Including Student Ability to Assess Learning with Other Assessment Tools

Zachariah Dietz[1], John Lovell[2], and Julia A. Norton[3]
[1]Probability Consultant, Rochester, New York 14620
[2]Department of Psychology, California State University, East Bay, Hayward, CA 94542
[3]Department of Statistics and Biostatistics, California State University, East Bay, Hayward, CA 94542

Abstract

Retrospective look at ten years of assessing introductory statistics courses over quarters. Introduces a "how sure are you of this answer" question. Since Fall Quarter 1999, the authors have collected data from a required common final in introductory statistics and some finals in introductory psychology courses. After ten years we wonder whether there is some relationship between correct response and an individual student's assessment of their ability to answer a particular question correctly. Our study considers continuity in exams and the usefulness of asking students to assess their own problem solving ability. For each of twenty questions on a common final in an introductory statistics course, students are asked to rate their personal ability to answer that particular question correctly. Responses are studied on a number of scales. One set of scales is designed to study particular topics in introductory classes. The second set of scales looks at the difficulty of the problems in terms of literacy, skill and reasoning required to answer. In an age requiring 'customer satisfaction' we ask whether students are able to correctly utilize basic course skills and assess personal learning.

Key Words: Departments of Statistics, Educational assessment, Learning, Teaching, Introductory Statistics

1. Introduction

Trends in assessment have turned lately to asking students how well they learned the material in a course. As with most publicly funded state colleges and universities, there is a continued effort to do more with less funding, to reduce or eliminate small programs, and to demand that all programs demonstrate that they provide something for the public good. Even before the current Assessment or Quality Control environment that we find ourselves in, the Department of Statistics and Biostatistics at then California State University, Hayward (now East Bay) tried to find some mechanisms for standing out in the eyes of the campus administrators and the state-wide university system (Norton 1997).

As student learning outcomes became mandated state-wide, the Statistic and Biostatistics Department embraced the opportunity to develop effective instruments to gather useful information about our programs and courses. Our first effort included a faculty wide report (summarized in Norton 1999) investigating our introductory statistics class, to assess the common information students possessed after a variety of instructors. We compared the similarities between these introductory non-statistics student results with those of students in a variety of our own programs such as the minor and bachelors in statistics.

Summaries of our results appear in several *Proceedings of the American Statistical Association Section on Statistical Education* as our data increased and the questions became more varied (Eudey and Norton with others between 2000 and 2006). These papers consider assessments in introductory courses and our statistics degree programs. The ideas discussed are consistent with the fundamental learning goals outlined in Garfield and Ben-Zvi (2007) and Norton and Lovell (1981). In 2006-2007 Norton served as Interim Director of Institutional Research, writing a broader survey of the assessment at the University (Norton 2007). Returning to teaching in 2007-2008 we collaborated with many faculty from all areas of the university in supporting assessment attempts (Norton, Zhou, and Ganjeizadeh 2008 and Eudey, Anand, Norton and Coulman 2009).

Seeking less controversial means of evaluation and ones perhaps less intrusive to the classroom, some suggest asking students directly about their learning experience in terms of what they had learned using a consumer model of assessment. Since our previous research concluded that common finals written by committee or by outside evaluators gave results that satisfied us, we wondered how a version of these new methods might work. We are not in favour of using student evaluations as assessments of class success. Therefore, we decided to associate the question of learning with the twenty questions already being used in the introductory statistics final.

2. Relationship Between Correct Response And Student Certainty

We wondered whether the correct and incorrect responses related to the degree to which students were certain of their answers. For each of the twenty questions, students indicated on a scale from 1 to 5 how certain they were of the answer that they had given. The scale was ranked from highest to lowest. Indicating 1 meant that the student was very sure that the response given was correct, while indicating 5 meant that the student was very unsure of the given response. The value 3 represented neutral on this scale, neither sure nor unsure about the response. Certainty of response and correct response are associated in Table 1. Table 1 lists the question type, the p-value associated with the chi-squared statistic for relationship, the linear associated p-value, and Spearman's correlation between response and certainty.

There is, with two exceptions, agreement between the expected direction of the association between getting a problem correct or incorrect and how sure students were of the results that they gave for the question. The first exception is the question on the sampling distribution of the mean. A situation is described with given mean and standard deviation for the population; a sample of size 400 or 900 or some other value with an easily computed square root is taken; and the student is asked to select the appropriate sampling distribution of the mean, normal with given μ and σ scaled by $\sqrt{n}$. This question is very straight forward if one recalls that there is such a thing as the sampling distribution for the mean. The information $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ is given in a formula sheet that every student has. Otherwise the exam is closed book. Looking at the table of counts for this particular question about the sampling distribution (below Table 2) there appears to be no apparent relationship between being correct and how certain the student actually is of the selected result.

Other questions where students have difficulty understanding what they don't know include problems 5, 7, 9, 16, 17, 18, and 20. Problem 4 is a problem that almost anyone who reads the question carefully gets right. In problem four there are 4 histograms shown: one is clearly normal, two very skew, and one is bimodal. Virtually all students are very to somewhat confident and at the same time, correct in identifying the normal.

