**PSY 600K Final Exam:**

This take-home exam consists of 5 essay questions, of which you should answer 4. Each response is worth up to 20 points, for a total of 80. In general, to get full credit, you must:

* Answer all parts of the question.
* Provide accurate interpretations and explanations of measurement concepts.
* Write and explain clearly, in language that is appropriate for an intelligent layperson (e.g., a PhD in another field). Please use complete sentences, proofread, etc. – I will overlook minor typos but sloppy writing makes it difficult for me to tell exactly what you mean.
* The deeper and more detailed your explanations, the better I can assess your understanding.
* Where the question asks you to address a specific research or practice area, make clear how the measurement concepts you are discussing apply in *that particular context*, not just in general. At the same time, the focus of your answer should be on the measurement concepts; give me enough background about the research/practice area to understand your answer, but it shouldn’t be the bulk of your answer.
* Support your arguments with citations, with an emphasis on primary sources (textbooks/primary readings) and not a sole reliance on lecture notes.

On average, I believe you should be able to answer each of these questions well in a few paragraphs (about a page or so but not more than two). Overly brief answers may not provide sufficient explanation and justification to earn full credit; overly long answers may provide more than is necessary and take time away from other things you could be doing.

**You may take advantage of your lecture notes, textbooks, readings, and any other class materials on this exam. If you choose to incorporate outside sources, please provide a reference list for any non-class material. Please remember that this is an individual exam and you should not discuss it with anyone other than me until all parties have turned it in**.

**This exam is due on Thursday, May 10th, at 11:59 PM in Canvas.**

**Background information of research:**

*My research is focused on the evaluation of an at-risk adolescent mentorship program called Campus Connections. We work with adolescents aged 11-18 years old, their paired mentor an adult aged 18 or older attending CSU, and the parents/caretakers of the adolescent. The college students and adolescent mentees work with each other once a week for 12 weeks throughout the CSU semester. We distribute a wide variety of surveys containing various measures for parents, adolescent mentees, and college mentors. Surveys are distributed before the start of the program (pre-intervention) Week 3, week 6, week 9, and the final week of intervention. More information about the program can be read here:* [*http://www.hdfs.chhs.colostate.edu/students/undergraduate/campusconnections/*](http://www.hdfs.chhs.colostate.edu/students/undergraduate/campusconnections/)

1. *Define a construct of your choice and then describe a nomological net, consisting of at least five other constructs or variables, around that construct. Explain where each construct falls relative to your construct on the continuum from convergent to discriminant validity.*

A construct I am always interested in, especially in the context of Campus Connections, is how much mentees feel a sense of *belongingness* to the program. This construct attempts to measure how inclusive one feels with the group. Items for this construct include:

1. I belong to this program
2. People would notice if I were not here
3. I feel like I am a part of this program
4. People value me here

There are many variables that may relate to the construct of belongingness’ nomological net. To start, let’s define what the nomological net is. The nomological net is the ‘web’ of variables and other constructs that relate (or don’t relate) to the construct of interest (Lecture 3; slide 7). Additionally, as R&M state (pg 4), the construct is defined by a certain set of behaviors, in other words, this helps to characterize the nomological net.

Now, to correctly define the nomological net for my belongingness construct, I need to proceed and process where our construct fits by assessing which variables and other constructs relate to my belongingness construct. Below is a list of 5 variables or constructs and how they fit into the nomological net.

1. **Loneliness:** I would expect loneliness and belongingness to have a high correlation with each other – a high *negative* correlation. This would help to establish *convergent validity.* For example, one example of a loneliness scale is the UCLA loneliness scale (Russell, Peplau & Ferguson, 1978). I would hope the correlation was very high in this, because I expect belongingness and loneliness to be very related. However, I do not Want this correlation to be too high of a correlation, or else there is no teasing out a distinction high loneliness and low belongingness or vice versa, which would make the scale pointless.
2. Next, my background research may indicate that desire for acceptance would relate to my construct of interest. As one desire for acceptance would likely relate to their ability to facilitate belongingness. Therefore, although it would be related, I would not expect the correlation to be as high as the loneliness correlation.
3. Next, hostility is another variable construct I thought of when defining my belongingness scale. It seems likely that belongingness and hostility may correlate negatively with each other. However, these are not the most related aspects in the world. This is the point in which I would expect a lower correlation to the belongingness, indicating that it is a but farther on the border of the nomological net – However, I feel there is reason to believe that hostility is in relation to the construct at hand. The reasoning behind this is that it is an expectation that youth that are more hostile have more chance of being outcasted. Therefore, bringing their ability to be in a group to go down, thus, belongingness to correlate somewhat with hostility.
4. Next, I would like to indicate evidence of discriminant validity. Here, I am looking for measures that would likely reveal a low relationship with my belongingness construct. As Devellis (pg. 69) notes, discriminant validity is the *absence* of correlation between two constructs. A measure I have chosen is extroversion from the personality scale. This isn’t necessarily related. The relationship it should have is indicated by either random error or the very looser relationship it holds on the nomological net.
5. Lastly, one more variable to help indicate discriminant validity. I have also chosen trait anxiety. For this measure, I really don’t expect to see much of a relationship at all. This falls the lowest in the relationship with belongingness. Although lower belongingness might cause anxiety, an individuals trait anxiety is a measure that is stable, regardless of an individuals belongingness to a group. Therefore, this would cause the most discriminant evidence to help define my construct and where and the relationship of variables within its nomological net
6. *Identify one important social consequence of testing that is relevant to your area of research or practice. This should be a consequence of using tests in the kind of work that you do, not a consequence of your work in general. Explain your consequence and then discuss, based on what you have learned from class, how you could either assess or mitigate this consequence, or both.*

