Example questions:

1) The *New York Times* recently reported that researchers reported finding a gene that determines language ability in humans. Critics, however, state that this gene determines much more than language ability. What specific type of validity concern are the critics raising?

2) What is the simple equation that serves as the cornerstone for 'classical test theory?'

3) A scale with 26 items has an alpha score of .95, with an average inter-item correlation of .20. What would be the alpha score for the scale if consisted of only 4 items, if they had the same average inter-item correlation of .20?

4) Explain how the alpha coefficient is related to split-half reliability

5) When examining an ROC curve, what statistic provides an overall score for the criterion validity of a scale?

6) What are the four steps to evaluate a multi-trait multi-method matrix?

7) Suppose a treatment procedure for people with high exposure to airborne Anthrax exists that is expensive and has high risk for negative side-effects. To help decide who should receive the treatment would you want a test with high sensitivity or high specificity?

8) If a scale has a reliability of .7, what is the theoretically highest correlation observable between it and a gold standard that contains no measurement error?

Question 1: Reliability. **Limit your answers to the space provided below.**

a) Name a reliability measure that is not sensitive to mean differences across two variables. In other words, which reliability measure would give you the same answer with and without a constant value added to all observations of one rater?

b) If more of the variance in an observed measure is due to error than is due to variance in the true score, what, specifically, will we know about the upper and lower bounds of the reliability coefficient?

c) What is the theoretical maximum correlation that can be observed between two variables if one has a reliability of .5 and the other has a reliability of .6?

Question 2: Validity/Factor Analysis. **Limit your answers to the space provided below.**

Questions a-b below refer to the article "A New Psychosocial Screening Instrument for Use with Cancer Patients" by Zabora et al. in the 2001 volume of *Psychosomatics*. This article was distributed in class during Dr. Zabora's talk; it is also available on-line through the Hopkins library system at http://psy.psychiatryonline.org/cgi/content/full/42/3/241.

a) In the article he presented in class, Zabora reported that the eigenvalue for the fourth factor in his principal component factor analysis was .98. What is the interpretation of a eigenvalue less than 1? What is the interpretation of an eigenvalue of exactly one?

d) Following is a multi-trait multi-method matrix from the article "Disability and Depression in Rheumatoid Arthritis; A Multi-Trait Multi-Method Investigation" by Peck et al. in *Arthritis and Rheumatism* (1989) 32:1100-1106. The sample for the project consisted of 107 married patients with complete information on (a) self-reported measures of disability and depression, (b) spouse-reported measures of disability, and (c) interviewer ratings of depression and disability.

	Disability			Depression
	Self-report	Observer	Spouse	Self-report
Disability				
Self-report				
Observer	.89			
Spouse	.87	.85		
Depression				
Self-report	.37	.31	.50	
Observer	.24	.17	.25	.69

1) What is the first step in the evaulation of a multi-trait multi-method matrix? Does this table provide the necessary information for this step?

2) What is the second step in the evaulation of a multi-trait multi-method matrix? What are the correlation values in the table that are pertinent to this second step?

3) What is the third step in the evaluation of a multi-trait multi-method matrix? What are the correlation values in the table that are pertinent to this third step?

4) What is the fourth step in the evaulation of a multi-trait multi-method matrix? What are the correlation values in the table that are pertinent to this fourth step?

5) Does this correlation matrix suggest that the convergent and discriminant validity of these measures is good, poor, or can it not determined from the information provided?

Question 3: Latent Class Analysis

The following is a latent class analysis applied to 14 symptoms used for diagnosing schizophrenia by the DSM-IV.

The authors were interested in describing schizophrenia in a more comprehensive manner, instead of just a simple "yes/no" diagnosis. A latent class analysis was applied to a clinical trial in schizophrenic patients, where there were two treatment groups: placebo and a new treatment. Note that each individual in this sample has been screened into the study by having a history of schizophrenia. A latent class analysis was performed on 14 symptoms, where individuals reported symptoms: 1 = symptom present, 0 = symptom absent. The symptoms of interest can be divided into two types: positive and negative, denoted by P and N in the first column of the table below. "Positive" symptoms refer to symptoms which add something to an individual's personality. For example, most people do not have hallucinations. As such, hallucinatory behavior "adds" something to an individual's behavior. Conversely, "negative" symptoms refer to symptoms which remove some aspect of normal personality. As an example, spontaneity is something that most people have (at least to some extent). But, in some schizophrenic individuals, the characteristic of "spontaneity" has been removed from their behavior.

