Statistical Literacy Skills Survey

To do: Minitab R-sq for correlation model. Cronbach's alpha for 30 and 35 question regression models.

Milo Schield, Augsburg College
W. M. Keck Statistical Literacy Project

Abstract

Although terms like numeracy, quantitative reasoning and statistical literacy lack generally-accepted definitions, the W. M. Keck Statistical Literacy Project has investigated the presence or absence of skills involved in reading the statistics commonly encountered in the everyday media. A 2002 survey indicated that students and faculty had difficulties relating simple mathematical concepts such as ratios with their expression in tables, graphs and ordinary English. Subsequent papers have reviewed the statistical elements in hundreds of news stories. On this basis, a 69 question assessment was given to 100 students at Augsburg College. This assessment has reliability (Cronbach alpha) of about .71. Models based on unconditioned and conditional correlations were generated and compared. As expected the multivariate regression model was generally superior. At present these instruments are not suitable for institutional use since they lack content validity. Suggestions for improvement are offered.

1. General terms

Terms and phrases such as literacy and reasoning have slightly different interpretations depending on the context as do quantitative and statistical. Thus, combinations of these words such as quantitative reasoning, statistical reasoning, quantitative literacy and statistical literacy have even more variability.

Yet from a general education perspective, these terms reflect skills or aptitudes that are desirable for college graduates. If colleges and universities are to measure their contribution to the education of their students, these terms need generally-accepted definitions and operational assessments.

A fundamental distinction underlying this discussion is the difference between observing and doing: between reading and writing, between listening and speaking, between understanding arithmetic and doing arithmetic, between interpreting a report on a statistical survey or experiment and doing a statistical survey or experiment.

In some cases, the doing may be implied by these terms as in the case of literacy. A literate person may be one who is able to read and write, to listen and speak. In other cases, the doing may be de-emphasized. A numerate person may need to be familiar – at home with – the interpretation of numbers, but may not need to be 'at home with' the generation of numbers.

In some cases, these terms are used to indicate a basic level of proficiency as compared with a more advanced level. For example in the UK, numeracy has come to mean 'school mathematics' such as arithmetic, algebra, geometry and trigonometry.

Quantitative courses may focus on the doing (design, execution, analysis and summarization of surveys, experiments or studies) or on the interpretation and evaluation of the numbers in reports or news summaries of these surveys, experiments or studies. This paper focuses on the latter.

These consumer-based quantitative courses are generally described as 'numeracy', 'statistical literacy' or 'quantitative literacy' (QL) courses.

Teaching a quantitative course based on evaluating numbers in the news has been recommended by educators in statistics. The American Statistical Association's GAISE college report (2006) suggests assessing statistical literacy by students "interpreting or critiquing articles in the news and graphs in media."

2. The 2002 Statistical Literacy Skills Survey

In 2002, an international survey of selected statistical literacy skills was conducted by Schield (2005).

Schield (2004) surveyed the prevalence of statistical words and phrases in USA Today, in Nature and in the Economist. Raymond and Schield (2008) surveyed the prevalence of statistical phrases and ideas in the media.

3. The 2008 Statistical Literacy Skills Survey

In 2008, a 69 question multiple-choice survey was compiled. A copy is in Appendix G. Objective format questions may seem unable to give authentic assessment. Cobb (1998) argues this is not necessarily so.

Of the 69 multiple choice questions, 40 are true false while 29 involve a richer set of answers. As shown in Appendix A, five involve a range of measures, 16 involve an A/B/Both/Neither, three involve an A/B/same, two involve A/B/C/same, one involves an A/B/C/more-than-one, and two involve A/B/Can't tell.

This survey was taken by 101 Augsburg students. The results were analyzed for data entry errors (values outside the range of answers for a given question). Questions having a high number of blank entries were closely examined along with those having a higher number of "Don't know" or "None of the above" entries. Blanks answers were treated as errors.

Blank answers were definitely associated with particular students. Of the 101 surveys taken, one student stopped after doing the first 10 questions. This survey was omitted. Of the 100 remaining surveys, 77 students left no blank answers, 15 left one question blank, four left two questions blank, while four students left five, seven, eight and 13 questions blank respectively.

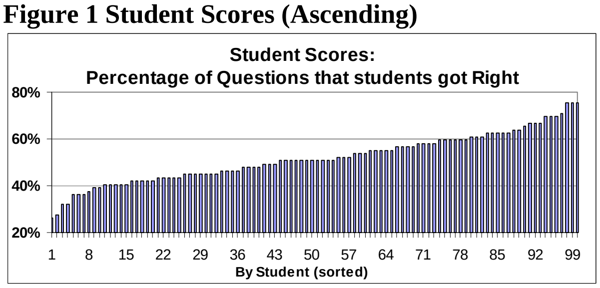
Blank answers were not strongly associated with particular questions. Of the 69 questions, 28 had no blank answers, 27 had one blank answer, nine had two blank answers and five had three blank answers.

The modal answer for each question was compared with the 'correct' answer. Of the 69 questions, the modal answer was the correct answer in 41. Of the 28 where the mode disagreed with the correct answer, 18 had a mode that was the majority while 17 had a mode/correct ratio of more than two. Of these, 15 disagreements involved both. These 15 questions were 15, 22, 27-32, 34, 38, 57, 61, 64, 76 and 77. These 15 were closely examined for errors or design flaws.

3.1. Analysis by Student

Overall scores as a percentage of the 69 questions that a student got correct ranged from 26% to 75% as shown in Figure 1. Scores had a mean of 35.7 (52%) and a standard deviation of 7.2 (10.5 percentage points).

Figure 1 Student Scores (Ascending)



A good instrument should result in student scores that are spread out – rather than being clustered together. A good instrument should avoid having the average be near the top (the ceiling) or the bottom (the floor).

The fact that no student scored below 25% or above 75% indicates there may be room for improvement in this instrument by eliminating some of the questions.

3.2. Student Comments

Student comments were both positive and negative.

Positive/Hard:

- Interesting survey
- This is like an IQ test.
- Didn't realize how in-depth surveys were!
- Hard!!!
- Loved completing this survey. Sometimes I wondered if there are trick questions? Thank you.
- Very cool survey!
- ITS HARD!
- Good luck!
- This was a very good survey. A good variety of situations for the statistics.

Negative/Problems/Easy

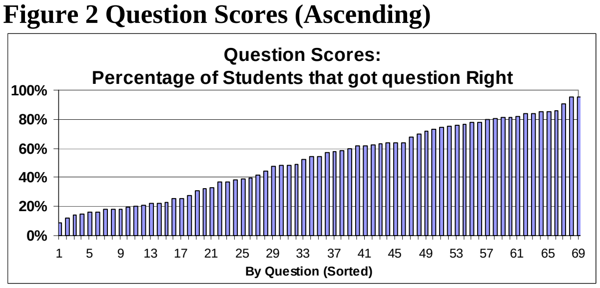
- Many questions worded poorly to the point of being ambiguous; good luck w. more useless stats.
- Questions not clear. Jumped around too much.
- I could do more percentages if I had a calculator.
- Are calculators allowed? I didn't use one.
- Don't care! Statistics should be acknowledged by the dept [of] one's major.
- The survey is interesting but too long. Surveyer may lose focus, concentration or patience. Therefore survey results may not be accurate.

Based on these comments, future survey instruments should note that using a calculator is OK. This survey is too long. But this is needed to identify which questions do the best job of predicting student scores so a future instrument can be shorter.

3.3. Analysis by Question

Appendix B ranks the questions by percentage correct and examines the ten questions students were most likely to miss. The percentage of students that got a question right ranged from 9% to 95% as shown in Figure 2. The mean was 52%; the standard deviation was 25 percentage points.

Figure 2 Question Scores (Ascending)



Teachers tend to focus on which questions students missed. This indicates materials that must be covered to help students improve. But the goal of this exercise is to generate a shorter instrument that accurately reflects student abilities.

If the same students were ranked based on their performance on the hardest questions, there is no guarantee that ranking would have any correlation with their score on the full survey. Percentage right does not necessarily indicate a correlation with overall student scores.

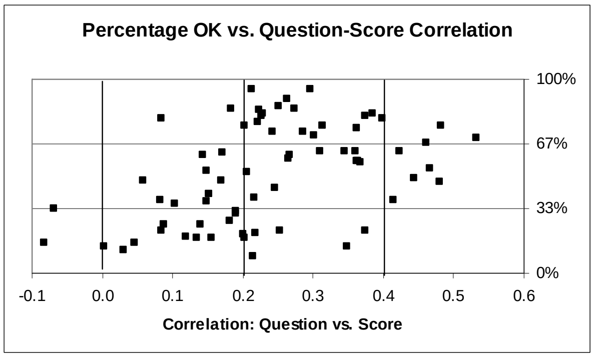
If above-average students got the easier ones right while the below-average students got the harder ones right, then picking the hardest questions would not identify the above-average students.