Question five asks students to select a true statement about the difference between statistics and parameters. Question seven asks students to look at a histogram and estimate s. Although most correctly identify a statement about the mean in problem six, and even though there is clearly only one practical answer, visualizing the standard deviation is more of a problem this early in their quantitative careers. Question nine shows a one-way probability table and asks about a combined probability. Most students get nine correct but are less sure in this problem than they are for the results in question four, the normal distribution shape. Questions 16 and 17 are hypothesis testing questions. Even though they are usually able to identify the alternative hypothesis in question 15, they struggle with the notion of a pair in 16 and with the correct statistic in 17 even though, again, the equation is given on the formula page. Questions 18 and 20 cover confidence interval definition (18) and applying the concept of changing either sample size or the level (20). The last six questions should be the freshest in each student's mind, but probably the ideas have not had time to jell into understanding.

Question	**Chi-squared p-value**	**Linear p-value**	**Spearman* correlation**
Q1: Percent within 2 sigma	0.000	0.000	-0.339
Q2: Median estimation (identify interval)	0.074	0.025	-0.209
Q3: Degrees of freedom in t test	0.000	0.000	-0.510
Q4: Identify normal histogram	0.402	0.313	-0.110
Q5: Statistic versus parameter	0.210	0.124	-0.110
Q6: Locate mean on histogram	0.000	0.000	-0.314
Q7: Estimate s from histogram	0.483	0.950	-0.006
Q8: Identify largest or smallest s	0.000	0.000	-0.330
Q9: Top 16% in normal curve	0.450	0.064	-0.161
Q10: Sampling distribution of mean	0.297	0.436	0.072
Q11: Definition of p-value	0.023	0.219	-0.104
Q12: Approximation of r from graph	0.009	0.000	-0.318
Q13: Independence and probability	0.031	0.003	-0.257
Q14: Probability in a one-way table	0.000	0.000	-0.406
Q15: Identify alternative hypothesis	0.076	0.006	-0.229
Q16: Identify a rejection region	0.363	0.061	-0.158
Q17: Identify a hypothesis pair	0.150	0.017	-0.189
Q18: Meaning of confidence interval	0.898	0.759	0.032
Q19: Choose interval computation	0.000	0.000	-0.398
Q20: Result of changing confidence or n	0.050	0.194	-.091

Table 1. Partial results of cross-tabulation between correct/incorrect for each of twenty questions and how certain the students were of the results selected (multiple choice questions.) One hundred thirty eight exams were used with all observations present. Rarely were the strict requirements of expectations greater than 5 not met; generally with expected greater than 1. Combined cells are considered in Table 3.

*The Gamma statistic (not listed above) is always in agreement with Spearman's correlation and the other statistics and tests; it is in the correct direction and always slightly larger in magnitude.

Q10 Sampling distribution of mean * Confidence for Q10 Cross-tabulation

Count		Confidence for Q10					Total
		Very Confident	Somewhat Confident	Neutral	Somewhat Unsure	Very Unsure	
Q10 Sampling distribution of mean	**Incorrect**	13	26	22	19	6	86
	Correct	11	8	12	16	5	52
Total		24	34	34	35	11	138

Table 2. Detailed results for question 10 on sampling distributions. Although we expect a negative association, clearly there is none. Historically, understanding sampling distributions is a difficult concept for beginners.

3. Summary of correct response by question

The authors have studied these types of questions for over ten years. To see whether the percent correct for these students followed past patterns, we summarized the questions by percent correct. Additionally, we indicated the percent of responses that went to the primary distracter for each question. Table 3 below gives the proportions for each question. These results show very similar patterns to past groups of students given similar questions.

As expected, identifying the normal among a set of graphs is answered correctly nearly all of the time. Other mostly correctly answered questions include question 6, locating the average on a histogram as being close to the median on a mound shaped curve; question 1, identifying the proportion of values expected to fall within one or two standard deviations of the mean; and question 14, computing a proportion for some property from counts in a one-way table.

Also following past patterns, the most missed problem was question 12, estimating the correlation from a scatterplot. The three distracters are not close to the correct answer. Usually two are of the incorrect sign and the third is either too close to 0 or 1 to be correct. This is a disappointing fact but fairly consistent, nonetheless. Other frequently missed problem types include problems 19 and 20 concerning confidence interval computations; problems 10 and 11 concerning p-values and hypothesis tests; and problem 13 concerning the notions of independence and joint probabilities.

As in past administrations of this type of assessment exam, there are consistent relationships between the total points on this exam and the total points that a students otherwise has in the course. Breaking the questions up into groups that relate to particular midterm exams, the pattern of responses is also consistent. That is, questions about descriptive statistics and results of the first midterm are similar as one would expect.

