EVALUATING COMPETENCES WITH COMPUTERIZED TESTS

Uldis Zandbergs, Janis Judrups
Baltic Computer Academy Ltd.
uldis.zandbergs@bda.lv

Abstract. The current employee competence evaluation methods typically used in companies rely on human experts and require a lot of time spent on conduction. Computerized tests address the problems with the time spent and possible subjectivity of the results. A computerized test has been used with employees of a training center to evaluate their customer service competences. The test results were compared with an established alternative method of evaluation. The use of computerized tests allowed for more precision in the evaluation but the development of the test requires a careful thought. A conclusion has been made that computerized tests are a suitable alternative for evaluation of employees in situations where the company requires a quick and objective evaluation option.

Keywords: competence, automated tests.

Introduction

Accounting of employee competences is a recognized approach in HRM that provides unified criteria for employee evaluation. Competence evaluation at a company is used in order to identify if there are missing competences that may leave an impact on employee working on the assigned tasks.Common methods of competence evaluation are structured interviews, monitoring employees, standardized simulations and role-play. These methods require involvement of experts, considerable time investment, and they involve some subjectivity, therefore they are used sparingly.An organization could save time and receive up-to-date information of employee competences if they used automated competence evaluation solutions fit for a large number of employees that provide reliable results.

Automated computer tests allow spending less time on competence evaluation and repeat them often. By using these tests an organization would need to rely less on the experts in competence evaluation. Such computer tests have been used for evaluation of technical skills. The purpose of this article is to show how computerized tests may be used for the evaluation of soft-skills and what benefits a company can expect by using the approach. Our goal is to show that computerized tests can be used for automated testing of competence.

In order to reach the goal several tasks have been identified:

- to create an example of automated competence measurement case;
- to evaluate a sample competence with the help of computerized tests;
- to evaluate the test results and to compare them with an alternative evaluation method.

Materials and methods

The Context

A study was conducted at a training center that specializes on classroom training with direct presence of a qualified trainer in two general training fields; information technologies (IT system administration, IT system and software solution development, IT security etc.) and soft skills (project management, business analysis, communication skills etc.). Training is conducted in small groups of students consisting of 5 to 12 participants. The target audience of the students is grown-ups and professionals working at their respective fields for several years.

In order to provide the service of high quality the training center requires the trainers to adjust to the training style that is fit for the grown-up groups and to comply with the customer service standards that are set by the training center. The quality of the trainers' work is measured at the end of the training by using course evaluation forms that are filled in by the students. These course evaluation forms contain several questions that are of interest to the training center but there are two questions that are specifically measuring the trainers. These questions are as follows.

- Did the trained conduct the training in a professional manner?
- Would I choose the same trainer again if I attended a new training course at the training center?

Measurement in the course evaluation forms is done using a scale from 1 (lowest) to 10 (highest). The evaluation results have a direct impact on the trainers' salary and the trainers are interested in receiving the best possible evaluation for their conducted training. The training center has set high standards for customer service and considers a score of 8 to be minimal for the trainers in order to continue working with the trainers in the long run. In case of new trainers that have not yet been evaluated for their customer service skills the training center runs a risk of the new trainers not reaching the company goals until sufficient data is gathered from the classes they are assigned to. The training center would benefit from the trainer evaluation method that would allow them to screen potentially underperforming trainers beforehand. Due to the nature of the company application of computerized tests to trainer evaluation has been chosen as a potential solution to the risk. But in order to gauge if the test results would be reliable, a study has been proposed where a sample computerized test would be offered to the existing trainers that had a valid alternative measurement of their competence which in this case was course evaluation forms.

20 trainers have been selected for the test. The trainers had been working for the company for various length of time and had been training different classroom courses. 12 trainers have been teaching IT professionals and 8 trainers have been conducting soft skills training. All of the trainers had an alternative evaluation result available for the comparison. Only these trainers were selected who had received a score from 8 to 10 in the course evaluation forms. The alternative evaluation results have been calculated as a combined average score of all evaluations that have been done for the trainers during the last year if the trainer has been working for the training center for at least a year or as an average score of all the evaluations available for a period of less than a year if the trainers have been working for the training center for a shorter period of time.

The customer service skills of the trainers have been described by using two competences: orientation towards the customers and orientation towards the results. The training center believes that the trainer ability to help the students reach their training goals has a direct relation to customer satisfaction; therefore it should be measured as part of customer service quality indicator. One of the questions that is asked in course evaluation forms (Did the trained conduct the training in a professional manner?) also includes elements of achievement of results. Therefore, all of the computerized test questions have been divided in 8 skill groups corresponding to two competences. The description of these skill groups is as follows:

- identification of the customer needs – trainer's ability to assess the customer needs and the ability to take the customer's point of view when analyzing the customer needs;
- helpfulness – showing interest in the needs of others and balancing their own needs with the customer needs;
- identification of difficult customer service situations – ability to differentiate between a common customer service situation and difficult customer service situation;
- overcoming communication problems – adaptation to the customer communication style;
- dealing with difficult customers – ability to solve difficult problems and to reach the consensus with the customer
- planning of the results – development of the action plan for reaching the goal;
- achievement of the results – execution of the action plan for reaching the goal;
- assessment of the results – evaluation of the results of the action plan for reaching the goal.

