

## **An Analysis of the Importance of Validity and Reliability as Measurement Standards of Selection Processes**

*Anisa Sultana\**

### **Abstract**

This paper shows that Validity and reliability of selection methods not only ensures the right candidate to get the job but also remain responsible to predict the job performance of the newly hired employees. The paper also examines the application and the co efficiency of the selection methods and further explains why sometimes both of these are lost. By discussing various traditional selection methods, this paper argues that a valid and reliable test, structured interview and unbiased interviewers are the most important elements of a successful selection strategy. In addition, it explicates that though other measures like Generalizability, Utility and Legality are sometimes considered to assess the selection outcome yet their importance as measurement standard of selection methods is not as high as reliability and validity .The paper ends with the discussion of the limitations and finally, further research directions are suggested.

**Key Words:** Reliability, Validity, Selection methods, Measurement standards

---

\* American International University Bangladesh (AIUB)

### **Introduction:**

Employers are always in a dilemma to ensure the appropriate selection methods to hire the right person for the organizations. They want to establish

the selection procedure to be flawless enough to predict the job performance of the newly hired employee. Hence 'Reliability' and 'Validity' are considered as two prime variables by the researchers and occupational psychologists to measure the effectiveness of selection process. Numerous researches have already been conducted on the importance of the validity and reliability of the selection procedure and a number of factors that may bias the outcomes associated with selection methods have also been identified. Assessment methods and tests of an employment selection process should have validity and reliability information to endorse their claims that the test is an appropriate measure. Reliability is a very important notion and works in tandem with Validity. A test can be reliable but not valid for a particular reason, however, a test cannot be valid if it is unreliable.

Mehrens and Lehman (1987) defines reliability 'as the degree to which interviews, tests, and other selection procedures yield comparable data over time and alternative measures, It is the measure of how stable, dependable, trustworthy, and consistent a test is in measuring the same aspect each time (Worthen et al., 1993)'.

However, losing reliability throughout the selection process is very common. One of the major reasons of it, probably, is pretty subjective and thus it exhibits inconsistency in measurement. For example, a candidate's test score may vary for many reasons: (a) the amount of the characteristics are being measured may change across time; (b) the particular questions asked in order to infer a person's knowledge could affect the score; (c) any change in directions, timing, or amount of rapport with the test administrator can cause score variability; (d) inaccuracies in scoring a test paper can affect the scores and finally (e) such things as health, motivation, degree of fatigue of the person, and good or bad luck in guessing can cause score variability.

On the other hand, validity in a selection process 'is the degree to which a test or selection procedure measures a person's attributes (Worthen *et al.*, 1993)'. 'It focuses on the extent to which certain inferences can be made from test scores or other measurement (Mehrens and Lehman, 1987)'. A selection process is valid if it helps employers to increase the chances of hiring the right applicant for the job. It is not valid on its own rather relative to a specific purpose and function. For example, the test that effectively predicts the work quality of a teacher will be useless for a database administrator. In

addition, a critical component of validity is reliability. Validity is characterized not only as the predictor of positive outcomes of a selection approach, but also embodies how *consistently* (i.e., reliably) it does so.

This paper reviews the importance of validity and reliability in selection process by analysing the applications of different selection methods in organizations. It also (a) discusses Thorndike’s (1971) model of reliability to assess the consistency of testing outcome; (b) present a case of an IT firm in UK that illustrates reliability of a test when repeated and inconsistency in marking; (c) identifies the correlations between predictive validity of selection methods; and (d) evaluates the role of other measurement standards like generalizability and utility and legality at the end. This paper has brought together a wide variety of studies which may provide us with avenues for future research.

**Thorndike’s (1971) model of Reliability:**

Thorndike (1971) proposed a model of test fairness that shows the consequences of the test, or the testing outcome. He suggested that selection based on test scores is unfair if the ratio of the proportion that would have been successful differs across sub-groups of test takers. Stated differently, Thorndike’s model suggests the use of a given test is fair if the ratio of two proportions is constant across two groups; that is, (a) the proportion of candidates (from one group) selected on the basis of test scores and (b) the proportion of candidates (from the same group) that would have been selected on the basis of successful job performance. If this ratio of proportions is constant across groups, the test is said to be unbiased and its use is fair.

