1. Define what we mean by the term *operational definition* and then cite a specific example of an operational definition from a study discussed in the Mook article. [5 pts]

In general, an operational definition is a clear and concrete statement of the process involved in the definition of a term. As Ray says, "…it redefines the (general) concept in terms of clearly observable operations that anyone can see and repeat. (p. 55)" To measure anxiety, Ray suggests that we might define anxiety from self-report measures, physiological measures, or observation of a specific behavior (hand wringing? ☺). "…we might define aggression as the number of times a child hits a toy… (Ray, p. 55)"

In the Mook article, Milgram operationally defined obedience as the participant's willingness to administer shocks to the "learner" as per the experimenter's instructions. You could use any of the other studies in the Mook article to answer the question.

2. Define both internal and external validity. Which type of validity is more important? Why? Using the Higgins and Marlatt study (drinking alcohol as a means of reducing tension) *and* the Argyle study (glasses and intelligence), tell me what Mook has to say about the importance of external validity. [15 pts]

Internal validity refers to the logical construction of the experiment. In essence, a study is internally valid if it is free from confounds, which would mean that the results of the study are not open to multiple interpretations. External validity refers to the extent to which the setting of the study is consistent the world outside (the "real world" ☺). In essence, a study is externally valid if one could readily generalize the results to a wide range of settings.

3. Hess has done work that suggests that pupil size is affected by emotional arousal. In order to ascertain if the type of arousal makes a difference, Dr. Bo Ring decides to measure pupil size under three different arousal conditions (Neutral, Pleasant, Aversive). Participants look at pictures that differ according to the condition (neutral = plain brick building, pleasant = young man and woman sharing an ice cream cone, aversive = graphic photograph of an automobile accident). The pupil size after viewing each photograph is measured in millimeters. The data are seen below:

Neutral	Pleasant	Aversive
4	8	3
3	7	7
2	7	3
3	8	4
3	6	7
5	7	4
4	8	4
3	7	6
5	6	5
2	7	6

a. Due to miscommunication between Dr. Ring and her research assistant Igor, she has no idea if the study was run as an independent groups design or as a repeated measures design. Igor has left the lab in a cloud of controversy amidst a whole range of allegations (involving improper use of lab alcohol, monkeys, and videotapes), but that's another story entirely. With no immediate hope of clarifying the way in which the data were collected, Dr. Ring decides that the most reasonable strategy would be to analyze the data as an independent groups design. Her logic should make sense to you, so explain to me why she would be smart to analyze the data as though they were collected in an independent groups design. [5 pts]

The repeated measures design is more powerful than the independent groups design. Thus, if you were unsure of the design, it would make more sense to analyze the data with an independent groups ANOVA. If the results were significant using the independent groups ANOVA, you would reasonably expect them to be significant with a repeated measures ANOVA. You should realize that in any case you couldn't do anything public with these data. It would be entirely too embarrassing to submit a paper for publication in which you admitted that you didn't know if the data came from a repeated measures design or an independent groups design.

b. OK, Dr. Ring analyzes the data using StatView, but she's a notorious slob and has dropped donut filling over portions of the source table. But she's not just a slob, she's a frugal slob. You've just been hired to replace her research assistant (so stay away from the alcohol, monkeys, etc.). She knows how sophisticated you are, so she wants you to figure out the ANOVA source table without running the data through the computer again. Of course, you can easily do so with the information remaining in the source table, so do so now. Then, to truly ingratiate yourself to your new employer, analyze the data as completely as you can. Impress the hell out of her! [20 pts]

ANOVA Table for Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Group	2	69.267	34.633	25.831	<.0001	51.663	1.000
Residual	27	36.200	1.341				

Means Table for Score
Effect: Group

	Count	Mean	Std. Dev.	Std. Err.
Aversive	10	4.900	1.524	.482
Neutral	10	3.400	1.075	.340
Pleasant	10	7.100	.738	.233

Because the P-value is < .05 (lots!), you would reject H_0. The next step would be to determine which of the groups differed. That involves computing the Tukey test. With q = 3.5 (3 treatments and 27 df_{Error}), HSD = 1.28. Thus, if two means differed by 1.28 or more, they would be considered significantly different. I would interpret the results of this study as indicating that people had significantly larger pupils when viewing pleasant stimuli compared to neutral or aversive stimuli. People also had larger pupils when viewing aversive stimuli compared to neutral stimuli.

c. As you noodle around the lab, having completed the analyses that Dr. Ring wanted, you stumble across the original data from the study and realize that Igor had actually run the study as a repeated measures design. This revelation strikes you as a combination of good news and bad news. What's the typical good news about running a repeated measures design? In this particular case, what might strike you as bad news in the design of this repeated measures study? [5 pts]

The good news (aside from being able to compute the appropriate analysis) is that the repeated measures design is more powerful than the independent groups design. The bad news is that only 10 participants were used. With 3 treatment levels, the study should have involved complete counterbalancing, which means that there should have been a multiple of 6 participants (6, 12, 18, etc.), and not 10!

