® ASSESSING STUDENTS' STATISTICAL LITERACY

BUDGETT, Stephanie, & PFANNKUCH, Maxine
The University of Auckland
New Zealand

This paper describes the methods and challenges in assessing an undergraduate course entitled Lies, Damned Lies, and Statistics. The course has three main aims: to instill in students the ability to "think statistically", to enable students to critically evaluate statistically based reports and to teach students to construct statistically sound reports. The assessment is examined in terms of these aims and the criteria developed by Gal (2002) for statistical literacy.

INTRODUCTION

In 2006 there was a major controversy in New Zealand on whether the government should fund the breast cancer drug Herceptin. There were many media reports such as: "Cancer drug too expensive: decision not to pay outrages patients and supporters" and statistics quoted.

After two years:

- Herceptin patients had a 32% lower risk of dying than the others.
- 7.6% of the Herceptin patients and 11% of the others had died.
- The relative risk of breast cancer recurrence was 36% less for the Herceptin group than the others (*NZ Herald*, 29/7/2006).

This controversy required a high level of statistical literacy in order to understand why the government did not purchase Herceptin despite women on TV pleading for the drug to save their lives. A statistically literate student would comprehend the differences between risk, relative risk, increased risk, and reduced risk, how risk statements can be misleading, and have knowledge about experimental design. The question is how to assess students' statistical literacy in such a situation.

THE COURSE

The University of Auckland's undergraduate paper titled Lies, Damned Lies, and Statistics is a course on statistical literacy designed to prepare everyone, regardless of statistical background, to become critical consumers of statistical information. In particular, the course is designed to attract aspiring journalists, politicians, sociologists, lawyers, health personnel, business people, and scientists. The course is based on topics that are often reported in the media. The uses, abuses, and limitations of statistical information in a variety of activities such as polling, public health, law, marketing and the environment are examined via media reports such as the one quoted above. The course aims to provide students with the skills to "think statistically", to critically evaluate statistically based reports, and to construct statistically sound reports. Content is delivered within six topics (Table 1) and is concentrated primarily on conceptual understanding rather than calculation.

Table 1
Content of Lies, Damned Lies, and Statistics undergraduate course

Topic	Content
Media Reports	Comprehending and evaluating reports and visual displays of quantitative information, critical questions, measurement issues, writing reports.
Surveys & Polls	Populations, samples, sampling variation, bias, variation in estimates, confidence statements, margin of error issues, survey design, ethics.
Experimentation	Observational studies, experiments, problems of different conclusions.
Risk	Types of risk, odds ratios, interpreting small probabilities, statistical and practical significance, false positives, causation/association.
Statistical Reasoning	Heuristics, variability, nature of randomness, expected values, coincidences.
Statistics & the Law	Prosecutor's fallacy, conditional probability, Bayes' Theorem, likelihood ratios.

The course is assessed using a combination of assignments (30%), tutorials (5%), term test (15%) and final exam (50%).

STATISTICAL LITERACY

The term statistical literacy has not gained a consensus about its meaning in the literature. Gal (2002) argues that statistical literacy applies to data consumers and describes people's ability to interpret and critically evaluate statistically based information from a wide range of sources and to formulate and communicate a reasoned opinion on such information. Watson (1997) initially developed a view of statistical literacy that centered on media reports and focused on the data consumer. She described a hierarchy of three tiers of statistical literacy: basic understanding of probabilistic and statistical terminology; understanding of statistical language and concepts embedded in wider social discussion; and challenging claims in the media. Recently she widened her definition of statistical literacy to include knowledge of how data are produced (Watson, 2006). Since a data consumer needs statistical knowledge to be statistically literate the widened definition that encompasses knowledge gained from experiencing statistical investigation is not inappropriate. But a question remains as to whether the knowledge gained from a typical undergraduate introductory statistics course is sufficient for statistical literacy. According to Gal (2002), if critically evaluating other people's statistically based reports is not explicitly taught then students will remain statistically illiterate. Furthermore, students also need to understand the statistical concepts, reasoning and data-based arguments that permeate their everyday life and appreciate how evidence-based thinking contributes to public and personal decisions. Awareness that many everyday events can be thought of from a statistical perspective, including the heuristics people use when reasoning, is also part of statistical literacy (Tversky & Kahneman, 1982; Pfannkuch & Wild, 2004; Utts, 2004).

