LEARNING THROUGH INDUCED ERRORS: A GARDEN-PATH APPROACH TO INTRODUCTORY STATISTICS

John Blake
Center for Language Research, University of Aizu
jblake@u-aizu.ac.jp

A garden-path teaching approach was adopted with postgraduate students. Participants were presented with exemplars from which rules were induced and then applied to different contexts. Scenarios to which application of the rule would result in a mistake were selected. The misapplication resulted in overgeneralization errors. Having been lulled into a false sense of security and then deceived, teachable moments are created in which participants are able to learn from their mistakes. Participants are hypothesized to focus more carefully on the features of and exceptions to the rule in order to avoid falling into a similar trap.

BACKGROUND

Researchers, journalists and politicians support their arguments with statistical evidence, but all evidence is not equal. In order to separate the wheat from the chaff, statistical literacy is a must. Statistical literacy is defined by Gal (2004, p.49) as the ability to interpret, critically evaluate, discuss and share opinions on statistical information and data-related arguments. This socially-situated view of literacy is shared by Watson (2006). This paper describes a novel approach to teaching introductory statistics in which teachable moments are created in a supportive yet challenging atmosphere. This course aims to develop statistical literacy, but differs from many introductory statistics courses in at least four ways. First, the course is interdisciplinary; combining statistics with logic (especially propositional logic) and critical thinking (especially reasoning processes and fallacies). Second, this course is for postgraduate students. Third, the aim is to develop awareness of statistical literacy by considering common statistical errors in research articles. Fourth, the course is delivered in English which is not the first language of the participants. One underlying assumption is that students who are able to identify errors in research articles written by different authors are more likely to avoid making the same errors in their research articles.

Context

The study was conducted at the Japan Advanced Institute of Science and Technology (JAIST), which is a university that serves only postgraduate students. Students study for a master's degree and/or a doctorate in information science, materials science or knowledge science. Although the majority of the student population at JAIST is Japanese, slightly under 40% (JAIST, 2017) are international students, mainly from other Asian countries especially China, Vietnam and Thailand.

Course

This study focusses on the development of statistical literacy within a two-credit-bearing course entitled Scientific discussions: logical and critical thinking. The course is designed to enable participants to critically evaluate scientific research papers by developing critical, logical and statistical literacies. This course consists of ten input sessions and five seminar discussions held over one quarter (8 weeks). This paper focuses on the statistical literacy development aspect within this course. Students are expected to be able to develop enough statistical literacy to be able to identify common statistical errors in research articles.

Assessment

The acquisition of these literacies is assessed using (1) in-class quizzes and seminar discussions, (2) online quizzes housed on a Moodle learning management system, and (3) submission of a final assignment. Knowledge of relevant terminology and concepts is tested in the online quizzes while in-class quizzes and seminar discussions are used to assess the ability to apply the knowledge. Prior to the seminar discussion, each student reads four different research articles on a pre-selected theme. In the seminar discussion, each participant presents a critical analysis of

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org [© 2018 ISI/IASE]

their chosen article. After the presentations, the group of four critical analyze the research articles. Participants discuss the claims, evidence, reasoning, assumptions, fallacies, statistical issues and so forth. Participants are expected to participate actively in all set activities during lectures and tutorials. Although there is no extrinsic reward for participation, there is a deterrent for lack of participation, namely loss of 1% of the total grade for the first instance, 2% for the second, 4% for third until the sixth instance when failure is automatic. A penalty card is handed to the student to ensure they are aware of their current status. Yellow penalty cards with the current deduction are handed directly to students emphasize the importance of participation.

Eureka effect

Eureka or "Aha!" moments occur when one is able to suddenly solve a problem or resolve a mystery. These moments are pleasurable and according to neurocognitive research on insight enhance neural networks (Jung-Beeman *et al*., 2004). Insight has been shown to stimulate the activity in the amygdala which plays a key role in long-term memory, and thus providing an optimal learning moment (Ludmer, Dudai and Rubin, 2011). In a classroom or lecture setting, significant planning and preparation is necessary to create an environment conducive to enabling learners to experience such Eureka moments.