Automated evaluation test

In order to evaluate the trainers' customer service competences a computerized test has been created by using the Moodle training management system. The test contained 32 multiple choice questions with 4 possible answers each. The questions were scenario-based describing the situation where the trainers needed to choose the appropriate answers for the situation in question. The answer would describe the behavior that can be expected from the trainers if they were exposed to a similar situation in real life environment. All of the trainers received exactly the same test questions. 35 minutes were allowed for the completion of the test and only 1 try was allowed for each trainer. The time assigned to the completion of the test was kept to the minimum in order to exclude the possible discussions with other trainers.

This test did not threaten to influence the trainers' relationship with the training center, therefore we conclude that the test produced reliable results for further evaluation of the test results.

Quality of the test

In order to produce a test that can be used as a preliminary evaluation method for the new evaluation of customer service skills of the new trainers before they have conducted their first training in the classroom the test results need to be trustworthy. The trustworthiness of the results depends on the quality of the test. The most important parameters of the test quality are validity and reliability [1]. The validity of the test shows how well the test results correspond to the entity that is measured, which in this case are customer service skills [2]. Some authors argue that there is no rigorous method to assess the content validity of the test [3]. At the same time there are several principles that may be followed in order to improve the content validity, one of them being the comparison of the results to other results [4].

For the test to be valid, high results in the test would correspond to high results in the alternative method of evaluation which in our case was course evaluation forms. As we have received high results from course evaluation forms for all the trainers, we can expect that the test results would all be at the same level, ideally at a high level in the terms of the test measurement system.

In order to prove the construct validity of the test [5], several steps have been made to prepare the test:

- The list of the necessary competences and skills has been prepared and aligned with the vision of the management of the training center about a good trainer;
- The list has been compared with the skill list for customer service skills that has been previously compiled by the HR experts;
- Situations that the trainer may encounter during their work have been identified;
- A set of test questions was prepared;
- The previously compiled set of questions has been given to the managers of the training center and human resource experts so that they could evaluate if the questions identify common situations in classroom training and if the correct answers indicate the expected behavior of the trainers;
- The questions were changed according to the information provided by the managers and experts.

A mix of inductive and deductive item generation principles has been used to generate the test questions to make them valid for the purpose of measuring the trainers' competences. In order to comply with the concurrent validity principle, all test results that were received were compared to the alternative evaluation results received from the course evaluation forms.

Test reliability, on the other hand, indicates if the test can truly measure the existing skill in the test subjects. Reliability depends on several factors: psychological and physiological state of the test subject, environment factors, form of the test, and the evaluators of the test [2]. A computerized test allows us to remove two of the factors from the equation, namely, the form of the test and the influence of the evaluator since there is a reasonable expectation that those would be the same for all of the test subjects.

Test reliability may be measured by parallel form and test-retest approaches [5]. During the study the test reliability has not been measured but the test-retest approach has been considered for the studies of the test group in the future.

Results and discussion

The competence test for customer service skills has been completed by 20 trainers of the training center. The time that was required for the trainers covered the interval of 13 to 35 minutes. The results ranged for 4.45 to 7.27 of the possible maximum of 10. The alternative method of course evaluation forms provided the results in the range from 8.58 to 9.87 out of the maximum of 10. Therefore, the average result of 6.28 from the computerized tests may be considered low. On the other hand, all of the trainers handled the test with the results that are close to each other which was expected of the

trainers that received results in close range with the alternative method. Several reasons have been indicated as to why results in thecomputerized tests are lower than expected:

- the trainers were not experienced in using tests for the evaluation of their customer service skills;
- the trainers did not have the necessary skills of working in the specific test environment used in this case;
- the trainers did not recognize the questions and answers as a representation of their everyday situations;
- so far the trainers have been solving the challenges occurring in their everyday life differently than is expected by the managers of the training center and although their behavior has led to high evaluation results from the customers, this behavior does not correspond to the plans of the management.

