

Lecture Notes for a Presentation to Burapha University  
Faculty of Nursing Students  
Larry R Nelson, PhD  
7 September 2018, with links updated 10 April 2021

Good morning! Thank you for attending my presentation today.

I am going to give a brief talk about the use of what I will call “assessment inventories”.

Many people might refer to “assessment inventories” as “questionnaires”. However, we are all working and studying at an advanced tertiary level, and the term “questionnaires” just does not sound as professional as “assessment inventories”. We should use a term more suitable to our level, and so we shall: “assessment inventories” it will be. (Smile.)

My speciality has to do with the development and evaluation of tests of student abilities and aptitudes; I also have had extensive experience with ~~questionnaires~~ assessment inventories, and how to identify good ones.

Today I will be talking about that, about “how to identify good ones”.

To begin, join me as I pretend to be a doctor of medicine in a family practice. The area I live is in a city that used to produce automobiles for the Australian market. Ford, Toyota, and Chevrolet all used to have manufacturing plants here, but in the past five years all of them have closed because now it is much cheaper for cars to be manufactured elsewhere, in countries where wages and taxes are not as high. This has created extensive unemployment. Many men and women have lost their jobs.

I suspect some of the patients coming to my practice these days may be experiencing depression.

I ask myself: *how to measure depression without incurring a tremendous expense?* Is there perhaps an instrument that is freely available as an internet download? Something I could at least use with my patients as an initial screening instrument for depression?

Yes. In fact there is. I found one called “Beck’s Depression Inventory”. There’s no copyright on it. I’m fairly sure there would be other instruments available, but maybe not copyright-free. I decided to have a good look at the “**BDI**”, Beck’s Depression Inventory.

I have copies of [the inventory](#) for us to experiment with. It consists of 21 statements, and we’re asked to indicate how well each statement describes us.

Before I distribute the copies, I should mention that some of the statements are (as you will see) really very personal ones. YOU DO NOT HAVE TO ANSWER THEM. But look at them, think about them. Answer them if you’d like, and score your answers using the instructions found at the end. WE WILL NOT BE COLLECTING YOUR ANSWERS.

Should we spread out, change our seats so that we are not right next to someone?

\*\*\* 15-minute pause while you look at the inventory \*\*\*

Okay, thank you.

Let’s now spend a little bit of time looking at [the responses](#) collected about eight years ago from 242 postgraduate students at a major Australian university.

\*\*\* 15-minute look at Australian results \*\*\*

Okay, thank you again.

The question now arises: what sort of reviews does this instrument get? So far I am somewhat impressed with the “BDI” -- its questions / statements are quite personal, but depression is, I am sure, a very personal thing.

I undertake an internet search using Google to look for “Reviews of the Beck Depression Inventory”.

It does not take long to find that, by and large, “Beck’s Depression Inventory”, seems to be quite well regarded.

Here’s what one reviewer wrote. I’ve picked this review as an example of an inadequate job:

**Beck Depression Inventory-II (BDI-II)**

The original version of the BDI was published in 1961. A revised version of the BDI was published in 1996 to correspond more closely with the DSM-IV criteria for major depressive disorder.<sup>[4]</sup> The BDI-II contains 21 multiple-choice items assessing symptoms of depression. Each item is a set of 4 statements reflecting increasing levels of symptom severity; thus, the scale consists of 84 statements. It takes 5-10 minutes to complete the scale. The scale has good internal consistency, item-scale correlations, and is sensitive to change. The BDI-II correlates highly with clinician assessments of depression severity. Total scores on the scale range from 0 to 63. Recommended severity score ranges are 0-13 (minimal depression), 14-19 (mild depression), 20-28 (moderate depression), and 29-63 (severe depression).

If I were to grade this review (the one above), I would give it just a **C**.

There are two common indices of inventory quality: “reliability” and “validity”.

I have given a **C** because the reviewer did not use either of these terms.

The reviewer wrote that the scale “... has good internal consistency ...”.

“Internal consistency” is commonly indexed by a statistic called “alpha”, also known as “coefficient alpha”, and “Cronbach’s alpha”. **This statistic is an industry standard.**

