Approximation answer and response similarity analyses: A practical approach

Russell W. Smith, Ph.D.

Alpine Testing Solutions

Introduction

This study describes a relatively easy and efficient method to evaluate pairwise collusion on tests. Specifically, it helps identify pairs of test takers who have an unlikely number of matching scores or responses. The method is meant to approximate answer or response similarity index (ASI or RSI) analyses (Eckerly, 2017; Maynes, 2016; Sideridis & Zopluoglu, 2018; Zopluoglu, 2017) to provide collusion analyses with relatively simple calculations. ASI and RSI analyses typically rely on Item Response Theory (IRT) and Nominal Response Models (NRM) (Wollack, 1997), respectively, to estimate the probability of an item score or response for a test taker of a given ability. The joint probability of matching scores or responses is calculated for each item, summed, and compared to the actual number of matching responses for each pair of test takers. These calculations can be computationally intense, require specialized software, and are based on relatively strong assumptions (Zopluoglu, 2019).

A practical simplified approach may be warranted when analyzing tests that are continuously available; have high volumes and computational power is limited; have low volumes where IRT or NRM may be inappropriate and/or not fit or converge; have items with either more than four or five options or items that are multiple select (which are problematic for NRMs); and short tests, such as classroom tests. By assuming that a test taker's percent correct score is a reasonable estimate of their average expected item score, SSI and RSI can be approximated with straight-forward computations, without specialized software, and in a less computationally intense manner.

The proposed method uses total scores to estimate the expected number of matching scores. The total percent correct is assumed to be an average expected value across items. Test takers' percent correct scores are used to estimate the joint probability across items for pairs of test takers:

$$\boldsymbol{E_{12} = n * [s_1 * s_2 + (1 - s_1) * (1 - s_2)]}$$ Equation 1

where $\mathbf{E_{12}}$ is the expected number of matching scores between test takers one and two, $\boldsymbol{n}$ is the number of items, and $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$ are the person percent correct scores for persons one and two, respectively.

Note that Equation 1 does not contain any item-level information. Consequently, this approach has the effect of assuming that all items are equally difficult. To the extent that the items differ in difficulty, this estimate is biased in that it underestimates the expected number of matching scores. The bias is most extreme when scores are close to each other <u>and</u> when the sum of the two scores are close to total score. The bias becomes negligible when test taker scores are far apart and/or the sum of their scores is far from the total score. For a 50-item test, for example, the expected number of matching scores for two people with scores of 25 is the most biased because the scores are as close to each other as possible and because their sum equals the total score.

An adjustment factor[1] can be added to the equation that can substantially reduce the bias caused by the assumption of no item difficulty variance. One component of this factor adjusts for how similar the scores are:

$$\frac{\boldsymbol{n - |n * s_1 - n * s_2|}}{\boldsymbol{n}}$$ Equation 2

---

[1] This was discovered primarily through trial and error. It was not intended to be a primary part of this study. There is likely a rational explanation, including a correction for continuity, which was not separately added, and item difficulty variance. Though it seems to work well, a mathematical explanation and more precision would be preferred and warrants more research.

while another correction adjusts for how close their sum is to the total:

$$\frac{(n-|n*s_1+n*s_2-n|)}{n}$$

Equation 3

The value of Equation 2 approaches one when the two test taker scores are very similar; when the test taker scores are very different, the value of Equation 2 approaches zero. The value of Equation 3 approaches zero when the sum of the two test taker scores is far from **n**; when the sum of the test taker scores approaches **n**, the value of Equation 3 approaches one. The possible values of these two components both range between zero and one, so the product of the two components also ranges from zero to one. This product can then be scaled to different values by multiplying the product by a constant, **b**:

$$E_{12}^* = n * [s_1 * s_2 + (1 - s_1) * (1 - s_2)] + b * (\frac{(n-|n*s_1-n*s_2|)}{n}) * (\frac{(n-|n*s_1+n*s_2-n|)}{n})$$

Equation 4

Therefore, the combination of the multiplier **b** and the adjustment factor (i.e., the product of Equation 3 and Equation 4) increases the estimate of the expected number of matches by at most **b** items in the most extreme case (i.e., when scores are close to each other <u>and</u> when the sum of the two scores are close to total score). The total amount of the adjustment decreases gradually to zero as test taker scores become farther apart and/or have a sum farther from the total score.

