Choose ONE of Schmidt & Hunter's (1986) scenarios that is reasonably applicable to the kind of research you do (e.g., similar constructs, similar research design).

(a) Summarize, in your own words, their argument for why correcting for measurement error is beneficial in this particular scenario.

(b) What counterarguments can you think of? Why might someone doubt or question whether such a correction is appropriate in this situation?

(c) Taking (a) and (b) together, what do you think is the most appropriate course of action in a situation like this?

Scenario #5.

1. Schmidt and Hunter (1986) point out that the researchers incorrectly accounted for the stability of the measure across a 1 year period. Surprisingly (to me at least) They find that in this situation you actually may manage to underestimate the reliability. Therefore, the measure that is tapping into the construct of negative affectivity actually is more stable than the researchers found. Additionally, by giving a parallel test a week apart helps to understand how negative affectivity changes in a smaller course of time, thus allowing for the researchers to better account the stability of the trait.
2. One argument I can think of is that not using the method they proposed is a more conservative estimate and helping us to not commit any type I errors with our measure. However, this idea is not suitable in helping to prevent Type II error. I also think someone is likely to doubt this method is because their method quite simply wasn’t what they were testing for. The researchers were obviously interested in the year difference between negative affectivity in this scenario, therefore, it does make sense to be testing for mainly that. Lastly, the idea that the construct stability is likely 0.94 (or even 1.00!) seems like a very bold conclusion overall.
3. In reality, I think Schmidt and Hunter (1986) really made a good case though. Researchers should be considering the methods to get at a more accurate reliability coefficient. I was a little bit confused if they were proposing to get a completely different measure for stability overall. But, it really does make sense to see how the stability acts in more timepoints than just one. This can help to make us understand if the change is more likely to just having an unstable construct or if our error fluctuates more so.