The fitted four class model estimates are below, which include the estimated class sizes and symptom probabilities for each class. The sample size of the data set was 813.

		Class 1	Class 2	Class 3	Class 4
P1	Delusions	0.28	0.47	0.93	0.97
P2	Conceptual Disorganization	0.10	0.39	0.75	0.93
P3	Hallucinatory Behavior	0.22	0.41	0.71	0.88
P4	Excitement	0.12	0.07	0.51	0.55
P5	Grandiosity	0.12	0.07	0.45	0.53
P6	Suspiciousness/persecution	0.19	0.42	0.81	0.87
P7	Hostility	0.03	0.13	0.36	0.48
N1	Blunted Affect	0.46	0.97	0.52	0.94
N2	Emotional Withdrawal	0.25	0.96	0.38	0.98
N3	Poor rapport	0.08	0.68	0.22	0.87
N4	Positive/Apathetic social withdraw	0.31	0.95	0.50	0.94
N5	Difficulty with abstract thinking	0.40	0.72	0.74	0.91
N6	lack of spontaneity	0.22	0.84	0.22	0.82
N7	Stereotyped thinking	0.13	0.33	0.74	0.90
	Class Size	0.11	0.18	0.26	0.45

Question 3 (continued)

a) Interpret the four class model in two to four sentences. Try to give a "psychiatric" as opposed to "statistical" interpretation of the classes.

b) The table below contains the -2Log-Likelihood statistics for the 2, 3, and 4 class models fit to the data set described above.

Model	-2Log-Likelihood
2 class model	9,479.73
3 class model	8,091.23
4 class model	7,742.31

Using these fit statistics (and any others that you may choose to evaluate), the background information provided, and assuming that all three models are well-identified, which model would you say is the most appropriate? Justify your answer in no more than 4 sentences using statistical and/or clinical or scientific rationale.

A researcher was referred to you because he heard that you have some expertise in factor analysis and are familiar with using Stata. He has some questions about a factor analysis that he is working on. He ran the ``factor'' command in Stata and got following result:

```
. factor  x1-x15, factor(4)

                Factor Loadings
    Variable |      1          2          3          4    Uniqueness
-------------+--------------------------------------------------------
          x1 |   0.48866   -0.34935    0.37694   -0.02681    0.34423
          x2 |  -0.57345    0.15749    0.10179    0.06215    0.41294
          x3 |  -0.26034    0.08658    0.46032    0.02642    0.90324
          x4 |   0.27456    0.44785    0.03550   -0.02083    0.88252
          x5 |   0.63653    0.34373    0.15758   -0.01865    0.45932
          x6 |   0.46811   -0.17559   -0.02744    0.02423    0.35424
          x7 |   0.57221   -0.01815   -0.05064    0.07975    0.45491
          x8 |   0.38532   -0.34935    0.16347   -0.25645    0.34555
          x9 |  -0.39523    0.15235    0.03742    0.42463    0.44213
         x10 |  -0.42365    0.34260    0.09367    0.17434    0.50012
         x11 |   0.51245    0.13563    0.37246   -0.24645    0.40134
         x12 |   0.34556    0.68734    0.14745   -0.09435    0.35981
         x13 |   0.30965   -0.35732   -0.47346    0.52466    0.31563
         x14 |   0.21463   -0.01815   -0.02574    0.03456    0.91135
         x15 |   0.46345   -0.36346   -0.27542    0.01445    0.39634
```

A. He was pleased that the eigenvalues (not shown here) suggested that there are four factors in the data. However, he was disappointed with the results in the factor loading matrix shown above. He thinks that making an interpretation is difficult based on the factor loadings produced from the factor analysis above. Specifically, he sees that all the variables seem to load on the first factor.
Explain to him why his results from the factor command are uninterpretable and how he can try to make them more interpretable.