Table 1

Competence test results for trainers

Trainers	Course evaluation (max 10)	Competence evaluation test (max 10)	1. Identification of customer needs, %	2. Helpfulness, %	3. Identification of difficult customer service situations, %	4. Overcoming communication problems, %	5. Dealing with difficult customers, %	6. Planning of results, %	7. Achievement of results, %	8. Assessment of results, %
1	**8.63**	**4.45**	31	75	25	100	0	25	0	100
2	**9.00**	**4.77**	31	25	25	25	25	75	75	100
3	**9.25**	**5.16**	38	75	25	50	25	75	75	50
4	**9.49**	**5.7**	56	75	50	50	50	100	50	25
5	**9.57**	**5.7**	31	75	50	25	75	75	75	50
6	**9.93**	**5.7**	31	50	50	50	25	100	75	75
7	**9.19**	**6.02**	56	100	50	50	50	75	50	50
8	**8.93**	**6.33**	6	100	50	100	50	75	75	50
9	**9.63**	**6.33**	6	100	50	100	75	50	75	50
10	**8.65**	**6.56**	0	100	75	100	50	75	100	25
11	**8.58**	**6.56**	0	100	75	75	50	75	75	75
12	**9.04**	**6.64**	31	75	50	75	50	75	100	75
13	**9.58**	**6.64**	56	50	75	100	50	75	50	75
14	**9.16**	**6.64**	31	100	25	100	75	75	25	100
15	**9.46**	**6.72**	13	75	100	100	50	75	75	50
16	**9.44**	**6.95**	56	100	50	75	25	75	100	75
17	**9.87**	**6.95**	6	100	75	75	75	75	75	75
18	**9.61**	**7.27**	56	100	75	75	75	75	50	75
19	**8.72**	**7.27**	56	50	75	75	100	75	100	50
20	**9.87**	**7.27**	56	100	75	75	50	75	50	100
Average	**9.28**	**6.28**	**33**	**81**	**56**	**74**	**51**	**74**	**68**	**66**

Although it is believed that the test itself should not be perceived as a threat for the trainers' position, the perception may not be entirely excluded leading to a possible influence to tests results due to negative emotions towards the test which may have led to difference in scores of evaluation forms and test results as a construct-irrelevant variance. There is also evidence in a study by Huff & Sireci that discrepancy between human raters (or evaluation forms in our case) and automated ratings (or tests in our case) is expected and tends to be solved closer to the automated ratings [6].

A simple unit score system has been used for test scoring without using intelligent scoring algorithms. This may have created an impact on scores and in the hindsight the study may have benefited from application of one of the aggregate scoring functions [7].

For more detailed analysis of the reason if and why trainers have been acting differently than expected, division of the test questions in 8 skill groups was used. The results in each group were expected to indicate the difference in the course of action chosen by the trainers and the expected behavior of the managers. The results are shown in Table 1.

The results according to the skill groups indicate that several groups have higher results for all of the trainers participating in the test, yet some groups have lower results for all of the trainers. The division of the results in skill groups provides the training center with a valuable feedback about the skills of the trainers and possible improvement of the course evaluation forms. The training center may use this information to indicate more precise competence gaps for individual trainers at a skill group level and change the possible skill development plans for the trainers.

A more detailed analysis of the first skill group revealed additional significant information. The statistical analysis of the questions showed a negative discrimination index (the measurements for these were taken from the tools provided by the environment in which the tests were conducted). Negative discrimination index gives evidence that a correct answer to this question does not correlate to the overall success in the test [8]. 3 of the 4 questions in the first skill group have shown a negative discrimination index of -3.39 %, -10.58 % and -15.93 %. This means that the results for the first skill group should be approached with caution and in case of improvement of the test validity these questions may be changed.

Also by analyzing the results a conclusion has been made that there is no correlation between the test results and the results received from the course evaluation forms (see Figure 1). For 20 measurements with an error of $\alpha = 0.05$ the minimal acceptable correlation coefficient is $r = 0.444$ [9]. The correlation coefficient in this study was $r(20) = 0.23$, $p > 0.05$. This is not a surprise given that the results from the course evaluation forms have been very close for all of the trainers but this shows that in order to use the test for the evaluation of the trainers in general additional work needs to be done for the test calibration.