Garden-path approach

One way to increase the likelihood of the occurrence of "Aha!" moments is to use a garden-path approach. This approach was introduced in the foreign language classroom by Tomasello and Herron (1988) who found that students who were enticed into overgeneralization of grammatical rules and received correction learned the exception to the rule better. In this approach students are presented with exemplars from which rules are induced and then applied to different contexts. Scenarios to which application of the rule would result in a mistake are selected. The misapplication results in overgeneralization errors. Having been lulled into a false sense of security and then deceived, teachable moments are created in which students are able to learn from their mistakes. Students are hypothesized to focus more carefully on the features of and exceptions to the statistical rule in order to avoid falling into a similar trap. This approach enables students to engage in critical questioning and reasoning, which is the highest level of construct of statistical literacy in the six-level hierarchical model of Watson and Callingham (2003).

CASE STUDY

Participants

There were 22 participants in this study. Fifteen students were enrolled in the two-credit course while seven were auditing the course. All participants were studying for a master's or doctoral degree. Nationalities represented in this cohort include Vietnamese, Indian, Pakistani, Chinese, Japanese, Taiwanese, Thai and Malay. The general level of English language competence was sufficient to understand the content and participate in discussions.

Course and syllabus

The course comprised ten ninety-minute lectures and five tutorials in which assessed seminar discussions were held. Each lecture began with a text analysis that harnessed a garden-path approach. There are three themes in the statistical syllabus, namely (1) misleading representation, (2) bias and (3) fallacies. Table 1 provides more details on some of the issues addressed in each theme.

Table 1: Themes and topics

Theme 1 Misleading representation	Theme 2 Bias	Theme 3 Fallacies
Insufficient statistics	Sample bias	Ad populum
Relative vs. absolute	Cherry picking	Equivocation
Missing details	Public vs. file draw effect	Appeal to statistics
Deceptive statistics	Data dredging	Gambler's fallacy

Three averages	False cause
Outlier effect	correlation ≠ causality
Simpson's paradox	Base rate fallacy
Misleading graphics	

Research articles frequently draw on statistics that are accurate, yet incorrectly represent what the statistics show. The topics in theme one, misleading representation, aim to arm students with some knowledge of how research results are presented, which are not fully supported (or even partially in some cases) by the statistics referred to. One simple example is how using bold colours and increasing the size of bars on a bar chart can visually trick readers into making false assumptions. Bias, the second theme, shows how taking a partial perspective can skew results. In published research, the skew is often in favour of the outcome that the researcher was striving to achieve. Fallacies that tie closely to misuse or statistics are dealt with in theme three. These fallacies rely on errors in reasoning, which less critical readers and those with little statistical knowledge may not question.

Materials and delivery

Ten short texts were extracted and, in some cases, adapted from various sources, including newspapers, textbooks, published and draft research articles. Short texts were chosen to enable the students who were less proficient in English to be able to read them in the allotted time. In a longer course or with more proficient users of English more detailed texts would be selected. These texts were used to contextualize various teaching points to raise the level of statistical literacy. Other potential sources of easily-available materials include advertisements and political speeches. The resource book "Teaching statistics: A bag of tricks" (Gelman and Nolan, 2002) contains numerous ideas and texts that could be used, particularly in chapter 10, which addresses statistical communication, aka. "lying with statistics".

The course content is structured in such a way so to help students identify statistical errors. This is achieved via a cyclical or spiral syllabus in which teaching points are revisited and refined in each iteration of the cycle. Texts were selected either as exemplars of good practice or as examples of the misuse or abuse of statistics. The texts were sequenced to create a learning progression beginning with more fundamental concepts.

Taking Text 1 as the starting point, students learn the importance of checking the source of the statistics. In the first text a statistic is pulled from thin air. The anticipated response of the students was to accept the truth value of the statistic without question. The teaching point was to check the source of any statistics provided.

Text 1. Power of statistics
The oft-quoted statement that 97% of statistics are made up should show the general public that research does not discover facts, but serves other purposes. Researchers who fabricate statistics are rarely challenged since the assumption is that the data reported in research articles exists. This default setting of "honest and ethical researchers" is rather naïve, given the positive effect that publishing in top-ranked journals can have on academic careers.

In the second text the source of the statistic was provided, and so students were expected to accept the truth value of the statistics, but this time there was a different issue. The statistics given were insufficient as relative percentages were used without absolute numbers.

