ODYSSEY: A JOURNEY TO LIFELONG STATISTICAL LITERACY

Milo Schield
Augsburg College, United States of America
Schield@Augsburg.edu

Statistical literacy is the ability to read and interpret everyday statistics in tables, charts, statements, surveys and studies. This skill needs to be developed in college so it can be sustained after college. This skill is more about evaluation and critical thinking than about calculation, derivation and proof. Forums can be used to help students develop this statistical literacy skill online. This paper introduces Odyssey: a new kind of forum where all participants are anonymous, everyone grades everyone, and the system tabulates average scores for each participant. This super-forum has two big advantages. (1) Students like Odyssey: They think it is easy to use; they like being anonymous, seeing how others think and getting immediate feedback and grading; (2) Odyssey is scalable to large-lecture classrooms. Results from using this unique forum are presented. By dealing repeatedly with everyday statistics in the news, students take their first step toward lifelong statistical literacy.

INTRODUCTION

Forty-five percent of college students show no significant improvement in the key measures of critical thinking, complex reasoning and writing by the end of their sophomore year (Arum & Roska, 2011). This finding ties in with this session: "How do we engage people in the learning of statistical literacy beyond a typical classroom setting? What are innovative methods by which the knowledge and concepts learned will be sustained as an integral part of lifelong learning?"

Moore (1998) distinguished "statistical literacy" (what every college graduate should know) from "statistical competence" (what we hope a business statistics student will retain five years later). Moore's literacy-competence dichotomy is quite different from the literacy-reasoning-thinking trichotomy analyzed by delMas (2002). The latter distinguishes cognitive activities as does Bloom (1956); the former distinguishes the statistical needs of the two groups of students.

Schield (1999, 2010) follows Moore (1998) and defined statistical literacy as "the ability to read and interpret data: the ability to use statistics as evidence in arguments", "the ability to think critically about statistics" and "the ability to read and interpret everyday statistics in tables, charts, statements, surveys and studies". Isaacson (2005) noted that statistical literacy involves hypothetical thinking: a type of thinking that most students have not developed. Statistical literacy is similar to quantitative literacy (Q/L) which the AAC&U (2009) described as a "habit of mind", a "competency, and comfort in working with numerical data".

Practice is required to develop and sustain a habit. Writing is an excellent way to demonstrate critical thinking about complex arguments. Evaluating student writing means grading. An increasing amount of student writing involves online forums. These fall into three groups:

1. Basic forums such as the Moodle Discussion board. [See for example, Schmid (2011) and Kreiner (2006).] Everson and Garfield (2008) provide extensive background along with using group feedback.
2. Participant-blinded forums (eliminates free-riding among those posting last). Teachers using Moodle's discussion board forums should consider upgrading to Moodle's Q&A forum.
3. Advanced forums in which all participants are anonymous (to minimize measurement bias), where everyone grades each other (to maximize different perspectives and minimize measurement error) and where the system uses grades to compute each participant's overall score (power).

THE ODYSSEY FORUM

Odysseys2Sense ("Odyssey": a game of lively discourse, http://odysseys2sense.com) is an advanced forum. [See for example, Schield & Copes, 2011; Schield, 2012a; 2012b for detailed description.] Instructors create *challenges*: analyze this new story, this survey or study. Each student submits their *response* to the challenge and then is required by the system to *review* and

grade three or four *responses* by others. Only then can they able see anything else. Late-posting students can't "free ride" by seeing the work of those who posted earlier.

The system calculates each participant's *power* based on (1) ratings received from others weighted by (2) the *power* of those giving the ratings. Additional power can be earned by participants who review additional responses. The program shows each individual their own power and the median power for all those in that Odyssey. By requiring evaluation with reasons, the peak of the Bloom (1956) taxonomy, Odyssey offsets statistics' traditional focus on computation (application).

Power in Odyssey is similar to *power* in games. Higher power gives you more status; your reviews have more weight. Gamers find this very familiar. [See Appendix A for details on power.]

Odyssey has been used at Augsburg 12 times by 170 students in 14 classes. Classes include traditional statistics (BUS379), statistical literacy for managers (BUS264) and statistical literacy for students in non-quantitative majors (GST200). The number of challenges per semester ranged from 8 to 24 with an average of 12. An improved version of Odyssey was introduced in fall 2012. Augsburg has used it five times since then: twice for statistical literacy and three times for traditional statistics. All of the data in this paper was obtained from the last two traditional-statistics classes (the first use tested new challenges that better matched the class content). Although Augsburg's class sizes are small, Augsburg's experience supports the idea that Odyssey can handle very large classes.