Again, rather than eliminate the hardest and easiest questions, a better approach is to use correlations between individual questions and the overall scores.

3.4. Student-Question Correlations

Appendix C presents the correlations between question score and overall score for each question. Figure 3 illustrates the results. Each question is a point. The vertical scale is the percentage of students that got that question correct. The horizontal scale is the correlation between a student's score on that question and a student's total score.

Figure 3 OK Percentage vs. Question Correlation



The 'lowest' correlations were negative. Although such correlations may seem inexplicable, a simple explanation is chance. With only 100 subjects, the Pearson correlation must be at least at least ± 0.2 before the difference from zero is statistically significant. See p. 181 in Wonnacott and Wonnacott (1981). Eliminating all questions that have a correlation of less than 0.2 eliminates 25 of the 69 questions.

Teachers would normally examine those questions that most students missed. If there was no ambiguity in the question and the teacher's answer was correct, then a low score would indicate that students didn't know this material.

To repeat, the goal of this exercise is not to identify student weaknesses. The goal is to create an instrument that accurately predicts a student's overall ability in this area with the minimum number of questions and with the highest possible accuracy and reliability.

So, correlation is important. The higher the correlation, the better a question is at predicting the total score.

The highest correlation was 0.53 which means taking that question (Q11) into account eliminated about a fourth of the initial variation. Obviously a single question is unable to do much in predicting a student's total score. But if the goal is to classify students into two groups (pass and fail), then a shorter assessment instrument may be quite adequate.

4. Total Correlation Models

The goal of this exercise is to identify those questions that had the highest predictive power – the highest correlations. These questions were used in decreasing order by unconditioned correlation to generate quizzes with varying numbers of questions. Appendix D presents the details. Table 1 illustrates the results.

Table 1 Compare Simple Correlation Models

	10Q	15Q	20Q	25Q	30Q	35Q	69Q
69Corr.	0.84	0.88	0.89	0.92	0.93	0.92	
Min	10%	13%	10%	16%	20%	23%	26%
Max	100%	100%	100%	96%	97%	94%	75%
Ave.	63%	61%	58%	62%	62%	64%	52%
StdDev	25%	22%	21%	19%	17%	17%	10%
CoefVar	0.39	0.36	0.36	0.30	0.28	0.26	0.20
MrgErr*	15.6	11.4	9.4	7.6	6.4	5.6	2.5
R-sq							

* 95% margin of error measured in percentage points.

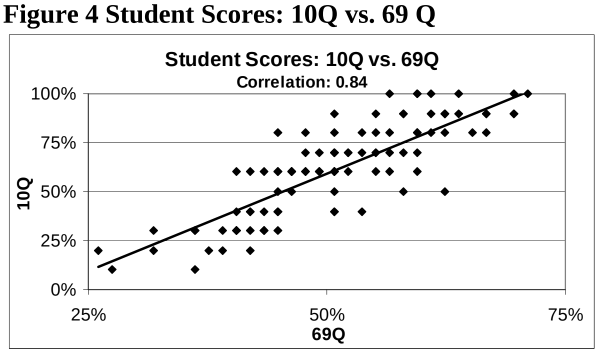
The label 69Corr indicates the correlation between two sets of scores: scores from the shorter instrument and scores from the 69 question survey. CoefVar is the standard deviation (StdDev) divided by the average. R-squared is the percentage of original variation eliminated by the model in question.

As expected, the standard deviation decreases and R-squared and the correlation between short quiz and long quiz scores generally increase as the number of questions increases.

At this level, the only indication of quality for a longer survey is the higher correlation (item discrimination) with the 69 question survey. But correlations (or R-squared) are not easily related to error rates.

An alternate approach is to use the percentage of students that are misclassified. To see that we need to compare student scores on the 69 question survey with those on the 10 question survey as shown in Figure 4.

Figure 4 Student Scores: 10Q vs. 69 Q



Notice how students getting 50% right on the 69 question are spread out from 40% to 80% on the 10 question survey. This provides an alternate approach to evaluating the quality of the shorter surveys.

4.1. Misclassification Error in Simple Models

If we assign students as above or below average in each survey, we can determine how many students are misclassified by using a shorter instrument.

Misclassification means a student who scores above (or below) average on the 69 question survey scores does not score above (or below) average on the survey in question. The misclassification rate is the percentage of all students that are misclassified by the model in question. Table 2 summarizes the results from Appendix D.

Table 2 Misclassification Rate in Correlation Models

Model	10Q	15Q	20Q	25Q	30Q	35Q	69Q
Rate	16%	20%	19%	17%	14%	14%	

One expects – and observes -- a decrease in this rate as the number of questions increases. The higher rate for 15 questions than for ten is an anomaly.

5. Collinearity

The foregoing analysis is quite straightforward and easy to understand. But what it omits is important. If all the questions measured different attributes or dimensions of statistical literacy, then each would be independently important. But if several of them measure the same thing, including more of them will not improve the quality of the prediction.

This is a classic problem in multiple regression called 'collinearity.' How can we select factors that are 'orthogonal' – that are "independent" from each other?

One way to select factors that are most orthogonal is do stepwise regression. Although there may be better ways of selecting these factors, stepwise regression has the advantage of being very explicit about what it does. In the additive model, the factor (question) having the highest correlation is added to the model first. Then the remaining factors (questions) are correlated with the residual variation (after including the first factor). Those that are highly correlated with that first factor will rank lower; those that are highly correlated with the outcome minus the first factor will rank higher. The remaining factor having the highest partial correlation is then added to the model, etc. At each step, R-squared – the total amount variation reduced by the more complex model – is calculated along with the p-value of the individual factors. At some point the process is halted based on a lack of statistical significance.

Appendix E presents the results of a step-wise regression by adding in the next most important factor – the factor having the greatest partial correlation after removing the influence of the factors previously included in the model.

6. Regression Models

In building a model, multiple regression uses partial correlations to take into account the influences of other factors already included. The adjective 'Regression' is used to distinguish these partial correlation models from the unconditioned 'correlation' models analyzed above. Appendix E presents various regression models. Cronbach's alpha was 0.71 on the 69 question survey.

Table 3 summarizes these results. Correlation is the correlation between the scores based on the shorter instrument and the scores on the 69 question survey. C-alpha is Cronbach's alpha – a measure of reliability. R-squared is the percentage of the original variation around the mean score that is eliminated by the model.

Table 3 Compare Regression Models by Size

	10Q	15Q	20Q	25Q	30Q	35Q	69Q
Corr.	0.86	0.90	0.92	0.92	0.95	0.95	
Min	0%	7%	15%	20%	20%	20%	26%
Max	100%	93%	90%	88%	87%	86%	75%
Average	65%	65%	58%	55%	54%	55%	52%
StdDev	19%	17%	15%	15%	13%	12%	10%
MrgErr*	12.1	8.7	6.8	5.8	4.7	4.2	2.5
C-Alpha	0.510	.549	.582	.585			0.71
R-sq	0.82	0.91	0.94	0.96	0.97	0.98	
Adj Rsq	0.81	0.89	0.93	0.95	0.96	0.97	

* 95% margin of error measured in percentage points.

As the size of the instrument increases, the average and standard deviation tend toward the 69 question values while the correlation with the 69 question score, R-squared and adjusted R-squared approach 100%.

Table 4 summarizes the misclassification rates from Appendix E for these regression-based models.

Table 4 Misclassification Rate in Regression Models

Model	10Q	15Q	20Q	25Q	30Q	35Q	69Q
Rate	18%	15%	13%%	15%	13%	10%	

As the number of questions increases, the misclassification rate tends to decrease although the decrease is much smaller after the first 15 questions.

7. Model Comparisons

Is a survey constructed from a multiple regression better than one constructed on just the raw correlations?

Compare the two approaches: individual correlation versus partial correlation regression. Both measure the association between question score and the total score for all students. Individual correlation measures the association without taking into account any previous questions included in a model. Partial correlation measures the association remaining after taking into account all previous questions included in a model.

The following tables compare results for two models: the simple raw-correlation model and the partial-correlation regression model. Models are compared based on the correlation of their scores with that in the 69 question survey (Table 5), based on their standard deviations (Table 6), their R-squared adjusted (Table 7) and their misclassification rates (Table 8).