For the purposes of our undergraduate course we define statistical literacy as applying to the data consumer and we extend Gal's definition in two ways. We include sources such as everyday events whereby people learn to view the world from a statistical perspective and we include people's ability to construct and synthesize statistically sound reports after analyzing original sources.

THE ASSESSMENT

Across the assessment tasks we ask students to evaluate media articles, journal articles, and technical reports on opinion polls, sample surveys, experiments and observational studies. Since the course is relatively new we are still in the process of developing assessment tasks. We have no research evidence that we are improving students' statistical literacy; we can only describe and analyze the assessment tasks with reference to the aims of the course and the statistical literacy literature for substantiation of the abilities that we may be assessing.

The First Assignment

Assignment One requires students to critically evaluate a statistically based opinion poll or sample survey study. Four journal articles or technical reports are presented, together with related media reports, from which they select one. They are expected to evaluate the study, give a judgment of the study, and critique the related media report. We now discuss what abilities we believe we are assessing in the assignment.

Assessing ability to evaluate. When students evaluate or analyze the study they use a set of "worry questions" or critical questions to prompt their thinking. These "worry questions" are framed in Utts (2004) seven step and seven critical components template for evaluating news articles. The hardest part for students is to use their own real world knowledge or think laterally to come up with issues that the authors have not addressed. There is an expectation by some students that the "answers" will be found in the study. But other students generate some really interesting insights on issues such as confounding variables and flaws in the design of the study.

Assessing ability to communicate. After students have completed the analysis they are assessed on their ability to provide a reasoned opinion on the study. Some students find it difficult to synthesize a judgment on the study and justify their judgment by selecting and summarizing key issues from their analysis of the study. Some justify their judgment by resorting to their own

beliefs or experiences rather than giving evidence-based justifications. From our experience we would agree with Murray and Gal (2002) that the analysis and synthesis of statistical information are two interrelated components of statistical literacy but both need to be specifically assessed.

Assessing ability to challenge claims in the media. After critically evaluating the study the students then consider the newspaper article on the study and determine whether the claims are justified. For example, an article on "Doctors' white coats a turn-off for patients" generalized the study to all New Zealanders when the study was conducted in only one city and was not true for older age groups. Furthermore, since the students had read the study they could raise many confounding variables that the study did not address, such as ethnicity and inconsistent use of color of clothing in the photographs of doctors, the time of year the study was conducted as well as a possible non-response bias. At a more detailed level students are assessed on their ability to challenge claims by media when the journalist, for instance, argues with point estimates. For example, for a survey of 800 people with a margin of error of 3.5% a quote from a newspaper article was: "Aucklanders were more in favour of hidden cameras than drivers elsewhere, by 63.2 per cent against 56.7 per cent in other regions". Students are expected to challenge such claims by roughly calculating the margin of error for these subgroups, based on their own knowledge that Auckland has one-third of the New Zealand population, and by calculating the 95% confidence interval for true population difference in percentages.

The Second Assignment

The goal of Assignment Two is to have the students construct a statistically sound report. Students are asked to find a newspaper article that reports a study that is of some personal interest. They are asked to track down the primary source and summarize the study and its findings. They then write their own newspaper article, together with a graph or table, and compose a letter to the editor of the newspaper stating whether the results of the research were meaningful. Finally, they state how and why their newspaper article differs from the original. Since students choose their own article they tend to select a study that is within their level of comprehensibility, thereby allowing them to work at their potential level. We now describe the abilities that are assessed in the assignment.

