	19.8	52.9	26.0	30.8	15.3	8.8
Arkansas	1,075.1	339.8	244.7	91.5	48.8	26.3	45.0	22.4	6.8	14.5	11.6
Pennsylvania	1,059.2	359.7	250.7	68.6	35.3	13.1	43.9	28.2	11.5	12.1	6.5
Missouri	1,021.9	345.3	230.7	72.9	43.5	20.6	46.1	23.4	8.8	13.5	8.9
California	709.8	216.3	162.8	51.4	29.3	14.1	34.2	16.2	20.4	11.7	11.6
Colorado	667.6	172.1	145.9	42.7	39.8	18.6	42.3	14.3	10.9	17.5	5.7
Hawaii	643.1	196.0	156.4	51.5	27.6	12.0	20.4	14.2	10.4	12.0	4.9
Utah	560.6	148.1	108.6	39.9	32.4	17.2	24.1	21.3	4.8	14.8	3.9
Alaska	423.0	90.6	95.1	24.0	56.2	16.1	17.7	9.3	5.0	17.1	8.9

ative clause – who received selected medical services – would be clearly indicated as the part in the percentage. A proper statement would be, "In 1995, 10% of [U.S.] women ages 40–44 received an HIV test." Since the title used "percent" rather than "percentage", students placed a numerical percentage – a number – before "percent" and then looked for a predicate to complete the incorrect statement, "10% of women, ages 14–44 who received an HIV test." All too many completed the incorrect statement with this predicate: "were 40–44."

Do the following statements *accurately describe* the 16.1 circled in Table 7? {Analysis: US rates are averages so states are wholes. Totals are sums so the columns are parts. "Death" modifies "rate" so "death" is a part along with the particular column heading.}

Q60. *In 1995, the death rate due to motor vehicle accidents was 16.1 per 100,000 Alaskans.* [Yes] Error rates: 21%, 18%, 8%, **17%**.

Q61. *In 1995, for those in motor vehicle accidents, the death rate was 16.1 per 100,000 Alaskans.* [No; "Motor vehicle accidents" is not a whole.] Errors: 61%, 73%, 15%, **57%**.

Q62. *In 1995, the rate of motor vehicle accidents was 16.1 per 100,000 Alaskans.* [No; the question omits the word "death."] Error rates: 59%, 55%, 28%, **37%**.

Q63. *In 1995 for Alaskans who were in motor vehicle accidents, the death rate was 16.1 per 100,000.* [No; "vehicle accidents" is not a whole.] Error rates: 67%, 82%, 18%, **70%**.

Note that the error rate was much lower for a true rate-statement (Q60) than for most true percentage statements (Q32, 34, 35, 45, 48, 51, 54 and 56).

References

[1] ASA Guidelines for Assessment and Instruction in Statistical Education, GAISE (2006). American Statistical Association. See www.amstat.org/education/gaise/.

[2] ASA Strategic Plan (2008). Strategic Plan for Education. American Statistical Association. See www.amstat.org/about/index.cfm?fuseaction=strategicplan.

[3] Australian Bureau of Statistics, Statistics Education Unit. See www.StatLit.org/pdf/2007ABS-Statistical_Literacy_Paper.pdf and www.StatLit.org/pdf/2007ABS-Statistical_Literacy_Competencies.pdf.

[4] J. Best, *More Damned Lies and Statistics*. Univ. of California Press, 2004.

[5] S. Budgett and M. Pfannkuch, Assessing students' statistical literacy in *Assessment Methods in Statistical Education: An International Perspective* edited by Bidgood, Hunt and Joliffe. Wiley, 2010, pp. 103–121.

[6] T. Burnham and M. Schield, An Online Grammar-parsing Program That Decodes Ordinary English Descriptions and Comparisons of Percentages and Rates. ASA Proceedings of the Section on Statistical Education, [CD-ROM], 2202–2208. See www.StatLit.org/pdf/2005BurnhamSchieldASA. pdf, 2005.

[7] T. Burnham and M. Schield, Introduction to an Online Ratio Statement Validator, IASSIST Workshop. See www.StatLit.org/pdf/2006BurnhamSchieldIASSIST.pdf, 2006.

[8] Cockcroft Report (1982). See www.dg.dial.pipex.com/documents/docs1/cockcroft.shtml.

[9] I. Gal, Functional demands of statistical literacy: Ability to read press releases from statistical agencies. IASE. At www.stat.auckland.ac.nz/∼iase/publications/3/3016.pdf, 2003.