Table 5 Comparing Models on Correlation

	10Q	15Q	20Q	25Q	30Q	35Q
Correlation	0.84	0.88	0.91	0.93	0.92	0.93
Regression	0.86	0.90	0.92	0.92	0.95	0.95

Table 6 Comparing Models on Standard Deviation

StdDev	10Q	15Q	20Q	25Q	Q30	Q35	69Q
Correlation	25%	22%	21%	19%	17%	17%	**10%**
Regression	19%	17%	15%	15%	13%	12%	**10%**

Table 7 Comparing Models on Adjusted R-squared

	10Q	15Q	20Q	25Q	Q30	Q35
Correlation						
Regression	0.81	0.89	0.93	0.95	0.96	0.97

Table 8 Comparing Models on Misclassifications

Questions	10Q	15Q	20Q	25Q	30Q	35Q
Correlation	**16%**	20%	19%	17%	14%	14%
Regression	18%	15%	13%	15%	13%	10%

Table 9 Comparing Models on Margin of Error*

Questions	10Q	15Q	20Q	25Q	30Q	35Q
Correlation	15.6	11.4	9.4	7.6	6.4	5.6
Regression	12.1	8.7	6.8	5.8	4.7	4.2

* 95% margin of error measured in percentage points.

When comparing models involving 10 questions and up, the regression model does the same job as the correlation model, but with

- two fewer questions for correlation.
- 15 fewer questions for standard deviation
- 15 fewer questions for misclassification
- Five fewer questions for margin of error

Appendix F shows compares the status of questions under these two models.

With one exception (the ten-question misclassification rate), the partial-correlation regression model is definitely better than the simple correlation model. At this point, let us reflect on whether these measures include all the major sources of error.

8. Content Validity

Content validity is the extent to which experts believe a measure represents a social concept such as depression, honesty or statistical literacy.

Content validity is similar to – but slightly different from – face validity – the extent to which non-experts think a measure represents a social concept. A question on a survey instrument may have high content validity and low face validity – or vice versa.

Until there is some measure of content validity among subject matter experts for a given instrument, any attempt to build an assessment instrument is premature.

9. Criterion Validity

Criterion validity is more objective than construct validity. Unfortunately there doesn't seem to be a good criterion on which to base QL.

Secondary forms of criterion validity are correlations with those factors that should contribute to greater quantitative literacy.

For example, Sundre (2008) reports on the QR test at James Madison University saying that the scores:

- correlate positively with grades in relevant courses.
- increase with greater relevant course exposure.
- discriminate between students who have completed their general education requirements and those who have not.

Demographic information was obtained for each student, but no analysis has been completed to see if scores correlate or discriminate as expected.

10. Institutional Assessment

Ultimately, each institution would be advised to create their own assessment instrument. Quantitative literacy is not a well defined construct. Having faculty involved in the process of creating an assessment instrument will help them come to grips with defining it.

In the mean time, adopting a well-tested assessment instrument, may give institutions a way to get started by getting data on their students. Obviously, generalizations to a class of students or to all students at a given college are impossible without reliable random samples as are comparisons of scores between institutions or between incoming students at different institutions.

11. Conclusion

A survey measuring statistical literacy skills has been developed and tested. The data will facilitate the design of a shorter instrument for broader use. However the skills assessed in this survey have not been reviewed for content validity by subject matter experts. Hopefully the approach used in this paper will be useful in future development of a survey instrument that is valid, reliable, meaningful and useful.

12. Future Work

The first step is to review other approaches to assessing QL such as:

- The National Assessment of Adult Literacy.[1]
- Dartmouth College Mathematics Across the Curriculum Survey. See Korey (2000).

[1] http://nces.ed.gov/Pubs2007/2007480.pdf

- the College Proficiency Exam (CPE) Part 2. See Crendall et al (2005).
- The Simpson Quantitative Literacy Competency exam.[2]
- The Q/R instrument at James Madison University. See Sundre (2008).

A second step is to create a list of Q/R skills and concepts. See Garfield (1995) and Utts (2003) for key statistical concepts. That list should be reviewed by skilled subject matter experts (SME). Once that list has been analyzed and surveys have been validated as measuring those skills, then surveys can be field tested.

Other activities are of a more detailed nature. This data could be analyzed more rigorously using item response theory (IRT). However Anderson (2002) argues that IRT does not provide much greater insight into 'true scores' than does this conventional analysis based on classic test theory. The decision as to which questions to retain might be better analyzed using other multivariate techniques than additive stepwise. The decision on how many questions are appropriate might also involve the tradeoff between student time (focus and motivation) and improved reliability (repeatability).

13. Acknowledgments

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." This talk was presented at a conference on assessing quantitative literacy held at Carleton College and sponsored by the Carleton QUIRK project and by project Kaleidoscope (PKAL). Thanks to Dr. Donna Sundre for the reference to Cobb (1998) and for the information on assessment on her web site.

14. References

Anderson, John (2002). Does Complex Analysis (IRT) pay any Dividends in Achievement Testing? Presented at *Measurement and Evaluation: Current and Future Directions for the New Millennium*, University of Alberta, October, 1998, Banff, Alberta. [3]

Cobb, George (1998). The Objective Format Question in Statistics: Dead horse, old bath water or overlooked baby? Invited paper, American Educational Research Association (AERA) Annual Meeting. [4]

Crandall, Gordon, James Giordano, Karlyn Koh and Janet Michello (2005). Across the Curriculum Strategies for Teaching Quantitative Literacy.[5]

Garfield, J. (1995), "How Students Learn Statistics," *International Statistical Review*, 63, 25-34.

Korey, J. (2000), "Dartmouth College of Mathematics Across the Curriculum Evaluation Summary: Mathematics and Humanities Courses," *Mathematics Across the Curriculum at Dartmouth College*[6]

Lutsky, Neil (2006). Quirks of Rhetoric: A Quantitative Analysis of Quantitative Reasoning in Student Writing *Proceedings of the ASA Section on Statistical Education*, pp. 2319-2322.[7]

Madison, Bernard L. (2006). Pedagogical Challenges of Quantitative Literacy. *Proceedings of the ASA Section on Statistical Education*, pp. 2323-2328.[8]

Raymond, Robert and Milo Schield (2008). *Numbers in the News: A Survey.* Proceedings of the ASA Section on Statistical Education.[9]

Schield, Milo (1998). *Statistical Literacy: Thinking Critically about Statistics.* Inaugural issue of the Journal "Of Significance" produced by the Association of Public Data Users (APDU). [10]

Schield, Milo (2004). *Statistical Literacy Curriculum Design.* IASE Curricular Development in Statistics Education Roundtable, pp. 54-74.[11]

Schield, Milo (2007). *Statistical Literacy: Teaching the Social Construction of Statistics.* Midwest Sociological Society.[12]

Schield, Milo and Cynthia Schield (2007). *Numbers in the News: A Survey.* Proceedings of the ASA Section on Statistical Education, pp. 2323-2328.[13]

Sundre, Donna (2008). Assessment Resources. [14]

Wonnacott, Thomas and Ronald J. Wonnacott (1981). *Regression: A Second Course in Statistics*. Krieger Publishing Co. p. 181 [95% confidence bands]

Utts, J. (2003), "What Educated Citizens Should Know About Statistics and Probability," *The American Statistician*, 57(2), 74-79.

Appendix A: Questions by Type

Of the 69 multiple choice questions, 40 are true false. Ten are two-answer true-false (24-25, 33-37, 40-41 and 61) and 30 are three-answer true/false/Don't know (9-23, 26-31, 54-55, 62-63, 65-68 and 75).

The remaining 29 involve a richer set of answers. Five involve a range of measures (32, 38-39, 64, and 69), 16 involve an A/B/Both/Neither (42-53 and 56-59), three involve an A/B/same (70-72), two involve A/B/C/same (73-74), and one involves an A/B/C/more than one (60) and two involve an A/B/Can't tell (76-77).

[2] www.simpson.edu/math/QLC.html
[3] www.educ.uvic.ca/epls/faculty/anderson/documents/Paper2.pdf
[4] www.jmu.edu/assessment/resources/resource_files/cobb_1998.pdf
[5] www.lagcc.cuny.edu/OpeningSessions/2005/workshops_I.htm
[6] www.math.dartmouth.edu/~matc/Evaluation/humeval.pdf
[7] www.StatLit.org/pdf/2006LutskyASA.pdf.
[8] www.StatLit.org/pdf/2006MadisonASA.pdf.
[9] www.StatLit.org/pdf/2008RaymondSchieldASA.pdf.
[10] www.StatLit.org/pdf/1998SchieldAPDU1.pdf
[11] www.StatLit.org/pdf/2004SchieldIASE.pdf
[12] www.StatLit.org/pdf/2007SchieldMSS.pdf
[13] www.StatLit.org/pdf/2007SchieldASA.pdf.
[14] www.jmu.edu/assessment/resources/Overview.htm

Appendix B: Questions by Percentage Right

Here are the questions (with the percentage of students that answered it correctly) – in question order.

