Lecture08 Psychometrics - LIVE!

I like to develop new measures (Strongly Agree - Strongly Disagree)

Dr Gordon Wright

# Overview

## The week ahead

The week ahead (week 8)

* PS52005C Design & Analysis Quiz 2 Due: 10am 16th December (week 10)
* W8 Personal Tutor session on feedback
* W9 Making the most of Goldsmiths Psych
* W10 Cognitive Essay Tutorial
* Strike day 3/3 - 30th (Wed)
* Labs - Ongoing task development and Ethics applications

## Who’d like to do some free-style Psychometrics?

Hands up for “Yes”

(in which case I’ll re-record the lecture)

**Will involve participation**

We can run this in-class exercise and if you want, I’ll apply for ethics to do the thing properly over the coming weeks. Could be publishable.

## Here’s an idea…

A psychological measure of

. . .

X

## If you are up for it, we can start the first two steps

1. Articulate the construct and the context
2. Assemble items and decide on response format
3. Data collection
4. Assess Psychometric properties

## Or let’s just do the lecture…

Your call!

## Psychometrics - Not as boring as you might think…

Psychological ‘constructs’ are tricky to pin down

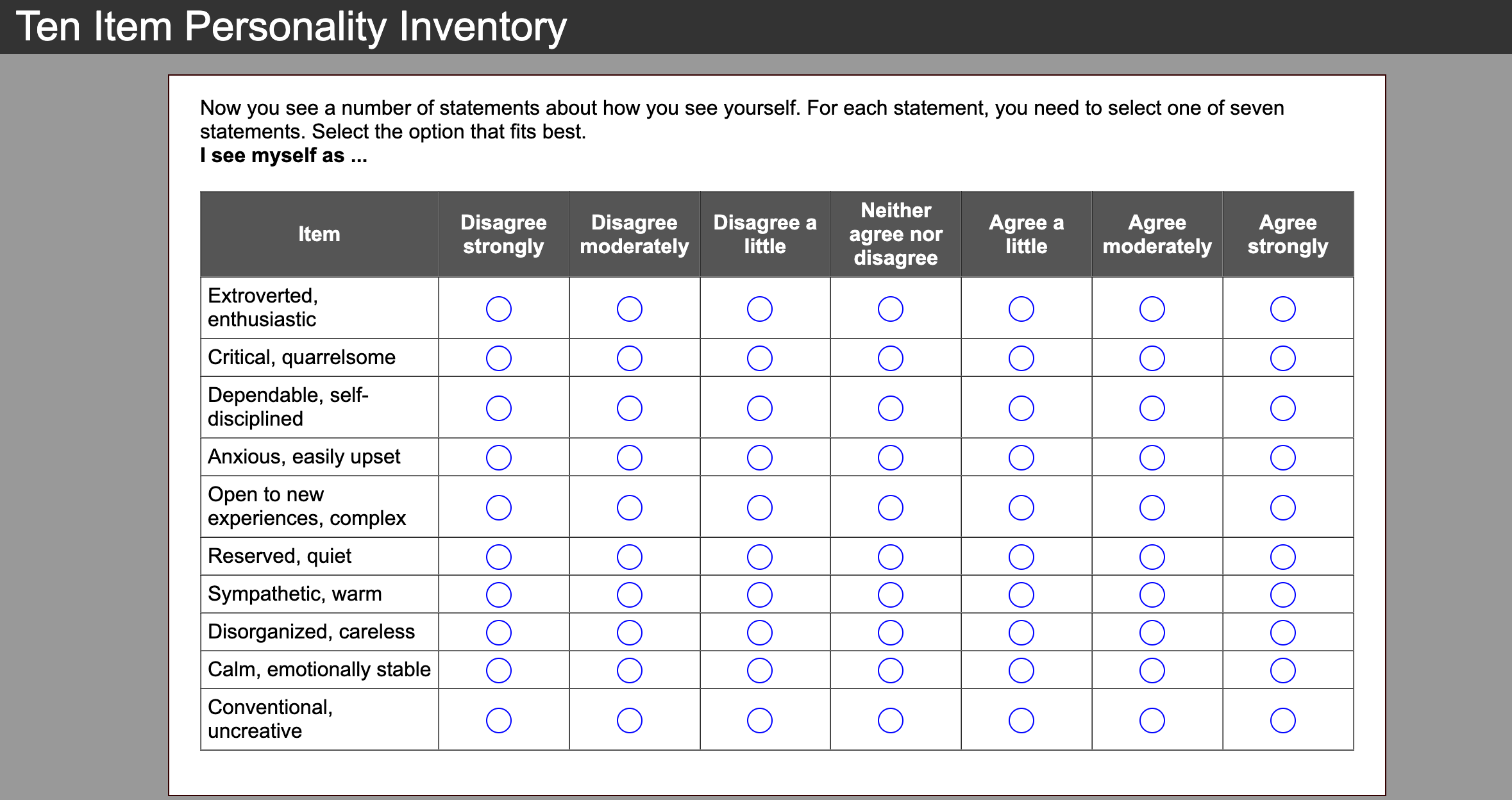
May differ from ‘naive’ understanding

* Consider ‘Intelligence’ and ‘Empathy’
* What about Personality?