Question	Primary distracter	% correct	% distracter
Q1: Percent within 1 or 2	Correct answer is 1 σ contains	83	15

sigma	68%; distracter is 2 σ percentage or 95% or vice versa		
Q2: Median estimation (identify interval)	Correct interval contains observation 51 out of 100; distracter is middle interval	49	42
Q3: Degrees of freedom in t-test	Distracter is n-2 rather than n-1 for one sample test	76	18
Q4: Identify normal histogram	Distracter positive skew	97	2
Q5: Statistic versus parameter	Parameter varies statistic fixed	36	28
Q6: Locate mean on histogram	Median less than mean or vice versa	88	6
Q7: Estimate s from histogram	Correct answer is 3.5; distracter is width of histogram interval, 1 unit	63	18
Q8: Identify largest or smallest s	Read as opposite	76	18
Q9: Top 16% in normal curve	Top 2.5% or bottom 2.5%	60	17
Q10: Sampling distribution of mean	Off by factor of 10 or fail too divide σ at all	38	32
Q11: Definition of p-value	Probability null is wrong or probability alternative is true	36	23
Q12: Approximation of r from graph	Incorrect sign; improbable value close to zero	26	28
Q13: Independence and probability	Confused disjoint and independent; use incorrect rule	40	21
Q14: Probability in a one-way table	Two results given for “or”	79	16
Q15: Identify alternative hypothesis	Statistic instead of parameter	59	14
Q16: Identify a rejection region	One sided region when two is appropriate; fail to note entire region	46	22
Q17: Identify a hypothesis pair	Statistics; wrong parameter	62	14
Q18: Meaning of confidence interval	Correct response is central limit theorem applies; incorrect definition with very specific values/parameters	53	21
Q19: Choose interval computation	Incorrect t; fail to divide by root n	33	32
Q20: Result of changing confidence or n	Larger versus smaller when opposite appropriate	30	26

Table 3 Individual question type indicates primary distracter, showing correct percent and percent for distracter

4. Scales, topics, and grades

In the past the authors have created two different types of scales using the percent of correct responses on either particular topics in introductory statistics or on assessment scales related to literacy, skill, and thinking. Assessment scales are based on the notion that questions are primarily ones of statistical literacy, skill at manipulation, or ability to think through to a solution. All scales are computed by summing the correct responses, averaging over the questions in the scale, and converting to percent correct for the scale. That is, a value of 100 indicates that all responses, related to this topic or scale, were correct or that there was 100% knowledge acquisition for the particular topic or scale.

Positively related scales for a student's confidence in a particular topic or assessment scale in statistical literacy, skill at manipulation, and ability to think through to a solution were created using the same basic definitions for the scales as used in formulating assessment measures. Positivity was achieved by inverting the confidence scale of the basic equations. Specifically, a literacy score was calculated based on the average percent correct for a series of eight questions that relied primarily on knowing the definitions and other basics of a topic. A skill score was computed as the average percent correct of six questions that relied primarily on the ability to apply a formula directly. A thinking score was calculated based on the average percent correct for six problems that rely on taking the definitions and formulas one step farther to obtain an answer. One way to think of these three scales is easy, medium, and difficult based on studying student learning in introductory statistics courses through these three assessment scales. The corresponding confidence scores were computed using the confidence in the correct response to the same scale questions, subtracted from the maximum value (no confidence), and scaled to 100. Table 4 below summarizes the correlation between the three assessment scales and their confidence counter parts. For all computations there were 138 data points.

	Literacy Assessment	**Skill Assessment**	**Thinking Assessment**	**Confidence Literacy**	**Confidence Skill**	**Confidence Thinking**
Literacy Assessment	1.000	0.746	0.299	0.462	0.025	0.380
Skill Assessment	0.746	1.000	0.384	0.424	0.218[1]	0.392
Thinking Assessment	0.299	0.384	1.000	0.159	0.143	0.304
Confidence Literacy	0.462	0.424	0.159	1.000	0.145	0.692
Confidence Skill	0.025	0.218	0.143	0.145	1.000	0.486
Confidence Thinking	0.380	0.392	0.304	0.692	0.486	1.000

Table 4. Correlations between assessment scales and confidence in assessment scales based on 138 observations [1](Correlations over 0.17 are .05 significant and over 0.22 at .01, roughly. Using Spearman's correlations does not change the results in an appreciable manner.)

Considering Table 4 leads us to conclude that the scales literacy and the ability to use equations are related, and that thinking is related to confidence in both literacy and skill while not necessarily highly related to the student's ability to correctly respond to literacy and skill questions. Confidence in literacy is most highly related to the literacy assessment scale and nearly as highly related to the skill related scale, but not very related to the ability to solve thinking related problems. The student's confidence in using equations (skill) seems not very related to any other type of problems except skill based questions and confidence in solving thinking problems.

The questions can be organized in a more topic oriented manner as can the confidence levels. That is, we can create scales that are the average percentage correct in the areas of descriptive statistics, normal curve, probability, hypothesis tests, graphs, and confidence intervals. Confidence scales are created the same way by averaging the confidence in the same questions, inverting the scale for positive correlations, and rescaling to range from 0 to 100. Table 5 below gives the correlations between these measures.

