Student Self-Evaluation of Learning with Other Assessment Tools, Part II

Zachariah Dietz[1], John Lovell[2], John Norton[3], 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

We undertake assessing introductory courses over quarters, grades, and the second introduction of a "how sure are you of this answer" question. Since Fall Quarter 1999, the authors have collected data from introductory Statistics and Psychology classes. After twelve years we wonder whether there is some relationship between correct response and an individual student assessment of their ability to answer a particular question correctly. Last year our study considered the usefulness of asking students to assess their own problem solving ability in statistics courses. This year we follow up on the past study and include a Psychology course. For each of twenty questions, 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: Educational assessment; Learning; Teaching

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 combine, reduce or eliminate small programs, and to demand that all programs demonstrate that they provide something for the public good. Dietz, Lovell and Norton have collaborated on a number of issues in educational assessment over the past ten years including learning in introductory psychology and statistics courses. (2000, 2005)

Summaries of our previous studies appear in several *Proceedings of the American Statistical Association Section on Statistical Education* as our data increased and the questions became more varied (Lovell, Dietz, 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 Norton 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 among university faculty and ones perhaps less intrusive into 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. Therefore, in 2011 we decided to associate the question of learning with the twenty questions already being used in the introductory statistics final. Achieving some indication of student ability to determine correctly how difficult a particular question was for an individual to answer, we decided to extend our investigation to introductory psychology courses as well. The results were mixed. In 2012 we included similar results from an Introductory Psychology course to see how the results varied.

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 question on the course final, 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 were associated. . We expected better students to have correlations, but we were unsure whether this was indeed realistic.

Comparing the point biserial correlations between whether a problem is correct or not and the level of assuredness a student says she/he has for that answer on the final over the course grades, an analysis of variance indicates that the groups are different, but the primary difference is between those failing the course and those passing the course (F=2.76, p=0,034 with 4, 77 df). Figure 1 visually indicates similar results.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*	*Standard error*
F	29	14.04469	0.484299	0.054096	0.043190064
D	16	9.449991	0.590624	0.03184	0.044609548
C	11	6.633603	0.603055	0.022577	0.04530369
B	8	4.986499	0.623312	0.046937	0.076597535
A	18	11.92789	0.662661	0.017691	0.03134995

Table 1. Introductory Psychology students (n=82) indicated their level of confidence on each of twenty questions in a course final. Point biserial correlations were computed for each student indicating the association between the correct response to each question and the level of confidence indicated (F=2.76, p=0,034 with 4, 77 df). Converting to Fisher's z did not alter the results significantly. Clearly the average correlation increases with grade.