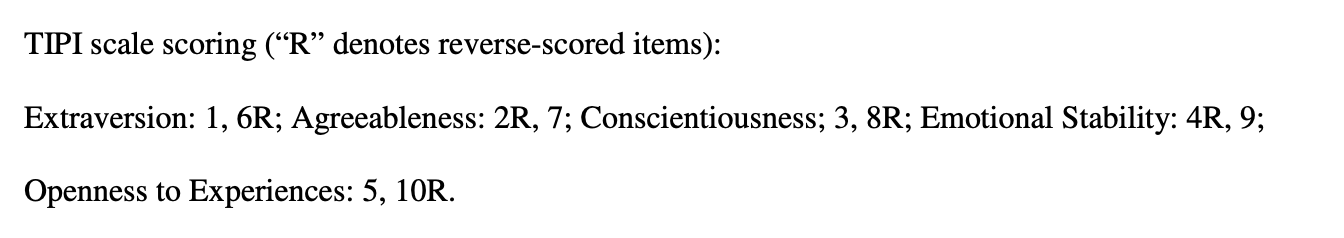
## Ten Item Personality Inventory (TIPI)

Gosling et al. (2003)

<https://gosling.psy.utexas.edu/scales-weve-developed/ten-item-personality-measure-tipi/>

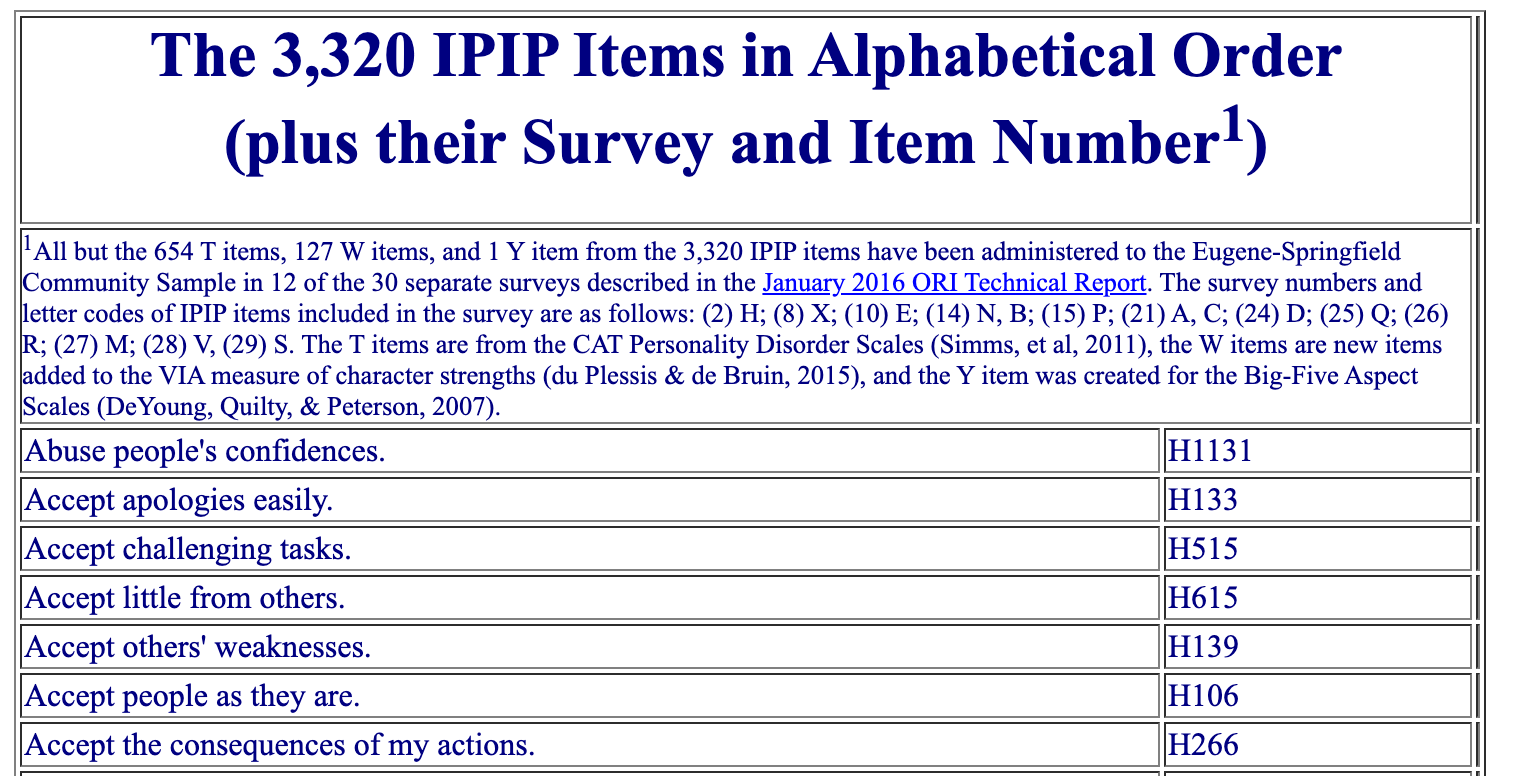


## TIPI Scoring

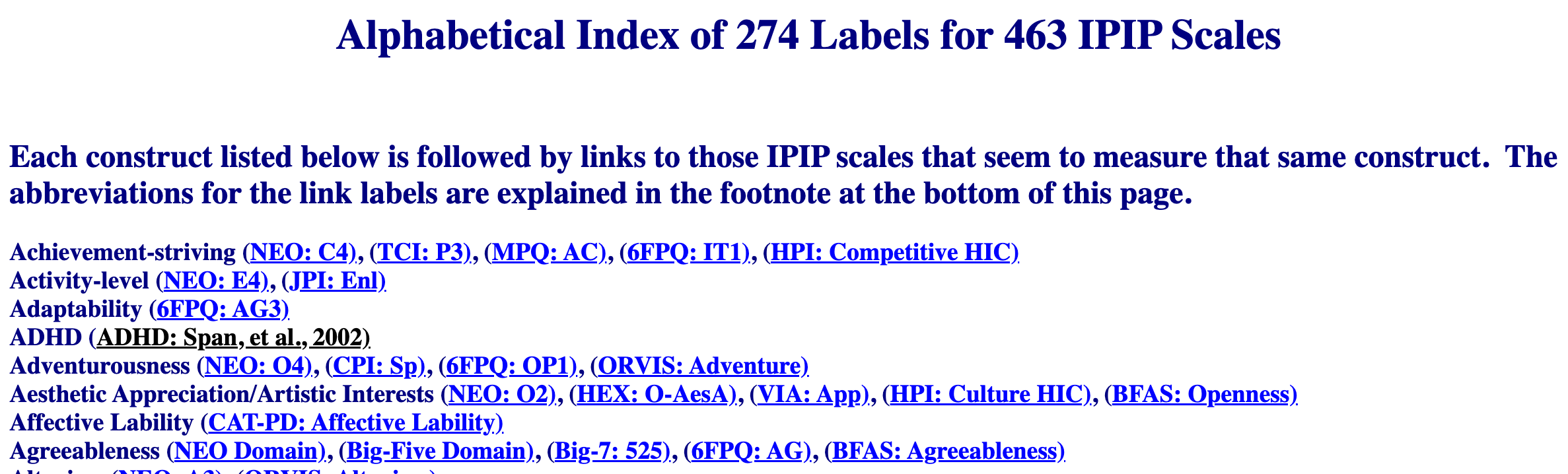


## The **International Personality Item Pool (**IPIP)

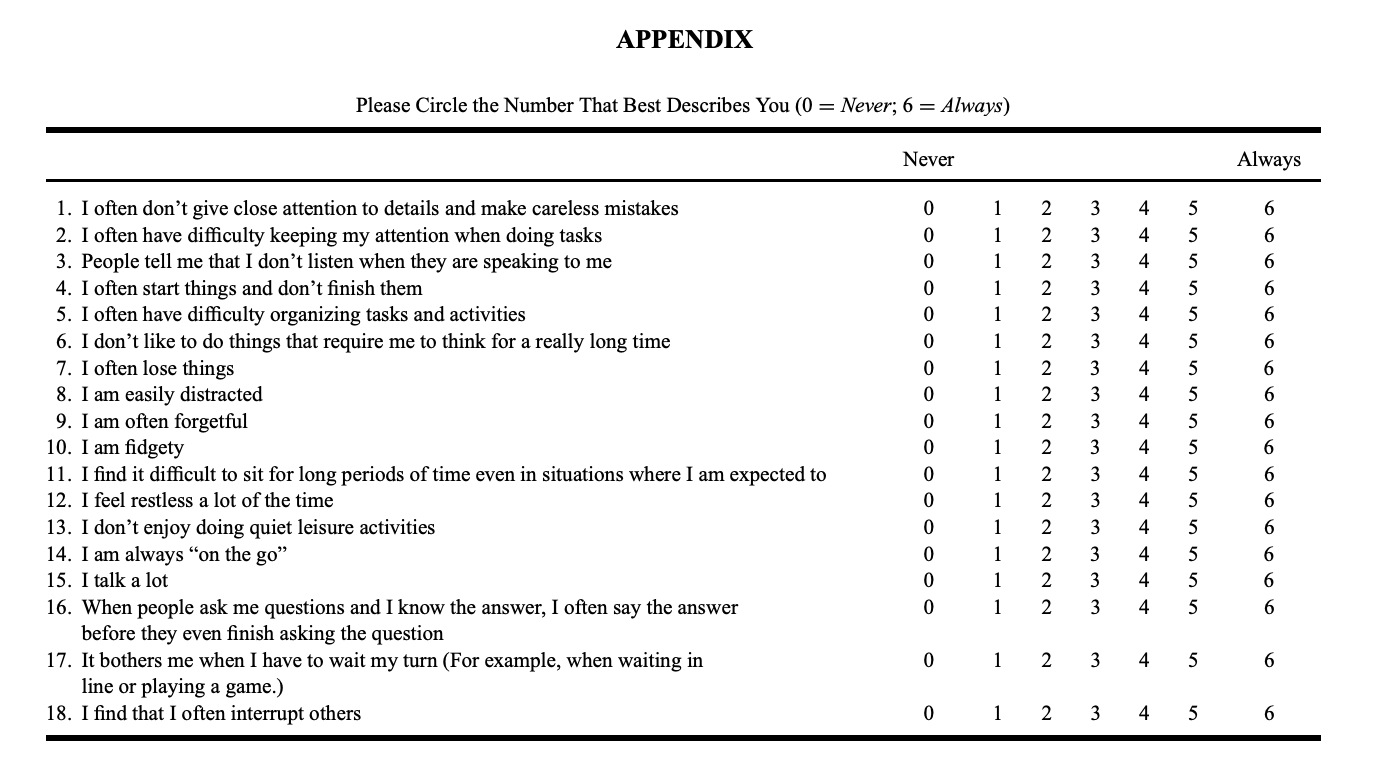
<https://ipip.ori.org/>

