Optimal Test Length:
How Many Pages do I Need to Grade (or Even Write)?

Meredith Dalrymple
Seton Hall University

Linnea Endersby
University of Tennessee

Kurt Rotthoff[1]
Seton Hall University

Abstract
Professors spend a lot of time writing and grading exams (even if you do not grade your own exams, you are spending someone's time to do it). This study looks at this time-consuming activity by analyzing the difference in averages over different test lengths. After finding myself writing (unnecessarily) long exams, we wanted to test the length of an exam that keeps the same average and standard deviation as a longer exam with more questions. When testing exams four to six pages long (with multiple questions per page, or a long question with multiple sections), we find that exams above four pages do not seem to add any precision to the averages or standard deviations. We find that shorter exams are still accurate on average, but the standard deviations shrink, making it harder to separate student quality.

Keywords: Exam Length, Optimal Grading, Assessment
JEL: A22, D84, I21

[1] Kurt Rotthoff at: Kurt.Rotthoff@shu.edu, Seton Hall University, JH 621, 400 South Orange Ave, South Orange, NJ 07079. Linnea Endersby is a graduate of Seton Hall and currently a Ph.D. student at the University of Tennessee and Meredith Dalrymple is an undergraduate student at Seton Hall University. A special thank you to Alonso A Arbulu Arenas for sparking the idea. Any mistakes are ours.

I. Introduction

Most faculty their time split between the many tasks that come with being an academic. When additional efficiencies are found, this can benefit other important academic duties, such as research. This study addresses one aspect of teaching that can be a time-suck: grading.

To assess the quality of student performance, exams with differing levels of questions are given, which often become unnecessarily long – in hopes of more accurate outcomes. I found myself writing exams five, six, or more pages long, only to end up debating with my students about grading exams. I admitted that the worst part of grading these long exams was that after about the second page, I knew what grades everyone would receive – yet I was obligated to grade the entire exam. These students challenged this statement, so I began collecting data on exam performance by page to test this hypothesis.

The question is: if the copier missed one of the pages on your exam, what would happen to the average and the standard deviation of the exam scores? Likewise, what would happen if you randomly ripped out one (or more) page(s) of the exam?

II. Literature

The purpose of having tests in a course is to measure the amount of information the students have learned and can apply (at least, that is the intent). Previous literature reports conflicting results about the impact of test length on student performance. Bell and Lumsden (1980) found that they could reduce test length by more than 60 percent without decreasing the validity. However, Jensen (2013) found that longer tests led to greater performance on the final exam (and did not result in fatiguing conditions). When Lee et al. (2014) looked at exam length, they looked at reducing a 3-hour exam to a shorter time with proportionately fewer questions (analyzing 2.5, 2, 1.5, 1, and 0.5-hour exams). Their study supported reducing exams was warranted.

Our study continues this research, but instead of looking at length in time we look at length in the total number of pages on an exam. A clear distinction between our work and the work of Lee et al. (2014) is that our data is comprised of exams given throughout the semester as opposed to final exams. Final exams may be even more stressful to students, impacting their performance. These exams were designed to be completed in the allocated 75-minute class period.

III. Data and Results

The number of questions on a given page can vary from one big question (worth many points) to multiple smaller questions. Thus, if one (or more) random pages were missing from the exam you wrote, what would be the impact on the average and standard deviation of the grades in your class? For this reason, we focus on the number of pages on a given exam, not the questions themselves.

Data is taken from 16 open-ended exams from six semesters, 2017 – 2020, with the data stopping before the COVID-19 pandemic. These exams come from two classes: Intermediate Microeconomics and Economics and Finance in Sports. The tests range from four to six pages in length (three six-page, eight five-page, and five four-page exams). The total exam points averaged 175 points (with a standard deviation of 34.72 points), with an average of 40 points per page and an average standard deviation of 10.4 points per page. Any bonus questions are small and insignificant in the overall grade.

The exams are different lengths over the years, and the questions and their location in the exam vary over the years.[2] To test how many exam pages matter, we look at how a given number of pages predict the actual outcome of the overall exam.

$$Overall\ Average_{ic} = \alpha + \beta_1(X - Page\ Exam\ Average)_{ic} + Y_i + S_i + \varepsilon \quad (1)$$

Where the Overall Average is the average score each class, *i*, scores on a given exam in each course, *c*. This is done for each possible sequence of the X-page exam combinations (so each different 1-page exam, then [separately] each 2-page exam, 3-page exam, etc.). Thus, the regression estimates the impact of a (or more) random page(s) being ripped out of a four-page exam. We examine the estimated probability that this new three-page exam accurately predicts the actual average. If this three-page exam average provides the same average as the four-page exam, then you can accomplish the same goal by only writing (and grading) three pages.