[10] J. Garfield, R. DelMas and A. Zieffler, Assessing important learning outcomes in introductory tertiary statistics courses in *Assessment Methods in Statistical Education: An International Perspective* edited by Bidgood, Hunt and Joliffe. Wiley, 2010, pp. 75–86.

[11] D.G. Haack, *Statistical Literacy*. Duxbury Press, 1979.

[12] M. Isaacson, Statistical Literacy – An Online Course at Capella University. 2005 American Statistical Association Proceedings of the Section on Statistical Education, 2244–2252. See www.StatLit.org/pdf/2005IsaacsonASA.pdf, 2005.

[13] International Association of Official Statistics Conference (2008). IAOS Conference on Reshaping Official Statistics, Shanghai. See www.iaos2008conf.cn/invitation.htm.

[14] B. Jenkinson, *Bureau of the Census Manual of Tabular Presentation*. U. S. Government. Extracts online at www.StatLit.org/pdf/2001SchieldASA2.pdf, 1949.

[15] P. Lehohla, Promoting Statistical Literacy: A South African Perspective. ICOTS-6, South Africa. See www.StatLit.org/pdf/2002LehohlaICOTS.pdf, 2002.

[16] J. Ridgway, J. Nicholson and S. McCusker, Reconceptualising 'Statistics' and 'Education'. IASE 2008 Monterrey, 2008.

[17] J. Sanchez, *Government Statistical Offices and Statistical Literacy*. The International Statistical Literacy Project (ISLP) of the International Statistical Institute (ISI). First edition. See www.stat.auckland.ac.nz/∼iase/ISLP/stats-offices-book, 2008.

[18] M. Schield, Statistical Literacy: Describing and Comparing Rates and Percentages. 2000 American Statistical Association Proceedings of the Section on Statistical Education, pp. 76–81. See www.StatLit.org/pdf/2000SchieldASA.pdf, 2000.

[19] M. Schield and T. Burnham, Statistical Literacy: Reading Tables of Rates and Percentages, 2001 American Statistical Association Proceedings of the Section on Statistical Education, [No pages numbers given]. See www.StatLit.org/pdf/2001SchieldASA.pdf, 2001.

[20] M. Schield and W.M. Keck, Statistical Literacy Survey. Copy at www.StatLit.org/pdf/2002StatLitSurvey.pdf, 2002.

[21] M. Schield, Statistical Literacy Survey Analysis: Reading Tables and Graphs of Rates and Percentages. International Conference on Teaching Statistics. See www.StatLit.org/pdf/2006SchieldICOTS.pdf, 2006.

[22] M. Schield, Statistical Literacy: Teaching the Social Construction of Statistics. Midwest Sociological Society. See www.StatLit.org/pdf/2007SchieldMSS.pdf, 2007.

[23] M. Schield, Statistical Literacy: Factual Assessment to Support Hypothetical Thinking. 2007 International Association of Statistical Educators (IASE), Guimarães, Portugal. See www.StatLit.org/pdf/2007SchieldIASE.pdf, 2007.

[24] M. Schield, Quantitative Graduation Requirements at US Four-Year Colleges. MAA JMM. See www.StatLit.org/pdf/2010SchieldJMM.pdf, 2010.

[25] M. Schield, Assessing Statistical Literacy: Take CARE in *Assessment Methods in Statistical Education: An International Perspective* edited by Bidgood, Hunt and Joliffe. Wiley, 2010, pp. 133–152.

[26] StatLit-Blog. *AP Creates Bogus Crime Wave*. See www.StatLitBlog.org.

[27] UK Office of National Statistics: www.ons.gov.uk/about/who-we-are/our-aims/index.html.

[28] United Nations Development Group (2005). *Indicators for Policy Management: A Guide for Enhancing the Statistical Capacity of Policy-makers for Effective Monitoring of the MDGs . . .* At www.undg.org/archive_docs/7759-Indicators_for_Policy_Management__Practical_Guide.pdf.

[29] United Nations Development Group (2008). *Indicators for Policy Management and Advocacy: MDGs and Statistical Literacy* www.unescap.org/stat/meet/sl/IPMA.01_Introduction.pdf.

[30] United Nations Development Group (2009). *Indicators for Monitoring the Millennium Development Goals*: http://milleniumindicators.un.org/unsd/mdg/default.aspx.

[31] U.S. Census Bureau (1998). *1998 U.S. Statistical Abstract*.

[32] U.S. Census Bureau. *1999 Financial Report* www.census.gov/prod/2000pubs/99finrpt.pdf.

[33] U.S. Census Bureau (2007). *U.S. Census Bureau Strategic Plan, FY 2007–2012*. June 2007. See www.census.gov/main/www/strategicplan/strategicplan.pdf.