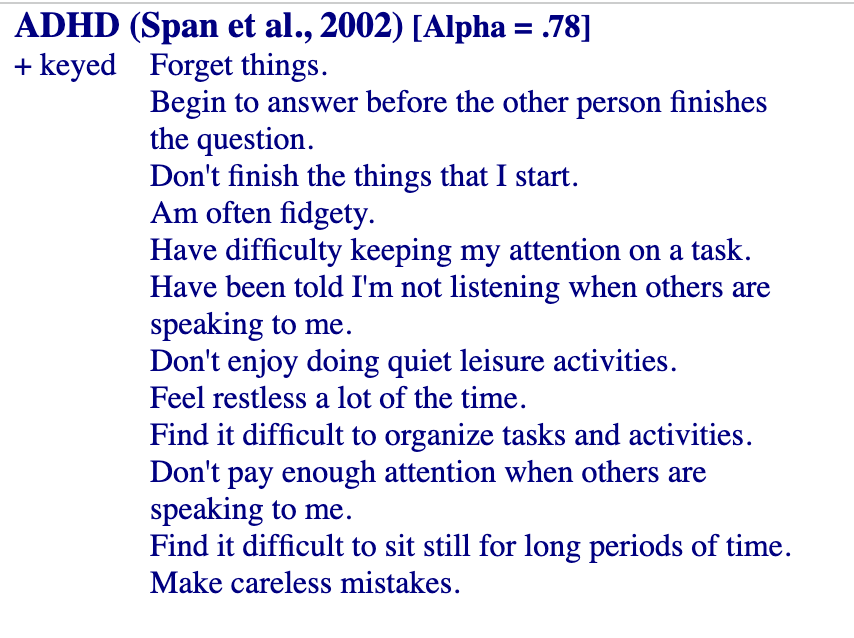


## Go crazy!



## Two measures for ADHD (Span et al., 2002 & IPIP)





Span et al. (2002)

An incredibly valuable resource and a lovely way in to new topic areas!