At Campus Connections, youth are often referred to the program by the juvenile corrections department as a deferment action that is an alternative to being placed in juvenile hall. Often, the program is expected to impose more positive outcomes among the participants on many measures. The purpose of the program is to improve the outcomes and we use measures to predict these. However, if the measures do not match the construct accurately, then the consequences of the measurement may be not correctly mitigated. These unassessed mitigations may have consequences toward everyone in society as the adolescent’s transition into adulthood and out go out into the real wordl. The precision of the scores on these measures helps to assess the outcomes. This relates heavily to predictive bias (lecture 17; slide 5). It is important that the tests that are being assessed are being interpreted correctly. It should also be noted that this is as at-risk population. So, although some measures may correspond directly to regular populations. It is possible that respondents in an at-risk category may not have comparable responses. Therefore, the fairness is not appropriate to our group and should be taken into consideration. In other words, there may be systematic bias within our testing population.

There are a few ways we are assessing for these consequences For one, looking into long term outcomes. By assessing the long-term outcomes, we can see how well the tests are in assessing the construct at hand. Additionally, we are mitigating this effect by using measures that have specifically been used on at-risk adolescent populations. However, this is not a guaranteed fix. If possible, it would be ideal to have a control group to understand this concept and see how less at-risk population fairs in comparison.

1. *In your line of work, are you more concerned about faking or cheating? Explain why, in a way that demonstrates that you understand the difference. Discuss two specific strategies you could use (again based on what you have learned from class) to address your concern, and compare the pros and cons of those strategies.*

Within my work, I am mostly concerned with faking. Specifically, as Schmitt & Oswald (2006) mention, the problem I see with my research is, “Fake Good” in which the youth are likely to indicate items that would likely to make them look better. However, I will say, in this context, unlike the Schmitt & Oswald (2006) article, we will not be ranking individuals for job outcomes, in which this will not have the same consequential actions. However, understanding youth, and their interest of social desirability, the effects will be skewed from this, even though responses are anonymous. So, two strategies have been imposed. The strategies are listed below, with the pros and cons explained:

1. *Enact a Social Desirability scale for the youth*
   1. As shown by Griffith et al., we can develop a social desirability scale. This would consist of items that would identify fakers. This measure consists of items that are (nearly) impossible to completely agree with. For example, from the Crown-Marlow social desirability scale: “My table manners at home are as good as when I eat out in a restaurant” in which those that are faking have a tendency to strongly agree.
      1. A *pro* of this is that we can indicate individuals that are faking. We can, at some, point make decisions to remove these individuals from the study. Or even control for their social desirability in order to account for the bias. Lastly, we can model the desirability directly into our model. The important thing here is that it is measured, and we can account for it.
      2. Unfortunately, this method comes with many more *cons*. For one, if we drop participants we are immediately losing power. Next, by controlling fro Socila desirability has the ability to actually reduce the criterion relationship between our variables of interested (piedmont et al., 2000). Therefore, when controlling for social desirability with self-reported delinquency, we will be accounting for variance that is literally associated with the outcome. For example, if we wanted to look at its relationship to crime records in the future as the criterion variable. Additionally, Peterson et al. (2011), also indicate that most social desirability tests did a bad job at indicating actual faking, so, unless a good scale is truly developed (which would be hard).
2. Obtaining actual grade reports and actual delinquency, remove the fakable measures entirely
   1. It is important that we avoid this all together. Or, we can even use the data we know is true (well, at least not faked) and compare it to the self-reported information. For example, if we obtain actual grade reports, and have the youth indicate how well they are doing in school, we can compare the two. In this event, the grade reports could serve as the criterion variable to understand the actual grades.
      1. A *pro* of this is that using the comparison, we are able to get insights as to how much individuals are likely to fake through a verifiable report (Piedmont et al.). Furthermore, this is much more verifiable evidence of the construct at hand (Grades would serve as an indicator of the construct *academic success*).
      2. The number of *cons* are much less. However, I would argue that 1.) There could be harder to obtain then getting actual self-report. And 2.) there is a certain amount of responsibility associated with obtaining grade reports and delinquency records. This is information that is personal and can be considered higher security. Therefore, it is both harder to obtain and is a trade-off.
3. *Define a situation in which you might plausibly be interested in comparing latent means for a measure across two groups (e.g., “Are Xs higher in Q than Ys?”). Explain what you would do to test measurement equivalence in this situation, and what evidence you would need to find at each step in order to reach the conclusion that you can appropriately compare those mean scores.*

