The Development of Test, Investigating the Accuracy of Professional Evaluation.

Dimitry V. Inozemtseva*

South Ural State University, Yznaya 4a-13, Chelyabinsk 454080, Russia

Abstract

The aim of this article is to describe test construction process. It could be used to estimate appraisers' performance. The authors used 5 scales of the 16 Personality Factors Raymond Cattell's test "Warmth", "Reasoning", "Emotional Stability", "Dominance" "Rule-Consciousness". Stimulus material for this test consists of a set of video records showing behavior of four respondents in three different situations: (interview), work in a pair (role playing game), work in small group (panel discussion). Each video record lasts from 4 to 7 minutes. Testee estimates people's behaviour on the video record. Test has two forms. Test was proved to have internal consistency, parallel-form reliability, test-retest reliability, criteria and construct validity.

The question of professional evaluation is quite new for Russian psychology, though it is widely discussed by foreign researchers [1], [2]. The main problem in this field of research is the lack of unified positions that regulate practice of appraisers [3] and the lack of standardized tests that can estimate appraisers' performance [4].

The aim of this study is developing the test, which can be used to select appraisers for assessment-center procedures (practice). This test basically includes the observation of one's behavior and assessment of personal traits using video records.

The authors based the test on the 16 Personality Factors Raymond Cattell's theory [5]. This choice was motivated by previous pilot studies, as well as the advantages of the theory.

Several experts selected among 16 traits descriptions those, which were used for estimation of personal traits. The experts ranked the 16 personal factors by several parameters: "internal consistency", "least of all expressed in behavior," "most of all expressed in behavior," and "the most significant for performance".

* Corresponding author. Tel.: +7-905-832-73-42.
E-mail address: info@dimino.me
The procedure of expert evaluation allowed selecting five personality traits in relation to which the experts had a consensus on the parameters "internal consistency" and "significance for performance". These five personality traits are "Warmth", "Reasoning", "Emotional Stability", "Dominance" "Rule-Consciousness".

One of the challenges of our research was to define a reference evaluations, which could be used as reference for results of the evaluation of one's behavior on video records. To solve this problem multi-method evaluation of personal traits was used. For each participant on all of five personal traits 4 different types of evaluation: expert opinions, evaluation by test 16-PF [5], peer review and external evaluation were collected. These estimates were checked for consistency with each other, inconsistent evaluations were excluded from the total sample. The authors used the median as the measure of accurate estimates (integral index).

The next step included development of stimulus materials, which consisted of video records of participant performing tasks that could reveal their personal characteristics. Situations for the tasks were selected based on typical activity at workplace: interviews, work in pairs, work in small group.

It is known, that video records can distort personal traits expressions, also the participants may not demonstrate their traits on video records. To take these biases into account it was decided to estimate these distortions using expert methods. Each of nine experts made his judgment on video records after that integral indicator of expert evaluations was calculated. Alpha-Chronbach indicator between expert estimations for every personal trait was overall .9.

Then the results of calculating the median (from 4 different types of evaluations) and expert estimations were compared with each other. It was 2 to 3 personal traits that did not differ between each other for each participant.

Thus stimulus material for this test consists of a set of video records with a showing behavior of four respondents in three different situations: (interview), work in a pair (role playing game), the work in small group (panel discussion). Each video record length was from 4 to 7 minutes.

As a measure of the accuracy of professional evaluation the difference between the value from 4 different types of evaluation and expert assessments and the value of the test were used. It is called the index of accuracy of professional evaluation. Test has two forms.

Next, psychometric features of test accuracy of professional evaluation were checked [6]. Spread of values in the test statistics ranged from 2 to 5, for 40% of the distribution of responses and from 1 to 5 for 60% of the distribution of responses. The mean in the distribution of responses is from 3.05 to 3.48, the standard deviation of .6 to .94.

The task of ensuring the internal consistency of the scale was fulfilled. Correlation coefficients for the parameter-consistency reliability for the first form test were .788, for the second form - .728. Then items were checked for the difficulty and discriminative. For the first form the value of the index range difficulties from .280 to .678, and for the second form - from .261 to .689. The data can be interpreted as follows: despite the fact that there is variation in the difficulty of questions. The test could be recommended for sample with average ability to evaluation personal traits. Discriminative indices for items of both forms of the diagnostic procedure were not lower than .3.

At the next step there was complete parallel-forms test reliability [6]. The correlation coefficient between the forms is .675. Further the two most important parameters of reliability test were identified: test-retest reliability and internal consistency reliability. To check the test-retest reliability of the diagnostic procedure 30 subjects were re-tested (interval in each case, from 6 months). The reliability coefficient was .602.

When checking internal consistency reliability used formulas Spearman-Brown prediction formula and the Kuder–Richardson Formula 20 (KR-20) [6]. The resulting coefficients were significant (their value in each case was greater than .6), indicating sufficient reliability of the test.

Further performed procedures determine the validity of test accuracy of professional evaluation. Ensuring content validity was conducted at the previous stages of the study: at the stage of determining the accuracy of the theoretical construct "accuracy of professional evaluation", at the stage of selection scales for assessment,
identify situations and selection of tasks for video recording, at the stage of obtain peer review and at the stage before and after the video records.

Method of contrasting groups was used to check the criteria validity. There are significant differences in the U-Mann-Whitney test (U = 34, p < .05) between a group of human resource managers (N= 30) and group of psychology students (N = 30).

To test the construct validity were investigated correlations between Accuracy of professional evaluation test and scales of other tests (Compact intelligent test [7], CPI, Evaluation Style and MSCEIT). See Fig 1.

Fig. 1. Correlations between Accuracy of professional evaluation test and scales of other tests (Compact intelligent test [3], CPI, Evaluation Style and MSCEIT);

Scale "Accuracy of professional evaluation" significantly associated with "CIT" scale "Awareness" (r = .306), "Evaluation style" scales "Variability" (r = .541) and "Polarization" (r = .355), "MSCEIT" scales «Understanding and analyzing emotions» (r = .239) and "Perception, evaluation and expression emotions" (r = .445) and "CPI" scales CPI Gi («Good Impression») (r = .631), Py («Psychological-Mindedness») (r = .572 ), Ie («Intellectual Efficiency») (r = .414).

Scale "Accuracy of professional evaluation" has negative correlations with scales CPI: Do («Dominance») (r = -.683), Ac ("Achievement vs. Conformance") (r = -.412).

Thus, the algorithm of creating the test for examining the accuracy of professional evaluation personal traits was developed, which allows the similar diagnostic test to estimate the accuracy of professional assessments of other objects.
Test meets the basic psychometric requirements for developed psychodiagnostic methods (reliability, the construct and criterion validity) and can be applied by specialists engaged in various practical problems that require professional evaluation of diagnostic accuracy, for example, the assessors participating in the assessment-center.

References