Assessing ability to construct statistically sound statements. The assessment schedule for the newspaper article assesses students' ability to include the main facets of the study for comprehensibility, to outline potential limitations of the study, and to create a statistically sound graphic and statements. The types of graphics and statements the students are required to use are not found in typical newspaper articles. For example, the graphics need to indicate variability in estimates by using confidence intervals and in the text confidence interval statements are expected to prevent over-interpretation of numerical information and to appreciate that sampling error is associated with almost all statistical summaries. Writing appropriate non-causal statements, defining the population on which inferences are made, differentiating between sample and populations statements, clearly defining the measures used, and so forth are checked for correctness in the assessment of their newspaper article.

Assessing ability to justify why statements are statistically sound. After writing the newspaper article the students analyze how and why their article differs from the journalist's article. This task assesses their ability to both critically evaluate the journalist's article and state the underlying statistical principles for why the statements in their newspaper article are sounder. From our experience this task identifies the students who have a better understanding about what makes a statement statistically correct.

The Test

In addition to assessing the same abilities in the test as described for the first assignment, we also assess knowledge that we believe is essential for being able to understand media reports. If students do not have some basic knowledge about representative samples, confidence intervals, significance, experimental and observational study designs, risk and so forth they will find it difficult to comprehend a study. The following two examples illustrate the type of knowledge and abilities we are assessing.

Example 1: *Assessing ability to ask further questions about a claim.*

For example, quotes are given to students to query such as: "since 2002 the unemployment rate has dropped by 20%", By excluding data collection issues we expect students to query the starting value, how unemployment has been defined, and whether the criteria for unemployment has changed in that period. Part of statistical literacy is to engender a healthy skepticism about claims made and to develop a feeling for what should be measured and compared to give a fair representation of a situation, and to encourage curiosity. Prompting an array of questions is part of developing an ability to take a critical stance.

Massage helps migraine sufferers

BY REBECCA WALSH
HEALTH REPORTER

A regular upper-body massage could be the key to reducing migraine attacks, an Auckland University study has found.

Research by doctoral student Sheleigh Lawler shows people who had a 45-minute massage once a week experienced fewer migraines than a control group.

Ms Lawler, who works in the university's psychology department, is now extending her research to see whether the addition of individual massage plans to a standard massage will help further reduce attacks.

Forty-seven migraine sufferers took part in the first study. Half the group received a 45-minute massage of the upper body and head region once a week for six weeks, in addition to their standard medication.

The control group continued their normal treatment.

The study found that in the four weeks before beginning the massage (developed by the New Zealand College of Massage), both groups experienced a similar number of migraines.

But six weeks after the massage treatment, the group receiving the massage suffered an average 3.5 migraines a month compared to 5.5 for the control group.

Those receiving the massage also appeared not to need to take medication as often.

Annette Hallam, national director of the New Zealand Migraine Sufferers Support Group, said any treatment that helped reduce migraines was welcomed by sufferers.

Ms Lawler is looking for 75 more migraine sufferers to take part in further research.

Figure 1. Newspaper article given in a test (*NZ Herald*, 31/7/2004, copyright permission granted)

Example 2: *Assessing statistical knowledge and ability to comprehend an article.*

> Read the above article (Figure 1) "Massage helps migraine sufferers" (*NZ Herald*, 31/7/2004) and answer the questions below:
>
> (a) What are the explanatory and response variables for this study?
> (b) Explain whether the experiment is double-blind, single-blind, or neither.
> (c) Explain why a control group was used in this study.
> (d) Some "difficulties and disasters" with experiments are:
> - confounding variables
> - interacting variables
> - placebo, Hawthorne, and experimenter effects
> - ecological validity and generalisability.
>
> Choose one of these "difficulties and disasters" and explain why it might be a problem in this experiment.

A newspaper article is given and students' knowledge about experiments is assessed. Implicit in the assessment is that the students require a certain level of literacy skills to make sense of the article together with their statistical knowledge – on experimental design, on averages which should have a measure of variability, on statistical significance which is not reported, on possible confounding variables where they need to draw on their real world knowledge – and their contextual knowledge about doctoral students, massage, and migraine, as well as their ability to understand statistical terminology. It should be noted that some students thought that because a student conducted the research the findings were suspect. By using a newspaper article we are assessing whether students can cope with the types of statistical information that everyone is exposed to all of the time. These include ideas of randomized

experiments and observational studies, and when we can and cannot conclude that a factor is the cause of an effect. We want to assess their ability to see weaknesses in the way that studies are performed so that they can discount sources of information which are obviously biased.