One approach for estimating a reasonable multiplier **b** is to calculate an expected value using an IRT model, if item-level data are available and assumptions are met, for two middle scores and subtract the unadjusted expected value in Equation 1. We have found using both empirical exam and simulated data that an adjustment of about 10% to 15% of the total score seems reasonable, depending on the amount of item difficulty variance. In other words, in practice, Equation 1 tends to underestimate the number of expected matches by approximately 10% to 15% in the most extreme case. So, choosing a value of **b** that is approximately 10% to 15% of **n** results in a much less biased estimate of the expected number of matches for pairs of test takers.

Thoughtful selection of the value of **b** can also help to minimize false positives or prioritize power. Higher values of **b** are more conservative (i.e., less likely to result in false positives) because they increase expected values. Lower values of **b** are less conservative (i.e., more likely to identify colluding pairs of test takers) because they decrease expected values.

Note that the Equation 4 relies on only four values: the two person scores, the length of the test (i.e., maximum possible score), and the user-selected multiplier. No IRT calibration is required, and no item-level data is used other than for calculating test taker scores.

With the expected number of matches now known for all test taker pairs, a score (or response) similarity index can be calculated. Specifically, a z-score can be calculated for each pair of test takers:

$$z_{12} = \frac{(M_{12} - E_{12}^*)}{\sqrt{npq}}$$

Equation 5

where the numerator is the difference between the empirical count of matching scores (or responses), M_{12}, and the expected count, E_{12}^*, and the denominator is the estimated standard deviation of a binomial distribution where **n** is the item count, $p = \frac{E_{12}^*}{n}$, and $q = 1 - p$. By calculating a z-score, an estimated pairwise probability of a given number of matching scores (or responses) given the estimated expected number of matching scores can

looked up using the unit normal table. Alternatively, the user could select a given probability a priori and then, using the unit normal table, find a z-score to be used as a critical value to identify anomalous pairs of test takers.

The expected number of matching <u>scores</u> can be considered a maximum expected number of matching <u>responses</u> because there is usually more than one way to answer a multiple-choice item incorrectly. Consider an item where the expected score for two test takers of the same given ability is 0.5. The probability of a matching score is 0.5 (i.e., 0.5 * 0.5 + [1 - 0.5] * [1 - 0.5]). If there is more than one way to answer the item incorrectly (e.g., a four-option multiple choice item with one correct answer), the probability of a matching incorrect response will be *no more than* 0.5. Therefore, using matching scores is a conservative, or most extreme, estimate of matching responses.

Given the number of pairs compared in collusion analyses, family-wise error rates can become problematic (Wollack, 2001). In practice, one approach to mitigating this is to 1) flag each pair of tests at a given critical value and then 2) count the number of times each individual test administration is flagged and compare that to critical number of flags criterion. For example, for a test administered to 1,000 people, there are 499,500 pairwise comparisons. Even at a critical value of 0.0001, we would expect to flag about 50 pairs of administrations just by chance. A secondary criterion can be used to identify administrations flagged with multiple other administrations (e.g., flagged with five or more other administrations). While it would be unlikely to flag a test taker without pre-knowledge once, it would be even less likely to flag this test taker five times or more.