In this study we also control for yearly fixed effects, *Y*, to capture any changes over time, and for the semester the course was offered, S is the spring semester (controlling for any potential learning loss, see Dills et. al, 2016). Additionally, we re-do

[2] Similar types are questions are often used, but typically not the same question across years.

these tests looking at the standard deviation of each exam. The overall average matters – but this is not always the best way to measure a test outcome (especially if a curve is warranted). Any estimated differences in standard deviations with random exam pages removed will analyze distributional differences.

$$Overall\ Std\ Dev_{ic} = \alpha + \beta_1(X - Page\ Exam\ Std\ Dev)_{ic} + Y_i + S_i + \varepsilon \qquad (2)$$

Both Equations 1 and 2 estimate the impact of removing one or more random pages of an exam to create a shorter exam, providing estimates on the impact on the average (1) and standard deviation (2) from the new shorter exam.

The summary statistics are presented in Table 1. The mean number of pages in an exam in our sample is 5.17, with an average score of 73.2 percent and a standard deviation of 18 percent. The average on the randomly selected x-page exams is close to the overall average of 73.17 but has a larger standard deviation. There are 160 different randomly drawn three-page exams with averages and standard deviations that are similar to the full sample. When looking at the 58 different random four-page exam combinations (these can only come from five- or six-page exams), we find the average and standard deviations are consistent with the full sample.

[Table 1]

Although we do this for all the different combinations, as an example: with a 4-page exam there are four different combinations of a randomly selected 3-page exam (1/2/3, 1/2/4, 1/3/4, and 2/3/4), with six combinations of two-page exams (1/2, 1/3, 1/4, 2/3, 2/4, and 3/4), and four combinations of one-page exams (1, 2, 3, and 4 – which are not presented for brevity). In Table 2, we find that the five pages of a six-page exam yield standard deviations that are the same as the six-page exam. The shorter exam average is

positively correlated with the full exam, and statistically significant. Likewise, the standard deviation is predicted to be the same when looking at any random four pages of a five- or six-page, and the average is positively correlated.

[Table 2A and 2B]

It seems clear from that sample that a six-page exam does not add value relative to any randomly drawn five-page or four-page exam from those samples. We continue to test this in more detail by looking specifically at exams totaling five pages (Table 3) and four pages (Table 4).

[Tables 3 and 4]

When looking at the exams that start as five-page exams, taking any random four-page exams is slightly different (although close) to one, but the standard deviations are positively correlated but slightly muted relative to the full exam. The averages are positively correlated as well. Then when looking only at exams that were initially four pages, the random three-page exams and less all show that they differ from one, although positively correlated with averages and standard deviations.

IV. Conclusion

While longer exams give students the chance to reveal more knowledge, they take time to create, grade, and to take. This study looked at exams that I have given over time to see if I would have randomly ripped one (or more) pages out of the exam given, would the average and standard deviation have changed? In general, removing one page from a six-page and a five-page exam would have no change. However, when moving to shorter

exams (four pages or less) the averages and standard deviations are positively correlated but muted. We encourage further research on optimal test length.

Works Cited

Bell, R., and Lumsden, J. (1980). Test Length and Validity. *Applied Psychological Measurement*, *4*(2), 165–170.

Dills, Angela K., Rey Hernández-Julián, and Kurt W. Rotthoff (2016) Knowledge Decay between Semesters *Economics of Education Review* Volume 50, February, Pages 63-74

Jensen, J., D. Berry, and T. Kummer (2013) Investigating the effects of exam length on performance and cognitive fatigue *PLoS ONE* vol. 8(8), pp. 1-9

Lee, Eric S., Connie Bygrave, Jordan Mahar, Naina Garg, and Suzanne Cottreau (2014) Can Exams Be Shortened? Using a New Empirical Approach to Test in Finance Courses *World Academy of Science, Engineering and Technology International Journal of Educational and Pedagogical Sciences* Vol:8, No:1

Table 1: Summary Statistics

Full Sample					
	Observations	Mean	Std. Dev.	Min	Max
Total Number of Pages	457	5.16849	0.668899	4	6
Test Average	457	0.732092	0.071571	0.553199	0.8157
Test Std. Dev.	457	0.180141	0.040628	0.1267	0.248152
Average On Select Pages	457	0.731771	0.099715	.432727	1.110317
Std. Dev. On Select Pages	457	0.202671	0.052155	0.069263	0.388533
Spring	457	0.592998	0.491814	0	1
(Random) Three-Page Tests Only					
	Observations	Mean	Std. Dev.	Min	Max
Total Number of Pages	160	5.25	0.663515	4	6
Test Average	160	0.735085	0.070925	0.553199	0.8157
Test Std. Dev.	160	0.178498	0.040511	0.1267	0.248152
Average On Select Pages	160	0.735496	0.084878	0.496061	0.946021
Std. Dev. On Select Pages	160	0.189771	0.045302	0.069263	0.343634
Spring	160	0.575	0.495895	0	1
(Random) Four-Page Tests Only					
	Observations	Mean	Std. Dev.	Min	Max
Total Number of Pages	58	5.310345	0.466675	5	6
Test Average	58	0.716033	0.07282	0.553199	0.7923
Test Std. Dev.	58	0.184614	0.041483	0.1295	0.248152
Average On Select Pages	58	0.712749	0.07834	0.527706	0.924916
Std. Dev. On Select Pages	58	0.192867	0.040782	0.126468	0.276765
Spring	58	0.62069	0.489453	0	1

Table 2A: Full Sample

VARIABLES	Average, 2-Page	Std. Dev., 2-Page	Average, 3-Page	Std. Dev., 3-Page
Example Average	0.49225***		0.64413***	
	(0.041)		(0.041)	
Example Std. Dev.		0.61997***		0.71637***
		(0.044)		(0.041)
Spring	-0.02069**	0.01253**	-0.01470*	0.00751*
	(0.009)	(0.005)	(0.008)	(0.004)
Yearly Fixed Effects	Yes	Yes	Yes	Yes
Constant	0.38327***	0.03442***	0.27014***	0.02934***
	(0.034)	(0.010)	(0.033)	(0.009)
P-Value that Estimate is = 1	0.0000	0.0000	0.0000	0.0000
Observations	155	155	160	160
R-squared	0.553	0.651	0.673	0.720

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2B: Full Sample

VARIABLES	4-Page	Std. Dev., 4-Page	Average, 5-Page	Std. Dev., 5-Page
Example Average	0.78036***		Average, 0.12578	
	-0.057		-0.06	
Example Std. Dev.		1.00108***		0.90681**
		-0.044		-0.112
Spring	-0.00431	-0.00452	0.00171	-0.00673
	-0.011	-0.004	-0.002	-0.006
Yearly Fixed Effects	Yes	Yes	Yes	Yes
Constant	0.16385***	-0.00051	0.69078***	0.0086
	-0.046	-0.009	-0.046	-0.02
P-Value that Estimate is = 1	0.0004	0.9803	0.0046	0.4915
Observations	58	58	6	6
R-squared	0.816	0.922	0.777	0.979

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Five-Page Exams Only

VARIABLES	Average, 2-Page	Std. Dev., 2-Page	Average, 3-Page	Std. Dev., 3-Page	Average, 4-Page	Std. Dev., 4-Page
Example Average	0.36363***		0.47283***		0.64299***	
	(0.058)		(0.058)		(0.073)	
Example Std. Dev.		0.38267***		0.32496***		0.81763***
		(0.054)		(0.066)		(0.074)
Spring	-0.02213*	0.01931***	-0.02211*	0.02022***	-0.01447	0.00476
	(0.013)	(0.005)	(0.012)	(0.006)	(0.013)	(0.005)
Yearly Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.48473***	0.06370***	0.40375***	0.07622***	0.27235***	0.01840
	(0.047)	(0.011)	(0.046)	(0.012)	(0.058)	(0.011)
P-Value that Estimate =1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0185
Observations	80	80	80	80	40	40
R-squared	0.493	0.821	0.588	0.775	0.761	0.934

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Four-Page Exams Only

VARIABLES	Average, 2-Page	Std. Dev., 2-page	Average, 3-Page	Std. Dev., 3-page
Example Average	0.02132		0.06564	
	(0.028)		(0.062)	
Example Std. Dev.		0.00452		0.01040
		(0.010)		(0.023)
Spring	-0.02780***	0.08262***	-0.02656***	0.08213***
	(0.005)	(0.001)	(0.006)	(0.002)
Yearly Fixed Effects	Yes	Yes	Yes	Yes
Constant	0.70848***	0.09965***	0.67640***	0.09907***
	(0.022)	(0.002)	(0.046)	(0.003)
P-Value that Estimate =1	0.0000	0.0000	0.0000	0.0000
Observations	30	30	20	20
R-squared	0.940	0.998	0.943	0.998

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1