	Confidence Descriptive	Confidence Normal	Confidence Probability	Confidence Hypothesis	Confidence Graphs	Confidence C Intervals
Assessment Descriptive	0.412	0.363	0.178	0.237	0.388	0.146
Assessment Normal	0.178	0.283	0.075	0.139	0.213	0.196
Assessment Probability	0.236	0.296	0.351	0.103	0.266	0.047
Assessment Hypothesis	0.254	0.297	0.132	0.318	0.257	0.170[1]
Assessment Graphs	0.277	0.275	0.158	0.179	0.325	0.128
Assessment C Interval	0.146	0.196	0.047	0.170	0.128	0.186

Table 5 Correlations between Assessment of Topic Learning and Confidence in Topic Learning based on 138 observations [1](Correlations over 0.17 are .05 significant and over 0.22 at .01, roughly. Using Spearman's correlations does not change the results in an appreciable manner.)

The topic-wise correlations in Table 5 are not very remarkable. The largest correlation was between descriptive statistics and its associated confidence score, followed by the correlation between descriptive statistics and the confidence score for the normal questions. These two topics recur repeatedly through-out an introductory class. The largest correlation with the normal assessment measure is the confidence in that topic. The largest correlation with probability measure is the confidence a student has in the probability questions. Similarly, for hypothesis tests and graphs, the assessment measure for the topic has the highest correlation with the confidence in those particular questions. Only the topic of confidence intervals varies slightly from this pattern. The correlation with normal question confidence is slightly higher than the correlation with the confidence interval confidence measure. Basically, students track their ability to answer a problem to a small degree.

Does this tracking ability vary by grade in the course? Considering the 40 students who earned A in the course, the same pattern follows except that confidence in the normal questions did not relate to any of the measures. For the 72 B students, the pattern persists except that a relationship between confidence in hypothesis tests and in confidence interval questions was not related to results. The 14 C students are such a small group that, the pattern while present is more questionable. The students who took all exams and still did not pass the course (D is considered not passing for general education credit so counted as no credit) was a similarly small group of 12 with no discernable pattern in correlations. The higher the student score, the better the student's ability was to assess how well they answered questions.

Using single overall measures, the total percent correct and a similarly computed level of confidence, results in a correlation of 0.425 for the 138 students. There definitely is a relationship. It is not a very strong one, however. Table 6 below shows fairly good tracking between the percentage of students who answer a question correctly and the percentage of students who respond very sure or sure of their answer (r=0.83), regardless of whether their answer is correct or not (original scale is inverted).

Question	Percent Correct	Percent Confidence
Q1: Percent within 1 or 2 sigma	83.3	86.2
Q2: Median estimation (identify interval)	48.6	82.6
Q3: Degrees of freedom in t test	75.4	76.1
Q4: Identify normal histogram	97.1	90.6
Q5: Statistic versus parameter	36.2	30.4
Q6: Locate mean on histogram	88.4	73.9
Q7: Estimate s from histogram	63.0	55.8
Q8: Identify largest or smallest s	76.1	68.1
Q9: Top 16% in normal curve	60.1	72.5
Q10: Sampling distribution of mean	37.7	42.0
Q11: Definition of p-value	35.5	34.1
Q12: Approximation of r from graph	26.1	42.8
Q13: Independence and probability	39.9	47.8
Q14: Probability in a one-way table	79.0	74.6
Q15: Identify alternative hypothesis	59.4	58.7
Q16: Identify a rejection region	45.7	38.4
Q17: Identify a hypothesis pair	61.6	56.5
Q18: Meaning of confidence interval	52.9	34.8
Q19: Choose interval computation	32.6	48.6
Q20: Result of changing confidence or n	29.7	28.3

Table 6 Original order of questions shows percent correct for each question and percent confidence (very sure or sure out of 5 possible responses) in the answer given for that question.

Instead of thinking about this problem in pieces, we can consider a repeated measures design for these data with grades as a between subjects variable. Figure 1 shows the level of knowledge acquisition (percent correct) on each of the student learning outcomes, literacy, skill, and thinking, interspersed with the corresponding confidence for each of these outcomes. The four separate lines are the estimated marginal means (percent correct) for the four grade groups: A, B, C, and NC (no credit.) The six learning measures for the 138 students are a significant part of the model as are the grades. There is significant interaction present in the model. B, C, and NC students are more confident than their ability in each of the three learning areas. This is also true for A students with the notable exception of skill where these students have more ability than reported confidence in their ability to answer the skill related questions correctly. Surprisingly, the C students were able to solve thinking problems at virtually the same percentage correct as the B students in this study.

After thinking about the psychological implications of the results, Lovell suggests that we consider Bandura’s ideas on self-efficacy. Albert Bandura coined the term "self-efficacy" to describe motivation. Self-efficacy is a person's belief in his or her ability and capability to solve a problem in any future situation (Bandura 1977, 1982, 1994). For example, if a person believes he is a brilliant scientist and can complete any scientific experiment, he has a high self-efficacy in science because he believes in his competency to perform a future experiment. Whether it is true that he is brilliant in science or not is not the same thing, it only matters what he believes.

The psychologist Bandura posited that self-efficacy influences motivation of a person's goals, actions, and successes (or failures) in life. For example, if your self-efficacy in an area is much lower than your ability, you will not be motivated to challenge yourself or improve. If your self-efficacy in an area is much higher than your ability, you may be motivated at first but then will set goals that are too high and fail which also leads to a decrease in motivation. The ideal self-efficacy is slightly above a person's ability: high enough to be challenging while still being realistic.