B. Realizing that this researcher lacks a proper understanding of factor analysis, you generously offer to help him by performing the factor analysis yourself. In doing so, you realize that of the 15 items in the analysis, three of them have "uniquenesses" of approximately 0.90, while the other items have uniquenesses of between 0.30 and 0.50.
Describe in words to the researcher what this means and what you would suggest for proceeding with the factor analysis in regards to the uniquenesses. Note that this researcher probably does NOT know what uniqueness means!

C. What do the high uniquenesses of three items imply about the reliability and/or validity of those items in regards to the latent construct?

D. Due to your eagerness to evaluate the results of the factor analysis, you neglected to do exploratory data analysis. Now, you have gone back to the dataset and realize that all of the 15 variables are binary.
Describe why this concerns you.

E. Now you decide that you should meet with the investigator again to talk about the appropriateness of factor analysis for his goals. In the course of your conversation, you introduce him to latent class analysis and he is very interested. Now he thinks that maybe he would prefer using latent class analysis to analyze the data instead. Assume that you have two datasets available to you: a 15 item dataset with the researcher's items in BINARY form, and another 15 item dataset with the items in CONTINUOUS form.

1. Given ONE example of a situation (i.e. a research goal) for which factor analysis would be a better approach than latent class analysis.

2. Given ONE example of a situation (i.e. a research goal) for which latent class analysis would be a better approach than factor analysis.

F. You've just left the researcher's office after agreeing to do both analyses (factor analysis and latent class analysis). Suddenly you realize that the sample size in the dataset is 152.

1. Why might the sample size worry you for the proposed analyses?

2. What characteristics of the dataset will affect needs for larger versus smaller sample sizes in factor analysis?

3. What characteristics of the dataset will affect needs for larger versus smaller sample sizes in latent class analysis?

Question 2 (continued). Varimax rotated factor loadings for two scales of psychological maltreatment of women by their male partners (adapted from Tolman, Richard. 1989. "The Development of a Measure of Psychological Maltreatment of Women by Their Male Partners." *Violence and Victims* 4:159-177)

Women were asked the following questions as they appear below. **For men, the pronouns and direction of abuse were altered so that the responses referred to their actions toward their partners** (e.g. women were asked to respond to the statement "My partner restricted my use of the telephone" while men were asked to respond to "I restricted my partner's use of the telephone")

	Female	Male
	1	2
Dominance-Isolation		
My partner monitored my time and made me account for where I was	.75	.06
My partner tried to keep me from doing things to help myself	.75	.29
My partner did not allow me to go out of the house when I wanted to go	.74	.14
My partner tried to keep me from seeing or talking to my family	.70	.16
My partner interfered in my relationships with other family members	.70	.17
My partner did not allow me to socialize with my female friends	.69	.14
My partner did not want me to go to school or other self-improvement activities	.68	.18
My partner restricted my use of the telephone	.67	.18
My partner was jealous or suspicious of my friends	.67	.13
My partner acted like I was his personal servant	.65	.35
Emotional-Verbal		
My partner blamed me for his problems	.24	.71
My partner blamed when he was upset about something	.36	.71
My partner blamed for causing his violent behavior	.27	.69
My partner said something to spite me	.19	.67
My partner did not let me talk about my feelings	.18	.65
My partner's moods changed radically, from calm to angry, or vice-versa	.29	.64
My partner treated me like an inferior	.36	.63
My partner told me my feelings were crazy or irrational	.34	.63
My partner withheld affection from me	-.03	.62
My partner tried to make me feel like I was crazy	.40	.61

Question 1: Reliability

a) Explain how to compute a split-half reliability for a scale.

b) Explain how the split-half reliability is related to the alpha coefficient. (Please do not give us any mathematical formulas in your answer – simply tell us how they are related in conceptual terms).