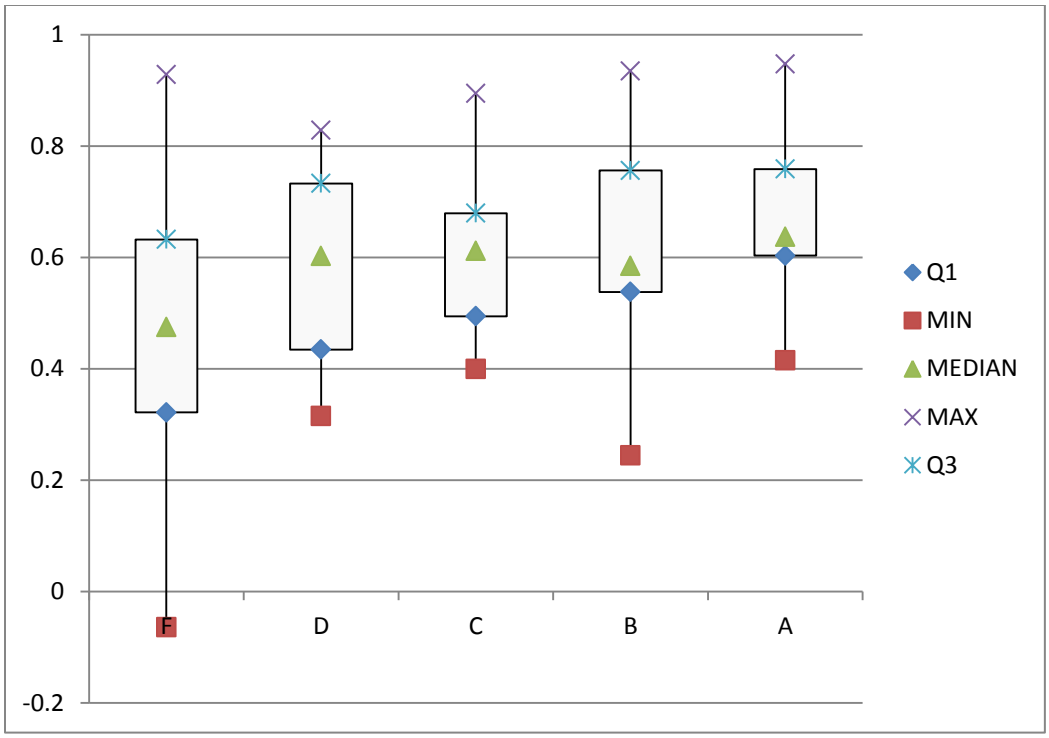

Figure 1. Box and Whisker Plot of point biserial correlations for each student, between self-assessment (assuredness) on each question and whether student got question correct or not, broken down by course grade, Winter 2012, Introductory Psychology.

Combining groups so that the samples are more nearly equal in size requires only that we combine A with B, C with D, so that there are 3 resulting groups. Table 2 below indicates an even starker difference among the three groups (F=5.51, p=0.006 with 2, 78 df). Figure 2 shows a visualization of this effect, that better students are better able to judge their performance, question by question.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*	*Standard error*
F	29	14.045	0.484	0.054	0.043
CD	27	16.084	0.596	0.027	0.032
BA	26	16.914	0.651	0.026	0.031

Table 2. Introductory Psychology students (n=82) indicated their level of confidence on each of twenty questions in a course final. Point biserial correlations were computed for each student indicating the association between the correct response to each question and the level of confidence indicated (F=5.51, p=0.006 with 2, 79 df). Converting to Fisher's z did not alter the results significantly. Clearly the average correlation between correct response and self-asssuredness increases with grade for psychology students.