Q9 (82%), Q10 (95%), Q11 (70%), Q12 (81%), Q13 (84%), Q14 (86%), Q15 (20%), Q16 (62%), Q17 (31%), Q18 (78%), Q19 (59%), Q20 (80%), Q21 (44%), Q22 (25%), Q23 (33%), Q24 (95%), Q25 (90%), Q26 (80%), Q27 (18%), Q28 (19%), Q29 (22%), Q30 (14%), Q31 (18%), Q32 (14%), Q33 (75%), Q34 (32%), Q35 (63%), Q36 (58%), Q37 (76%), Q38 (22%), Q39 (25%), Q40 (85%), Q41 (63%), Q42 (52%), Q43 (85%), Q44 (22%), Q45 (82%), Q46 (63%), Q47 (67%), Q48 (76%), Q49 (47%), Q50 (58%), Q51 (54%), Q52 (73%), Q53 (61%), Q54 (71%), Q55 (48%), Q56 (63%), Q57 (9%), Q58 (38%), Q59 (16%), Q60 (21%), Q61 (27%), Q62 (39%), Q63 (41%), Q64 (16%), Q65 (61%), Q66 (81%), Q67 (37%), Q68 (48%), Q69 (38%), Q70 (73%), Q71 (57%), Q72 (49%), Q73 (36%), Q74 (53%), Q75 (76%), Q76 (12%), Q77 (18%).

Here is the same information sorted by percent correct in descending order.

Q10 (95%), Q24 (95%), Q25 (90%), Q14 (86%), Q43 (85%), Q40 (85%), Q13 (84%), Q9 (82%), Q45 (82%), Q12 (81%), Q66 (81%), Q20 (80%), Q26 (80%), Q18 (78%), Q48 (76%), Q75 (76%), Q37 (76%), Q33 (75%), Q52 (73%), Q70 (73%), Q54 (71%), Q11 (70%), Q47 (67%), Q46 (63%), Q35 (63%), Q56 (63%), Q41 (63%), Q16 (62%), Q53 (61%), Q65 (61%), Q19 (59%), Q36 (58%), Q50 (58%), Q71 (57%), Q51 (54%), Q74 (53%), Q42 (52%), Q72 (49%), Q55 (48%), Q68 (48%), Q49 (47%), Q21 (44%), Q63 (41%), Q62 (39%), Q69 (38%), Q58 (38%), Q67 (37%), Q73 (36%), Q23 (33%), Q34 (32%), Q17 (31%), Q61 (27%), Q39 (25%), Q22 (25%), Q38 (22%), Q44 (22%), Q29 (22%), Q60 (21%), Q15 (20%), Q28 (19%), Q31 (18%), Q77 (18%), Q27 (18%), Q64 (16%), Q59 (16%), Q32 (14%), Q30 (14%), Q76 (12%), Q57 (9%).

First consider the questions most often gotten wrong.

Q57. [9%] Which describes the comparison in the data?

Adults	SEX		
RACE	Men	Women	ALL
Black	75%	25%	100%
White	51%	49%	100%
Other	40%	60%	100%
ALL	52%	48%	100%

a. Among adults, women are more likely among whites than among blacks. [9%] b. Among adults, blacks are more likely among men than among women. [25%] c. either of the above (both are OK). [53%] d. None of the above [8%] e. Don't know. [5%]
Answer: A. Sex (not race) is the common part.
Analysis: Missing this may reflect difficulty in reading a simple 100% table, in seeing what percentages can be compared or in using ordinary English to state an appropriate comparison of two percentages.

Q76 [12%] Which ending is necessarily implied by the introductory statement? More doctors recommend Crest a. than any other toothpaste. [78%] b. than nurses recommend Crest. [6%] c. Can't tell from the statement [12%] d. Don't know or not sure [3%]
Answer: C. Neither A or B are 'necessarily implied.'
Analysis: Missing this may reflect an inability to see ambiguity in context, to distinguish what is often intended by those using the phrase from what is actually asserted by the syntax.

Q30. [14%] Eight is three times more than two.
a. True [14%] b. False [82%] c. Don't know [3%]
Answer: A. Four times as much is three time more than.
Analysis: Unaware that this means "eight is three times (two) more than two, or 'eight is six more than two.

Q32. [14%] If incomes of rich and poor people both increase at the same rate, the income gap between rich and poor… a. will decrease [0] b. will stay the same [79%] c. will increase [19%] d. Don't know. [4%]
Answer: C. \$90K versus \$30K is a \$60K difference. Doubling both incomes doubles the difference.
Analysis: Missing this question may reflect an inability to distinguish percentage change (multiplication) from additive change (subtraction).

Q59. [16%] Which of the following is an accurate comparison using this data?

Adults	SEX		
RACE	Men	Women	ALL
Black	35%	70%	55%
White	50%	25%	35%
Other	15%	5%	10%
ALL	100%	100%	100%

a. Among adults, women are more likely among blacks than among whites. [22%] b. Among adults, blacks are more likely among women than among men. [16%] c. either of the above (both are OK). [39%] d. None of the above [8%] e. Don't know. [14%]
Answer: B. Race is the common part – not gender.
Analysis: Same inability as in Q57. May be redundant.

Q64. [16%] A medical test for HIV has 95% accuracy: 95% of those with HIV test positive. Suppose that most of the subjects are like you. They don't engage in risky sex practices and they don't do intravenous drugs. Most of them don't have HIV. If you test positive, what is the chance that you really have HIV?

a. Less than 95% [16%] b. 95% [58%] c. More than 95% [10%] d. Don't know or Not sure [16%]
Answer: A. Would be 95% if 50% of group had HIV
Analysis: Unaware of the prosecutor's fallacy.

Q77, [18%] Which ending is necessarily implied by the introductory statement? People who smoke have lower IQ … a. than other people who don't [68%] b. than these same people had before they started smoking [9%] c. Can't tell from the statement [18%] d. Don't know or not sure [4%]. No Answer [1%]
Answer: C. Note 'necessarily implied' in the question.
Analysis: See analysis for Q76.

Now consider the questions typically gotten right.

Q10 (95%), Q24 (95%), Q25 (90%), Q14 (86%), Q43 (85%), Q40 (85%), Q13 (84%), Q9 (82%), Q45 (82%), Q12 (81%), Q66 (81%), Q20 (80%), Q26 (80%),

Q10. [95%] 40% of smokers are Protestants. See pie chart on next page for Q11.

Q24. [95%] If Jay is older than Tom and Nan is older than Jay, than Nan must be older than Tom.

Q25. [90%] If Jay is older than Tom and Nan is younger than Jay, than Nan must be older than Tom.

Q14. [86%] Among smokers, the percentage of Protestants is 40%.

Q43. [85%] The budget was cut from $2 million to $1.5 million. What was the percent change? a. 75% decrease [1%] b. 25% decrease [85%] c. 75% increase [1%] d. 25% increase [9%] e. None of these [4%].

Q40. [85%] If "Most smokers are non-runners" then "Most non-runners must be smokers."
Analysis: The "must be" in the answer may have influenced students to say "No" since 'must' or "always" statements are generally false.

Q13. [84%] The percentage of smokers who are Protestants is 40%. See Pie chart

Q9. [82%] Among Protestants, 40% are smokers

Q45. [82%] Today is Nov. 3, 2008. A police report says the victim was born on April 21, 1948. How old was the victim? a. 48 [1%] b. 52 [0%] c. 58 [1%] d. 60 [8%2] e. 70 [4%] f. None of these [10%] Blank [2%]
Answer: 60. Analysis: 12% for none of these and blank. Did the presence of month and day make rounding seem inappropriate?

Q12. [81%] Among Protestants, the percentage of smokers is 40%. See single pie chart.

Q20. [80%] Protestants are more likely to be smokers than to be non-smokers. Compare two pie charts.

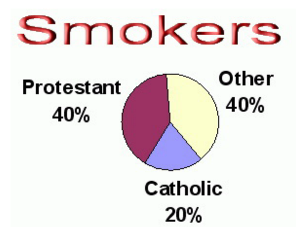
Q26. [80%] Eight is four times as much as two.

Appendix C: Questions by Individual Correlation

Here are the 69 questions in descending order by their individual correlation with student scores.
Q11 (.53), Q48 (.48), Q49 (.48), Q51 (.47), Q47 (.46), Q72 (.44), Q46 (.42), Q69 (.41), Q20 (.40), Q9 (.38), Q12 (.37), Q38 (.37), Q71 (.37), Q50 (.36), Q33 (.36), Q36 (.36), Q35 (.36), Q32 (.35), Q56 (.35), Q75 (.31), Q41 (.31), Q54 (.30), Q10 (.30), Q52 (.29), Q43 (.27), Q53 (.27), Q19 (.26), Q25 (.26), Q44 (.25), Q14 (.25), Q21 (.25), Q70 (.24), Q45 (.23), Q66 (.23), Q13 (.22), Q18 (.22), Q60 (.22), Q62 (.22), Q57 (.21), Q24 (.21), Q42 (.21), Q31 (.20), Q37 (.20), Q15 (.20), Q17 (.19), Q34 (.19), Q40 (.18), Q61 (.18), Q16 (.17), Q55 (.17), Q77 (.16), Q63 (.15), Q74 (.15), Q67 (.15), Q65 (.14), Q39 (.14), Q27 (.13), Q28 (.12), Q73 (.10), Q22 (.09), Q26 (.08), Q29 (.08), Q58 (.08), Q68 (.06), Q64 (.05), Q76 (.03), Q30 (.00), Q23 (-.07), Q59 (-.08).

