A convention for interpreting validity coefficients

Henk Elffers*

Validity

Whenever we are considering measurements for concepts in a theoretical well-developed field, we usually consider, next to reliability and content validity, the degree of *criterium validity* with respect to a given criterium variable of interest. For instance, when we propose an *egoism scale, E,* (cf. Van Giels et al. 1992, Adams & Webley, 1996) within a research project on crime, we tend to relate this scale to an index of engaging in crime, *C*, and are happy if indeed, in accordance with that theory, E and C are correlated. A commonly used index to gauge the success of E with respect to C is the amount of variance explained. We are even more happy when we are able to demonstrate that E indeed is able to augment to the variance explained in C, when another standard concept from the theory has been already incorporated into the explanation of C. E.g. Gottfredson & Hirschi (1990) proposed to use an *index of low self control, S,* to explain crime. We look, then, at the amount of variance explained in C by E *and* this standard test S together, in comparison with the contribution of the standard test S *alone*. If the variance explained is rising 'enough', we say that E is useful as having *discriminant* or *divergent validity* next to S for C. Moreover, in most cases we also hope that E and S are sufficiently related themselves as well, as this shows, in a well-developed theoretical field, that E is not a complete stranger within the theory: we demand that E and T share variance, or, to formulate it in the jargon, that E and S have *concurrent validity* as well. An overview of the usage of various terms in validity theory is given in Kerlinger (1986), see also Lewis-Beck (1994). We aim for concurrent validity, as, new though our concept may be, it will be related - in most theoretical contexts - to other concepts. We aim for discriminant validity, as otherwise the new concept would not contribute to our understanding next to the existing concepts.

* Erasmus Centre for Sociolegal Tax Research
Faculteit der Rechtsgeleerdheid
Erasmus Universiteit Rotterdam
Postbus 1738
3000 DR Rotterdam
The Netherlands
phone: +31 10 408 1613 / 7
e-mail: Elffers@soc.frg.eur.nl

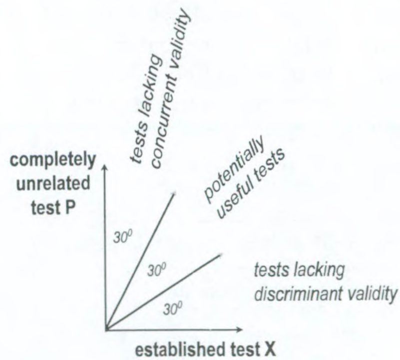
Context free discriminant validity
Now, while the concept of concurrent validity between E and S is independent of other variables, the above argument on discriminant validity is tainted by the occurrence of the criterion variable C. Of course, E and S may have discriminant validity with respect to some criteria C_1, C_2 , ..., and not with respect to other criteria D_1, D_2 , ...This means that no absolute discriminant validity exists. However, we would like to have such a concept available, and we propose here a convention as to when two tests have simultaneously satisfactory concurrent validity as well as (context free) discriminant validity.

A convention for validity coefficients
If X is a standard well established test for a well established concept in a field, and Y is a newly developed test for a new concept, thought to be clearly different from X, but at the other hand, in the same field as X, what do we expect of the validity coefficient of Y w.r.t. X, i.e. the product moment correlation coefficient $\rho = \rho(X,Y)$? Because Y should be clearly different from X, we hope that ρ is not close to 1, which would mean that Y and X are almost indiscernable, i.e. lack *discriminant validity*. Say that we demand ρ to be less than a threshold c_{disc} . On the other hand, if X and Y have something in common, we hope that ρ is not close to 0 either, which would mean that either error is overshadowing Y, or X and Y have almost nothing in common, i.e. X and Y lack *concurrent validity*. Say we like ρ to surpass a threshold c_{conc} .
Together we have the following demands:

a new scale Y in the realm of an established scale X can be useful only if

$$\mathbf{c_{conc} < \rho(X,Y) < c_{disc}}$$

What now are prudent choices for the threshold values c_{conc} and c_{disc} ?
To a large extent this is an arbitrary matter, and certainly dependent on the field of research. We do, however, propose now a convention, based on the *geometrical representation* of variates in a linear space. We can represent variates as vectors, their association being reflected in the angle that they make. Associated



variates are represented by vectors with a small angle between each other, variates that are less associated are represented by vectors making an angle of nearly 90°, and independent variates are represented by perpendicular vectors. Given this representation, we propose to divide the area between a standard test X and a completely unrelated perpendicular test P in three equal angles of 30° each. The area nearest to X is the area of variates too close to X (lacking discriminant validity), the area near the perpendicular vector P is the area of vectors lacking concurrent validity), and the middle area is the area of interesting variates. That is, we propose to take the vectors bordering these areas are the ones that have just c_{disc} and c_{conc} as correlations with X.

Just a convention

We are well aware of the fact that the proposed convention is indeed a convention, in the same way as for example thresholds for reliability coefficients are conventionally being used. Indeed, if in a given context somebody sees fit to use tests that are disqualified by our convention, he should not hesitate to argue his case. The existence of a convention, though, fixes language usage, and makes it easy to discuss the reasons why one is not satisfied with the proposed classification.

The convention translated into correlation coefficients

The correlation between variates is equal to the cosine of the angle between the vectors representing them, so the vectors bordering these areas do correspond with correlations, on the one hand, $\rho = \cos(30°) = .87$, and $\rho = \cos(60°) = 0.5$ on the other hand, that is, $c_{conc} = 0.50$ and $c_{disc} = 0.87$.
The proposed convention, then, boils down to:

> **a new test Y offering something new in the field of an established test X, with which it ought to be related, is worth considering only if**
>
> $$\mathbf{0.50 < \rho(X,Y) < 0.87}$$

Literature

Adams, C. & Webley, P. (1996). The role of economic and psychological variables in parking violations, *Psychology, Crime & Law,* 3, p. 111-133.

Giels, B. van, Hessing, D.J. & Elffers, H. (1992), *De kwaliteit van de SSO-schaal.* Rotterdam: Erasmus Centre for Sociolegal Tax Research.

Gottfredson, M.R. & Hirschi, T. (1990). *A general theory of crime.* Stanford: Stanford University Press.

Kerlinger, Fred N. (1986), *Foundations of Behavioral Research*, 3rd edition. Fort Worth: Holt, Rinehart and Winston.

Lewis-Beck, M.S. (ed., 1994). *Research Practice. Vol. 6, International handbook of Quantitative Applications in the Social Sciences.* London: Sage.