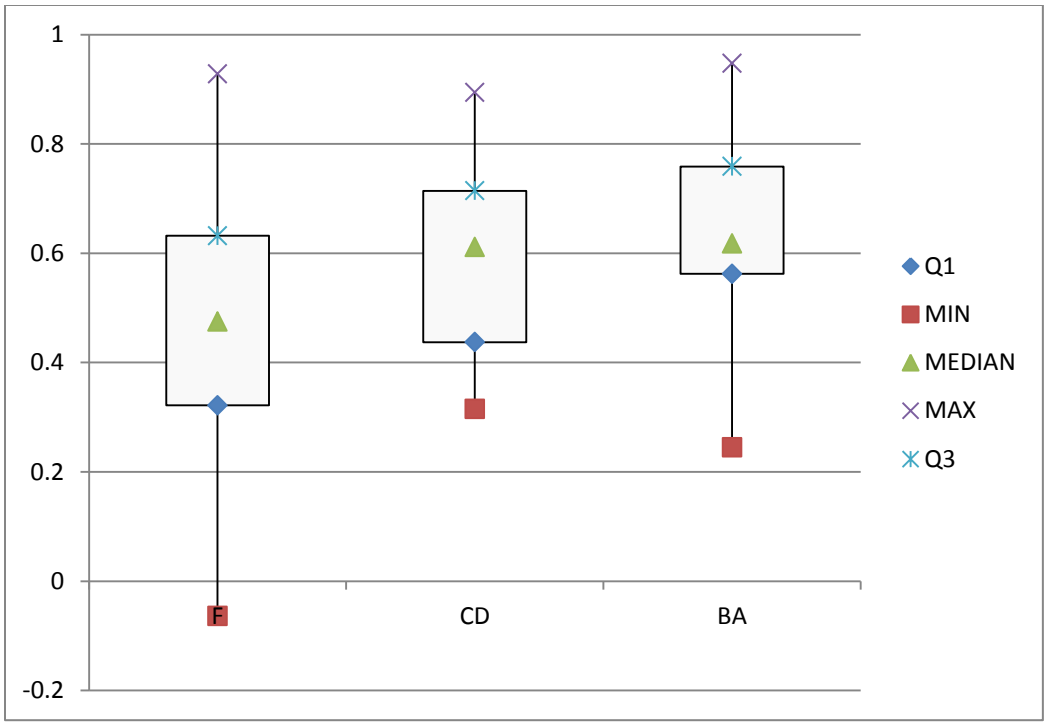

Figure 2. Box and Whisker Plot of Point biserial correlations on each student, between self-assessment (assuredness) of each question and whether student got question correct or not, broken down by course grade; A and B grades combined, C and D grades combined, Winter 2012, Introductory Psychology.

We wondered whether statistics students and psychology students were different in their ability to assess how likely the response to a question was correct. Using a simple unequal variance t-test to compare introductory psychology students and introductory statistics students from Winter 2012, we found significant differences between the two groups ($t=6.7$, $p<.001$, approximate $df=120$ with $n_{stat}=63$ and $n_{psyc}=82$). Figure 3 visually shows that Psyc students were better able to assess their knowledge of the subject matter than were the statistics students.

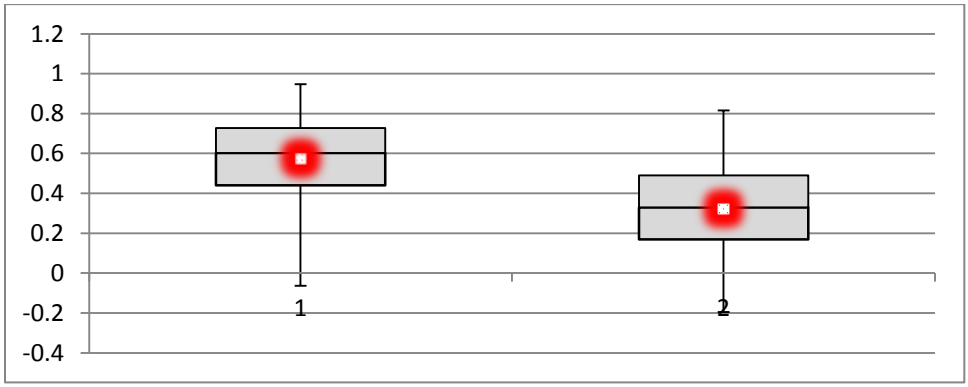

Figure 3. Introductory Psychology students Winter 2012 (n=82) on the left have higher correlations than Introductory statistics students on the right (n=63).

We considered whether students in statistics or psychology were better able to assess their progress required that we compare data from the two courses. We decided to use data from winter 2012 for both courses. Using a two way unbalanced ANOVA resulted in the surprising result that the only significant difference was between courses (See Table 4 below). In order to compare the data, we had to convert to Fisher's F statistic and rescale for comparable standard error. This seemed important since there were different numbers of questions evaluated in the two courses. Since this is a larger test than either of the previous ones, it is reasonable to assume that there might be a difference in the ability to evaluate based on course that we had not anticipated. In fact, the results from the Winter 2012 statistics course was in a different direction with A students being less sure of their results than other students. Figure 4 indicates this pattern.

