Second Assignment: A Graded Discussion – Creating a Valid Instrument

On this board, I want you to do something a bit less formal than usual, and if you want to make it so, a bit more fun.

- In your write-up, I want you to pretend that you are a publisher of assessment instruments. And not just typical educational instruments which assess intellect, language or academic achievement. No, instead you've got a specialized company which custom-creates assessments to serve the needs of your customers, some of whom are a bit odd in their requests.
- I'll want you to describe first in your write-up a construct that a customer might want you to assess, i.e., describe a purpose for an assessment, and then discuss the kinds of evidence you would provide to support the validity of your assessment process.
  - o An example: I might have a customer who wants to create a boy band, and hopefully make millions of dollars, and he's asked me to create an assessment process to identify potential candidates to join the boy band. So I'll create the "Boy Band Beefcake Battery," a comprehensive assessment system which will assess candidates on things such as:
    - good looks, cheesy grin, ability to wink on cue
    - motor skills/dance ability
    - oral-motor reflexes (to assure they can lip sync to their backing track without excessively embarrassing themselves more than once a week)
- I've now outlined the basic elements that make up my construct which I'll call "Boy Band-ability," and have created an assessment process to use with the excessively good-looking artists who show up for tryouts. I'll be creating a series of scales which the boy band producers complete on each person who auditions, including a physical attractiveness scale, a motor response timed task (e.g., each time they hear a syllable over the loudspeaker, they have .2 seconds to move their mouths in sync with it), the ability to gyrate their hips rhythmically yet somewhat asexually, etc. .... The details of how one might assess these things aren't really relevant to your write-up so go crazy on this part if you wish. What's particularly relevant in judging assessments is the technical data that I should provide the producers, to show them that I have an assessment process which is worth the astronomical sum I charge. So in my case...
  - o I've provided them with predictive validity evidence, e.g., others who score high on my assessment tend to have good album sales several years later.
  - o I also have good concurrent validity evidence: I ran my assessment process on several folks during blind tryouts, and my process successfully identified those who were already professional boy band members, while filtering out the riff-raff wannabees.
  - o I have good construct validity as well: I obtained this by showing my criteria to Simon Cowell, Randy Jackson and Paula Abdul formerly of "American Idol" fame, and the judges from "The Voice" (Cee Lo Green excluded of course, because, well, you know). I asked whether these experts agreed that my criteria identify the kinds of things needed to be a famous boy band member: they agreed that my theory of "Boy-Bandability" was a solid one, and these experts even offered content validity: they each did a rating of the extent to which each of the tasks I had the boys do belonged on my scale, and I tossed out any tasks which weren't highly-agreed to by all of them. They agreed that lip-synching was an important part (i.e., construct validity), but thought the .2 second delay I'd been using was unacceptably slow (i.e., content validity), so suggested I change it to .1 seconds.

Persinger, J. (2020). *A graded discussion: Creating a valid instrument.* [Assignment description]. Emporia, KS. Department of Psychology, Emporia State University.

- o Finally, I established interrater reliability by having multiple judges use the assessment on the same boy band wannabees at their auditions, and each judge obtained virtually the same ratings, independent of each other.
- o So, I've given you an example... Now it's your turn! Per my example above, I'll want you to describe a construct that a customer might want you to assess, and then discuss the kinds of evidence you would provide to support the technical characteristics of your assessment process. Label and describe <u>at least</u> two distinctly-different kinds of validity-related evidence, and one kind of reliability-related evidence. If you think you have some really good evidence but aren't sure what to label it, provide the evidence and let's see if your peers can identify what "type" it is. <u>So long as you do provide at least three distinctly-different types of evidence which you think support the technical characteristics of your assessment, Dr. Persinger will be satisfied, even if you aren't sure how to label it or even if you label it incorrectly.</u> There are subtle distinctions between construct and content validity, for example, and you needn't know those in order to do a good job on this.
- o <u>In your follow-up replies</u> include a technical critique of at least two of your peers' proposed assessment instruments/schemes, as if you were the customer offering suggestions for ways to improve the tool. "You said you'd offer content validity for this Boy-Band BeefCake Scale," the customer might say, "by having the judges from American Idol agree with what it measures. Why not have members of successful boy bands describe what characteristics they believe should be included in the assessment?" So, suggest other kinds of evidence the writer could have presented to support its quality. Those kind of follow-ups are a great way to demonstrate to the instructor that you comprehend this material.
- Remember that until you make your own posting describing your instrument, you won't be able to look at postings and make replies to others!
- Try to have fun and be creative with this. In the past, students have created scales which assess spouse suitability, potential for a puppy to become a racing champion, dragon-slayer recruitment process, country-lifestyle scale, teenage angst scale, a Chuck Norris "chuckability" scale which included a mustache-dimensions measure, and more. You could do it with a math test example, but why not live a little?

Persinger, J. (2020). *A graded discussion: Creating a valid instrument.* [Assignment description]. Emporia, KS. Department of Psychology, Emporia State University.