It seems possible that increases in self efficacy might be a primary benefit of education. If a person becomes more likely to take on challenging work in the future, it may lead them to having a chance at greater accomplishments. On the other hand, irrelevant factual knowledge may have little benefit other than to promote the individual's self efficacy.

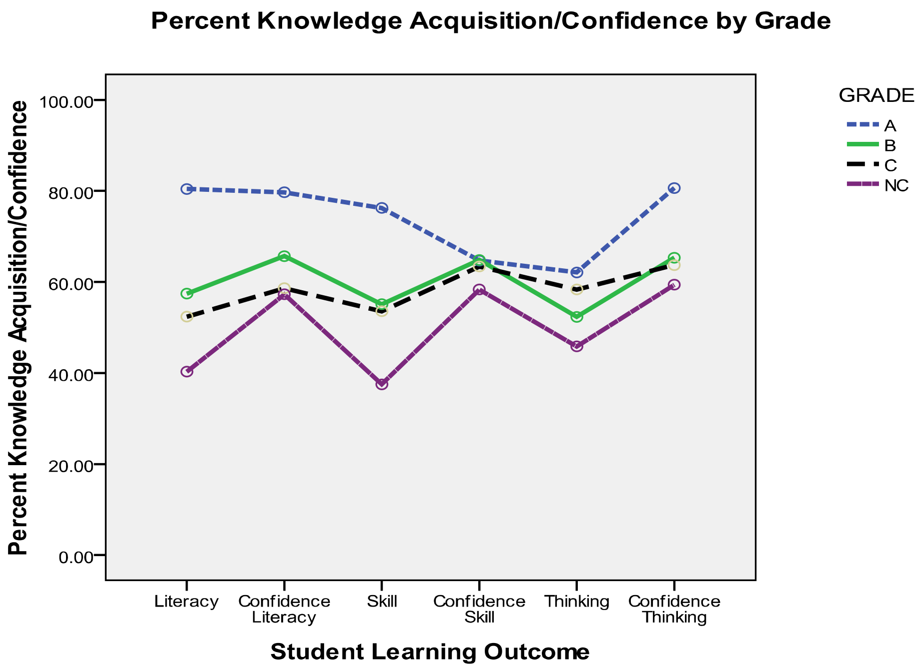

Figure 1. Profile Plot by Grade in Course for Learning Outcomes, Literacy, Skill, and Thinking and Reported Confidence in Answering Those Associated Questions Correctly for 138 Students in Beginning Statistics Courses Fall and Winter 2010-2011

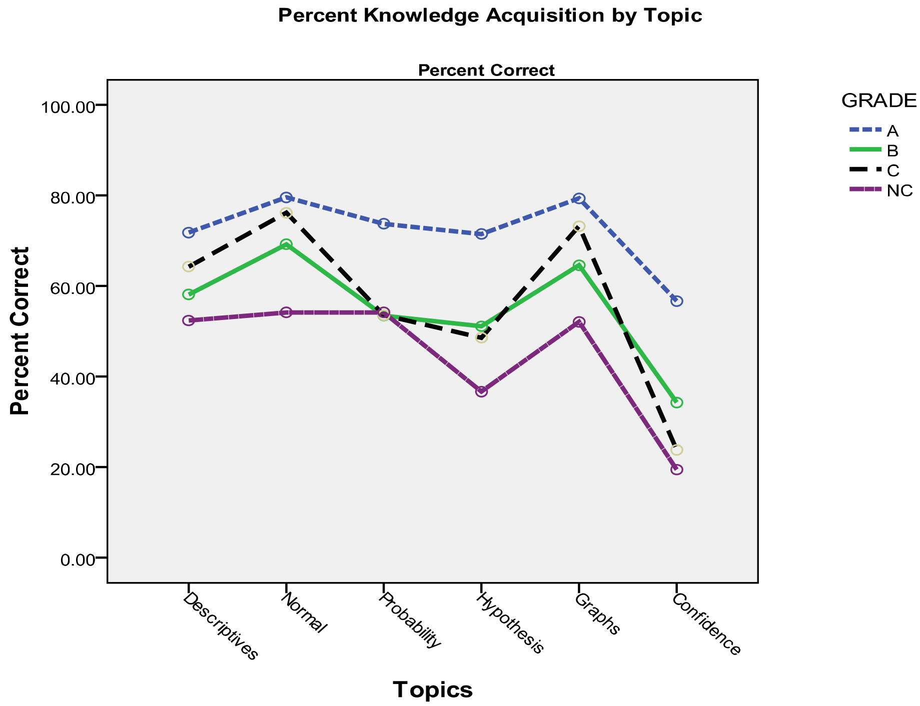

Figure 2. Profile Plot of Grade in Course by Topics covered for 138 Students in Beginning Statistics Courses Fall and Winter 2010-2011. Measurement is average percent correct for each of the six topics on the final.

