# Simulation Study

The false positive rate (FPR) and true positive rate (TPR) of the proposed Z-test with conventional fixed-form tests were investigated using Monte Carlo simulation. The FPR is the proportion of simulees (simulated examinees) who had no change in their latent trait profile, but incorrectly determined to have significant change by the hypothesis test. The TPR is the proportion of examinees who had a specified level of change and were correctly determined to have significant change.

The simulated testing situation is as follows: At Time 1, simulees were given 3 conventional fixed-form tests, one test for each of three separate unidimensional latent traits. At Time 2, simulees were given three tests parallel to those at Time 1. The three separate tests can alternatively be conceived of as one long test, with three separate unidimensional subscales (no within-item multidimensionality).

The test information for all tests was peaked around the initial average latent trait score. This means that in situations where change occurs, the standard errors at Time 2 may be greater on average than at Time 1, as the latent trait scores at Time 2 should be further from the peak of the test information function. This situation was chosen to follow a realistic fixed-form testing scenario, where the direction and magnitude of change would be unknown, so tests at Time 2 could not easily be targeted toward the average trait level at Time 2. The item parameters of the three tests were simulated the same way for each test, following a 3PL IRT model. The discrimination parameters were simulated form a normal distribution with mean of 1 and standard deviation of 0.15, the difficulty parameters were simulated from a standard normal distribution, and the guessing parameters were simulated from a uniform distribution with minimum of 0 and maximum of 0.20 (Wang & Weiss, 2018). The test length was 40 items. The test information function and conditional standard error of measurement function (which is just the reciprocal square root of the test information) for each test (identified by dimension) are displayed in Figure 1.

**Figure 1**

*Test Information and Conditional Standard Error of Measurement for Each Dimension*



The latent trait profiles for 10,000 simulees at Time 1 were simulated from a multivariate normal distribution with a mean vector of 0s and a covariance matrix with 1s along the diagonals and *r* along the off-diagonals (Wang & Weiss, 2018). The off-diagonals are the correlation between pairs of latent traits and for simplicity were assumed to be the same for all pairs. Two values for the correlations were used: *r* = 0 and *r* = 0.5. The value of zero was chosen to investigate the simple situation of no correlation, while 0.5 was chosen to follow Wang and Weiss (2018), who chose it as a value that is seen in real testing situations. Wang et al. (2020) found that the level of correlation did not affect several omnibus hypothesis tests of multivariate change. This result is given a preliminary re-examination here for the elevation Z-test, as one of its potential uses is as a post-hoc test to one of the omnibus tests, and so it is desired to see if the same property holds.

Seven different change conditions were simulated for Time 2, chosen based on prior adaptive measurement of change literature (Finkelman et al., 2010; Wang et al., 2020; Wang & Weiss, 2018). To evaluate the FPR, a no-change condition was simulated, in which . To evaluate the TPR, six different change vectors were added to the latent trait profiles of Time 1. For the first four conditions, a uniform change vector was added, with four different magnitudes of change: , , , (Wang & Weiss, 2018).These vectors change the elevation of the profile without changing its scatter or its shape. To give preliminary exploration to non-constant change patterns, in which the profiles scatter or shape also change, the last two conditions used the following change vectors: and . These conditions represent the same average change in elevation as M2 (0.5), but distributed differently. For each timepoint and condition, maximum likelihood was used to estimate , with the estimates bounded in the range .

# Results

The results are presented in Table 1. The FPR is evaluated from the No Change condition, while the TPR is evaluated from conditions M1-M6. The FPR for both r = 0 and r = 0.50 was below the nominal 0.05 level. The FPR in the r = 0.50 condition was higher than in the r = 0 condition. As expected, the TPR increased in the uniform-change conditions (M1-M4) as the magnitude of change increased. The non-uniform change conditions (M5 and M6) had lower TPRs than the M2 condition, which had the same average level of change. The TPRs in the r = 0.50 condition were slightly higher than the TPRs of the r = 0 condition for all change conditions.

**Table 1**

*Proportion of Significant Z-tests for all Conditions*

|  |  |  |
| --- | --- | --- |
| Change Condition | r = 0 | r = 0.50 |
| No Change | 0.037 | 0.045 |
|  | 0.159 | 0.170 |
|  | 0.485 | 0.500 |
|  | 0.761 | 0.778 |
|  | 0.870 | 0.877 |
|  | 0.461 | 0.475 |
|  | 0.402 | 0.417 |

Because the test information functions were peaked near , but the average in change conditions was not at , the SEMs were investigated. The SEMs for each condition and each dimension are presented in Table 2. The SEMs did not differ significantly between the r = 0 and r = 0.50 conditions. The SEMs increased in Dimensions 1 and 2 as the magnitude of change increased. The test for Dimension 3 provided higher information at higher compared to the other two dimensions, and the SEM did not increase nearly as much. For conditions M5 and M6, the average SEM across the three dimensions was higher than that of M2, the condition with the equivalent average change. This was to be expected, as the non-uniform change in M5 and M6 resulted in s further from the peak of the test information functions for the dimensions in which the change was concentrated.

**Table 2**

*Standard Error of Measurements for Each Dimension and Each Condition*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | r = 0 | | | r = 0.50 | | |
| Condition | Dim. 1 | Dim. 2 | Dim. 3 | Dim. 1 | Dim. 2 | Dim. 3 |
| Time 1 | 0.325 | 0.343 | 0.319 | 0.325 | 0.333 | 0.324 |
| No Change | 0.330 | 0.342 | 0.319 | 0.331 | 0.332 | 0.321 |
| M1 | 0.337 | 0.349 | 0.315 | 0.333 | 0.345 | 0.318 |
| M2 | 0.359 | 0.366 | 0.322 | 0.360 | 0.364 | 0.319 |
| M3 | 0.401 | 0.411 | 0.334 | 0.396 | 0.405 | 0.332 |
| M4 | 0.453 | 0.475 | 0.351 | 0.451 | 0.472 | 0.348 |
| M5 | 0.400 | 0.416 | 0.321 | 0.400 | 0.410 | 0.324 |
| M6 | 0.613 | 0.339 | 0.321 | 0.608 | 0.336 | 0.324 |