Text 2. Downtown
Large businesses are destroying the small town feel of our "downtown" area. Just last year, the number of large businesses in the city has increased by 75%. (adapted from Brown & Keeley, 2011, p.171)

The accumulative result is that for the third text students are expected to check the source of statistics and assess whether the statistics are sufficient. This continues throughout the series of texts. Texts 3 and 4 deal with chi-square hypothesis testing, p-values and degrees of freedom. In

the fifth text, an appeal to statistics is used to mislead readers into believing a claim that is not supported by the data.

Text 5. GoFat
A new weight-loss drug, GoFat, is effective in helping obese people lose weight. In a clinical trial, only 6 out of 100 people on GoFat reported any side effects with taking the drug. The company manufacturing the drug argues "With 94% of people having positive results with GoFat, it is safe to say our pill is one of the most effective weight-loss pills in the market".
(adapted from Brown & Keeley, 2011, p.171)

The text states that six people of the 100 in the sample had side effects, but it does not report the number of people who actually lost any weight. Instead, it focuses on the fact that 94% did not report side effects. The claim that it is the most effective weight-loss pill is based on no evidence at all.

At this stage students should be familiar with hypothesis testing, chi-square, degrees of freedom and p values, and so Text 6 gives them a chance to use this newly acquired knowledge. Students are expected to check the statistics and fall hook, line and sinker into believing the conclusion. The main teaching point is the equivocation of significant, which gives the teacher the opportunity to discuss the differences between practical and statistical significance.

Text 6. Water
The data (see Table A below) give a chi-square test statistic of 18.89 with 3 degrees of freedom and a p-value of 0.000288.

Table A. Data for the Chi-Square Goodness-of-Fit Test for First Preference in the Double-Blind Test

Type of Water	Tap	Aquafina	Fiji	Sam's Choice	Total
Observed	12	27	44	26	109
Expected	27.25	27.25	27.25	27.25	109
Contribution to Test Statistic	8.534	0.002	10.296	0.057	18.889

Because this is a significant result, the local government needs to ascertain the reasons why tap water is less popular than the other choices.

Texts 7 to 9 develop students understanding of potential problems using t tests and one-tailed tests. In the final tenth text, students are expected to check the statistics and reach the same conclusion as presented in the text. Naturally, that conclusion is false, and so provides the teacher the perfect opportunity to explain the Simpson's paradox.

Text 10.Simpson's paradox
The University of California, Berkley was sued for bias against women when it was discovered that 11% fewer women were admitted than men (men, 44%: women, 35%). This disparity shows clearly that men are 25% more likely to be accepted to admitted than women. (Bickel, Hammel, & O'Connell, 1975).

METHOD

A test-teach-test format was used. In each of the ten input sessions participants were given up to ten minutes to read one of the ten texts and complete a short one-question formative assessment. Each text was selected to increase the likelihood of "Aha! Moments". Teaching points were identified, and student responses anticipated. The instruction on the formative assessment was usually to write a critical evaluation of the text. Students then discussed their answers in pairs or small groups. Afterwards, student responses were elicited and feedback was provided to the whole class, emphasizing the teaching point. Feedback was given in both oral and written forms. The oral

feedback centered around the elicited responses and was pitched at a level appropriate to the students. The written feedback was more comprehensive and was targeted at students who wanted a more in-depth understanding of statistical concepts.

Participants' knowledge of and ability to apply the statistical concepts were tested in the end-of-course summative online assessment that was based on ten different texts, but contained similar statistical errors to the texts used in class work. There were ten questions in total in both the formative and summative assessments.

RESULTS

Formative and summative assessment scores

The number of students able to identify the statistical errors in each formative assessment was extremely low with at most two students being able to pinpoint the problem. Ten students completed all the assessments. The results are given in Table 2. The mean formative assessment score for the 10 students was 8% while the mean summative assessment score for the 10 students was 79%.

Table 2: Number of students (n=10) able to identify statistical errors in each assessment question

	Q1	Q1	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
Formative assessment	0	2	0	2	0	1	1	1	1	0	8
Summative assessment	7	10	7	8	6	8	9	10	7	7	79

Participant feedback

Qualitative responses from participants were overwhelmingly supportive of this approach. In the university-administered end-of-course questionnaire there is opportunity for students to write their comments. Some students commented on their learning, e.g. "I learnt a lot." [student 1], "I did not know much about statistics before." [student 2] "It was fun. It was easy to learn". [student 5]. Other students commented (without prompting) on the garden-path approach, e.g. "The professor tricked us and we fell into the trap" [student 8], "I knew my answer was going to be wrong. I knew it was a trap, but I could not escape" [student 10].