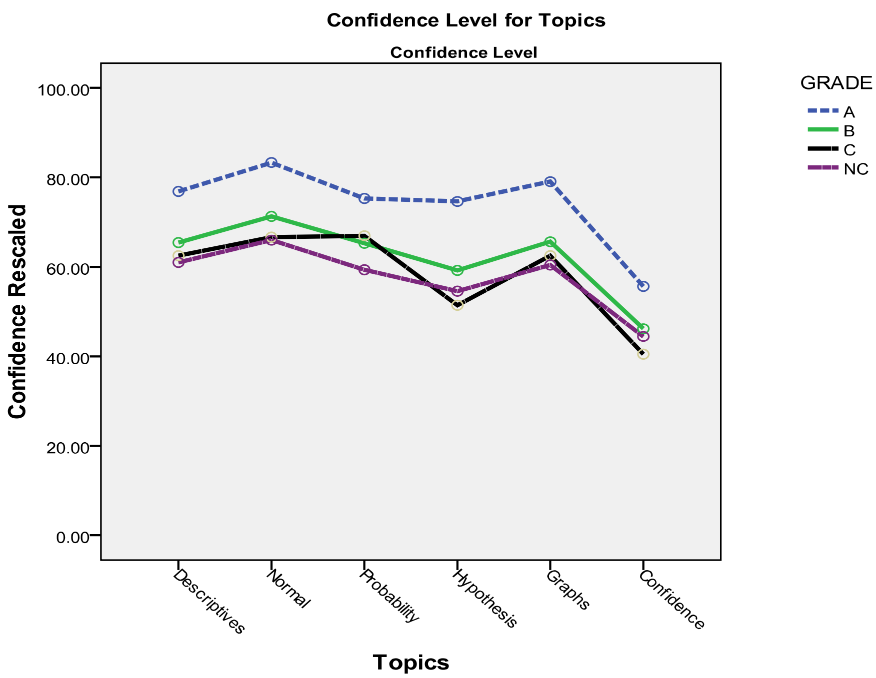

Figure 3. Profile Plot of Grade in Course by Topics covered for 138 Students in Beginning Statistics Courses Fall and Winter 2010-2011. Measurement is rescaled confidence expressed in each of the six topics on the final.

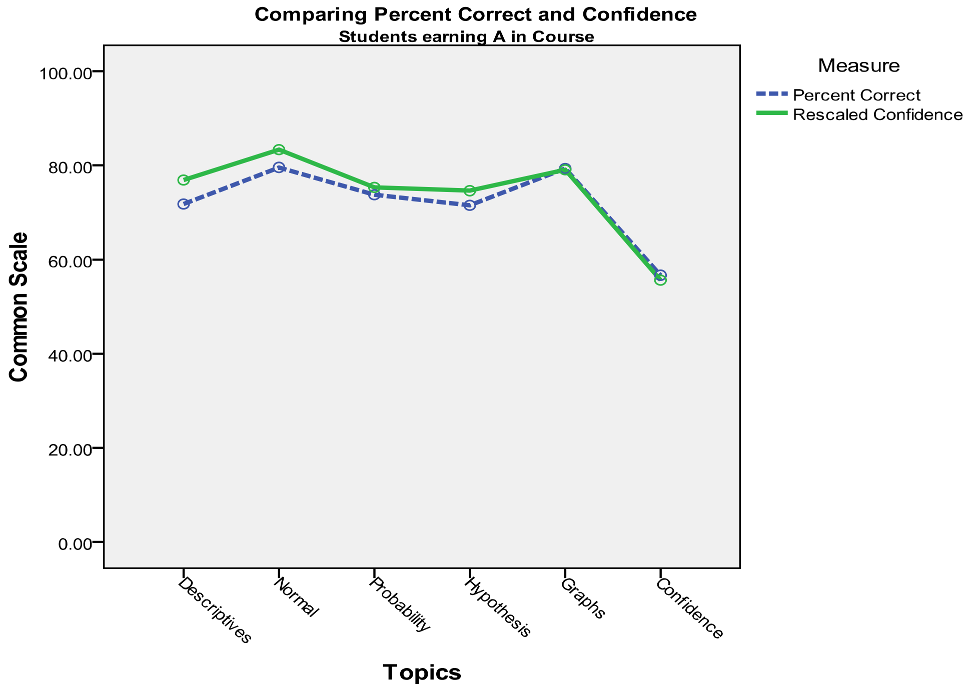

Figure 4. Profile Plot of Topics covered on a common scale for n=40 Students earning A in Beginning Statistics Courses Fall and Winter 2010-2011. Measurements are Percent Correct and Rescaled Confidence expressed in each of the six topics on the final.