## The challenge of measurement

* To do it requires a clear understanding of the concept in question
* You need technical skills and some friends
* You need to be creative and thoughtful
* It’s research in itself, OR, a step required to do novel research!
* Fascinating concepts *when seen in practice!*

## 4 key steps to scale construction

1. Articulate the construct and the context
2. Assemble items and decide on response format
3. Data collection
4. Assess Psychometric properties

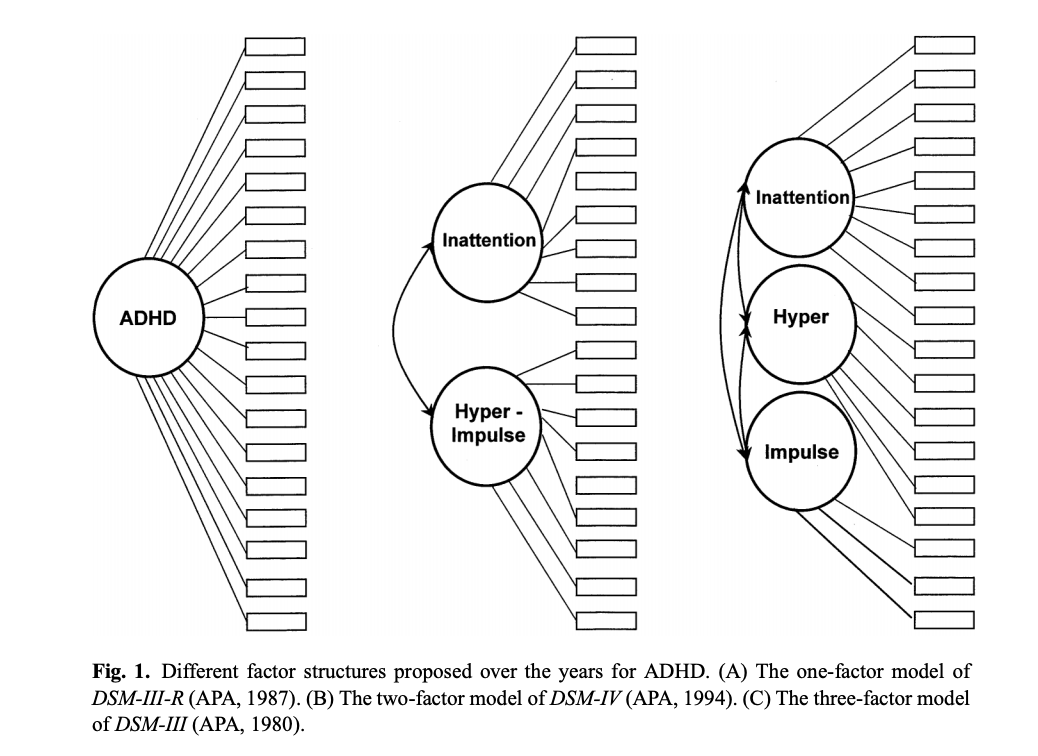
## 1. Articulate the construct and the context

What are we measuring and how do we conceptualise it?

In what context is the construct displayed?

Is it ‘unitary’ or ‘multidimensional’?

## 1.b Dimensionality (or Factors, Sub-scales, Facets)



## 2a. Assemble items

Either write items or ‘find’ items.

* e.g. The lexical approach to personality
* or the Delphi Technique (Expert Opinions e.g. Clinicians)
* Previous scales or items (other research
* Talk to the target population/observation