Trainer feedback

The trainer comments were equally positive. Students with no or little prior statistical training invariably walked down the proverbial garden path and to their chagrin had the gate slammed in their face. The trainer was able to see eyes light up when students realized they had overgeneralized or had missed some key point which with the benefit of retrospect was obvious; but, at the time, had evaded them. The classroom delivery of the garden-path section was straight forward, encouraged active participation of students and gave the trainer the chance to assess which students were likely to need more support and which would benefit from more challenging activities.

DISCUSSION

Critical analysis of texts contextualizing the use, misuse or abuse of statistical evidence raised awareness of the importance of statistical literacy. Ten texts were sequenced to enable learners to apply knowledge gained from earlier texts to later texts. According to the positive feedback, the garden-path approach was well received by both the students and the trainer. Summative assessment scores were invariably higher than the formative scores, indicating that knowledge had been acquired and was able to be applied.

A potential criticism of this approach could be that participants may be left with the idea that statistics used in research articles are never reliable. This is a valid point if the only exposure to statistics was in the ten short texts, which contained errors. However, an abridged version of an exemplary research article written by Peltokorpi, Allen and Froese (2015) was shared and discussed at length in the critical literacy phase of this course. This article drew extensively on statistics to support its claims, and had the added advantage that the first author was a professor at the same institution.

Compared to the formative assessment, summative assessment scores increased almost tenfold. Learning tends to occur more rapidly lower down the learning curve, and so a substantial increase is unsurprising in any introductory course. The mean gain of 71% may be attributed to numerous factors, one of which is dropout bias since students who complete the course may not be representative of all students. Various other confounding factors no doubt played a role. Although the causality of the learning was not established and the small sample size prohibits robust statistical analysis, it is hypothesized that the textual analysis, garden-path approach and subsequent discussions contributed to the overall improvement in mean score.

This garden-path approach can be used in a lecture format with a large class, and works especially well when there is a student response system, such as clickers so students can anonymously share their views with the rest of the class.

An alternative to using written texts as the vehicle to introduce the scenario and its associated garden-path is to use video extracts. Television commercials, such as those selling hair and beauty products, which frequently use statistics to persuade viewers. Videoed extracts of newscasters referring to misleading visuals are easily found. Interviews with politicians and even conversations in a popular serial drama are all fruitful sources.

This pilot study lays the groundwork for future work, which could use multivariate analysis in conjunction with pre-test and post-test data to compare the efficacy of adopting a garden-path approach with other teaching approaches.

REFERENCES

Bickel, P.J., Hammel, E.A. & O'Connell, J.W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science*, *187*(4175), 398–404.

Brown, M. N., & Keeley, S. M. (2011). *Asking the right questions with readings*. NJ: Pearson.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In J. B. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht: Kluwer.

Gelman, A., & Nolan, D. (2002). *Teaching statistics: A bag of tricks*. New York: Oxford University Press.

Jung-Beeman M, Bowden EM, Haberman J, Frymiare JL, Arambel-Liu S, Greenblatt R, et al. (2004) Neural Activity When People Solve Verbal Problems with Insight. *PLoS Biol 2*(4): e97. https://doi.org/10.1371/journal.pbio.0020097

Ludmer, R., Dudai, Y., & Rubin, N. (2011, March 10). Uncovering Camouflage: Amygdala Activation Predicts Long-Term Memory of Induced Perceptual Insight. *Neuron*, *69*(5), 1002–1014.

JAIST (Japan Advanced Institute of Science and Technology). (2017). Quota / Actual number of students. https://www.jaist.ac.jp/english/about/outline/student.html

Peltokorpi V., Allen D. G., & Froese F. (2015), Organizational embeddedness, turnover intentions, and voluntary turnover: The moderating effects of employee demographic characteristics and value orientations, *Journal of Organizational Behavior*, *36*, 292–312.

Tomasello, M., & Herron, C. (1988). Down the garden path: Inducing and correcting overgeneralization errors in the foreign language classroom. *Applied Psycholinguistics*, *9* (3), 237–246. https://doi.org/10.1017/S0142716400007827

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

Watson, J. M. & Callingham, R. (2003). *Statistical literacy: A complex hierarchical construct. Statistics Education Research Journal, 2*(2), 3–46.