Personally, I am interested in gender and how the underlying variables of measures of delinquency differ across the groups. I would personally hypothesize that adolescent boys are on a different part of the continuum for delinquency measures. Therefore, to do this, I would focus on performing a Multiple Group Confirmatory Factor Analysis (MGCFA). For all portions talking about MGCFA, I will be referencing Vandenberg and Lance (2000). Performing and MGCFA is similar to that of running a regular factor analysis, however, we will be comparing the two model, but constraining all of the parameters we are assessing across the two groups to assess the model fit. If the model falls apart in one of our group CFA’s, there is evidence that our group’s latent means are not equal across groups. Thus, indicating a group difference on the construct.

In order to test this we will need to run through our order of operations using the observed data and constraining free and constrained parameters. Therefore, if we were looking at a measure of delinquency across groups. We would constrain all of the factor loadings across both groups, run the CFA model, allowing all of the other factors to vary freely. This is whe we are assessing model fit across the two groups. Lack fo model fit in one group, but not the other will indicate group differences. There are 5 steps to doing this.

* **Step 1:** We fully restrict our model in which we allow nothing to vary across groups compared to that of which everything varies. If we find that everything correlates nicely and just about equal, we can stop and assume no group differences. However, if this is not the case, we continue to step 2. Therefore, if self-reported delinquency is the same here, we are saying that boys and girs measure equally on the constructs latent variable.
* **Step 2:** Now we start testing for configural invariance, this is the step that we try to test the same models to both groups. If we find a difference between the boy and girl groups the we are looking for differences and how much they vary. However, in the event that there is a very large difference, then we stop and it would imply that the latent variable isn’t even comparable across boys and girls. I personally don’t think this would be the case though. Chances are they are at least comparable.
* **Step 3:** Finally, we really start to compare the factor loadings. This is where the individual items comparisons come in. If we find that all of the loadings are different between boys and girls then the two are completely invariant, meaning they are measured on completely different levels of the latent variable. However, if they are not *all* completely different between the boy and girl groups, we can claim partial invariance. Personally, I feel like the latter is the case in the measuring of a delinquency construct between adolescent boys and girls.

1. *Explain, in your own words, what each of the letters a, b, c, and D represent in the 3-parameter logistic model. Indicate which correspond to elements of the common factor (and how they compare) and which are new contributions of the item response theory framework.*

The 3PL model is a logistic representation of the Ogive model. The Ogive is the representation of Cumulative Distribution Function (R&M, pg. 276). The 3PL model provides a relatively good estimate of the actual Ogive model, which helps to give the probability that a response will be chosen (This is our Theta). In order to properly estimate this model, we require parts a, b, c, and D. Below is the definition and explanation of each letters in the 3PL model.

* a = *item Discrimination*
  + When this is graphed, it literally helps to define the *steepness* of the curve in the 3PL model (R&M, 278). Therefore, the larger *a* is, the more distinct an item is. This has important implications when thinking of IRT, such as the ability to identify those that will choose a certain response based on their level of the construct.
* b = *item difficulty*
  + This can also be called the difficulty parameter. Graphically, this will indicate where the curve lies on the plot. As *b* increases, the probability of selecting a certain item, given the level of the construct, becomes less likely (R&M, 279).
* c = *guessing parameter*
  + As with any measure or test, the ability to guess should always be considered. *C* is the probability that a respondent will identify a response given very low value of theta (R&M 295). Theta, of course being the proportion of respondents that will identify given their ability, or level of the construct (R&M, 277). When visualized graphically, the *c* response helps to control the y-axis of the IRT curves. If a respondent has a 25% chance of correctly guessing an item, then the IRT curve will not go below 0.25, as no matter what, someone has *at least* a 25% chance of identifying that item.
* D = *scaling constant of the Ogive model;* ~1.701 (R&M, pg. 293-294)
  + Quite plainly, as stated in the definition, this is a requirement as we convert to the Ogive model to the 3PL model. It is important to understand that the 3PL model is an *estimate* of the Ogive model, which is the true function that is providing the theta values of the response item. Without this scaling constant, there would be consistently biased results.