## 2b. Response format

There are options

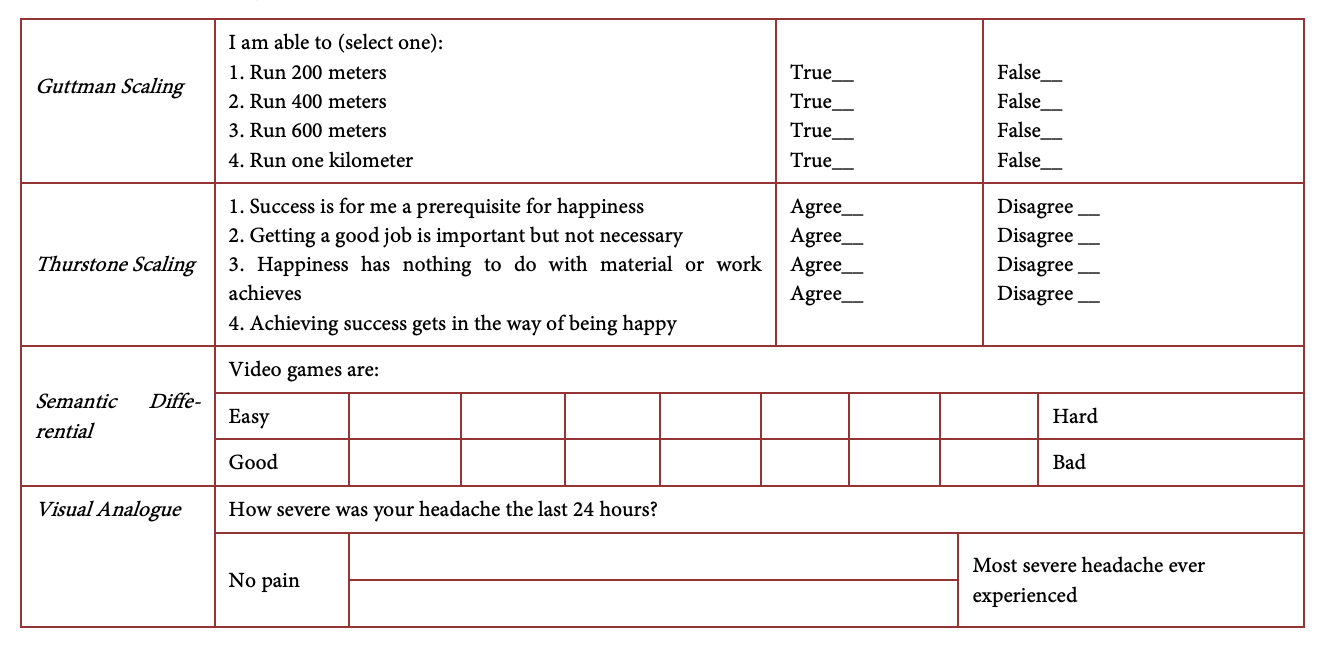
* **Likert Scales** (To what extent do you agree with SD - NAND -SA)
* **Rating Scale** (On a scale of 1-7 how x are you?)
* **Forced Choice** (a - Modesty doesn’t become me. b - I am essentially a modest person)
* **Semantic differential** (Healthy —— Unhealthy)

## 2.c Response choices

* **Options** (True False, 7 point)
* **Labels/anchors** (What is the midpoint?)
* **Mid-points** (4 item or 5, Don’t know, N/A, No opinion)
* **Consistency of response across items**

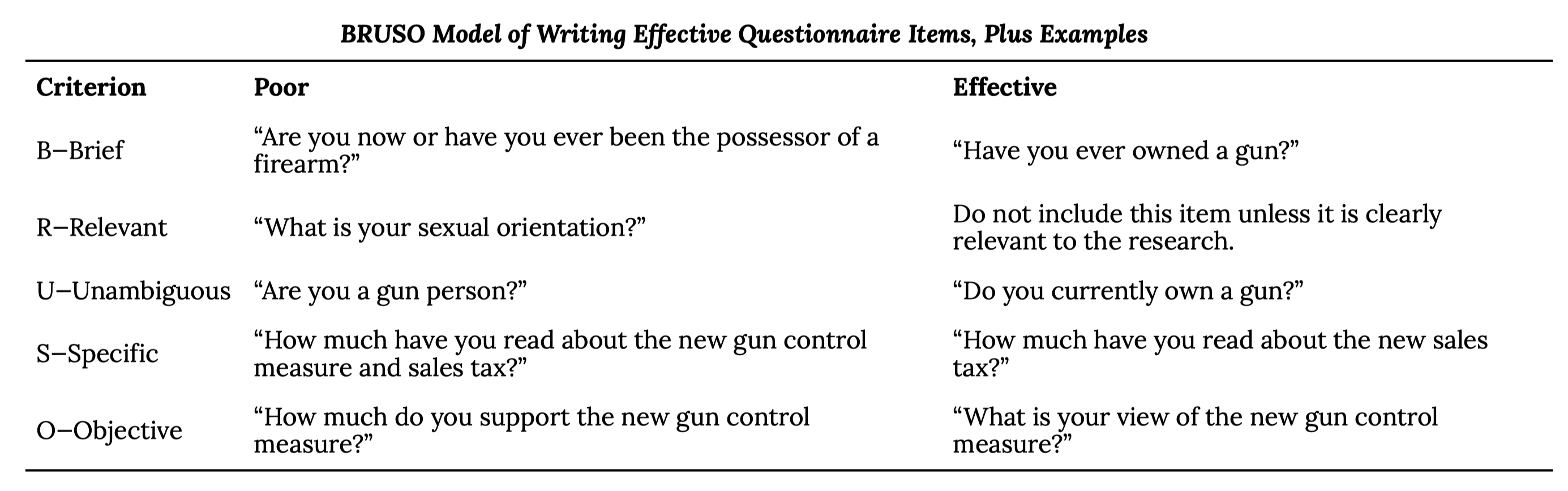
## Likert Scales (1) (Lick-ert) and Likert-type rating scales (2&3)