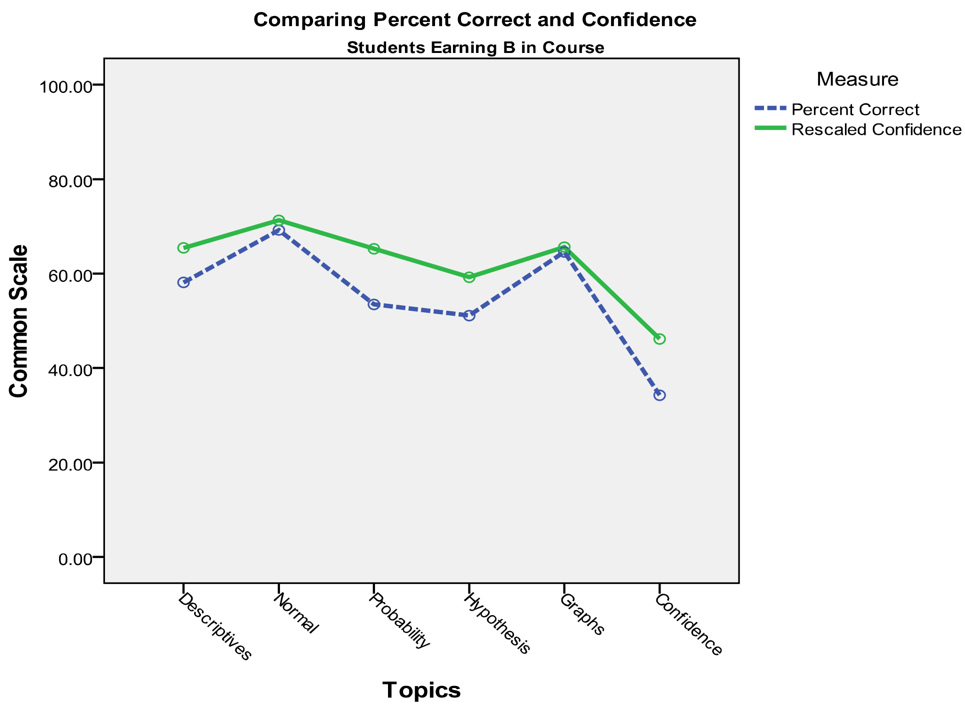

Figure 5. Profile Plot of Topics covered on a common scale for n=72 Students earning B in Beginning Statistics Courses Fall and Winter 2010-2011. Measurements are Percent Correct and Rescaled Confidence expressed in each of the six topics on the final.

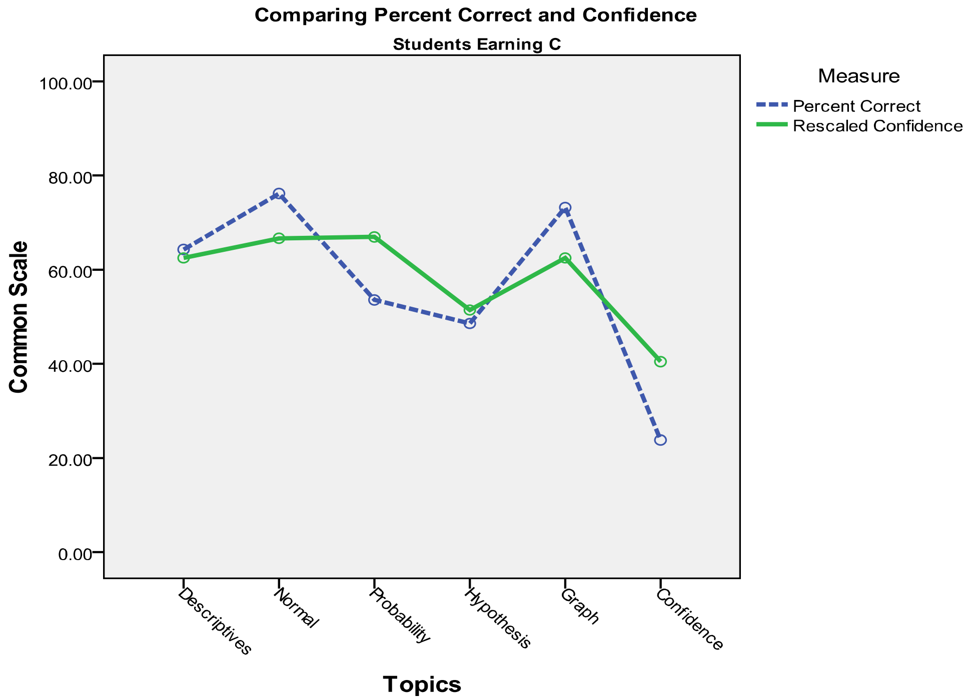

Figure 6. Profile Plot of Topics covered on a common scale for n=14 Students earning C in Beginning Statistics Courses Fall and Winter 2010-2011. Measurements are Percent Correct and Rescaled Confidence expressed in each of the six topics on the final.