As noted earlier, randomness is one explanation for the negative correlations. Consider the questions having the highest individual correlations with student scores.

Q11. [R=0.53, 70% OK] The percentage of Protestants who are smokers is 40%. a. Yes [22%] b. No [70%] c. Don't know or Not sure [8%]



Answer: B. Religion (Protestant) is part – not whole.
Analysis: Missing this may reflect an inability to read a pie chart or to use ordinary English to describe a simple ratio.

Q48. [R=0.48, 78% OK] An oil tanker hit a rock offshore from your city and spilled about 180,000 gallons of crude oil. How many swimming pools would the spilled oil fill if a swimming pool holds 12,000 gallons? a. 0.07 [0%] b. 1.5 [2%] c. 7 [8%] d. 15 [76%] e. 70 [0%] f. 168,000 [2%] g. None of these [9%]. Didn't answer (2%).
Answer: D: 180,000 / 12,000 = 15.
Analysis: Missing this may reflect the lack of a calculator which in turn makes the question a test of ability to work division mentally. Calculators were not excluded, but some students commented they would have done better with a calculator.

Q49. [R=0.48, 47% OK] Your city has a property tax rate of $10 per $1,000 assessed value (but $50,000 of the home value can't be taxed because of the state's homestead exemption.) What would be the tax bill for a typical homeowner if the median assessed value of homes is $250,000? a. $200 [9%] b. $500 [0%] c. $2,000 [47%] d. $2,500 [22%] e. $20,000 [16%] f.. $250,000 [0%] g. None of these [5%]. no answer [1%]. Answer C. Analysis: One more step than Q48.

Q51. [0.47; 54% OK] Last year's budget was $4 million. This year it's $4.4 million. What was the percentage change?
a. $400,000 increase [10%] b. 4% increase [21%]
c. 10% increase [54%] d. 40% increase [8%]
e. None of these [6%]. Didn't answer [1%]
Answer: C. % change= (4.4-4.0)/4.0 = 0.1 = 10%.
Analysis: The 10% answering A didn't distinguish percentage change from a simple difference. Those answering B, D or E are harder to explain.

Q47. [R=0.46, 67% OK] Your city of 150,000 had a total of 75 murders last year. What was the murder rate per 100,000? a. 30 [7%] b. 50 [67%] c. 75 [6%] d. 150 [4%] e. 500 [0%] f. None of these [14%].
Answer: B. 75/150K = X/100K; X = 75*(100K/150K)
Analysis: In ability to handle a classic three-term problem. May reflect not having/using a calculator.

Q72. [R=0.44, 49% OK] Which definition of "heat-wave deaths" gives the largest number?
a. deaths caused by a heat wave [28%]
b. deaths occurring during a heat-wave [49%]
c. No difference [17%] d. Don't know/not sure [6%]
Answer: B. Note: Most students missed this question.
Analysis: Missing this question may reflect ignorance about the difference between association and causation or how this difference is expressed in ordinary English.

Q46. [R=0.42, 63% OK] One out of 12 residents of your city speaks Spanish. About what percent of the population speaks Spanish? a. 1% [9%] b. 5% [5%] c. 8% [63%] d. 12% [10%] e. 15% [2%] f. 80% [0%] g. None of these [9%]
Answer: C. 1/12 = .0825 = 8.25% which is about 8%.
Analysis: Having "g. None of these" seems inappropriate when the question asks for "about". Answering "d" may reflect an ignorance of reciprocals.

Q69. [R=0.41, 38% OK] Suppose Jim finishes 20% of his job and Jan finishes 60% of her job. Suppose that Jim's part is 80% of the project and Jan's part is 20%. How much of the total project is done?
a. 20% [3%] b. 28% [38%] c. 32% [13%]
d. 40% [14%] e. 52% [8%] f. None of these [4%]
g. Don't know; Not sure [13%]
Answer: B (0.2*0.6)+(0.8*0.2) = 0.12+0.16 = 0.28
Analysis: This is a hard problem; most students (87%) got it wrong. Note that 13% of the students said they didn't know or weren't sure.

Q20. [R=0.41, 13% OK] If a stock earnings drops by $600,000 from +$300,000 to -$300,000, that is a 200% decrease.
a. True [13%] b. False [80%] c. Don't know [7%]
Answer: A. [-300 – 300)/300 = -600/300 = -2 = -200%.
Analysis: Missing this may reflect ignorance on percent change – especially in unusual circumstances.

Appendix D: Correlation Models

One way to select questions for a shorter instrument is to pick those that individually have the highest correlation with the outcome of interest: student scores. Here are the questions with their correlation in parenthesis.

Q11 (.53), Q48 (.48), Q49 (.48), Q51 (.47), Q47 (.46), Q72 (.44), Q46 (.42), Q69 (.41), Q20 (.40), Q9 (.38), Q12 (.37), Q38 (.37), Q71 (.37), Q50 (.36), Q33 (.36), Q36 (.36), Q35 (.36), Q32 (.35), Q56 (.35), Q75 (.31), Q41 (.31), Q54 (.30), Q10 (.30), Q52 (.29), Q43 (.27), Q53 (.27), Q19 (.26), Q25 (.26), Q44 (.25), Q14 (.25), Q21 (.25), Q70 (.24), Q45 (.23), Q66 (.23), Q13 (.22),

A ten question instrument would use the first ten questions: Q11, Q48, Q49, Q51, Q47, Q72, Q46, Q69, Q20 and Q9. In order by number, these ten are Q9, Q11, Q20, Q46, Q47, Q48, Q49, Q51, Q69 and Q72.

A 15 question instrument would use these ten questions and Q12, Q38, Q71, Q50 and Q33. In order by number, there additional five are Q12, Q33, Q38, Q50 and Q71.

A 20 question instrument would use these 15 plus these questions: Q36, Q35, Q32, Q56 and Q75. In order by number, these additional five are Q32, Q35, Q36, Q50 and Q71.

A 25 question instrument would use these 20 plus these five questions: Q41, Q54, Q10, Q52 and Q43. In order by number, these additional five are Q10, Q41, Q43, Q52 and Q54.

A 30 question instrument would use these 25 plus these fie questions: Q53, Q19, Q25, Q44 and Q14. In order by number these additional five are Q14, Q19, Q25, Q44 and Q53.

A 35 question instrument would use these 30 plus these five questions: Q21, Q70, Q45, Q66 and Q13. In order by number these additional five are Q13, Q21, Q45, Q66 and Q70.

The following tables present the number of students who scored above average in each survey and on both surveys. From this one can deduce the number of students who are misclassified.

Table 10 compares the above/below average on the 10 question survey with that on the 69 question survey:

Table 10 Misclassifications using a 10 question survey

Correlation	69 Question		
10 Q	Below Ave.	Above Ave	ALL
Below Ave.	45	7	52
Above Ave.	9	**39**	**48**
ALL	54	**46**	**100**

Of the 100 students, 16 (16%) would be misclassified using a 10 question survey.

Table 11 compares the above/below average on the 20 question survey with that on the 69 question survey:

Table 11 Misclassifications using 20 question survey

Correlation	**69 Question**		
20 Q	Below Ave.	Above Ave	ALL
Below Ave,	42	7	49
Above Ave	12	**39**	**51**
ALL	54	**46**	**100**

Of the 100 students, 19 (19%) would be misclassified using a 20 question survey.

Table 12 compares the above/below average on the 30 question survey with that on the 69 question survey:

Table 12 Misclassifications using 30 question survey

Correlation	**69 Question**		
30 Q	Below Ave	Above Ave	ALL
Below Ave	43	3	46
Above Ave	11	**43**	**54**
ALL	54	**46**	**100**

Of the 100 students, 14 (14%) would be misclassified using a 30 question survey.

Table 13 compares the above/below average on the 40 question survey with that on the 69 question survey:

Table 13 Misclassifications using 35 question survey

Correlation	**69 Question**		
35 Q	Below Ave	Above Ave	ALL
Below Ave	44	4	48
Above Ave	10	**42**	**52**
ALL	54	**46**	**100**

Of the 100 students, 14 (14%) would be misclassified using a 40 question survey.

Appendix E: Regression Models

A much better way to select questions for a shorter instrument is to use those that have the highest partial correlation with student scores after taking into account the influence of all previously included questions. These results were obtained using SPSS 16.

