Is Cronbach's alpha relevant indicator for quality of educational measurements?

Volodymyr Kovtunets
Institute of Higher Education of National Academy of Education Science
Kyiv, Ukraine
vkovtunets@ukr.net

Nadiia Polishchuk
Kyiv National I. K. Karpenko-Karyi Theatre, Cinema and Television University
Kyiv, Ukraine
n.polishchuk2000@ukr.net

Abstract—**Mathematical model of criteria learning outcomes assessment shows that well known Cronbach's alpha may not be appropriate indicator of assessment of professional qualifications. This indicator is good for area of learning outcomes having only one main factor. Professional standards include a list of professional functions some of them form the main factor in a sense of factor analysis.**

Keywords—assessment; learning outcomes; professional qualifications; Cronbach's alpha; validity; reliability.

I. INTRODUCTION

In [1] it was published mathematical model of content validity of criteria assessment. But well known Cronbach's alpha usually is used to evaluate a reliability of testing (assessment) and apply it as evaluation of the assessment quality. In many cases it may be right, for example if MOODLE recommends to evaluate a quality of formative assessment in testing form (it recommends not use test with alpha less then 0.70) because in teaching process assessment usually assesses instructing of narrow set of learning outcomes. Factor analysis of such assessments really shows one main factor.

But in case of more wide content set it is not true. For example, a content set of practical psychologist in social work professional standard [2] includes learning outcomes (knowledge and skills) concerning behind psychology also law, social work, communication etc. It implies that set main factors includes at least three elements.

Multisubject test for school graduates used in Ukraine in 2019-2023 shall have at least two main factors linked to language and mathematics [3].

Authors also observed case with negative alpha in 2011 when ability test was piloting for university students selection. Nevertheless this test added to subject test score increased predictive validity of the score for selection of students.

In this paper we'll explain borders of successful using Cronbach's alpha and suggest modernized model of assessment content validity.

II. MAIN RESULTS

American Psychological Association (APA) recommends three criteria of assessment quality:

- validity;
- reliability;
- fairness [4].

[1] suggests some method for evaluating validity of assessment.

Cronbach's alpha is developed to evaluate reliability of assessment. Parallel tests approach is the origin of this indicator as well as for other approach to evaluate reliability of the test.

Let us consider geometrical model of assessment by testing process.

Having some test X with n test items and N persons who passed the test we consider vectors $x_i=(x_{i1}, x_{i2}, ..., x_{iN})$, $i=1,2,...., n$, where x_{ij} denotes score of j-th person gotten at i-th item.

As it follows from the definitions

$$r(x_i, x_j) = \cos(x_i - \overline{x_i}, x_j - \overline{x_j}), \qquad (1)$$

where $r(x_i, x_j)$ means Pearson correlation between test scores of i-th and j-th test item. Positive correlation means that the angle between two vectors x_i and x_j in some N-dimensional Euclidean space is less then $90°$.

Cronbach's alpha may be calculated by formula

$$\alpha = \frac{n}{n-1}\left(1 - 2\frac{\sum_{i,j=1, i \neq j}^{n} r(x_i, x_j)}{\sigma^2(X)}\right), \qquad (2)$$

where $\sigma^2(X)$ is a dispersion of the test X (see [4], [5]).

Therefore one may make a conclusion that alpha corresponds to level of closeness of test items near some common (average) direction (factor).

Its known that L. Cronbach early considered also such interpretation of alpha [4].

Its easy to show that having data for evaluation of content validity [1] one may get value of alpha. But mentioned data doesn't confirm test validity. Some additional research is needed to confirm test validity. For sure parallel test techniques also may be used.

So Cronbach's alpha does not guarantee content validity of the test. It guarantees only closeness of test items to some direction (factor) which may not be strongly identified.

Therefore statements on quality of test based on Cronbach's alpha value in general aren't right.

For example test assessing learning outcomes of psychologist in social work qualification seekers usually has two principal factors. Of course Cronbach's alpha is at least less then 0.7.

The same situation we observed while studying assessment of learning outcome of high school graduates. Formally one may tell these tests assess some common construct – school graduates achievement and there are reasons to apply Cronbach's alpha as relevant indicator for the test quality. But factor analysis discovers that test has at least two principal factors linked to languages and mathematics [6] and it implies decreasing of Cronbach's alpha.

All mentioned above doesn't mean denial of Cronbach's alpha. It is recognized indicator of test reliability. And its good at least in psychological tests.

But reliability is useful only in the case when test is contently valid. That is Cronbach's alpha shouldn't be applied as universal indicator of test quality. Especially it concerns assessment of professional qualifications.

The question is what may be a strategy for evaluation of test quality. Let omit fairness, it should be study by some sociological methods. Let consider validity and reliability.

With regard to mentioned above factor analysis should be performed at the first stage. This discovers number of dimensions of space of results. It allows to consider space of all test items (see [1]) as a compact set in Euclidean space and apply Euclidean metrics instead one suggested in [1].

If parallel tests items were added to the test its validity may be evaluated.

In case of one-dimensional space where single principal factor corresponds to content for assessment Cronbach's alpha is perfect indicator of test quality.

In case of more dimensions small value of Cronbach's alpha isn't indicator of bad test. If professional standard anticipates essentially different working functions alpha may be close to zero or even negative.

Different approach may be applied to evaluate test reliability. The simplest one anticipates dividing test on two or more parts and evaluation of reliability of each.

The most fruitful method is based on studying of predictive validity. This method if combined with mentioned above may submit reasons for improvement of professional standards.

III. CONCLUSIONS

Quality of assessment of learning outcomes includes three indicators. Reliability defined by Cronbach's alpha is one of them. Test validity is also important indicator. Cronbach's alpha far from 1 does not mean that test is of low quality. Only complex study including validity and reliability gives right conclusion. Model of test validity study is applicable to study validity and reliability. Factor analysis is powerful method to ensure right complex study of both indicators.

REFERENCES

[1] V. Kovtunets, "Mathematical Model of Criterion Assessment," Proceedings of III International scientific and practical conference Modeling, Control and Information Technologies, 2019, p. 145–149.

[2] "Practical psychologist (social work). Professional standard." [Online]. Available: https://register.nqa.gov.ua/profstandart/prakticnij-psiholog-socialna-sfera. [Accessed: August 11, 2023]. (in Ukrainian).

[3] "Ukranian Center of Education Quality Assessment (UCEQA)." [Online]. Available: https://testportal.gov.ua/en/. [Accessed: August 11, 2023].

[4] "Standards for Educational and Psychological Testing." [Online]. Available: https://kupdf.net/queue/aera-apa-amp-ncme-2014-standars-for-educational-and-psychological-testing_5a68c4f8e2b6f5341181b776_pdf?queue_id=-1&x=1630046340&z=MzcuNTcuMC4yMTQ. Accessed: August 11, 2023].

[5] E. Cho, and S. Chun, "Fixing a broken clock: A historical review of the originators of reliability coefficients including Cronbach's alpha," Survey Research, vol. 19(2), 2018, pp. 23–54.

[6] Ковальчук Ю.О. Теорія освітніх вимірювань. Ніжин: ПП Лисенко М.М., 2012. 200 с.