ODYSSEY CHALLENGES

The key to good forums is the same as that for good discussions (Everson & Garfield, 2008). Challenges should be open-ended (as opposed to right-wrong), have multiple right answers, encourage multiple approaches and require analysis, synthesis and evaluation. To date, more than 40 challenges have been tested in teaching statistical literacy and traditional statistics online, hybrid and face-to-face. Twenty of these challenges (many featured in traditional statistics) are listed here:

Critical thinking exercises
1. Can Critical Thinking be taught?
2. How Much Math do we really Need?
3. Damned Lies and Statistics: Joel Best
4. Coincidence? Lottery winners bad luck

Reading and interpreting tables and graphs
1. Per-Person Spending: Married vs. Single
2. US Dropout-Rates by race, ESL, etc.

Reading and interpreting surveys
1. 1 in 50 US Kids is homeless: study
2. Halloween Consumer Survey (2012)

Statistical Inference Topics
1. Survey: 95% Margin of Error
2. Statistically Insignificant=No Difference?
3. Statistical Tie = Statistical Dead Heat

Explaining Data Patterns
1. Excess of Males in the SAT tails
2. Are Heights Normally-Distributed?

Observational studies: Cross-sectional
1. Spanking Lowers IQ
2. Women Who Drink Tend to Be Thinner

Observational studies: Longitudinal
1. Fewer Boys Following 9/11
2. Women on the Pill Live Longer: Study
3. High gas Prices drive down Car Fatalities

Randomized Experiments
1. Booze + Diet Soda = Big Buzz?
2. Bigger Tableware Widens Waistlines
3. Giving Criminals Money after Release Cuts Recidivism?

Teachers may find that writing good challenges about current news stories can be difficult. Teaching students how to evaluate and grade their peers can be even more challenging.

EXAMPLES OF CHALLENGES
- *Statistical Tie*: Challenge: Do some web research on "statistical tie" and "statistical dead heat" in describing a close race between two candidates. Comment on (1) the advantages and (2) disadvantages of these two phrases. Number the two parts of your answer. *Evaluation*: The challenge requires some knowledge of statistical inference and statistical significance. Knowing the difference between a criminal trial and a civil trial is helpful.
- *Bigger Tableware Widens Waistlines*. Read the article at www.StatLit.org/CP/2006-Bigger-Tableware-Helps-Widen-Waistlines.pdf). What opportunities are there for bias in this study? 2)

Evaluate the quality of the argument. How strongly do these statistics support the point of the story? Any plausible confounders? Give your reasons.
Evaluation: Few students realized that randomly assigning plates eliminated almost all the plausible confounders that might influence the observed association between bigger plates and bigger portions – even though they had studied that same kind of situation previously.

GRADING CHALLENGES
When grading responses, Odyssey reviewers grade three specific items:

- *Responsiveness*: Acceptable, comprehensive answers to ALL questions (+1); Partially acceptable or less than comprehensive answers to SOME questions (+1/2); Minimally acceptable or unacceptable or missing answers to some questions (0).
- *Explanation*: Valid, comprehensive AND insightful justification/explanation for ALL answers (+1); Inadequate, not fully comprehensive OR not very insightful justification/explanation for SOME answers (+1/2); Minimal or missing justification/explanation for SOME answers (0).
- *Extension*: Thought-provoking extension question or connect (+1); Superficial extension question or connection (+1/2); No extension question or connection (0).

A rating of 0 is given with no response (1 when a response is given but hasn't been reviewed by anyone). The maximum rating is 4: a one point maximum for each item plus one point for answering. The three dimensions follow Perry (1970): answering (dualism) = 1; comprehensively responding (multiplicity) = 2; explaining opinions (relativism) = 3; integrating ideas (extension) = 4. The prior version allowed grading on a continuous scale. Students did not like it; they felt it led to greater variation in grades from different graders. Those adept at grading might prefer a scale with better resolution and discrimination. Most students have never graded and are not that adept.

ANALYSIS
Appendix B summarizes results of using Odyssey in spring and fall 2013 by 46 working-adult Business majors taking traditional Statistics at Augsburg College. As they used the system, participants' perceived value increased as did their ease of use, although their enjoyment decreased. When asked if Odyssey should be used in future classes (Question #9), 22% strongly agreed, 43% agreed, 24% were neutral, 11% disagreed and no one strongly disagreed.