Here are the variables entered into a model using a stepwise additive approach:

Q11, Q51, Q20, Q48, Q35, Q13, Q32, Q70, Q41, Q47, Q67, Q43, Q15, Q54, Q72, Q21, Q22, Q27, Q69, Q53, Q17, Q16, Q61, Q59, Q52, Q36, Q18, Q37, Q23, Q50, Q29, Q25, Q63, Q68, Q19, Q74, Q38, Q49, Q57 and Q58. This list of 40 questions leaves out 29 questions as lacking in statistical significance.

A 10 question instrument would use questions Q11, Q51, Q20, Q48, Q35, Q13, Q32, Q70, Q41 and Q47. In order by number, there are Q11, Q13, Q20, Q32, Q35, Q41, Q47, Q48, Q51, and Q70.

A 15 question instrument would use the preceding 10 questions and these five: Q67, Q43, Q15, Q54 and Q72. In order, these five are Q15, Q43, Q54, Q67 and Q72.

A 20 question instrument would use the preceding 15 questions and these five: Q21, Q22, Q27, Q69 and Q53. In order, these five are Q21, Q22, Q27, Q53 and Q69.

A 25 question instrument would use the preceding 20 questions and these five: Q17, Q16, Q61, Q59 and Q52. In order, these five are Q16, Q17, Q52, Q59 and Q61.

A 30 question instrument would use the preceding 25, questions, eliminate one, Q22, and add these six: Q36, Q18, Q37, Q23, Q50 and Q29. In order, these six are Q18, Q23, Q29, Q36, Q37 and Q50.

A 35 question instrument would use the preceding 30 questions and these five: Q25, Q63, Q68, Q19 and Q74. In order, these five are Q19, Q25, Q63, Q68 and Q74.

Table 14 compares the above/below average on the 10 question survey with that on the 69 question survey:

Table 14 Misclassifications using 10 question survey

Regression	69 Question		
10 Q	Below Ave.	Above Ave	ALL
Below Ave.	42	6	48
Above Ave.	12	**40**	**52**
ALL	54	**46**	**100**

Of the 100 students, 18 (18%) would be misclassified using a 10 question survey.

Table 15 compares the above/below average on the 20 question survey with that on the 69 question survey:

Table 15 Misclassifications using 20 question survey

Regression	69 Question		
20 Q	Below Ave.	Above Ave	ALL
Below Ave,	43	2	45
Above Ave	11	**44**	**55**
ALL	54	**46**	**100**

Of the 100 students, 13 (13%) would be misclassified using a 20 question survey.

Table 16 compares the above/below average on the 30 question survey with that on the 69 question survey:

Table 16 Misclassifications using 30 question survey

Regression	69 Question		
30 Q	Below Ave	Above Ave	ALL
Below Ave	49	8	57
Above Ave	5	**38**	**43**
ALL	54	**46**	**100**

Of the 100 students, 13 (13%) would be misclassified using a 30 question survey.

Table 17 compares the above/below average on the 35 question survey with that on the 69 question survey:

Table 17 Misclassifications using 35 question survey

Regression	69 Question		
35 Q	Below Ave	Above Ave	ALL
Below Ave	49	5	54
Above Ave	5	**41**	**46**
ALL	54	**46**	**100**

Of the 100 students, 10 (10%) would be misclassified using a 40 question survey.

Appendix F: Comparing Questions by Approach

The two different approaches – simple correlation and partial correlation regression – rank the questions differently as shown in the following list. In each triplet, the first number is rank in the partial correlation model, the second item is the question, and the third number (in parentheses) is the rank in the simple correlation model.

1 Q11(1), 2 Q51(4), 3 Q20(9), 4 Q48(2), 5 Q35(17), 6 Q13(35), 7 Q32(18), 8 Q70(32), 9 Q41(21), 10 Q47(5), 11 Q67(54), 12 Q43(25), 13 Q15(44), 14 Q54(22), 15 Q72(6), 16 Q21(31), 17 Q22(60), 18 Q27(57), 19 Q69(8), 20 Q53(26), 21 Q17(45), 22 Q16(49), 23 Q61(48), 24 Q59(69), 25 Q52(24), 26 Q36(16), 27 Q18(36), 28 Q37(43), 29 Q23(68), 30 Q50(14), 31 Q29(62), 32 Q25(28), 33 Q63(52), 34 Q68(64), 35 Q19(27), 36 Q74(53), 37 Q38(12), 38 Q49(3), 39 Q57(39), 40 Q58(63)

As expected, the first question is the same in both approaches. As expected, there are some important differences between the two lists. Question 29 jumps 29 ranks from 35th on the simple correlation list to 6th on the partial correlation list. Question 15 jumps 31 ranks, 27 jumps 39 ranks, 22 and 67 jump 43 ranks, and 59 jumps 45 ranks.

Some of the questions that jumped ahead were statistically insignificant when taken separately. But when tested based on their partial correlation they were statistically significant. These questions and their individual correlations are: Q67 (.15), Q22 (.09), Q27 (.13), Q17 (.19), Q16 (.17), Q61 (.18), Q59 (-.08), Q23 (-.07), Q29 (.08), Q63 (.15) and Q58 (.08).

Some of the questions that were statistically significant when correlated individually became statistically insignificant based on their partial correlations. These questions and their individual correlations are: Q46 (.42), Q09 (.38), Q12 (.37), Q71 (.37), Q33 (.36), Q56 (.35), Q75 (.31), Q10 (.30), Q44 (.25), Q14 (.25), Q45 (.23), Q66 (.23), Q60 (.22), Q62 (.22), Q24 (.21) and Q42 (.21).

As expected, some of the questions that were not statistically significant based on their individual correlations remained statistically insignificant based on their partial correlations. These questions and their individual correlations are: Q31 (.20), Q34 (.19), Q40 (.18), Q55 (.17), Q77 (.16), Q65 (.14), Q39 (.14), Q28 (.12), Q73 (.10), Q26 (.08), Q64 (.05), Q76 (.03), and Q30 (.00)

Appendix G

2008 Statistical Literacy Skills Survey

Sponsored by the

W. M. Keck Statistical Literacy Project

Invitation to Participate

This survey has been commissioned by Augsburg College W. M. Keck Statistical Literacy Project. It focuses on the general use of numbers and informal statistics in everyday situations. This survey seeks to identify *the general level of understanding* in reading and interpreting rates, percentages and basic conditional probabilities.

You have received this invitation because you are a member of the relevant population for this study and because we believe you are motivated to help improve quantitative literacy. Your opinions and beliefs on the topic are extremely important to us. ***The information you supply in this survey will be presented to educators for their use in shaping various curricula in quantitative literacy in schools and colleges throughout the world.***

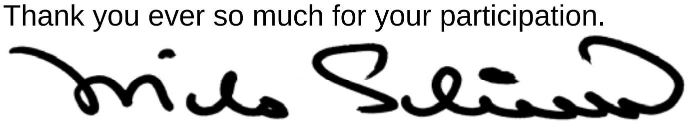
We are asking you to complete this survey. This survey (7 pages) will take 40 to 50 minutes. We apologize for the length, but this detail is needed to help us accurately identify the levels of statistical literacy. Your participation is completely voluntary. There are no direct monetary payments to induce you to take this survey but you will help Augsburg maintain its position as the leader in Statistical Literacy.

If you don't know the answer to a question after a reasonable time, just say "Don't know" and move on. Do not talk with others who are working on their survey. You may withdraw at any time. If you decide to withdraw simply turn in your survey.

You need not identify yourself on the questionnaire. The data obtained in this survey will be publicly available and may be used by researchers, educators and authors in various ways. . Any personal identifications will be omitted from the publicly available data. Your submission of your survey will signify your informed consent to participate.

Please enter your comments on the last page of your survey. If you have any questions or concerns, please forward them to us via the email address shown below.

Thank you ever so much for your participation.



Milo Schield, Ph.D. Professor of Business Administration, Augsburg College
Director of the W. M. Keck Statistical Literacy Project
schield@augsburg.edu
Augsburg College IRB #2002-19-3

Demographics

Instructions: Circle the answer you believe is best or correct.

1. **What best describes <u>how comfortable you are in dealing with formal statistics and algebra (e.g., chance, probability, sampling distributions, confidence intervals, etc.)</u>?** *(Select only one)*

a. extremely comfortable   b. somewhat comfortable   c. somewhat uncomfortable   d. extremely uncomfortable.

2. **What best describes <u>how comfortable you are in dealing with informal statistics and numbers (e.g., reading and interpreting tables and graphs containing rates and percentages)</u>?** (*Select only one*)

a. extremely comfortable   b. somewhat comfortable   c. somewhat uncomfortable   d. extremely uncomfortable.