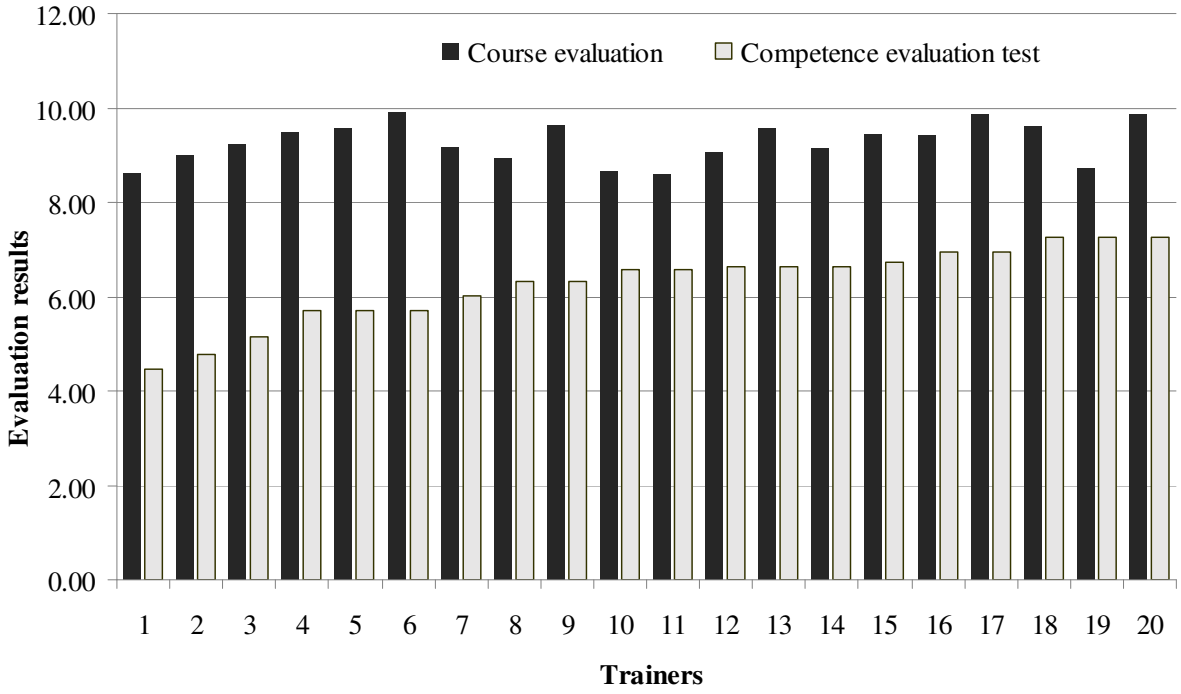

Fig. 1. **Comparison of computerized test results and course evaluation form results**

Conclusions

Several conclusions have been drawn from the study:

1. The trainers that have received similarly high results in the course evaluation form receive similar results in the computerized test.
2. Lower results in general for in the test compared to higher results in the course evaluation indicate that the test questions should be reevaluated and improved in order to use the computerized tests for employee evaluation.
3. The test needs to be calibrated additionally so that it can be used for both trainers with high course evaluation results and low course evaluation results.
4. No correlation between the low results in the test and the low results in the course evaluation forms has been detected because of the target group used in the studies.
5. Some skill groups showed lower results than the test in general. Evaluation of the questions in these skill groups is advised.
6. The test results indicate that course evaluation forms may need revision since their level of detail is possibly lower that the achievable by computerized tests.
7. The test results received were at a higher detail level than those received from other sources and provided the managers with additional information that could be used for the creation of skill development plans for the employees.
8. Computerized tests may be used for employee evaluation if sufficient quality of the tests can be provided.

Acknowledgement

Research is a part of the project "Competence Centre of Information and CommunicationTechnologies" run by IT Competence Centre Ltd., contract No. L-KC-11-0003, co-financed by European Regional Development Fund.

References

1. Understanding Test Quality-Concepts of Reliability and Validity [online] [19.03.2015] Available at: http://www.hr-guide.com/data/G362.htm
2. Testing and Assessment: A Guide to Good Practices for Workforce Investment Professionals. [online] [19.03.2015] Available at: http://www.onetcenter.org/dl_files/proTestAsse.pdf
3. Brahma S. S. Assessment of Construct Validity in Management Research. Journal of Management Research, 9, 2009, pp. 59-71.
4. Golafshani N. Understanding reliability and validity in qualitative research. The Qualitative Report, 8(4), 2003, pp. 597-606. [online] [19.03.2015] Available at: http://www.nova.edu /ssss/QR/QR8-4/golafshani.pdf
5. Building High Quality Examination Programms. [online] [19.03.2015] Available at: http://www.proftesting.com/test_topics/test_quality_validity.php
6. Huff K.L., Sireci S.G. (2001). Validity Issues in Computer-Based Testing. Educational Measurement: Issues and Practice, 20(3), pp. 16-25. [online] [23.03.2015] Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.2001.tb00066.x/abstract
7. Luecht R. M. Operational Issues in Computer-Based Testing. Computer-Based Testing and the Internet: Issues and Advances (pp. 91-114). John Wiley & Sons, Ltd., 2008
8. Moodle Docs. Quiz Statistics Report [online] [19.03.2015] Available at: https://docs.moodle.org/ 27/en/Quiz_statistics_report
9. Is there a relationship (difference) or isn't there a relationship (difference)? [online] [19.03.2015] Available at: http://www.gifted.uconn.edu/siegle/research/correlation/alphaleve.htm