**Table 1:** Reliability as Stability overtime

Applicants	High Reliability		Low Reliability	
	Test	Retest	Test	Retest
Robert	90	93	90	68
Ahmed	65	62	65	88
Lotoya	110	105	110	67
Chan	80	78	80	111

**Table 2: Reliability as Consistency (Inter-rater Reliability)**

Applicants	High Reliability			Low Reliability		
	Rater 1	Rater 2	Rater 3	Rater 1	Rater 2	Rater 3
Robert	9	8	8	9	5	6
Ahmed	5	6	5	5	9	4
Latoya	4	5	5	4	2	7
Chan	8	8	8	8	4	2

A study on an IT firm in East London, UK (2009) shows how Thorndike’s reliability model actually measures the consistency and inconsistency of Test results during the selection process. The tables above present information about the written test results of five applicants that stand for the reliability as stability overtime (whether high or Low) and consistency in inter-rater reliability (High or Low). Table 1 shows that as the test has taken twice, the upper part of the score board demonstrates high reliability as all the applicants are consistent in their score at the retest. However, the bottom part of the table shows low reliability as some of the scores are inconsistent at the retest time. On the other hand, table 2 recorded the marks of the same applicants rated by three examiners and there is a consistency between the evaluators in marking at the first instance, but the bottom of the table shows that applicants did not receive same or similar marks during the second exam rather there is a huge inconsistency in evaluation by the evaluators. Hence, if the aptitude level of Mr. Robert (the first candidate) is 90, he is supposed to obtain more or a little less than 90 at the retest but as he obtains less than 68, that does not reflect his ability and making an employment decision based on that is certainly a big mistake. Similarly, when two evaluators rate him equally and the third and fourth evaluator’s marking varies highly, the result is not valid and reliable.

**Validity in Selection Process:**

According to Stephen Pilbeam and Majorie Corbridge, the main concern of the recruiters is the predictive validity of selection methods: how effective is an interview, a test or an assessment centre in predicting the eventual job performance of a candidate. The predictive validity of selection methods can be compared by using a correlation coefficient to measure the probability that the selection process that has been executed will predict performance in a job. A correlation coefficient of 1.0 represents certain prediction; a correlation coefficient of 0.5 approximates to 50 percent chance that the selection method will predict performance, and a correlation coefficient of 0.0 indicates that no connection between the selection method rating and job performance.

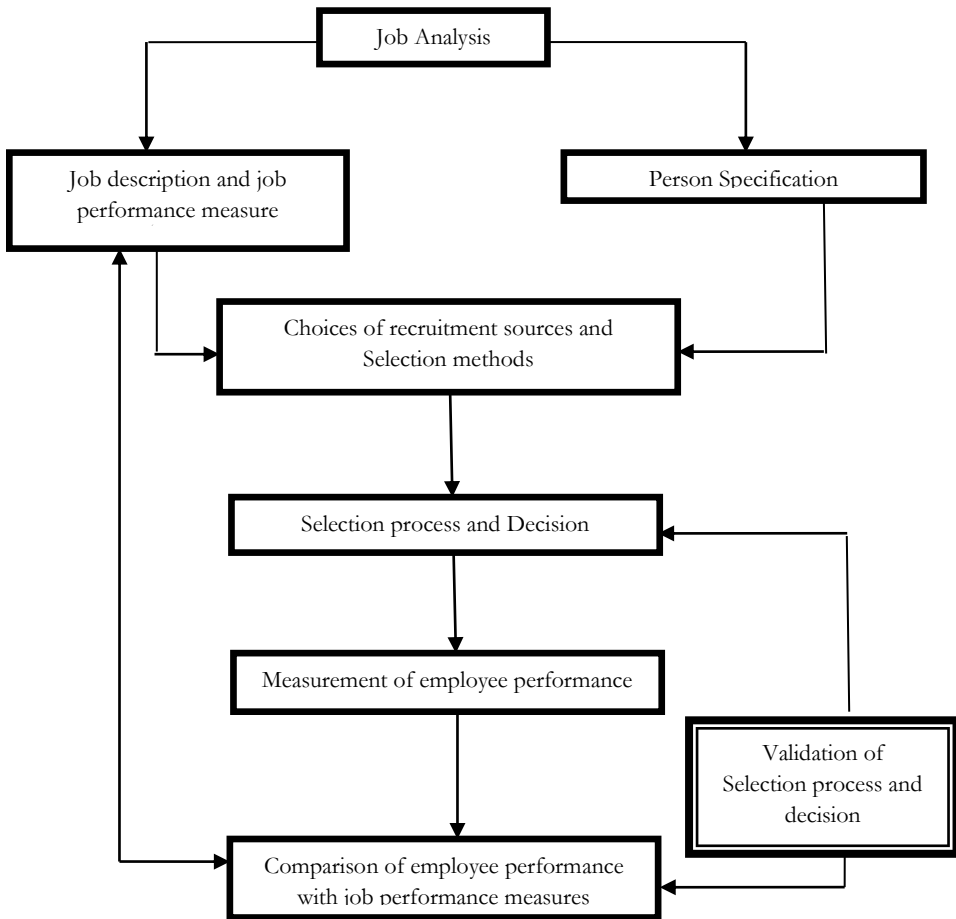
**Table3<sup>1</sup>:** Predictive validity of selection methods: a summary of correlations.