3. **What best describes <u>your occupation?</u>** *(Select only one.). [If retired, use your prior occupation.)*

   a. high-school student
   b. college student
   c. teacher, elementary (K - 8)
   d. teacher, secondary (9 -12)
   e. teacher, college
   f. other profession (E.g., Administration, law, social work, accounting)
   g. other (E.g., visual or performing arts, trade, service industry)

4. **What best describes <u>your highest level of schooling completed?</u>** *(Select only one.)*

   a. Not applicable
   b. Secondary school / high school
   c. Two-year college (Associates degree)
   d. Four year college (Bachelor's degree)
   e. Graduate degree (Master's or Ph.D.) or post-doctorate

5. **What best describes <u>your fluency in English?</u>** *(Select only one.)*

   a. English was a native language by primary school
   b. Became fluent in speaking and reading English after primary school
   c. Not yet fluent in speaking and reading English

6. **Which best describe the highest-level mathematics <u>you are taking or have taken</u>**?

   a. Elementary/Middle-school mathematics
   b. High school algebra I
   c. High-school algebra II
   d. College algebra, Math for Liberal Arts, discrete/finite mathematics or pre-calculus.
   e. College level mathematics or statistics beyond those in the preceding answer.

7. **How many college-level (non-remedial) mathematics courses have you completed?** *(Select only one.)*

   a. None   b. 1   c. 2   d. 3   e. 4   f. more than 4

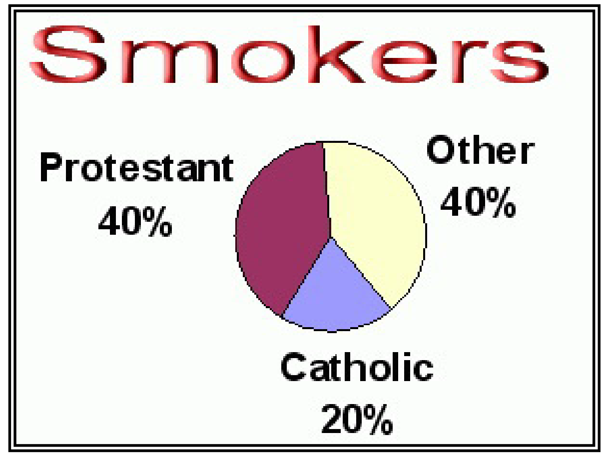
8. **What best describes <u>your major (or likely major) in college?</u>** *(Select only one.)*

   a. Not applicable
   b. Education: Primary, Special Ed, or Secondary with a non-quantitative emphasis (English, History).
   c. Education: Secondary with a quantitative emphasis (Math or science).
   d. Mathematics or a science major.
   e. Other Quantitative major (E.g., business, psychology, sociology, social work, economics)
   f. Other non-quantitative major (E.g., Visual and performing arts, Communications, Journalism, Physical or Sports Education, English, History, Political Science, Religion or Philosophy)

A Graphs and Charts A

Circle only one of the answers for each problem.

Do these statements accurately describe the data in this pie chart?



9. **Among Protestants, 40% are smokers**
   a. Yes b. No c. Don't know or Not sure

10. **40% of smokers are Protestants.**
    a. Yes b. No c. Don't know or Not sure

11. **The percentage of Protestants who are smokers is 40%.**
    a. Yes b. No c. Don't know or Not sure

12. **Among Protestants, the percentage of smokers is 40%.**
    a. Yes b. No c. Don't know or Not sure

13. **The percentage of smokers who are Protestants is 40%.**
    a. Yes b. No c. Don't know or Not sure

14. **Among smokers, the percentage of Protestants is 40%.**
    a. Yes b. No c. Don't know or Not sure

Do these statements accurately describe the comparisons within this pie chart?

15. **Smokers are more likely among Protestants than among Catholics.**
    a. Yes b. No c. Don't know or Not sure

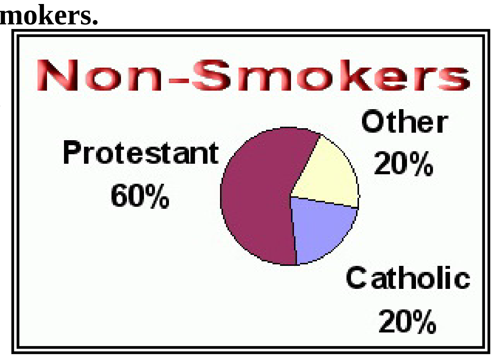
16. **Protestants are more likely among smokers than are Catholics.**
    a. Yes b. No c. Don't know or Not sure

17. **Protestants are more likely to be smokers than are Catholics.**
    a. Yes b. No c. Don't know or Not sure

18. **Smokers are more likely to be Protestants than to be Catholics.**
    a. Yes b. No c. Don't know or Not sure

Do these statements accurately compare the data in this chart with that in the previous chart

19. **Protestants are more prevalent among non-smokers than among smokers.**
    a. Yes b. No c. Don't know or Not sure



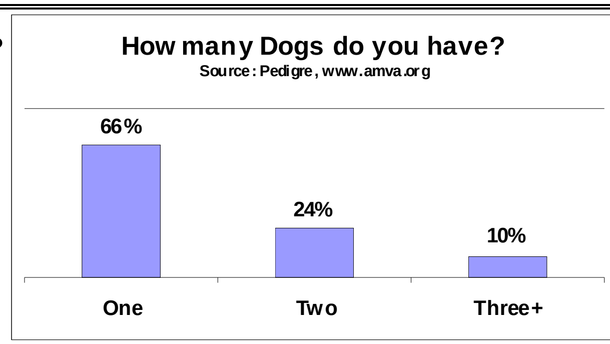
20. **Protestants are more likely to be smokers than to be non-smokers.**
    a. Yes b. No c. Don't know or Not sure

21. **Non-smokers are more likely to be Protestants than are smokers.**
    a. Yes b. No c. Don't know or Not sure

22. **Non-smokers are more prevalent among Protestants than among Catholics.**
    a. Yes b. No c. Don't know or Not sure

Does this statement accurately describe the data in this graph?

23. **34% of adults own two or more dogs.**
    a. Yes b. No c. Don't know or Not sure



Comparisons: Qualitative and Quantitative

B

QUALITATIVE COMPARISONS:

24. **T F** If Jay is older than Tom and Nan is older than Jay, than Nan must be older than Tom.

25. **T F** If Jay is older than Tom and Nan is younger than Jay, than Nan must be older than Tom.

QUANTITATIVE COMPARISONS OF NUMBERS

26. a. True b. False c. Don't know **Eight is four times as much as two.**

27. a. True b. False c. Don't know **Eight is 300% more than two.**

28. a. True b. False c. Don't know **Eight is four times more than two.**

29. a. True b. False c. Don't know **Two is four times less than eight.**

30. a. True b. False c. Don't know **Eight is three times more than two.**

31. a. True b. False c. Don't know **8% is 2% more than 6%.**

QUANTITATIVE COMPARISONS OF VARIABLES

32. **If incomes of rich and poor people both increase at the same rate, the income gap between rich and poor**
 a. will decrease b. will stay the same c. will increase d. Don't know

33. T F **If Jay is half as old as Tom and if Tom is twice as old as Nan, then Jay and Nan are the same age..**

34. T F **If Jay is 50% younger than Tom, Tom is 50% older than Nan, then Jay and Nan are same age**..

35. T F **If a stock decreases 50% and then increases by 50%, it will be back to its original value.**

36. T F **If a stock increases 50% and then decreases by 50%, it will be back to its original value.**

37. T F **If a stock drops from $300 to zero, that is a 300% decrease.**

---

38. **A company has a 30% market share in the Western US and a 10% market share in the Eastern US. What is the company's overall market share in the entire US?**
 a. Between 10% and 30%. B. 30% c. Between 30% and 40% d. 40% e. Over 40%

39. **If the incomes of rich and poor people both double over 30 years, the income gap between rich and poor**
 a. will stay the same b. will double c. will increase but need not double d. Don't know

40. T F **If "Most smokers are non-runners" then "Most non-runners must be smokers."**

41. T F **If a stock earnings drops by $600,000 from +$300,000 to -$300,000, that is a 200% decrease.**

42. Last year's budget was $8 million. This year the budget will be cut by 5 percent. What is this year's budget?
a. $800,000 decrease b. $400,000 decrease c. $7.2 million d. $7.6 million e. None of these

43. The budget was cut from $2 million to $1.5 million. What was the percent change?
a. 75% decrease b. 25% decrease c. 75% increase d. 25% increase e. None of these

44. In a loose crowd, each person takes up about 9 square feet. What is the best estimate of how many people in a loose crowd will fill a plaza that measures 100 yards by 40 yards?
a. 400 b. 500 c. 3,600 d. 4,000 e. 40,000 f. None of these