Method

To evaluate the effectiveness of the proposed similarity approximation, a simulation study was conducted. Additionally, as part of an NCME symposium, the approximation methodology was applied to a set of real data common to each paper; the intent was to allow a comparison of the methodologies proposed by each author. For both the simulation and real data studies, true score similarity index (True SSI) analyses were run using the Rasch model; the joint probabilities of test takers getting the same score on each item were calculated, compared to the actual number of matching scores, and then the probability using the generalized binomial model was calculated (van der Linden & Sotaridon, 2006). The approximation score similarity index (Approx SSI) was also calculated using the method described above. For the real data, in addition to True and Approx SSI, the approximation of *response* similarity was applied by simply using the actual number of matching responses in place of the actual number of matching scores. The results between the two methods are compared.

Simulation

Thirty sets of data were simulated based on the following categories and variables:

- number of test items (50 and 100)
- person score distribution shape (left skewed, uniform, normal)
- percent of item "exposed" (60, 70, 80, 90, and 100%).

The item parameters from the empirical study were used to simulate the fifteen 50-item conditions. For the fifteen 100-item conditions, the distribution of Rasch item difficulty measures was simulated targeting the same mean and standard deviation of the item difficulties as the 50-item test (-0.44, and 0.92, respectively).

Table 1 provides the distribution moments of the 800 simulees without pre-knowledge for each distribution shape. For each of the 30 conditions, the item score matrix was created by comparing the Rasch probability of each of 1,000 simulees getting an item correct to a random number. Then, item scores for 20% of the simulees were stochastically re-estimated by adding 3.0 to each of their Rasch measures, comparing to a random number, and updating their item score. This was repeated for each of the percent of items "exposed" conditions. In other words, data were generated assuming that simulees with access to exposed items had a 3-logit advantage on those items. The 80% of the simulees without pre-knowledge had the same scores for each percent of items within each test length and person score distribution shape conditions.

Table 1. Rasch Ability Measures: Distribution Moments

	Skewed	Uniform	Normal
Mean	0.99	0.35	0.45
St Dev	1.67	2.44	1.07
Skewness	-0.55	0.07	-0.14
Kurtosis	0.52	-0.72	0.67

The working assumption is that 20% of the test takers had access to a given percent of the same item content. The rationale for re-estimating scores stochastically is that not all exposed content is perfect or complete, and test takers do not recall or use the exposed content perfectly. A test taker who has pre-knowledge will increase their likelihood of a correct response, but a correct response is not guaranteed. Further, the 20% of simulees for

whom scores were upwardly adjusted were randomly selected, meaning that their distribution reflected the total distribution of scores. The assumption is that gaining pre-knowledge is independent of ability.

For each of the 30 simulations, the updated score distribution was used to recalibrate Rasch person and item measures. True SSI was calculated including pairwise probabilities. The number of pairs and individuals flagged were counted for each of five critical values: 0.000001, 0.00001, 0.0001, 0.001, and 0.01. Probabilities were estimated using the Approx SSI method described above using a b value of 12.5% of the test length, and the number of pairs and individual were counted for each of the same five critical values. The counts and percent of pairs flagged, individuals flagged at least once, and individuals flagged at least five times are reported.

Real Data

The empirical data are based on 11,888 administrations of a single form of an Information Technology (IT) certification exam. The test consists of multiple forms as well as pre-test items. A concurrent Rasch calibration was run across the multiple forms and included the unscored pre-test items. The Rasch person and item measures were used to calculate the expected number of matching scores for a single form. True SSI including the number of pairs and individuals flagged were counted for each of six critical values: 0.000000001, 0.00000001, 0.0000001, 0.000001, 0.00001, 0.0001, and 0.001. Approx SSI was calculated, including counts and percent of pairs and individuals flagged for each of the same critical values using the method described above. Additionally, by simply replacing the count of matching responses for the count of matching scores in the z-score equation (equation 5), approximation of response similarity index (Approx RSI) was estimated including the counts and percent of pairs and individuals flagged.