<b>Correlation Coefficient</b>	<b>Certain Prediction</b>
1	×
0.9	×
0.8	×
0.7	Assessment centres for development
0.6	Skilful and structured interviews, Ability Tests, including numerical and verbal reasoning.
0.5	Work sampling
0.4	Assessment centres for job performance, Bio data, personality assessment
0.3	Unstructured interviews
0.2	×
0.1	references, Interests, Years of job experience
0	Graphology, Astrology, Age

Nevertheless, one must comprehend the process of selection to assess the importance of reliability and validity and at the same time when both of these attributes could be lost, needed to be ascertained. A hiring decision is not only taken after the written test or interview; the process starts much before the recruitment decision is taken in a company. Hence it should be understood that ensuring reliability and validity only during selection process does not warrant the right person to be hired. Therefore to get the right employee performing in the organization, validity and reliability of job analysis up to decision making must be ensured.

---

<sup>1</sup> Table 3 contains a summary of rounded figures (Smith and Robertson,1993;Shackleton and Newell,1991;Smith,1994;Arnold et al.,1998;Conway et al ,1995;Fowler,1997,1998;CIPD,2004a,2005; Barclay, 2001; Robertson and Smith, 2001; Harel et al., 2003)



**Figure 1:** Validating stage of the selection process and decision (Source: Pilbeam S & Corbridge M, 2006: People Resourcing)

Experts and professionals advise that job analysis, job description and job specification are prerequisites of an effective selection method. But during these processes, there are many ways in which ‘Validity’ and ‘Reliability’, either or both can be lost. As job analysis is an information gathering process, it generates information which is converted into the tangible outputs of a job description and a person specification. Achieving ‘Validity’ and the ‘Reliability’ in job analysis process is a must as it plots the ground for proper execution of Selection methods. However, as it is almost subjective and vast, sometimes job analysis cannot identify the job requirements, working conditions,

responsibilities precisely and it cannot ensure the validity and reliability of job description and job specification.

In the case of Selection Methods, there are some methods which also have high validity and reliability and some do have moderate and low. Before examining the degree of reliability and validity, we need to see the proportion of the selection methods that are being used and consequently achieved much popularity according to CIPD Survey 2004b.