The Examination

Approximately half of the examination paper is constructed around a newspaper article and its associated journal article. The remainder of the examination consists of both short and long answer questions. The following two examples illustrate the types of abilities and knowledge we are assessing. For Example 3 the newspaper article was "Study links older fathers to autism" (*NZ Herald*, 6/9/2006) and the journal article was "Advancing Paternal Age and Autism" (Reichenberg, Bresnehan, Rabinowitz, Lubin, & Davidson, 2006). Examples 3a and 3b use information from Figure 2.

Table 1. Associations Between Paternal Age and Risk of Autism Spectrum Disorder (ASD) in the Smaller Subset With Data on Both Paternal and Maternal Age

Paternal Age Group, y	Non-ASD Cohort (n = 132 161)	ASD Cases (n = 110)	Risk	Unadjusted OR (95% CI)	*P* Value	Adjusted OR (95% CI)*	*P* Value
15-29†	60 654	34	6:10 000	1.00	...	1.00	...
30-39	67 211	62	9:10 000	1.64 (1.08-2.50)	.02	1.62 (0.99-2.65)	.06
40-49	4106	13	32:10 000	5.65 (2.98-10.71)	<.001	5.75 (2.65-12.46)‡	<.001
≥50	190	1	52:10 000	9.39 (1.28-68.94)	.03	...	...

Abbreviations: CI, confidence interval; OR, odds ratio.
*Adjusted for year of birth, socioeconomic status, and maternal age.
†Reference group.
‡Includes oldest age group.

Figure 2. Data from journal article (Reichenberg, et al., 2006, p. 1029)

Example 3a: *Assessing statistical knowledge.*

(i) The title of the newspaper article is: "Study links older fathers to autism" (A). Another newspaper article title was: "Older dads have children with autism" (B).

Which title is statistically more appropriate, A or B? Explain why.

(ii) The newspaper article reports that:
Men over 40 are more than five times more likely to father a child with autism than their younger peers, a study has found.

What baseline group does this statement compare "men over 40" to? State why this comparison is misleading.

These items are designed to prompt students to think about how statements in the newspaper may be misleading and to give a reason based on statistical knowledge. We believe that we are assessing their statistical knowledge, how statistical argumentation is constructed, and are indirectly developing the disposition to take a critical stance.

Example 3b: *Assessing mathematical knowledge.*

The newspaper article states that:
The results showed that among fathers aged 15 to 29 when their child was born, the risk of autism was six in every 10,000 children. Among fathers aged 30 to 39, nine in 10,000 children had autism (50 per cent higher) …

(i) Describe or show the calculations for how "50 per cent higher" was obtained.
(ii) Using Figure 2 explain, showing your working, how the risk of 6 in every 10,000 children was obtained.

(iii) Why are the risks in this study expressed as rates per 10,000?

Basic mathematical knowledge is needed to be statistically literate. Anecdotal evidence in the tutorials alerted us to the fact that students did not understand an increased risk such as 100% and had misconceptions about rates by thinking that the base rate was the number of people sampled. Therefore this item tests whether students understand how figures were obtained and why very small percentages are expressed as rates and the necessity, if rates are used, to have the same base rate for comparison purposes.

Example 3c: *Assessing construction of statistically sound statements.*

> Using Table 2 [not shown here but similar to Figure 2], write a brief statistically sound paragraph for a newspaper article. Your paragraph should not misrepresent statistics about risk and should compare paternal age with the risks of having male and female offspring with autism.

This item assesses students' ability to construct statistically sound statements and that when they write, for example, a relative risk statement they should include the baseline risk.