Results

Simulations

Table 2 shows the results of one the 30 simulated conditions. Specifically, it shows the results of a normal distribution of person scores, with simulated pre-knowledge of 80% of the items, and on the 50-item test. With 1,000 simulees, there are 499,500 pairs of test takers. Using a critical value of 0.001, True SSI properly flagged 35 of the 19,900 pair where each of the two simulees had pre-knowledge and 57 of the 479,600 pair where at least one of the simulees did not have pre-knowledge. Using the same critical value of 0.001, the approximation approach flagged 13 of 19,900 pair where each of the two simulees had pre-knowledge and 19 of 479,600 pair where at least one of the simulees did not have pre-knowledge. Further, at the same critical value, True SSI properly flagged 41 simulees with pre-knowledge at least once and two at least five times where the approximation approach flagged 15 simulees with pre-knowledge at least once and none more than five times. Appendix A includes the same information for all 30 simulated conditions.

Table 2. Simulation of a 50-item test, with a normal distribution of person scores, and 80% of the items considered to be exposed

| | Critical value | True SSI | | | | Approximation SSI | | | |
| | | Pre-knowledge | | At least one without Pre-Knolwedge | | Pre-knowledge | | At least one without Pre-Knolwedge | |
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
	0.00001	1	0.010%	1	0.000%	0	0.000%	1	0.000%
	0.0001	5	0.030%	5	0.001%	1	0.005%	1	0.000%
	0.001	35	0.180%	57	0.012%	13	0.065%	19	0.004%
	0.01	350	1.760%	780	0.163%	320	1.608%	513	0.107%
Individuals flagged at least once	0.000001	0	0.000%	2	0.250%	0	0.000%	0	0.000%
	0.00001	2	1.000%	2	0.250%	0	0.000%	2	0.250%
	0.0001	7	3.500%	8	1.000%	2	1.000%	2	0.250%
	0.001	41	20.500%	85	10.630%	15	7.500%	31	3.880%
	0.01	122	61.000%	442	55.250%	108	54.000%	376	47.000%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	2	1.000%	1	0.130%	0	0.000%	0	0.000%
	0.01	45	22.500%	71	8.880%	42	21.000%	35	4.380%

To show how the percent of items exposed impacts the results, Figure 1 shows for a 50-item test, with a normal score distribution, and a critical value of 0.01 the percent of appropriately flagged simulees with pre-knowledge (power) as well as the percent of test takers without pre-knowledge who were flagged (Type I Error). The four values plotted in Figure 1 at 80% of items exposed can be found in the last row of Table 2. For nearly all simulated conditions, the approximation has the same or less power than True SSI. At the same time, the approximation has less Type I error than True SSI. Appendix B includes the same information as Figure 1 for all 30 simulated conditions for critical values of 0.01 and 0.001.
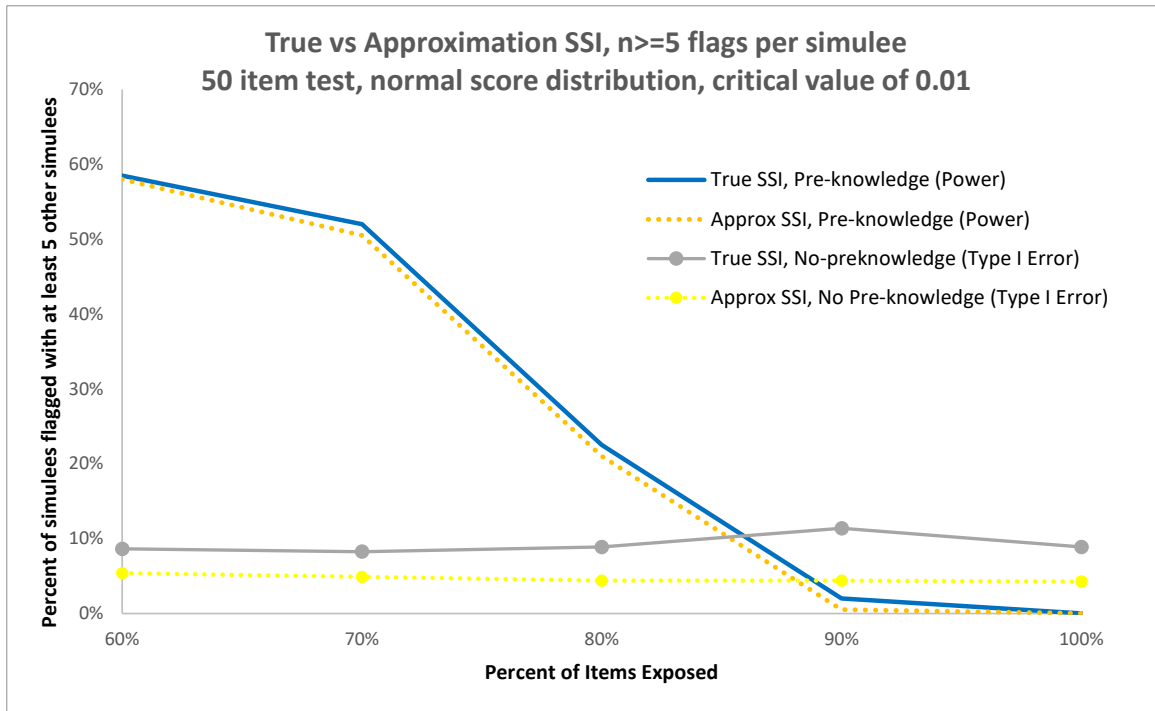
Figure 1. Comparing True SSI to Approximation SSI across differing amount of item exposure for a 50-item test, a normal distribution, and a critical value of 0.01