c) The authors report that the alpha coefficient for an expanded dominance-isolation scale that consists of 26 items is .95. If the average inter-item correlation of the scale is .20, what would be the alpha reliability of the scale if it consisted of only 4 items, if they had the same average inter-item correlation of .20?

d) The developers of the scale decide that 26 items require too much time on the part of the respondents and desire to trim it. With an average inter-item correlation of .20, how many items would they need to retain in order to have an alpha coefficient of at least .70? (We will provide a window of + or - 1 for your answer; e.g., if the answer is 27, we will give credit to people who answer 26, 27, or 28).

e) If the trimmed 10 item scale has a reliability of .70, what is the theoretically highest correlation we can observe between the scale and a variable that contains no measurement error?

f) If the items have marked floor or ceiling effects, what mathematical assumption does this violate? Briefly explain and provide an example(s) of how this violation either leads to inflated or attenuated reliability estimates (or both).

Question 2: Validity/Factor Analysis

Use the results of the factor analysis on the following page to answer the following questions

a) Which type of validity would be most appropriate to assess the validity of Dominance-Isolation and Emotional-Verbal scales: criterion validity, construct validity, or content validity? Why? Explain why the answer you choose is more appropriate than the other two types of validity. (In your answer be sure to include a sentences or phrases that clearly demonstrate that you understand each type of validity)

b) Factor analysis is most informative for what two types of validity?

d) Do these scales take acquiescence bias into account? If so, how; if not, how could they?

e) Note that these two factor analyses have been rotated using a varimax rotation. What does this imply about the relationship of factor 1 and factor 2 in each of the analyses? How is an oblique rotation different and why might we prefer it in some cases?

Question 3: Latent Class Analysis

The following latent class analysis has been adapted from Kendler, Karkowski, and Walsh (1998) *Archives of General Psychiatry*.

The authors were interested in the nosologic structure of psychotic illness. A latent class analysis was applied to detailed symptomatic and outcome assessments of 6502 individuals with broadly defined schizophrenia and affective illness ascertained from a population-based psychiatric registry in Roscommon County, Ireland. Note that each individual in this sample has been diagnosed as having a mental disorder. A latent class analysis was performed on 9 symptom groups (see table) and a 4 class model was chosen as most appropriate. The estimated class sizes and symptom probabilities for each class are shown in the table below.

	Class 1 Classic Schizophrenia	Class 2 Major Depression	Class 3 Schizophreniform Disorder	Class 4 Schizodepression
Affective Symptoms Predominate	0.30	0.93	0.18	0.14
Reduced Affect	0.95	0.13	0.64	0.95
Persecutory Delusions	0.63	0.09	0.64	0.92
Auditory Hallucinations	0.64	0.04	0.51	0.89
Deterioration from premorbid condition	0.98	0.02	0.10	0.91
Elevated mood	0.01	0.02	0.35	0.01
Dysphoria	0.07	0.95	0.19	0.94
Psychomotor Change	0.02	0.70	0.09	0.76
Fatigue	0.06	0.84	0.03	0.65
Class Size	0.35	0.40	0.20	0.05

Question 3 (continued)

a) Choose one of the four classes identified in the model. Describe the class relative to the other classes in terms of size/prevalence and symptom patterns/profile. Repeat this for one other class.

b) What if I now tell you that the sample size is NOT actually 6502, but is really only 343. What concerns do you have about interpreting the model given the sample size? Why is it important to have a large sample in latent class analysis and how can we check to see that there is enough data to estimate all of the parameters?

1) An investigator is developing a scale on stigma experienced by drug users, not having found a similar scale in the literature. She wishes to examine the internal consistency of the scale, which has 40 questions. She correlates question 1 through 20 with questions 21 through 40, and gets a r = .67.

a) From the information available, what method of testing internal consistency can you use?

b) Calculate the internal consistency using this method.

c) What alternative methods to calculate internal consistency could the investigator use?

d) Interpret the value that you calculated in part b), meaning, how good is the internal consistency?

e) If the internal consistency was poor, what are things you might consider doing to improve it?