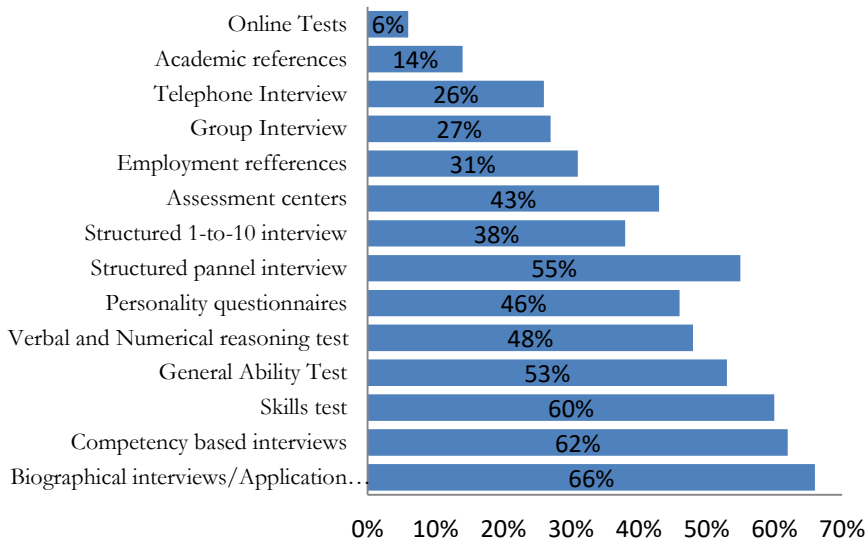


Figure 2: Popularity of Selection methods (percentage of surveyed organizations: CIPD, 2004b)

The bar chart clearly shows the popularity of selection methods in UK organisations. Stephen Pilbeam and Marjorie Corbridge provide another comparative dimension, with three broad groupings identifiable. Interviews, references and application forms, termed the classic trio, have almost universal popularity despite evidence of low predictive validity and lack of reliability in practice.

## Considering Interview:

Reliability also looses at the time of interview. There is low reliability among interviewers regarding what questions should be asked of applicants and how applicants should be evaluated. Applicants' appearances including their external beauty, facial attractiveness, cosmetics, attire, gesture, posture and expressions bias interviewer's assessment. A recent study of appearance bias in the interview was provided by an experiment finding that moderately obese applicants (especially female applicants) were much less likely to be recommended for selection (Pingatore, Dugoni, Tindale, & Spring, 1994). Non-verbal cues (eye contact, smiling, etc.) also bias interviewer ratings (Dipboye, 1992). Even if information is obtained prior to the interview or during its early stages, it dominates interviewer judgments. Research has shown that on average, interviewers make a selection decision within only 4 min of a 30-min interview. Interviewers with positive first impressions of an applicant sell the applicant more on the company, do more recruiting, and tell them more about the company (Dougherty, Turban, & Callender, 1994).

Interviews are highly expected by both candidates and managers because it provides information to predict performance, and also gives an opportunity for the interviewer and interviewee to meet face to face and exchange information. Thus, usually, an interview has the predictive and face validity. Yet research evidence highlights the limitations of the traditional interview as a poor predictor of a candidate's performance in the job. Anderson and Shackleton (1993) while drawing the reasons of loosing reliability and validity in an interview, include the reasons, such as:

- The self-fulfilling prophecy effect: Interviewers may ask questions designed to confirm initial impressions of candidates gained either before the interview or in its early stages.
- The stereotyping effect: Interviewers sometimes assume that particular characteristics are typical of members of a particular group. In the case of sex, race, disability, marital status or ex-offenders, decisions made on this basis are often illegal. However, the effect occurs in the case of all kinds of social groups.
- The halo and horns effect: Interviewers sometimes rate candidates as 'good' or 'bad' across the board and thus reach very unbalanced decisions.



- The contrast effect: Interviewers can allow the experience of interviewing one candidate to affect the way they interview others who are seen later in the selection process.
- The similar-to-me effect: Interviewers sometimes give preference to candidates they perceive as having a similar background, career history, personality or attitudes to themselves.
- The personal liking effect: Interviewers may make decisions on the basis of whether they personally like or dislike the candidate.

### **Validity of Interviewer not Interview is required:**

Research has consistently demonstrated that interviewers differ widely in their ability to accurately forecast job performance. This has caused some researchers to wonder whether we should focus on the validity of the interview or on the validity of the interviewer. It has even been argued that since interviewers differ in their evaluations and use different parts of the rating scale, aggregation of ratings across interviewers systematically understates the ability of interviewer ratings to predict job performance (Dreher, Ash, & Hancock, 1988). Recently, Pulakos, Schmitt, Whitney, and Smith (1996) completed the most comprehensive investigation of individual differences in interviewer validity to date.