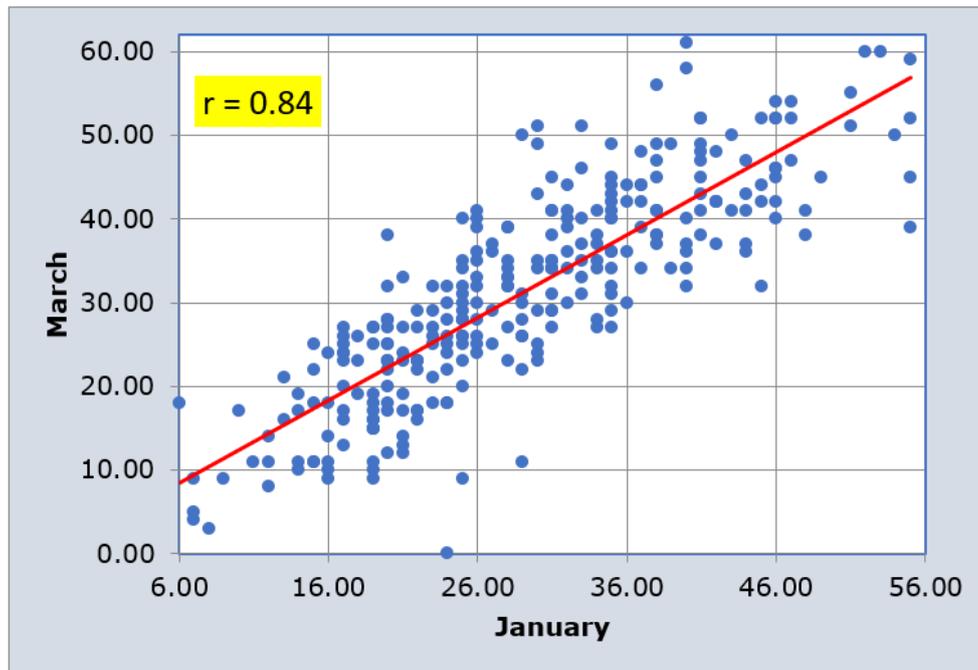
The reviewer should have said, for example “... has good internal consistency, with a coefficient alpha value of 0.78 ...”.

Alpha’s maximum possible value is 1.00 – alpha should be at least 0.70 for an instrument to be classified as “good” – alpha over 0.80 is “sound” – over 0.85 is “very good”.

Alpha is a measure of test “reliability”. It is an indication of what might happen if we were able to get people to answer the inventory twice.

Imagine that we could. Let’s say we had people answer the BDI in January and then again in March. We get a computer program like Excel to plot the scores.

Suppose, for example, that the plot looked like this:



There's a pattern. People with low scores in January tended to also have low scores in March, likewise with high scores. When a test or scale has results like this, we say that we have evidence of "reliability" – repeated testing of the same group of people (should that be possible) will show that their scores don't change very much, there is consistency in their answers.

Scale **validity** means that the instrument measures what it claims to. If we have people answer the BDI and the NDI (Nelson Depression Inventory), do the two instruments agree?

Validity is very often measured by plotting two scores, just as above, except that "January" might be "BDI", and "March" might be "NDI".

The reviewer quoted above reported that "The BDI correlates highly with clinician assessments of depression severity".

A statistic called "**r**", the "**correlation coefficient**", is often used to indicate the degree of correlation, with r of 0.70 or more considered good.

The reviewer should have said, for example "The BDI correlates highly with clinician assessments of depression severity (r = 0.75)".

I gave the reviewer a **C** because s/he failed to use the terms expected. The review has a positive tone to it ("*good internal consistency ... correlates highly*"), but, in the end, the tone is not enough – imagine a car dealer saying "*this car gets good gas mileage*" – **we want to have a figure**, like "it uses just 2 litres of petrol for every 100 kilometres of travel".

Here's what another reviewer wrote of the BDI; **notice that actual figures have been used**:

## Results

Go to:

The Cronbach's alpha for the BDI-II total score was 0.89. The correlation between the BDI-II and the PHQ-9 was strong (r=0.75), and anxiety-related measures were 0.68 and 0.71, which were also in the high range. Among the five different factor structures, the modified three-factor model demonstrated the best overall fit.

I would give this reviewer an **B**. He or she would get a **B+** from me for a review like this:

Reliability, as measured by Cronbach's alpha, was 0.89. Validity was measured by correlating BDI scores with the PHQ-9 ( $r=0.75$ ), and with two selected anxiety-related measures ( $r=0.68$ ,  $r=0.71$ ).

I may have given the reviewer an **A** had PHQ-9 been spelled out, and had the names of the "anxiety-related measures" been given.

Incidentally, PHQ is:

The PHQ-9 is a multipurpose instrument for screening, diagnosing, monitoring and measuring the severity of depression: n The PHQ-9 incorporates **DSM-IV depression** diagnostic criteria with other leading major depressive symptoms into a brief self-report tool.

### IN SUMMARY

It turned out to be easy to find a depression inventory of good quality – the Beck Depression Inventory was among the first hits that I got from Google. I was able to get a copy of the actual instrument as a pdf file, and it wasn't copyrighted. It took a bit of digging, but it was even possible to get some actual results. And, with a bit more help from Dr Google, I was able to find numerous reviews of the BDI, and, by and large, they were quite positive.

If ever you have to do something like this, remember to look for two things: a statement regarding reliability (usually coefficient alpha), and another statement about validity (often  $r$ ).

Alpha should be at least 0.75, and  $r$  values for validity should be about the same, say at least 0.70.

Feel free to write to me with any questions you might have: [l.nelson@curtin.edu.au](mailto:l.nelson@curtin.edu.au)

---

\*\*\* THANK YOU VERY MUCH FOR YOUR ATTENTION TODAY \*\*\*  
(I hope my talk has not been depressing.)