## There are other scale types!



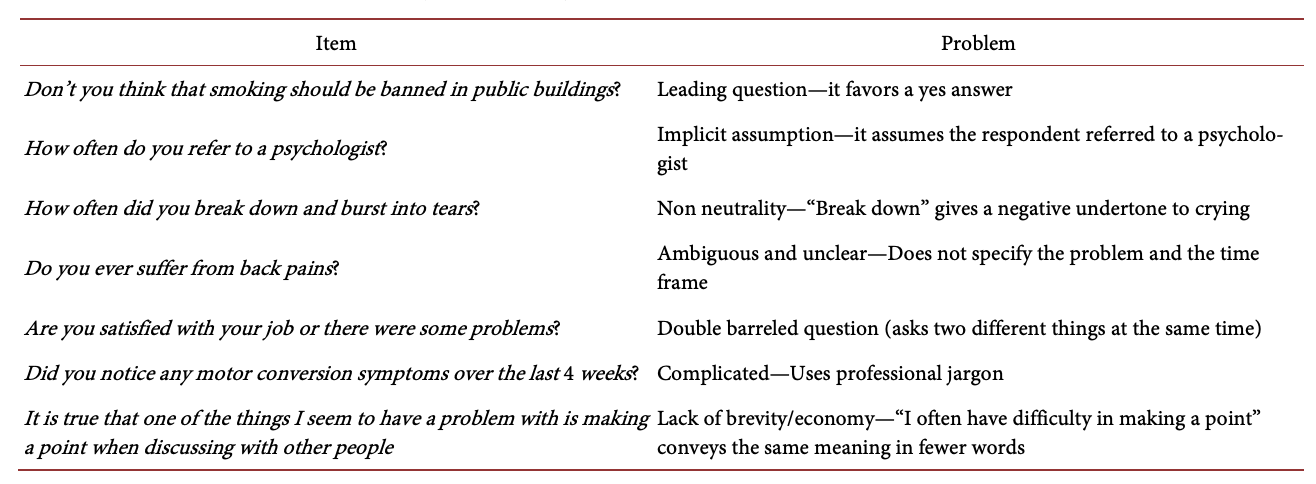
## **BRUSO Model of Writing Effective Questionnaire Items**

An acronym that stands for “Brief,” “Relevant,” “Unambiguous,” “Specific,” and “Objective,”



Peterson (2000)

## Unsuccessful items

Adapted from Barker et al., 2016: pp. 111-112; DeVellis, 2017: p. 101.

## “Stress” Simple, huh?

The **Social Readjustment Rating Scale (**Holmes and Rahe, 1967) is a self-report questionnaire on which people identify stressful events that they have experienced in the past year and assigns points for each one depending on its severity. For example, a man who has been divorced (73 points), changed jobs (36 points), and had a change in sleeping habits (16 points) in the past year would have a total score of 125.

. . .

The **Daily Hassles and Uplifts Scale** (Delongis et al., 1982)is similar but focuses on everyday stressors like misplacing things and being concerned about one’s weight.

. . .

The **Perceived Stress Scale** (Cohen et al., 1994)is another self-report measure that focuses on people’s feelings of stress (e.g., “How often have you felt nervous and stressed?”).

. . .

Researchers have also operationally defined stress in terms of several physiological variables including blood pressure and levels of the stress hormone cortisol.