Appendix C shows the relation between student power and how they answered the survey questions for the 21 students in spring 2013. When converting the five ordinal levels into a ratio scale from zero to four, the average score on Q9 was 3.1 for those with above median power (2.1 with those at or below the median power). Among those with above median power, 45% strongly agreed that Odyssey should be used in the future (only 8% among those with median or below median power). None of the above-median students disagreed with using Odyssey in the future (all of those who disagreed had at or below median power). This gives additional support to continue using Odyssey.

Appendix D summarizes what students like most and least about Odyssey when given only five choices (based on prior students' comments). Being anonymous was what 56% liked best; grading their peers was what 40% liked least (36% wanted more teacher feedback on a challenge after it was due). From a teaching perspective, grading peers is more than offloading the teacher's job onto the students. Being able to grade others should improve one's ability to evaluate one's own writing and thinking. More work is needed to help students become critical thinkers who can analyze, synthesize and evaluate arguments clearly, succinctly and helpfully.

Appendix E lists students' open-ended comments on what they liked best about Odyssey. Students tend to see less value after taking statistics than they did before (Millar & Schau, 2010). Non-quantitative students tend to find statistics most challenging. Students using Odyssey report that extensive use of this forum allowed them to use their verbal skills which reduced anxiety and improved their motivation. Focusing on numbers in the news helped them see statistics as valuable.

Given student discomfort in giving grades, materials are needed to help them improve the quality of their grading and their reasons. Research is required to see if Odyssey improves student understanding, to see if Odyssey students see more value in statistics after the course than before,

and to see if Odyssey students can be trained to give grades that are accurate and helpful.

CONCLUSION

Students like analyzing everyday statistics in graphs, tables, stories and studies. Odyssey is the vehicle that takes them on that journey. Students like being anonymous; they like getting immediate feedback provided the grade is accompanied by reasons; they like receiving multiple reviews on their response; they like being able to critique a rating they think is unjustified; and they think Odyssey should be used in the future (especially the students who did well using Odyssey).

Teachers like having students who are better able to think critically: to see weakness and strength in a given response; they like being able to give students current news stories that involve the materials in the course; and they like having students who find statistics enjoyable and relevant.

Odyssey takes online education to a new level. It provides new opportunities – and places new demands – on teachers and students. When used in teaching statistical literacy, Odyssey requires students to think critically about everyday numbers in the media. Statistical educators should use advanced forums such as Odyssey to promote lifelong statistical literacy: critical thinking about everyday statistics as evidence.

ACKNOWLEDGEMENTS:

REFERENCES

AAC&U (2009). Quantitative Literacy VALUE Rubric. [VALUE: Valid Assessment of Learning in Undergraduate Education] www.aacu.org/value/rubrics/pdf/QuantitativeLiteracy.pdf

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses.* University of Chicago Press.

Bloom, B. S. et al.; (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook I: Cognitive domain*. New York: David McKay Company.

del Maas, Robert (2002). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education. 10(*3). www.amstat.org/publications/jse/v10n3/delmas_discussion.html

Everson, M., & Garfield, J. (2008). An innovative approach to teaching online statistics. *Technology Innovation in Statistics Education*. www.escholarship.org/uc/item/2v6124xr

Isaacson, M. (2005). Statistical literacy – An online course at Capella University. *Proceedings of the Section on Statistical Education*. See www.statlit.org/pdf/2005IsaacsonASA.pdf

Kreiner, D. (2006). A mastery-based approach to teaching statistics online. *International Journal of Instructional Media*, *33*(1), 73. Win 2006. http://eric.ed.gov/?id=EJ749765

Millar, A., & Schau, C. (2010). Assessing students' attitudes: The good, the bad, and the ugly. ASA. www.statlit.org/pdf/2010MillarSchauASA.pdf

Moore, D. (1998). *Statistical literacy and statistical competence in the 21st century*. MSMESB Univ. Iowa. Abstract only. See http://www.statlit.org/pdf/1998MooreMSMESB.pdf

Perry, W. (1970). *Forms of intellectual and ethical development in college years: A scheme*.