Between-Subjects Factors

		N
COURSE	PSYC	82
	STAT	63
GRADE	A	34
	B	40
	C	21
	F	50

Tests of Between-Subjects Effects

Dependent Variable: fisher_z_scaled_by root n-3 to obtain constant standard error

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	36.873[a]	7	5.268	4.549	.000
Intercept	260.345	1	260.345	224.816	.000
COURSE	21.847	1	21.847	18.866	.000
GRADE	1.511	3	.504	.435	.728
COURSE * GRADE	4.029	3	1.343	1.160	.328
Error	158.651	137	1.158		
Total	620.114	145			
Corrected Total	195.524	144			

a. R Squared = .189 (Adjusted R Squared = .147)

Table 4. A two factor ANOVA indicates that course differences is the only significant factor unlike the t-test cited above. Even though a t-test finds these results between grades significant for Psychology, the F test does not support that finding. The surprising result is in figure 5 that the statistics students behave much differently than the psychology students.

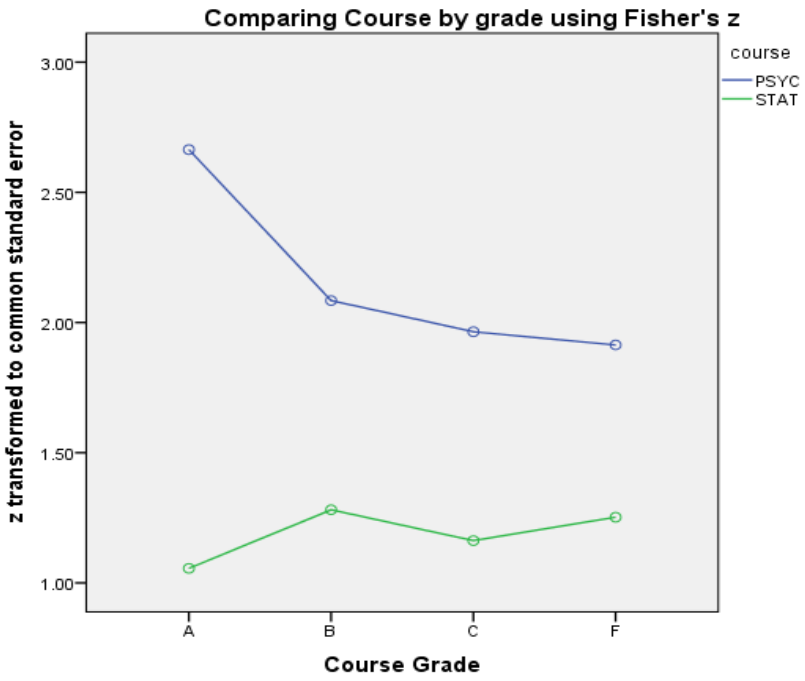

Figure 5. Graphing the averages by course and grade shows a surprising trend for statistics students earning a grade of A in this quarter to underestimate their ability to solve problems correctly.

These results are not in agreement with the pattern that we observed in statistics students in Winter 2011. Figure 6 below is a repeat of the information found last time.