Analyzing the decisions of 62 interviewers, who completed an average of 25 structured interviews, Pulakos et al. (1996) showed differences in interviewer validity (correlation between an individual interviewer's ratings and job performance for the interviewees who were hired) ranging from  $r = 0.10$  to  $0.65$ . Contrary to Dreher et al.'s (1988) arguments, they found that aggregation across interviewers did not lower the estimated validity of the interview. As Pulakos et al. (1996) note, this difference may have been due to the fact that Dreher et al. (1988) studied unstructured interviews while Pulakos et al. (1996) studied structured interviews. Thus, there appear to be vast differences in interviewer validity, which suggests the somewhat ironic (though obvious) point that those who do the selecting (interviewers) need to be selected in a careful manner. What background characteristics lead to individual differences in validity is an obvious question for future research.

### **Structuring the interview:**

Structuring the interview helps improve its ability to predict performance in the job and a growing number of employers are now taking this approach. A 'structured interview' assures that questions are planned carefully before the interview (applicable both for face to face and telephonic interview), all candidates are asked the same questions, answers are scored using a rating system and questions focus on the attributes and behaviours needed for the job.

### **Considering Psychometric Test:**

Psychometric Tests are used to measure individual differences like personality, aptitude, ability, attainment or intelligence. Though, Construct validity is particularly important in psychometric assessment, it is a recent practice to consider Content Validity (refers how well a test samples the kinds of behavior it is seeking to test) and Criterion Validity (refers how well a test relates to other external measures) of the process. Some concerns of psychometric testing deal primarily with measurement issues, such as determining how accurately or honestly a person has responded to a test. To deal with these concerns, it is to be remembered that the characteristics of the test should be known to be satisfactory through data that show an adequate range and variance. Thus, the test must be neither too hard nor too easy, and there must be adequate variability.

### **Considering Reference:**

The purpose of references is to obtain information about a candidate's employment history, qualifications, experience and/or an assessment of the candidate's suitability for the post in question. It is estimated that one in eight people exaggerate or falsify their qualifications! CIPD 2008 Recruitment, retention and turnover survey also shows that 25% of the employers withdrew job offers during 2007 because someone lied or misrepresented their application.

Employers are highly required to remember that because a job applicant is being asked to give the names of referees, they are highly unlikely to choose anyone who will give them a bad reference and therefore potential inaccuracies may arise. As bad references are very rare and previous employers may have gone out of business, they may not show interest and incapable of judging performance or communicating their opinion. Where a reference appears inadequate or incomplete, prospective employers should check,

preferably by telephone, to probe further. References tend towards ‘leniency errors’ and, like interviews, offer a great deal of scope for both direct and indirect bias. Employers can improve validity and reliability by using a highly-structured rating form. References tend to be more reliable if the old and new job is very similar in content. Questions on the reference form should be based on candidates’ job analysis and facts rather than merely focusing on opinions of the referee. Moreover, before seeking for referral, employer must provide with relevant information about the job vacancy and make sure that the referee has had an opportunity to directly observe the requested form and identified competences of the candidate.

Other methods like Ability tests, Personality assessment and Assessment centres have medium, yet increasing popularity and bio data, graphology and astrology have low popularity (Shackleton and Newell, 1993; CIPD, 2004b, CIPD 2005).

**Table-4<sup>2</sup>** : Validity and validity Co-efficiency of Selection Methods with predictors

<b>Predictor</b>	<b>Selection Method</b>	<b>Validity</b>	<b>Validity coefficient</b>
Assessment centres	<i>Assessment Centres</i>	High	0.41-0.43
Work samples	<i>Work samples</i>	High	0.38-0.54
Peer evaluations		High	
Physical ability		Moderate-High	
Cognitive ability and special aptitude	<i>Cognitive Tests</i>	Moderate	0.25-0.45
Biographical information		Moderate	
Self assessments		Moderate	
Interviews	<i>Interviews</i>	Low	0.14-0.23
Personality	<i>Personality Tests</i>	Low	0.15
Interest		Low	
Seniority		Low	
reference checks		Low	
Academic performance		Low	

**Other Selection Measures:**

---

<sup>2</sup> Table 4 compiled by Wendy Banfield (2000) from data in plain type in (Neal, Schmitt & Raymon Noe, ‘Personnel Selection and Equal Employment opportunity’ in International review of Industrial and Organizational Psychology, eds, Cooper Cary L. & Robertson, Ivan T.