Schield, M., & Copes, L. (2011). "Odysseys: Teach Critical Thinking" Poster. Augsburg College. www.statlit.org/pdf/2011SchieldCopesPosterOdysseys.pdf

Schield, M. (1999). Statistical literacy: Thinking critically about statistics. APDU *Of Significance, 1*(1), 15-20. www.StatLit.org/pdf/1999SchieldAPDU.pdf

Schield, M. (2010). Assessing statistical literacy: Take CARE. Ch 11 in *Assessment Methods in Statistical Education* (pp. 133-152). Wiley.

Schield, M. (2012a) Odyssey user guide. At www.StatLit.org/pdf/2012-Odysseys-Guide-1up.pdf

Schield, M. (2012b) Odyssey systems guide. www.StatLit.org/pdf/2012-Odysseys-Guide2-1up.pdf

Schmid, K. (2011). Virtual discussion for real understanding. *ASA Proceedings of the Section on Statistical Education* (pp. 2108-2115). See www.statlit.org/pdf/2011Schmid-JSM.pdf

APPENDIX A: *POWER* IN ODYSSEY

Power has a different meaning in statistics than in Odyssey but that technical meaning is not encountered the end of the course and is seldom confused with the common notion of ability or status.

Power can be used to assign grades. One way is to standardize power relative to the class median: A (>105% of the class median), B (100-105%), C (80-100%), D (60-80%) and F (<60%). The system notes if a student response is late so teachers can deduct power points if they choose.

A participant's *power* reflects the quantity and quality of their responses. For example, getting ratings of 3 from all your reviewers on all your responses and posting two reviews per challenge (25 points max) will increase your power by about 350 points per challenge: 3,500 after 10 challenges.

Students can "game" the system by giving more reviews (25 points max). But we want students to spend more time on task. Reviews can be graded; bad reviews can get bad ratings and lower power.

Since Odyssey is anonymous, instructors can grade whenever they want. This negates student being nice (give high ratings) or colluding. Grading everyone initially helps establish good practice.

Odyssey keeps instructor grades separate from participant grades. This allows instructors to see how their instructor grade compares with the participant grades for a given student and challenge.

APPENDIX B: PARTICIPANT FEEDBACK ON ODYSSEY FORUMS (2013)

46 respondents: 21 in spring; 25 in fall. 100% response rate from all classes

QUESTIONS	**CHOICES (Multiple choice)**
Q1: How easy was Odyssey to use at first	0 Very difficult; 2 OK/mixed; 4 Very easy
Q2: How easy is Odyssey to use now	0 Very difficult; 2 OK/mixed; 4 Very easy
Q3: How valuable was Odyssey at first?	0 Almost worthless; 2 OK; 4 Very valuable
Q4: How valuable is Odyssey now?	0 Almost worthless; 2 OK; 4 Very valuable
Q5: How enjoyable were the challenges for you at first?	0 Almost painful; 2 OK; 4 Very enjoyable
Q6: How enjoyable are the challenges now?	0 Almost painful; 2 OK; 4 Very enjoyable
Q7: How helpful were the comments you received?	0 Almost worthless; 2 OK; 4 Very helpful
Q8: How accurate were the scores you received?	0 Almost worthless; 2 OK; 4 Very accurate
Q9: Odyssey should be used in future classes.	0 Strongly disagree; 2 Neutral; 4 Strongly agree
Q10: Odyssey could be improved with better challenges.	0 Strongly disagree; 2 Neutral; 4 Strongly agree

Class	Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	ALL
Spring	Average*	2.5	**3.1**	**2.2**	2.6	2.6	**2.1**	**2.0**	**2.2**	2.6	2.4	2.4
	Mode	3	**4**	2	2	2	**1**	2	2	2	3	2
Fall	Average*	2.5	**3.0**	2.3	**2.9**	2.5	2.7	2.4	2.8	**2.9**	2.0	2.6
	Mode	3	**3**	2	**3.5**	3	3	2	3	3	2	3
ALL	Average*	2.5	**3.1**	**2.2**	2.7	2.5	2.4	2.2	2.5	2.8	2.2	2.5
	Mode	3	3	2	2	3	3	2	3	3	2	3

* Average after treating ordinal data as ratio. Note: Q6 is bimodal; all others have a single mode