d. Now that you know that the data were collected as a repeated measures design, you re-compute the analyses appropriately, as seen below. [Notice that the means, etc., aren't displayed below, because they would not change from the earlier analysis, right? Note, also, that other parts of the source table don't change as you move to a repeated measures analysis.] Once again, analyze these data as completely as you can. Given what you know about the relationship between independent groups analyses and repeated measures analyses, the results might surprise you somewhat. However, you can readily explain the anomaly, right? [10 pts]

ANOVA Table for Stimulus Type

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	5.467	.607				
Category for Stimulus Type	2	69.267	34.633	20.284	<.0001	40.568	1.000
Category for Stimulus Type * Subject	18	30.733	1.707				

With a significant result, the next step would be to compute HSD. With q = 3.61 (3 treatments and df_{Error} = 18), HSD = 1.49. Thus, the interpretation would not change when analyzing the results as a repeated measures design. People had significantly larger pupils when viewing pleasant stimuli compared to neutral or aversive stimuli. People also had larger pupils when viewing aversive stimuli compared to neutral stimuli.

You should expect to have a larger F-ratio for this analysis, but you don't (F = 20.2 compared to the F = 25.8 for the independent groups ANOVA). Huh? How did that happen? The reason is that you've lost more df in your error term than you've gained back by reducing the SS_{Error}. In essence, there is too little variability among your participants (relatively few individual differences). Only in such circumstances would you find a smaller F-ratio when computing a repeated measures ANOVA. Of course, there's nothing you can do about the result. You can't compute the inappropriate independent groups ANOVA just because it will give you a larger F-ratio.

4. Suppose that you analyze the results of an independent groups experiment and find that the resulting F has a p-value of .09. After the appropriate cursing, what might you consider doing? (Looking for a new career is not an option.) You should be able to briefly list a number of options available to you. [5 pts]

Your problem, of course, is one of power. How could you gain more power in your study? Several options should make sense to you. First of all, you could re-run your experiment with more participants. You could also redesign the study so that your treatment effect was larger (give a larger dosage, etc.) or the random variability was smaller (clearer instructions, etc.). You could also consider doing the study as a repeated measures design, if the experiment lends itself to such a design.

5. Hunt and his colleagues (1975) have conducted several studies examining the relationship between intelligence and the speed of basic mental processes. In a typical experiment, Hunt

measured reaction time for a simple mental task (e.g., determining whether two letters are the same or different) and then computed the correlation between the participants' reaction times and their IQ scores. Hypothetical data representing the results of one experiment are analyzed below. Interpret the results as completely as you can. If you knew that a person had an IQ of 110, what would you predict that person's reaction time to be? [5 pts]

Count:	R:	R-squared:	Adj. R-squared:	RMS Residual:
6	.334	.111	-.111	14.671

Analysis of Variance Table

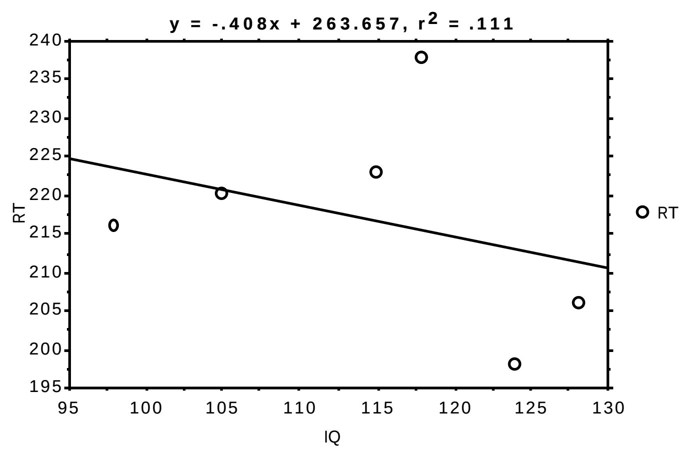
Source	DF:	Sum Squares:	Mean Square:	F-test:
REGRESSION	1	107.938	107.938	.502
RESIDUAL	4	860.895	215.224	p = .5179
TOTAL	5	968.833		

Beta Coefficient Table

Variable:	Coefficient:	Std. Err.:	Std. Coeff.:	t-Value:	Probability:
INTERCEPT	263.657				
SLOPE	-.408	.577	-.334	.708	.5179

Confidence Intervals Table

Variable:	95% Lower:	95% Upper:	90% Lower:	90% Upper:
MEAN (X,Y)	200.205	233.462	204.065	229.601
SLOPE	-2.009	1.193	-1.638	.821



First of all, you should note that the results are not significant (p = .5179 > .05). Thus, the r = .334 is not significant, so there doesn't appear to be a linear relationship between IQ and RT in this particular (small) set of data. [Again, potentially a power problem.] Of course, in the absence of a significant linear relationship, you cannot use the regression equation to make predictions. Thus, regardless of a person's IQ, your best bet of RT would be the mean RT overall.