In practice, sometimes the predictive validity correlations are of limited value because they cannot be generalised to particular organisational situations. Other measurement standards like Generalizability, Utility and Legality are then or simultaneously considered. Generalizability is the extended standard of validity. If the validity or trustworthiness can be maximized or tested, more “credible and defensible result” (Johnson, 1997, p. 283) may lead to generalizability’. But it is obvious that generalizability cannot be achieved without validity as it is the degree to which one can extend it to other contexts. Utility on the other hand is the degree to which information from selection method enhances bottom line effectiveness. It also searches the answer of the question: ‘does the test increase the accuracy of the selection process?’

## Conclusion

Implementing the reliability and validity in the tests during the selection process requires obtaining information from all parts of the recruitment process. This is how errors (if any) are identified quickly to take necessary steps. In addition, it is a necessity to make reviews with all staffs who are involved in the selection process, including new recruits if possible (predictive validation). Optimum use of reliability and validity in selection strategy makes selection process effective and trust worthy, save time and money, and help to implement quality and talented minds across an organization. In practice, other dynamics like generalizability, Utility and Legality are being reasonably considered but do not have equal or greater importance than reliability and validity.

However, this study is subject to few limitations. First, it has explained the importance of reliability and validity in selection process but not included the role of Equal Employment Opportunity (EEO). It could have considered this legal issue that is closely connected to validity: employment discrimination. Moreover, other measuring aspects like generalizability, utility and legality have not been explained and distinguished enough to establish the greater effectiveness of reliability and validity. Finally, more examples of the cases when reliability and validity are lost paired with appropriate theories could also be incorporated.

Future research could investigate the role of ‘discrimination’ in losing reliability and validity in a selection process. In addition, comparative study on all measurement standards of selection process and their application would be a worthy area of upcoming research.

## References

- Adkins, C. L., Russel, C. J., & Werbel, J. D.* (1994): Judgments of fit in the selection process: The role of work value congruence, *Personnel Psychology*, Volume 47, pp. 605±623.
- Aibright, L., Kenny, D. A., & Malloy, T. E.* (1988): Consensus in personality judgments at zero acquaintance, *Journal of Personality and Social Psychology*, Volume 55, pp. 387±395.
- Alderfer, C. P., Mccord, C. G.* (1970): Personal and situational factors in the recruitment interview, *Journal of Applied Psychology*, Volume 54, pp. 377±385.
- Anderson, N. and Shackleton V.* (1993): *Successful selection interviewing*, Oxford: Blackwell.
- Arkin, A.* (2005): Burden of proof. *People Management*, Volume 11, No 4, 24 February, pp. 30-32
- Brewster, C.* (1995): Towards a 'European' Model of Human Resource Management, *Journal of International Business Studies*, Vol. 26, No. 1 pp. 1-21
- Fox, G. and Taylor, D.* (2000): *The complete recruitment and selection toolkit*, London: Chartered Institute of Personnel and Development.
- Ghiselli, E. E.* (1973): Personal psychology the validity of aptitude tests in personal selection, volume 26, pp. 461-477
- Hackett, P.J.* (1998): *The selection interview. Management shapers*, London: Institute of Personnel and Development.
- Hoynes, M.* (2005): How to interview effectively, *People Management*, Volume 11, No. 24, 8 December, pp. 40-41.
- Judy, P., Michael, A., McDaniel, Deborah and Whetzel,* (1994): The Validity of Employment Interviews: A Comprehensive Journal of Applied Psychology, Volume 79, No. 4, pp. 599-616

- Murphy, N.* (2007): Job interviewing: an overview of IRS's research, IRS Employment Review, No 785, 18 June.
- Pilbeam, S. and Corbridge, M.* (2006): People Resourcing 3<sup>rd</sup> Person Education.
- Reeve, C. L. and Bonaccio, S.* (March 2009): Measurement Reliability, the Spearman–Jensen Effect and The Revised Thorndike Model of Test Bias, International Journal of Selection and Assessment, Volume 17 Number
- Suff, R.* (2008): Background checks in recruitment: employers' current methods. IRS Employment Review, No 896, 17 April.
- Super, E. D. and Bohn J. M.* (1971): Occupational psychology, Behavioral Science In Industry Series, Tavistock Publications.
- Taylor, S.* (2005): People resourcing. 3<sup>rd</sup> ed. London: Chartered Institute of Personnel and Development.