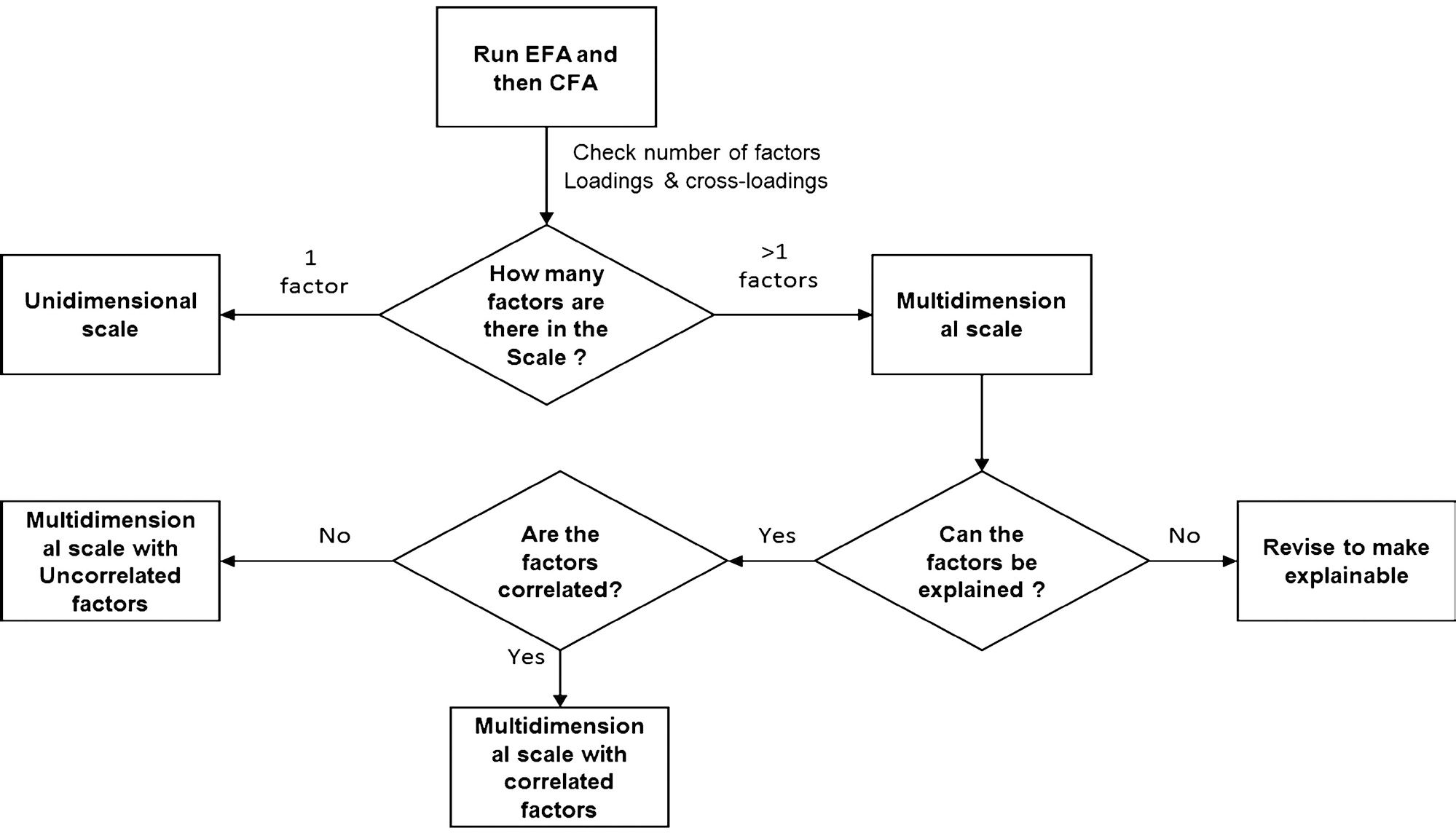
(Cohen et al., 1994; DeLongis et al., 1982; Holmes & Rahe, 1967)

## Blirt - Blirtatiousness (Swann et al., 2001)

<https://labs.la.utexas.edu/swann/the-blirt/>



## Factor Analysis



## Would you like to have a go?

If you would like to, how about we develop a scale of our own.

Everyone who wants to have a go, come to the front.

## Glossary (Validity & Reliability)

**Discriminant validity**

the relationship between some traits that should have weak or no relationship

**Face validity**

Does the scale appear to be measuring the variable? to researcher AND participant! e.g. Psychopathy at work scales. Cannot be TOO ‘criminal’ or ‘deviant’

**Content validity**

Do the items on the scale represent the various (and full) aspects of the variable being measured?

**Construct validity**

Does the scale actually measure the intended variable?

**Concurrent validity**

Does the scale relate to a relevant outcome or behavior that was measured at the same time?

**Predictive validity**

Does the scale relate to a relevant outcome or behavior that occurs in the future, after the scale is completed?

**Convergent validity**

the relationship between traits that are similar to (but not identical to) the trait being measured

**Criterion validity**

the relationship between some measure and some real-world outcome

**Discriminant validity**

the relationship between some traits that should have weak or no relationship

**Inter-rater reliability**

Inter-rater reliability agreement demonstrates consistent results between different examiners.

**Test-retest reliability**

reliability between tests conducted over short periods of time.

**Split-half reliability**

the correlation between to halves of the same proposed measure. Doing this for all the possible halves and averaging the result it Cohen’s Alpha!

## References

Cohen, S., Kamarck, T., Mermelstein, R., et al. (1994). Perceived stress scale. *Measuring Stress: A Guide for Health and Social Scientists*, *10*(2), 1–2.

DeLongis, A., Coyne, J. C., Dakof, G., Folkman, S., & Lazarus, R. S. (1982). Relationship of daily hassles, uplifts, and major life events to health status. *Health Psychology*, *1*(2), 119.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, *11*(2), 213–218.

Peterson, R. (2000). *Constructing effective questionnaires*. SAGE Publications, Inc.

Span, S. A., Earleywine, M., & Strybel, T. Z. (2002). Confirming the factor structure of attention deficit hyperactivity disorder symptoms in adult, nonclinical samples. *Journal of Psychopathology and Behavioral Assessment*, *24*(2), 129–136.