Change in average for all 46 respondents from the start of course (then) to the end (now):
Q1 vs. Q2: Easy: then vs. now: 2.5 vs. 3.1.   Odyssey got easier as time passed.
Q3 vs. Q4: Valuable: then vs. now: 2.2 vs. 2.7   Odyssey got more valuable as time passed.
Q5 vs. Q6: Enjoyable: then vs. now: 2.5 vs. 2.4   Odyssey got slightly less enjoyable.
Note: In spring, three students completed less than four of the 9 challenges. Their scores are included.
Q9 Rating (#): 4(10), 3(20), 2(11), 1(5)   30/46 agree: "Odyssey should be used in future classes"

APPENDIX C: AVERAGE SCORES AS A FUNCTION OF POWER (21 students in spring)

Correlation of Answers with Power for 17 Spring Participants

Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Correlation	0.56	**0.62**	0.54	0.34	0.39	0.52	0.25	0.14	**0.63**	**⁻0.61**

Positive correlations mean high-power students agreed more than low-power students.
The negative correlation for Q10 means high-power students agreed less than low-power students.

Average Scores Classified by whether students had above or below average power.

Average	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Above	2.8	3.4	2.4	2.6	2.9	2.4	2.1	**2.1**	**3.1**	2.2
Below	2.2	**2.9**	2.0	2.5	2.4	**1.8**	1.9	2.3	2.1	2.6

Diff	0.6	0.5	0.4	0.1	0.5	0.6	0.2	-0.2	**1.0**	-0.4
%diff	28%	17%	20%	2%	23%	32%	10%	-8%	**48%**	-17%

Q2 had the highest scores for each group. Odyssey is very easy to by the end of the course.
Q9 had the biggest difference between students based on above vs. below average power.
Q6 had the second biggest difference in average scores based on above vs. below average power

APPENDIX D: WHAT 25 STUDENTS LIKED BEST & WORST ABOUT ODYSSEY

#1	#2	**Choice of Best**
4	10	Being totally anonymous
1	1	Getting immediate feedback
2	5	Use of critical thinking
2	0	Reviewing/grading peers
9	16	Total

#1	#2	**Choice of Worst**
1	0	Being totally anonymous
5	4	Lack of teacher feedback
1	4	Being graded by your peers
2	**8**	Reviewing/grading peers
9	16	Total

The #1 and #2 reflect separate fall classes working in the same Odyssey with the same challenges.

APPENDIX E: WHAT 25 STUDENTS LIKED BEST ABOUT ODYSSEY (anonymous essay)

- The Odyssey challenges helped me to think about and apply what I learned from the book and in class sessions, as well as the feedback received by peers. It would have been great to analyze the challenges more in class.
- I like being anonymous to review others work. It was nice being able to review with the fear of criticism of our reviews.
- The best quality about Odyssey was the immediate feedback I received.
- It helped me read and understand graphs and articles better using my critical reasoning skills. Doing the odysseys helped me develop more confidence in my ability to critically evaluate articles and graphs and charts.
- I liked creating a response to the questions before having the opportunity to read what others had replied. I felt my initial response was independent.
- It allowed me to see how others considered and contemplated the challenges. That is what I was most looked forward to in the class.
- I was able to think in many different aspects. there's no right nor wrong answers.
- I enjoyed having to think and evaluate the case studies and statistics in each example. It also allowed us to see how others answered the questions and open up some new possibilities ….
- I liked seeing all the responses after submitting and reviewing the required subset. There was usually several different factors focused on in each Odyssey.
- It's an easy system to navigate. The challenges went with exactly on what we learned that work so it helped with the homework.
- Outside sources for review
- I thought it was an interesting way to have students put in their input.
- Learning from each other.
- It was focused on the next chapter we were reviewing. The first few were the most difficult as we did not really understand what we were doing. I suggest leaving it as is, because it forces you to think outside the box, review the next chapter and then get an opportunity after posting your response to validate you were on track.
- I like that Odyssey uses real life situations and articles to emphasis how advertisers use statistics to their advantage and also how to look through the message that the advertiser is presenting to find the truth.
- Defining and interpreting data.
- I enjoyed not having to worry about upsetting someone.
- You were able to express your opinion on the subject
- Got to complete the responses on my own time, rather than having a specific time to be online and interact with the rest of the class.
- I like the premise of voting on the quality of posts.
- I like using Odyssey in conjunction with my chapter studies because it helps me confirm what I've learned. It also adds more insight by seeing how other students consider the questions.
- I thought Odyssey was a great format for this course. To me, not being able to see whose post it was, was great. I really enjoyed this. I wish I would have put in more effort for these to be honest.