Retesting Policies for Training and Education Programs

Eric T. Vincent
CEO & President
Industrial-Organizational Psychologist

When a student fails an exam from a professional training program, they usually want a "do-over", a retest opportunity to try to pass the test. Though such requests are common, decisions related to providing retest opportunities involve much more than simply giving the student another try. For example, should the same test form be used again for the same student? How many times can a student retest? Does there need to be a waiting period before the test can be taken again? Shouldn't the same retesting policies apply to all testing programs? Without meaning to, programs may forego careful consideration of these issues and inadvertently expose the use of the test result to validity and reliability threats. Thorough answers to these questions should to be determined before retesting occurs. The purpose of this white paper is to provide an overview of the primary issues related to developing adequate and useful retesting policies and procedures for professional training and education programs.

Though it would be easier and tempting to simply mimic policies of other programs, determination of the best practices to implement must focus on the relevant psychometric, personal, and contextual factors that come from an organization's own program (American Educational Research Association [AERA], et al., 2014, p. 4). At issue is the fact that organizations develop their own tests. They aren't using another existing test. Typically, they develop tests that measure knowledge, skills, or abilities that change via training. In some cases, the focus of training is to change values, attitudes, beliefs, and sometimes personality. But regardless of what is measured, when results are used to make an employment decision (e.g., hiring, promotion, advancement) or to protect the public (e.g., safety), adherence to professional standards becomes critical (AERA, et al., 2014, p. 3).

Because such programs develop their tests, they have the responsibility to follow generally acceptable professional practices in the educational, training, and testing industry. The industry, however, does not issue absolutes regarding policies and procedures, such as test retakes. Instead, qualified professionals are expected to use their education and training to consider all relevant issues, such as published professional standards, relevant laws, and recent research when setting such policies. Much of this information is taught in graduate level educational measurement and psychometric courses. Consequently, it is challenging for educational and professional training programs to develop sound practices on their own and to understand the varying underlying issues that could challenge their test results.

Of primary importance in establishing these policies is to prevent threats to the evidence that the test measures achievement of learning the **course** content, *not* the test content. In other words, the point of the test is to measure the learning that was achieved, and the retesting policies should support this purpose. As a result, effective retesting policies should address the following key issues:

- Restrict overexposure to the exam (e.g., learning the test items instead of achieving its related learning outcome).
- Avoid test security breaches (e.g., test fraud).
- Reduce classification errors (e.g., "passing" those who haven't achieved the learning outcomes and "failing" those who have).

When test results from these programs indicate a failing score, it means that the learner has not demonstrated an acceptable level of achievement of the intended learning outcomes. This conclusion is

appropriate and assumes that the test already has adequate reliability and validity evidence for its use. If the test does not have this evidence, retesting with it is misdirected and ill-advised. In such cases, immediate efforts should be made to prevent any further use of the test until it is fixed. Programs should ensure that adequate evidence is achieved *prior* to using a test. If evidence emerges that there are issues that undermine its valid use, actions should be taken as soon as possible to remedy the problem.

But if the test is a sound instrument, and there are likely situations that could affect the test result, it may be appropriate to allow a retake. For example, learners may be sick or too nervous, or there may be unplanned interruptions to the administration of the test (e.g., fire alarm, computer malfunction). Such situations arise, and a program may find it is in their best interest to prepare for an opportunity to retake a test without undermining the actual test program. When creating retest policies, consider the following to avoid undermining the valid use of the test result:

- **Number of retesting attempts** – When a quality test instrument is used, it is typical to allow one "do over", called a "retest". Procedures for requesting issuing a retest should be clearly documented. These policies and procedures should be communicated to learners throughout the course and after taking the exam. In some limited cases, it may be appropriate to provide a second retest attempt, which is a third administration of the test to the same examinee. Such situations should be carefully considered and the reasons for three attempts without training remediation should be documented.
- **Number of items on the exam** – Small numbers of items on a test form are easy for learners to remember from test to test. Using a form that simply changes the order of presentation of the items does not eliminate this problem. Professional practices advise using alternate and unique forms (without repeating items) or to impose adequate time between administrations of the same test with a small number of test items to limit the influence of prior testing on the next test score. Programs should consider whether there are enough test items that the result can adequately establish attainment of the learning outcomes.
- **Waiting period between the exam** – In cases when only one test form exists or there is significant overlap of items between active test forms, a time delay should be considered prior to retesting. This is done to reduce the influence of a learner's retention of items from the first test experience. Programs should consider the amount of time between test administrations that would be enough to reduce the influence of experience with the items, but that would also avoid decay of the learning that was achieved in training. A 24-hour period may not be enough, but 30 days may be too much. This will likely be influenced by the nature of the content learned and the available access to learning materials between test administrations.
- **Number of alternate forms of the exam** – A program with only one test form is especially vulnerable to the test items being released into the public. Once the items are available to the public, the program can no longer ensure that the test is measuring what it is intended to measure without being contaminated with foreknowledge of the test. They are in a much better position to defend the valid use of their test if they have alternate and parallel forms of the exam that have unique sets of test items. Programs with only one form of a test should avoid immediate retesting and should impose adequate time between the test administrations to limit the influence of taking the test the first time. In addition, programs with immediate retesting should also examine their test forms to prevent items on the first form teaching people answers to items on the subsequent test administration without any training of the underlying learning outcomes.
- **Complexity and variety of content** – When tests contain simple items that are easy to remember, memorize or research, the test content is easy to remember for a retest and to communicate to others. Simple items measure lower levels of cognitive targets such as recall, define, describe, and recognize. Retesting policies in these cases should be sure to use an alternate form with unique items. In addition, tests that focus on only a few similar content areas are also more vulnerable to being easily remembered. In addition to using an alternate form with unique items, programs should consider creating items that cover the breadth of the training content more completely so a more robust set of alternate forms can be developed.
- **Test security practices** – Securing the exam items from being communicated to others is essential to ensuring that the test is measuring what it intends to measure. When examinees

have seen the items before they take the test, the result isn't based solely on training program achievement. There is particular risk for tests that are unproctored and computer-based if the computer isn't "locked-down" from storage or accessing other software. In these cases, immediate retesting using the same form should be discouraged. If other software and internet programs are available during the testing, access to other information and to distributing items won't be adequately controlled.

Considering these issues will help programs develop effective and appropriate retest policies and procedures. It should be noted that programs can reference professional literature for guidance and references that support their retesting practices. They should find resources that are most applicable to their own program's relevant factors. For example, resources that use a criterion-referenced pass/fail cut score to determine an individual's attainment of learning outcomes after a learning event are generally most relevant. Results may have limited applicability to professional training programs, however, when they use norm-referenced standardized distributions to compare an individual's performance to other people, or when they measure competence that is not linked to a particular learning event or learning outcomes. Finding the right resources and competent advice will help programs avoid inadvertently undermining the valid use of their test results.

REFERENCE

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.