Real Data

Table 3 shows the results of True SSI, Approx SSI, and Approx RSI for the real data. As expected, Approx SSI flags fewer pairs and people than True SSI and Approx RSI flags fewer pairs and people than Approx SSI. Further, based on these results, the amount of pre-knowledge appears to be egregious. Using a critical value of 0.001, True SSI flagged 28% of the test takers at least five times, Approx SSI flagged 27% of the test takers at least five times, and Approx RSI flagged 16% of the test takes at least five times. By way of comparison, in the simulation of a 50-item test, a normal distribution of person scores, with 80% of the items exposed, and where 20% of the simulees had pre-knowledge, at the same critical value True SSI flagged 1% of the simulees with pre-knowledge at least five times and Approx SSI did not flag any. In the simulation, because nefarious simulees were known, the percents are within groups; for the real data percents are across all test takers because reality is unknown.

Table 3. Real data, a 50-item IT test administered to 11,888 test takers

	Crit value	Real data, 50 items, 11,888 administrations					
		True SSI		Approx SSI		Approx RSI	
		Count	Percent	Count	Percent	Count	Percent
Pairwise	0.00000001	216502	0.31%	26928	0.04%	14639	0.02%
	0.0000001	316400	0.45%	61647	0.09%	38682	0.05%
	0.000001	453310	0.64%	105051	0.15%	88471	0.13%
	0.00001	615714	0.87%	154572	0.22%	142349	0.20%
	0.0001	779834	1.10%	291390	0.41%	264028	0.37%
	0.001	1021938	1.45%	415916	0.59%	334568	0.47%
Individuals flagged at least once	0.00000001	1965	16.53%	448	3.77%	360	3.03%
	0.0000001	2203	18.53%	802	6.75%	690	5.80%
	0.000001	2410	20.27%	1168	9.83%	1083	9.11%
	0.00001	2617	22.01%	1983	16.68%	1854	15.60%
	0.0001	3366	28.31%	2619	22.03%	2145	18.04%
	0.001	7016	59.02%	6147	51.71%	2416	20.32%
Individuals flagged at least 5 times	0.00000001	1538	12.94%	370	3.11%	167	1.40%
	0.0000001	1959	16.48%	688	5.79%	296	2.49%
	0.000001	2228	18.74%	1047	8.81%	517	4.35%
	0.00001	2419	20.35%	1811	15.23%	1104	9.29%
	0.0001	2555	21.49%	2176	18.30%	1577	13.27%
	0.001	3348	28.16%	3199	26.91%	1932	16.25%