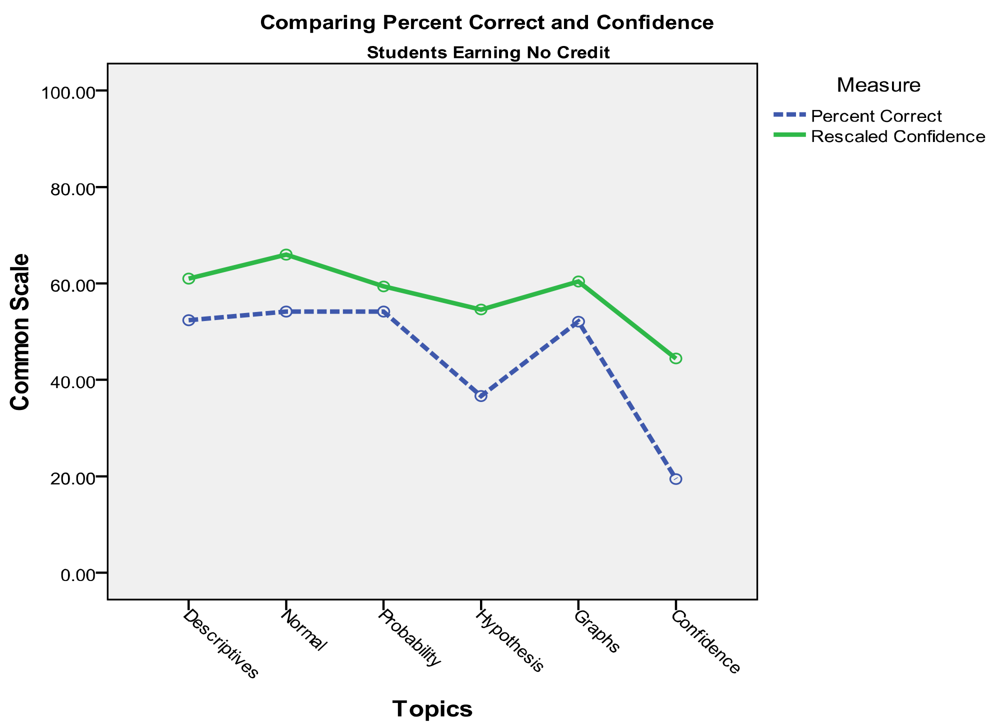

Figure 7. Profile Plot of Topics covered on a common scale for n=12 Students earning NC in Beginning Statistics Courses Fall and Winter 2010-2011. Measurements are Percent Correct and Rescaled Confidence expressed in each of the six topics on the final.

5. Summary

If educators are considering the self-study model of asking students how sure they are of the knowledge that they have obtained, at least in this setting of assigned surety to individual problems, we found that there is a similar pattern and association between correct responses and student confidence in a particular answer. While these results are only from two classes of introductory statistics, we were surprised to find so much agreement between achievement and certainty of achievement.

References

1. Bandura, A. (1977). "Self-Efficacy: Toward a Unifying Theory of Behavioral Change", *Psychological Reviews, v. 84*, pp 191-215.
2. Bandura, A. (1982). "Self-Efficacy Mechanism in Human Agency", *American Psychologist, v. 37.*
3. Bandura, A. (1994) "Self-efficacy", *Encyclopedia of Human Behavior, v. 4*, pp 71-81.
4. Garfield, J. B. and D. Ben-Zvi (2007). Developing students' statistical reasoning: connecting research and teaching practice. Emeryville, CA: Key College Publishing.
5. Lovell, J. D. and Norton, J. A. (2002) "Percent Content Mastery Testing Of Content In College Courses", *Proceedings of the American Statistical Association Section on Statistical Education.*
6. Norton, J. A. (2007), "Assessment Meets Institutional Research," *American Statistical Association 2007 Proceedings of the Section on Statistical Education.*
7. Norton, J. A., and Eudey, T. L. (2006), "Program Assessment in Statistics at the Masters Level," *American Statistical Association 2006 Proceedings of the Section on Statistical Education*.
8. Norton, J with Statistics faculty and staff: S. Fan, D. Fearn, J. Kwon, E. Nebenzahl, M. Orkin, E. Suess, B. Trumbo, M. Watnik, W. Rodriguez, C. Sugahara, A. Cambra, and P. Towers (2005) "Self-Review Document, Department of Statistics, California State University, East Bay. Internal Document.
9. Norton, J. A. (2001) "Assessment of Introductory Statistics: Phase II", *Proceedings of the American Statistical Association Section on Statistical Education.*
10. Norton, J. A. and J. D. Lovell (2000) "Repeated Measures Design In Assessment: Added Value of Instruction", *American Statistical Association 2000 Proceedings of the Section on Government Statistics and Social Science*, pp. 282-283.
11. Norton, J., J. Lovell, E. Suess, B. Trumbo, C. Sugahara, W. Rodriguez, J. Fowler, E. Worth, R. Young, T. Grube, V. Sue. (1999). "Statistics Program Assessment," *American Statistical Association 1999 Proceedings of the Section on Statistical Education.* 298-303.
12. Norton, J. A. (1997) "Strategies for Small Departments of Statistics Facing Uncertain Times", *American Statistical Association 1997 Proceedings of the Section on Statistical Education.*

13. Norton, J. A., and Lovell, J. D. (1981). "Statistics as an appropriate subject for satellite-based distance teaching", *Education for the Eighties: The Impact of the New Communications Technology*. Deakin University Press, 151-154.
14. Norton, J. A., Zhou, Y., and Ganjeizadeh, F. (2008). "Better Features in Teaching Introductory Statistics", *American Statistical Association 2008 Proceedings of the Section on Statistical Education.*