Statistical Detection of Cheating

Thomas O'Neill
Jerry Gorham
Michelle Reynolds
Weiwei Liu

Introduction & Rationale

Test publishers and testing services are always concerned with irregular behavior, especially with regard to how it will impact the reliability and validity of the test scores. Irregular behavior can range from minor deviations from the standardized conditions to more significant types of behavior, such as, copying from another test-taker, accessing unauthorized resources (books, notes, cheat sheets, etc.), or attempting to steal the item content. Although test publishers are quite vigilant, it is difficult to catch people doing something that they wish to keep secret. Therefore test publishers and test services typically employ a battery of procedures to discourage the unauthorized behavior and facilitate the detection of violations.

Using the cheating detection classification categories, observational and statistical, as proposed by Cizek (1999), it seems that the greatest successes in proving cheating have come from the observational camp (Dwyer & Hecht, 1994). Observational procedures include training test center staff to be on the look out for violations of the standardized procedures, such as proxy testers, accessing unauthorized materials during the test, and removing/recording item content. These processes can often be enhanced through the use of recording technology. Examinees can be video taped while they test; they can be fingerprinted and photographed upon sign-in. When the observational detection and documentation is done well and assuming there has been the appropriate legal language regarding the conditions under which the examinee agrees to test, it provides very powerful evidence that permits test publishers to take action against cheaters. Nevertheless, even the best training, the best check-in procedures and the best recording of the testing session can be defeated by a proctor that is working against the security procedures. Less nefariously, it can also be that some of the irregular behavior is going on outside of the test center, which is not under the test center staffs' control. Accessing "stolen items" before the test or "brain-dumping" items after the test would fall into this category.

seems necessary to consider each individual test center's change in pass rates to see if anything unexpected is occurring. Perhaps a test center has historically had higher pass rates due to such differences as educational programs in the area. Investigating the average change in pass rate from 2004 to 2005 resulted in flagging a total of 18 test centers. The average change in pass rates for all test centers was 2.48% (SD 3.40%). Four test centers were flagged for having increases in pass rates above the 95% threshold (9.15%), and an additional 14 were flagged for changes above the 80% threshold (6.84%). (See Table 1, Map 3)

The two test centers flagged for the 80% threshold in 2005 and 2004 are in close proximity in Ohio. However, these two test centers where not flagged for a pass increase at either the 80% or 95% threshold when comparing the difference between 2004 and 2005 pass rates. This indicates that these test centers may on average have a higher pass rate than the population of test centers, thus have a higher baseline for which to make comparisons in changes to their pass rates.

Center Deviation from Previous Center Pass Rate. This procedure identified 18 centers at the 80% confidence level and another 1 at the 95% confidence level. These results looked somewhat similar to the

Map 4

Difference in pass rates from 2004 to 2005 flagged **(using Joint Standard Error)** test centers (Yellow – 80%, Red 95%)*

* Guam's test center was also flagged at the 80% threshold.

Discussion

The first point worth mentioning is that there is a very common practice, examining pass rates by test center, that is not very well documented in the literature. There are probably several reasons for this. First, is that test publishers are probably a little shy about discussing the specific details of their vulnerabilities. This is understandable from both a business and a legal pe