Discussion

This paper proposes simplified approximations of score and response similarity analyses. It is intended to address a very specific situation where test takers have unintended pre-knowledge of test content. It is meant to provide an alternative to True SSI or RSI requiring less complicated calculations. The method requires only a count of matching scores or responses and test taker percent correct scores. No knowledge of the item difficulty, classic test theory or IRT, is required. No specialized software is required. This makes Approx SSI/RSI more accessible than True SSI/RSI.

The Approx SSI method appears to accurately identify test takers with pre-knowledge compared to True SSI for a 50-item test under similar conditions. It does not perform quite as well for the 100-item test. Understanding why Approx and True SSI perform different with varying test lengths is worth consideration. No method accurately identifies all test takers with pre-knowledge, at least not with a reasonable Type I error rate. For a 50-item test, a normal score distribution, 60% of the items exposed, and at a critical value of 0.001, True SSI correctly identifies 16% of pre-knowledge test takers while Approx SSI correctly identifies 11% of the pre-knowledge test takers. The Approx method identifying 11% is reasonable compared to 16% for the True method, but neither are especially impressive. Several conditions appear to make identifying test takers with pre-knowledge quite difficult using such methods, including when people have knowledge of most or all content. SSI methods, Approx or True, seem to work well only for a narrow set of conditions.

When comparing across the different distribution shapes, the results were not surprising. Both True and Approx SSI seem to work better for normal score distributions, followed by uniform score distributions, and then skewed distributions when holding other variables constant. SSI seems to work better identifying test takers with a moderate or low ability to start. A left skewed distribution has fewer such test takers. A uniform distribution has more low ability, but also has more high-ability test takers when compared to a normal distribution. The narrow conditions for which True and Approx SSI works, then, is when test takers of moderate or low ability have access to some, but not all test content.

Approx SSI has a lower rate of Type I error rate for the simulated conditions. Both the power and the Type I error rate are impacted by the multiplier used in the Approx method. The user can control how conservative the estimates of the expected values are which would have a direct impact on power and Type I error. More research is needed to understand how and when to use what value. More research is also needed to better understand the nature of Approx SSI and RSI especially with respect to test length, percent of people with pre-knowledge, and percent of content exposed as well as the interaction effects. The results show that as the percent of exposed content increases from 60% to 100%, the ability of each method to detect test takers with pre-knowledge diminishes. The simulation here used a 3-logit benefit to estimate how test takers respond to exposed content. This resulted in stochastic behaviors; that is, test takers were more likely, but not 100% likely, to answer correctly on exposed content. If we assumed test takers followed an exposed bank (with both correct and incorrect exposed keys) with 100% fidelity, results might have differed and power at high exposure rates may have improved.

Another specific recommendation for future research is to simulate conditions with fewer exposed items and evaluate and compare the different methods. If very few people have pre-knowledge, there is less impact on the IRT calibration used by True SSI. Therefore, it is possible that True SSI may be more sensitive to detecting pre-knowledge than Approx SSI when fewer people have pre-knowledge. It is also possible that because Approx SSI is not impacted by item-level statistics, it may function better, or at least different, than True SSI when a large percent of test takers has pre-knowledge.

In short, for a limited set of circumstances, Approx SSI seems to do a reasonable job of identifying test takers with pre-knowledge when compared to True SSI. It can provide estimates of pairwise probabilities, a desirable trait for forensic statistics, with relatively simple calculations, very little information, and without specialized software. While much more research is warranted, the method has promise for some specific situations.