Example 3d: *Assessing ability to communicate a reasoned opinion on a study.*

> Write a serious letter (about 200 words) to the *NZ Herald* editor in response to the newspaper article "Study links older fathers to autism". State you have read the original source. Outline any potential limitations of the study. Include in your letter whether you are willing to believe the findings or whether you need more convincing. Justify your judgment and any other statements that you make. End your letter with your recommendations and advice to potential parents in New Zealand.

We ask them to write a letter to the editor, which again assesses the ability of students to construct a reasoned opinion on statistical information. Scaffolding in the examination questions assists students in the production of a letter since they have been asked to think of limitations, confounding variables, and worries about the study. But in the letter they must synthesize the information and rationalize whether the study should be taken notice of by parents. Such a task includes students using their own contextual knowledge about Israel and autism. With the requirements for a recommendation and willingness to believe the findings, the beliefs and attitudes of the students tend to be exposed and hence the ability to supply evidence-based rationales is also assessed.

Example 4: *Assessing ability to reason statistically about everyday events, identify fallacies and construct correct statements.*

> (a) Explain how real estate agents or TV advertisers may try to use *anchoring* to exploit potential customers.
> (b) In June 2005 a bolt of lightning killed a Waikato racehorse named Rainhailorshine. The trainer of the horse was shocked and amazed at such an unusual event (*NZ Herald*, 4/6/2005).
> Explain to the racehorse trainer why this coincidence is not so amazing.
> (c) From a given newspaper article the students extract an incorrect statement such as: "Most of the blood samples matched Mr Boylan's DNA, and were 30 million million times more likely to be his than any other male in New Zealand, she said." (*The Daily News*, 23/6/2005). They then explain why the statement is incorrect and rewrite it correctly.

Students are assessed about their awareness of heuristics, common misconceptions and fallacies in statistical thinking and reasoning that occur in everyday life.

DISCUSSION

Gal (2002, p. 3) defines at a cognitive level that a critical evaluation of statistically based information is predicated on the joint activation of "a knowledge component (comprised of five cognitive elements: literacy skills, statistical knowledge, mathematical knowledge, context knowledge, and critical questions) and a dispositional component (comprised of two elements: critical stance, and beliefs and attitudes)." Our assessment draws on all the elements from Gal's knowledge component. From our experience, if students are to give a reasoned opinion they need not only the knowledge component but also a realization that their reasoning must be evidence based not based on their personal experience and pre-existing opinions. So therefore we would add a reasoning component to Gal's definition (comprised of two elements: statistical argumentation knowledge, and viewing everyday events from a statistical perspective). Argumentation knowledge would include statistical inferential reasoning and constructing statistically sound statements and graphics. Everyday event knowledge would include awareness of heuristics people use for reasoning and viewing and reasoning about a multitude of everyday life events from a statistical perspective. The dispositional component is not explicitly assessed but we believe that our assessment prompts the ability to take a critical stance and challenges students' beliefs.

Watson's (1997) three-tier hierarchy is also applicable to our assessment as students are required to understand statistical terminology and recognize statistical ideas embedded in may different contexts and to challenge claims. We believe the assessment is assisting students to become critical consumers of statistically based information whether from a media report, original source, or an everyday life event.

REFERENCES

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1-51.

Murray, S., & Gal, I. (2002). Preparing for diversity in statistics literacy: Institutional and educational implications. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

Pfannkuch, M. & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, (pp. 17-46). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Reichenberg, A., Bresnehan, M., Rabinowitz, J., Lubin, G., & Davidson, M. (2006). Advancing paternal age and autism. *Archives of General Psychiatry*, *63*, 1026-1032.

Tversky, A. & Kahneman, D. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic & A. Tversky (Eds.) *Judgment under Uncertainty: Heuristics and Biases* (pp. 3-20). New York: Press Syndicate of the University of Cambridge (originally published in *Science* (1974), 185, 1124-1131).

Utts, J. (2004). *Seeing through statistics*. Belmont, CA: Thomson Brooks/Cole.

Watson, J. M. (1997). Assessing statistical thinking using the media. In I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education*. (pp. 107-121). Amsterdam, The Netherlands: IOS Press.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, N.J.: Erlbaum Associates.