Measuring reliability and validity of functional competencies

Developing people capability is the paramount challenge for any organization. The need of technically expert employees has been never as much before as it is now when the war for talent is so obvious. We need employees who are functionally competent. Whether it is R &D, manufacturing, IT, finance, HR or Sales, we need skill proficiency at all functional levels. But how do we develop our employees to have the required level of technical knowledge and skills?

Many organizations have attempted to develop a functional competency model to address this issue. They chart out required levels of knowledge/skills at different levels of expertise and measure the employees on the same by using a proficiency scale. For example, a Sales functional competency model will have all the technical competencies listed down required for a superior workplace performance for beginner-level, middle-level and expert-level. Usually, the competency model is developed by learning and development experts (usually HR) in collaboration with the business line managers. But how can we be sure that such a model will result into the desired objective of developing good performers? Can we carry out statistical tests of reliability and validity on these competencies?

The answer is yes, we can. Note: I would be assuming that we already have a functional competency model and my objective in this article will only be to test the reliability and validity of the model.

What to test?

Before we start to perform any test we want to establish what it is that we are trying to validate or check reliability of. We are trying to validate that the expertise on the functional competency will have a direct impact on the performance of the employee. For example, if the employee is good in the ‘product knowledge’ he would be good at selling it to the customers. When we say we are measuring reliability of competencies we mean that if the same employee is measured for a competency by different individuals using a proficiency scale, both will end up giving same score to the individual.

How to test?

Reliability - In order to test the reliability, we can perform an inter-rater reliability test. Subject matter experts can score functional expertise of employees using a proficiency scale (1-Beginner, 5-Expert). This score can then be tested on the correlation between various SMEs. If the correlation is more than 0.6 we can say that the experts are in agreement and the model is statistically significant.

Validity – Here we can perform 3 type of tests on the competency model.

1. **Face/Content Validity** – We can score the competency model on ‘How essential the identified competencies are for superior job performance’ on a scale of-
   - **Negative 1** – if not essential
   - **Zero** – if not relevant,
   - **Positive 1** – if essential
   This question can be addressed to all the employees(Face) and 7-10 Subject matter experts(Content) and calculate the Aiken V (Aiken 1980) values for each case. If the value is more than 0.7 we can say that they are statistically significant and the content mentioned in the model (or as defined by the competencies) will result into good performance.
2. **Criterion Validity** – Here we calculate correlation between employee job performance and his functional expertise. The idea is that if the employee is an expert at all the required competencies then he will produce superior performance at the workplace. To perform this test we can take scores of current expertise (Scale: 1-beginner, 5-expert), on all the required competencies for the function, and performance scores (Annual ratings) of all the employees in the function. Then we perform a Spearmen’s rank correlation test and calculate its coefficient. If the coefficient is more than 0.7 we can say that the functional capability will result into good performance at the job.

3. **Differentiability Test** – The competencies identified for a function, if are able to differentiate between a good and an average performer then we can be sure of the validity of the competency model. We can create two groups of employees based on their past performance- GOOD and AVERAGE. Then we can calculate score (Scale: 1-beginner, 5-expert) of their expertise with the help of supervisors. Then we perform a Student t-test to check if the means score of expertise really are statistically differentiating (Zhang 2012). It the results are significant (p-value < 0.05) then we can say our model is robust enough to differentiate a good and an average performer.

Why do we need to be so worried about validity and reliability of the competency model, which already has taken so much effort to come up with? The answer lies in the question itself. We want to reduce time taken to develop functional competency model. This can be done by delegating task and developing the model in a decentralized manner. But in order to be successful in performing this task in such a decentralized manner, we need control over the whole affair. This test provides that control. If the model developed by different teams is valid and reliable we will be more confident of achieving the desired state of developing functionally competent employees at workplace.

**Reference**
