

Retesting—What is the Right Time Period?

Authors:

Matthew O'Connell, Ph.D., Select International
Matthew Reeder, M.A., Michigan State University
Dennis Doverspike, Ph.D., University of Akron

Background

Administering a test costs time and money, so giving a test multiple times when it will not significantly change the outcome for a candidate is a waste of both time and money. Moreover, there is always the possibility that practice effects could become a factor if the test is taken again within a short period of time, and therefore raise the score of the individual without any “true” change in the attributes that you are testing for in the first place.

A basic property of a good test, or of the scores generated by a good test, is adequate levels of reliability or consistency. Test-retest reliability is the measure of a test’s temporal consistency when the test is taken by the same individual or group of individuals at least twice. Traditionally, test-retest reliability is demonstrated by correlating the individual scores from Time 1 with the individual scores from Time 2. Test reliability is never perfect. There are always things that affect a person’s score from one time to another. There tends to be a fair amount of “noise”, which in testing terminology is referred to as “random error” that affects scores. It is important to try to separate this noise from true changes in abilities, characteristics, knowledge or whatever the tests are designed to measure (Schmidt, Le, & Ilies, R., 2003; Schuerger, & Witt, 1989).

“Noise” can be reduced by trying to maintain consistency in the environment in which the tests are administered. For instance, if the first administration was proctored, then ideally the second administration should be proctored. If the first administration was administered in a web- or computer-based format then the second administration should also be delivered in a similar medium. Additionally, if alternative forms of the assessment are used in subsequent administrations, then those alternate forms should be created to be as equivalent as possible. In other words, it is important to remove extraneous factors and attempt to maximize consistency across administrations.

From a practical perspective, the question is not whether or not you should retest, but when you should retest. Most organizations allow candidates to reapply or retest after a certain period of time has expired. This allows candidates who, for one reason or another, were not ready to meet the standards the first time they applied, to try and improve themselves and test themselves at a later time.

Although there are no clearly defined standards for retesting, the SIOF **Principles** provide some guidance on the topic. *“Generally, employers should provide opportunities for reassessment and reconsidering applicants whenever technically and administratively feasible. In some situations, as in one-time examinations, reassessment may not be a viable option. In order to facilitate consistency of treatment, the administration guide should clearly explain whether applicants may be reassessed and how reassessment will take place. In some organizations, specific time intervals must elapse. In others, although difficult to evaluate, significant developmental activities must have occurred prior to reassessment.”* (p. 57)

Recommendation

Based on research looking at test score changes over time leads us to conclude that retesting periods should be set at one year or longer. Retesting after shorter intervals appears to maximize the amount of “noise” in test scores. Multiple studies of various types of tests, including cognitive-type measures (e.g., verbal and numeric reasoning), information processing type measures (e.g., simulations), as well as self-report personality measures, indicate that score changes within twelve months are substantial and then tend to stabilize off after one year. This suggests that those early changes are more based on short-term practice effects and other extraneous factors, i.e. noise, than they are on true changes in the underlying traits or attributes.

Consider the following results of a study that evaluated the impact of different retesting periods (Reeder, Doverspike, & O'Connell, 2008). The study looked at the same assessment taken twice by candidates for the same position. Candidates either took the test a second time: (a) within six months of the first administration; (b) between six-twelve months; or (c) greater than twelve months. Of the nine competency areas measured, the greatest change from the first administration, almost 80% occurred within one year. Only 20% occurred after one year. While that is admittedly a gross measure of change, it certainly speaks to the extreme variability that occurs within the first twelve months.

Even though the majority of variability occurs within the first twelve months, the differences, in general, are actually quite small. In the study mentioned above, the average standardized mean difference (SMD) across all administrations for non-cognitive measures was $d=.20$, and $d=.22$ for cognitive measures. This is consistent with findings from a meta-analysis on retesting which found that the average standardized mean difference to be $d=.21$ (Hausknecht, Halpert, Di Paolo, & Moriarty, 2007). According to Cohen (1988), who actually defined the standardized mean difference as a scale-free means of explaining effect size, standardized mean differences of 0.2 or less are insubstantial, trivial or otherwise not worth worrying about. Therefore, while people do clearly change over time, there is a substantial amount of consistency, at least in terms of the type of attributes that we are typically interested in from an employment testing perspective. Real change is important and should be identified. Clearly though, we should not place great emphasis on relatively trivial and insignificant change based mostly on practice effects and noise. Therefore, it seems quite reasonable to establish twelve months as a minimum time period for retesting in most situations.

Lastly, as is commonly the case in selection systems involving multiple test batteries, if a candidate has passed a given test but failed another test in the selection process, the candidate should not be made to re-take the particular phase or test he/she has already passed. The assumption here is the test given was a "good test", as described above. Mainly, there is simply greater risk in retesting a candidate who already passed a test, given the possibility that he/she might fail the second time. The point being, unless a substantial amount of time has passed, the job requirements have changed, or the candidate's knowledge or attributes have changed, the original test score should be sufficient.

Selected References

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., and Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*, 373-385.
- Reeder, M., Doverspike, D., & O'Connell, M.S. (2008). *Temporal stability and retest effects across personnel selection methods*. Paper presented at the 23rd annual meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Schmidt, F. L., Le, H., and Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of Individual differences constructs. *Psychological Methods, 8*, 206-224.
- Schuerger, J.M., & Witt, A.C. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology, 45*, 294-302.
- Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the validation and use of personnel selection procedures*. (Fourth Edition), College Park, MD: Author.