45. Today is Nov. 3, 2008. A police report says the victim was born on April 21, 1948. How old was the victim?
a. 48 b. 52 c. 58 d. 60 e. 70 f. None of these

46. One out of 12 residents of your city speaks Spanish. About what percent of the population speaks Spanish?
a. 1% b. 5% c. 8% d. 12% e. 15% f. 80% g. None of these

47. Your city of 150,000 had a total of 75 murders last year. What was the murder rate per 100,000
a. 30 b. 50 c. 75 d. 150 e. 500 f. None of these

48. An oil tanker hit a rock offshore from your city and spilled about 180,000 gallons of crude oil. How many swimming pools would the spilled oil fill if a swimming pool holds 12,000 gallons.
a. 0.07 b. 1.5 c. 7 d. 15 e. 70 f. 168,000 g. None of these

49. Your city has a property tax rate of $10 per $1,000 assessed value (but $50,000 of the home value can't be taxed because of the state's homestead exemption.) What would be the tax bill for a typical homeowner if the median assessed value of homes is $250,000
a. $200 b. $500 c. $2,000 d. $2,500 e. $20,000 f.. $250,000 g. None of these

50. Last year your small town had three murders. This year there were none. How would you describe the change?
a. 300% decrease b. 100% decrease c. 30% decrease d. 3% increase e. 1% decrease f. None of these

51. Last year's budget was $4 million. This year it's $4.4 million. What was the percentage change?
a. $400,000 increase b. 4% increase c. 10% increase d. 40% increase e. None of these

52. The city managers in five towns are paid $75,000, $80,000, $65,000, $120,000 and $60,000 a year. What is their average salary?
a. $65,000 b. $75,000 c. $80,000 d. $85,000 e. None of these

53. In the previous question, what is their median salary?
a. $65,000 b. $75,000 c. $80,000 d. $85,000 e. None of these

54. A study shows that doing yoga cuts the frequency of migraines. Could this study have been done as a classic before-after experiment where the number of migraines per month before doing yoga is compared with the number of migraines per month after doing yoga for the same group of people?
a. Yes b. No c. Don’t know or can’t tell

55. A study shows that eating nuts cuts the risk of developing cancer. Could this study have been done as a classic before-after experiment where the cancer risk before eating nuts is compared with the cancer risk after eating nuts for the same group of people?
a. Yes b. No c. Don’t know or can’t tell

Reading Tables

Adults	SEX		
RACE	Men	Women	ALL
Black	75%	25%	100%
White	51%	49%	100%
Other	40%	60%	100%
ALL	52%	48%	100%

56. Which of the following describes the circled 25%?
    a. 25% of adult women are blacks.
    b. 25% of adult blacks are women.
    c. either of the above (both are OK).
    d. None of the above
    e. Don't know.
57. Which of the following describes the comparison using this data?
    a. Among adults, women are more likely among whites than among blacks.
    b. Among adults, blacks are more likely among men than among women.
    c. either of the above (both are OK).
    d. None of the above
    e. Don't know.

Adults	SEX		
RACE	Men	Women	ALL
Black	35%	70%	55%
White	50%	25%	35%
Other	15%	5%	10%
ALL	100%	100%	100%

58. Which of the following describes the circled 70% in this table?
    a. 70% of adult women are blacks.
    b. 70% of adult blacks are women.
    c. either of the above (both are OK).
    d. None of the above
    e. Don't know.
59. Which of the following is an accurate comparison using this data?
    a. Among adults, women are more likely among blacks than among whites.
    b. Among adults, blacks are more likely among women than among men.
    c. either of the above (both are OK).
    d. None of the above
    e. Don't know.

US Unemployment Rates by Educational Attainment & Race (Percent unemployed of working adults)					
		--- Highest Grade Achieved ---			
RACE	**Total**	**< 12**	**12**	**13-15**	**> 15**
Total:	4.4	10.4	5.1	3.8	2.0
White:	3.9	9.4	4.6	3.4	1.8
Black:	8.1	16.6	**8.2**	6.1	4.4
Other	7.3	9.6	7.5	5.5	3.0

60. **Describe the circled 8.2% shown in the Black row, 12th grade column.**

    a. 8.2% of unemployed working adults whose highest grade level achieved was 12th grade are black
    b. 8.2% of unemployed working adults who are black had 12th grade as their highest grade level achieved.
    c. 8.2% of black working adults whose highest grade achieved was 12th grade are unemployed.
    d. More than one of these
    e. None of these
    f. Don't know or can't tell

61. **T F In any group, the "percentage of married adult <u>smokers</u> who are female" is always greater than "the percentage of married adults who are female <u>smokers</u>."**

62. **In German districts in the 1930s, the higher the percentage of Catholics, the higher the percentage of votes for Hitler. Does this show that Catholics were more likely to vote for Hitler than non-Catholics?**
a. Yes b. No c. Don't know or Not sure

63. **Among suicides, suppose that widows are more likely than widowers. Does this mean that widows are more likely to commit suicide than widowers?**
a. Yes b. No c. Don't know or Not sure

---

64. **A medical test for HIV has a 95% accuracy: 95% of those with HIV test positive. Suppose that most of the subjects are like you. They don't engage in risky sex practices and they don't do intravenous drugs. Most of them don't have HIV. If you test positive, what is the chance that you really have HIV?**
a. Less than 95% b. 95% c. More than 95% d. Don't know or Not sure

65. **A researcher is determining whether patients feel better after taking a drug. If the researcher knows which patients received the drug and which did not, could this bias the results of the study?**
a. Yes b. No c. Don't know or Not sure

66. **Patients are assigned to either a treatment group or a control group. The treatment group gets the active treatment; the control group gets a passive placebo (fake drug or sugar water). If the patient knows which group they are in, could this bias the results of the study?**
a. Yes b. No c. Don't know or Not sure

67. **A poll of 400 random sampled voters shows the two candidates at 46% and 54%. Is this difference statistically significant?**
a. Yes b. No c. Don't know or Not sure

68. **In 2000, a *research hospital* had a <u>higher overall</u> death rate than a *rural hospital*. Each patient's condition was classified as "fair" or "poor."**
The research hospital said the same data showed that they had a <u>lower</u> death rate than the *rural hospital* for patients in "fair" condition – AND – for patients in "poor" condition?
Q. Is their claim logically <u>possible</u>?
a. Yes/Possible b. No/Impossible c. Don't know or Not sure

69. **Suppose Jim finishes 20% of his job and Jan finishes 60% of her job. Suppose that Jim's part is 80% of the project and Jan's part is 20%. How much of the total project is done?**
a. 20% b. 28% c. 32% d. 40% e. 52% f. None of these g. Don't know; Not sure

==

Table 1: US Women (in millions) who had a child in 2004 by family income

< 10K	10K-19.9K	20K-24.9K	25K-29.9K	30K-34.9K	35K-49.9K	50-74.9K	75K and up	TOTAL
4.2	6.2	3.4	3.8	3.6	8.9	10.6	12.5	53.2

Source 2006 US Statistical Abstract. Table 88.

Based on family income, who had the most babies: rich moms or poor moms?

70. **Suppose we define 'Rich' as $35K and up and define 'Poor' as under $35K.**
a. Rich moms b. Poor Moms c. The same d. Don't know or Not sure

71. **Suppose we define 'Rich' as $75K and up and define 'Poor' as under $25K.**
a. Rich moms b. Poor Moms c. The same d. Don't know or Not sure

Constructing Groups and Measures

<u>Circle only one of the answers for each problem.</u>

How does the definition influence the size of the group?

72. Which definition of "heat-wave deaths" gives the largest number:
    a. deaths caused by a heat wave
    b. deaths occurring during a heat-wave.
    c. No difference
    d. Don't know or not sure

73. Which definition of workers gives the smallest number:
    a. Full time (40 hours/week) employees
    b. Permanent (year round) employees
    c. Full-time (40 hours/week), permanent (year-round) employees
    d. No difference
    e. Don't know or not sure

Which group would be expected to have the higher average?

74. **Average Income**
    a. Full-time (40 hours per week) employees
    b. Permanent (year-round) employees
    c. Full-time (40 hours/week), permanent (year round) employees.
    d. No difference
    e. Don't know or not sure

==

75. **In a famous ink-blot test, most schizophrenics selected a particular shape.**
    Does this mean that most people who select that ink-blot shape are schizophrenic?
    a. Yes b. No c. Don't know or Not sure

Q. Which ending is necessarily implied by the introductory statement?

76. **More doctors recommend Crest**
    a. than any other toothpaste.
    b. than nurses recommend Crest.
    c. Can't tell from the statement
    d. Don't know or not sure

77. **People who smoke have lower IQ …**
    a. than other people who don't
    b. than these same people had before they started smoking
    c. Can't tell from the statement
    d. Don't know or not sure

ENTER ANY COMMENTS HERE: