1. (From G&W) There is some evidence to suggest that high school students justify cheating in class on the basis of the teacher's skills or stated concern about cheating (Murdock, Miller, & Kohlhardt, 2004). Thus, students appear to rationalize their illicit behavior on perceptions of how their teachers view cheating. Poor teachers are thought not to know or care whether or not students cheat, so cheating in their classes is viewed as acceptable. Good teachers, on the other hand, do care and are alert to cheating, so students tend not to cheat in their classes. Below is a partially completed source table and summary statistics that are consistent with the findings of Murdock et al. The scores represent judgments of the acceptability of cheating for students in each sample. Complete the source table below and interpret the data as completely as you can. What is your best estimate of the population variance (σ^2)? [15 pts]

Descriptives

Acceptability of Cheating

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Good Teacher	20	2.1000	.85224	.19057	1.7011	2.4989	1.00	4.00
Poor Teacher	20	6.0500	1.27630	.28539	5.4527	6.6473	3.00	8.00
Total	40	4.0750	2.26894	.35875	3.3494	4.8006	1.00	8.00

ANOVA

Acceptability of Cheating

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	156.2	1	156.2	132.4	.000
Within Groups	44.8	38	1.18		
Total	201	39			

The best estimate of σ^2 is MS_{Within} = 1.18.

$F_{Max} = \frac{1.629}{.726} = 2.2$, **so with $F_{\text{Max Critical}}$ = 2.5, you would conclude that there is little concern about heterogeneity of variance. You might also note that the probability attached to the *F* obtained in your ANOVA is so large that it would be significant even if you were to set α = .01.**

Decision: Reject H0, because p < .001.
Conclusion: Students with poor teachers rate cheating as more acceptable (*M* = 6.05) than students with good teachers (*M* = 2.1). Of course, this is not a true experiment, because students are not randomly assigned to classes. Thus, it may be that students who seek out good teachers are less likely to cheat.

2. (From G&W) In order to study cardiovascular responses to embarrassment, Harris (2001) had people sing the *Star Spangled Banner* in front of a video camera while she recorded their heart rate and blood pressure. She found that blood pressure increases steadily for two minutes before gradually returning to normal. What about the heart rate data? Below is a partially completed source table for these heart-rate data. Complete the table and analyze/interpret the results as completely as possible. Is the pattern for heart rate similar to that for blood pressure? [15 pts]

Descriptive Statistics

	Mean	Std. Deviation	N
Baseline Heart Rate	76.9167	1.72986	12
Heart Rate at 1 Min	89.2500	1.86474	12
Heart Rate at 2 Min	78.0833	1.78164	12

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Time	Sphericity Assumed	1112.9	2	556.4	333.800	.000	.968	667.600	1.000
	Greenhouse-Geisser					.000	.968	655.253	1.000
	Huynh-Feldt					.000	.968	667.600	1.000
	Lower-bound					.000	.968	333.800	1.000
Error(Time)	Sphericity Assumed	36.7	22	1.667					
	Greenhouse-Geisser								
	Huynh-Feldt								
	Lower-bound								

a. Computed using alpha = .05

Decision: Reject H_0, with $p < .001$

Post Hoc Test: $HSD = 3.55\sqrt{\frac{1.67}{12}} = 1.32$

	Baseline	1 Minute	2 Minutes
Baseline	---		
1 Minute	**12.3***	---	
2 Minutes	**1.18**	**11.2***	---

Conclusion: At 1 minute the mean heart rate (*M* = 89.25) is greater than both baseline (*M* = 76.912) and at 2 minutes (*M* = 78.083). Thus, the heart rate data don't follow the same pattern as the blood pressure data.

Even though this is a repeated measures design, no counterbalancing is possible. Briefly explain why not.

Because the study is looking at the effects of time, counterbalancing wouldn't make sense. That is, you couldn't have the 2-minute condition come before the baseline condition.

3. As you know, when designing a repeated measures experiment, or an experiment with a repeated measures factor, one must counterbalance. Why? In other words, if one did not counterbalance, why would the experiment be confounded? Be *very* explicit! An example might help. [5 pts.]

You must counterbalance to avoid the confound that would be introduced if you had order or carry-over effects. For example, suppose that there is a practice effect (increased performance over time). In a simple repeated measures experiment with two conditions, that would mean that performance on the second task would improve (e.g., +5). Thus, without counterbalancing, every subject would get Treatment 1 first and Treatment 2 second. Thus, it might look as though Treatment 2 is better than Treatment 1:

Subject	Treatment 1	Treatment 2
1	X1 + 0	X1 + 5
2	X2 + 0	X2 + 5
3	X3 + 0	X3 + 5
4	X4 + 0	X4 + 5

However, with appropriate counterbalancing (*S*s 1 and 2 get T1->T2, but *S*s 3 and 4 get T2->T1), there would be no advantage for the second treatment, so any difference found would be due to an actual difference between the treatments:

Subject	Treatment 1	Treatment 2
1	**X1 + 0**	**X1 + 5**
2	**X2 + 0**	**X2 + 5**
3	**X3 + 5**	**X3 + 0**
4	**X4 + 5**	**X4 + 0**

4. Briefly define Type I Error, Type II Error, and power. Then, tell me at least three ways you can to increase power. [5 pts]

Type I Error is incorrectly rejecting H_0. Type II is incorrectly retaining H_0. Power is correctly rejecting H_0. You can increase power by:
• increasing the sample size (n)
• increasing the treatment effect (e.g., giving dosages that differ to a greater extent)
• decreasing individual differences (e.g., using only subjects with IQ above 120)
• decreasing random variability (e.g., ensuring that the experimental setting is optimal and identical for all)

5. Two researchers were interested in studying the effects of reward magnitude on performance. Both researchers used introductory psychology students as participants, the same total number of participants (21), the same type of reward and reward magnitudes ($1, $5, $20), the same apparatus, the same task, and the same performance measure (DV). One researcher used an independent groups design and, on the basis of the results, cannot reject the null hypothesis (that reward has no effect on performance). The other researcher used a repeated measures design and found a statistically significant effect of reward magnitude — larger rewards lead to better performance. Assume that neither study has a major flaw (e.g., repeated measures design is properly counterbalanced, random assignment to conditions). There are two fundamental reasons why the two researchers might have reached different conclusions. One reason concerns the sensitivity of the test of the null hypothesis. The other reason concerns the nature of the participant's experience in the two studies. Complete the source tables for the two experimenters seen below. Then provide a clear explanation of the two reasons for the different results that the two researchers obtained.

Independent Groups Design ($F_{Crit} = 3.55$):

Source	*SS*	*df*	*MS*	*F*
Treatment	28	**2**	**14**	**3.5**
Error	**72**	**18**	**4**	
Total	100	**20**		

Repeated Measures Design ($F_{Crit} = 3.23$):

Source	*SS*	*df*	*MS*	*F*
Treatment	20	**2**	**10**	**5**
Within	**180**	**60**		
Subject	100	**20**		
Error (Subj x Treat)	80	**40**	**2**	
Total	200	**62**		

1. The repeated measures design is more powerful than the independent groups design. As a result, it should come as no surprise that a repeated measures design would yield a significant result when a similar independent groups design does not. Note, for example, that with the same number of subjects (*n* = 21), the independent groups design would yield 21 pieces of data, while the repeated measures design would yield 63 pieces of data. Another change that results from the greater amount of data is the F_{Crit} is lower for the repeated measures design.

2. Another, less satisfying, explanation for the difference is that shit happens. That is, in one case we might have a correct decision and in the other case, we have an error (either Type I, for the repeated measures design, or Type II, for the independent groups design).

3. However, it may also be important to consider the subject's experience. In the independent groups design, a person is told about only the one reward that he or she will receive in the study. On the other hand, people in the repeated measures design are first given one reward level (e.g., $20) and then, after completing the first trial, given a second reward level (e.g., $5) for the second trial, and a third reward level (e.g., $1) for the third trial. Under these conditions, a subject may either intuit that the reward level is important, and modify her or his responses to be consistent with that expectation. It may also be that a subject decides that after being paid a lot (e.g., $20), there's no reason to work as hard for a much smaller reward (e.g., $1). In the end, it may make more sense to run such a study as an independent groups design, even though it would have less power.