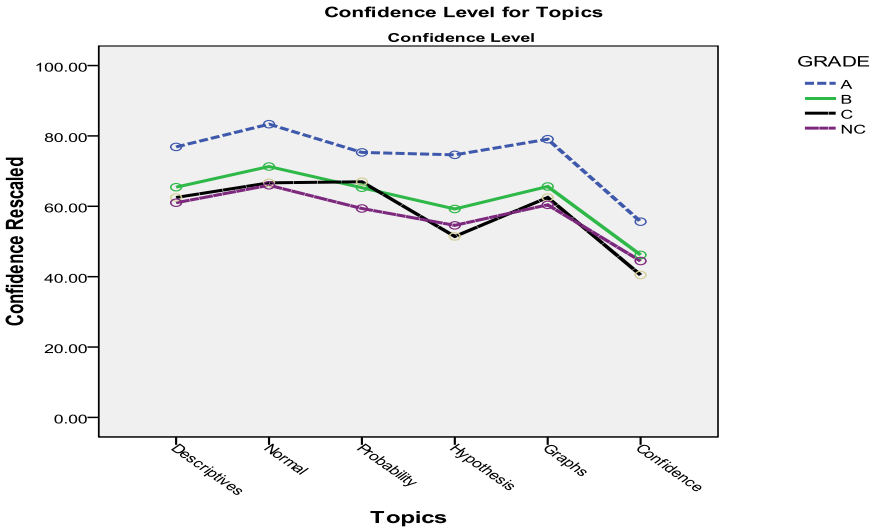

Figure 6. Profile Plot of Grade in Course by Topics covered for 138 Students in Beginning Statistics Courses Fall and Winter 2010-2011. Measurement is rescaled confidence expressed in each of the six topics on the final.

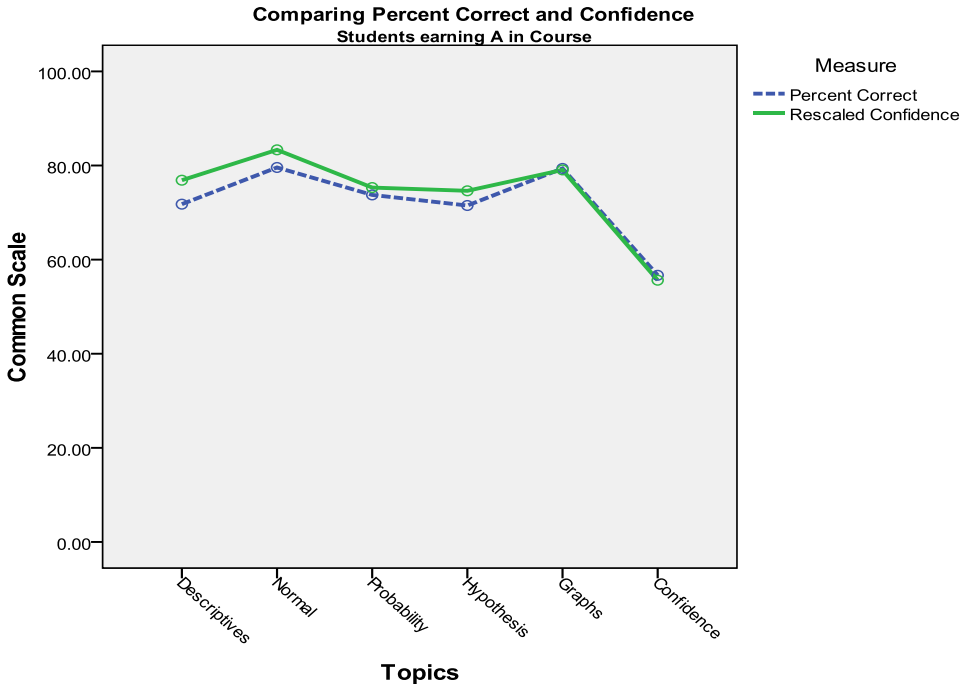

Figure 7. Profile Plot of Topics covered on a common scale for n=40 Students earning A in Beginning Statistics Courses Fall and Winter 2010-2011. Measurements are Percent Correct and Rescaled Confidence expressed in each of the six topics on the final.

3. 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 last year and this year in an introductory psychology course, but different for an introductory statistics course this past winter. While these results are only from two years of introductory statistics and one from introductory psychology, we were surprised to find so much agreement between achievement and certainty of achievement and yet have more questions to investigate.

References

1. Dietz, Zachariah, Lovell, J. D., and Norton, J. A. (2011) " Including Student Ability to Assess Learning with Other Assessment Tools", *American Statistical Association 2011 Proceedings of the Section on Statistical Education*
2. 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.*

3. Norton, J. A. and Lovell, J. D. (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..
4. 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.*