References

Eckerly, C. (2017). Detecting preknowledge and item compromise: Understanding the status quo. In Cizek, G. J., Wollack, J. A. (Eds.), *Handbook of quantitative methods for detecting cheating on tests* (pp. 101-123). New York, NY: Routledge.

Maynes, D. (2016). Detecting potential collusion among individual examinees using similarity analysis. In Cizek, G. J., Wollack, J. A. (Eds.), *Handbook of quantitative methods for detecting cheating on tests* (pp. 101-123). New York, NY: Routledge.

Sideridis, G., & Zopluoglu, C. (2018). Validation of Response Similarity Analysis for the detection of academic cheating: An experimental study. *Journal of Applied Measurement*, *19*.

Wollack, J. A., Cohen, A. S., & Serlin, R. C. (2001). Defining error rates and power for detecting answer copying. *Applied Psychological Measurement*, 25(4), 385–404.

Wollack, J. A. (1997). A nominal response model approach for detecting answer copying. *Applied Psychological Measurement*, *21*, 307-320.

van der Linden, W. J., & Sotaridona, L. (2006). Detecting answer copying when the regular response process follows a known response model. *Journal of Educational and Behavioral Statistics, 31*(3), 283-304.

Zopluoglu, C. (2017). Similarity, answer copying, and aberrance: Understanding the status quo. In Cizek, G. J., Wollack, J. A. (Eds.), *Handbook of quantitative methods for detecting cheating on tests* (pp. 25–46). New York, NY: Routledge.

Zopluoglu, C. (2019). Computation of the Response Similarity Index M4 in R under the Dichotomous and Nominal Item Response Models. *International Journal of Assessment Tools in Education 6*, 1-19. 10.21449/ijate.527299

Appendix A

Simulation Results, Normal Score Distribution, 50 items

50-item Test, Normal Score Distribution, 60% of items

	Critical value	True SSI				Approximation SSI			
		Pre-knowledge		At least one without Pre-Knolwedge		Pre-knowledge		At least one without Pre-Knolwedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
	0.00001	2	0.010%	1	0.000%	0	0.000%	1	0.000%
	0.0001	39	0.196%	3	0.001%	19	0.095%	1	0.000%
	0.001	241	1.211%	49	0.010%	189	0.950%	21	0.004%
	0.01	1275	6.407%	769	0.160%	1322	6.643%	593	0.124%
Individuals flagged at least once	0.000001	0	0.000%	2	0.250%	0	0.000%	0	0.000%
	0.00001	4	2.000%	2	0.250%	0	0.000%	2	0.250%
	0.0001	36	18.000%	5	0.625%	21	10.500%	2	0.250%
	0.001	88	44.000%	72	9.000%	82	41.000%	31	3.875%
	0.01	153	76.500%	426	53.250%	151	75.500%	379	47.375%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	1	0.500%	0	0.000%	0	0.000%	0	0.000%
	0.001	32	16.000%	0	0.000%	22	11.000%	0	0.000%
	0.01	117	58.500%	69	8.625%	116	58.000%	43	5.375%

50-item Test, Normal Score Distribution, 70% of items

	Critical value	True SSI				Approximation SSI			
		Pre-knowledge		At least one without Pre-Knolwedge		Pre-knowledge		At least one without Pre-Knolwedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
	0.00001	4	0.020%	1	0.000%	0	0.000%	1	0.000%
	0.0001	21	0.110%	3	0.001%	7	0.035%	1	0.000%
	0.001	149	0.750%	44	0.009%	106	0.533%	19	0.004%
	0.01	975	4.900%	730	0.152%	1031	5.181%	541	0.113%
Individuals flagged at least once	0.000001	0	0.000%	2	0.250%	0	0.000%	0	0.000%
	0.00001	7	3.500%	2	0.250%	0	0.000%	2	0.250%
	0.0001	28	14.000%	5	0.630%	11	5.500%	2	0.250%
	0.001	88	44.000%	68	8.500%	66	33.000%	31	3.880%
	0.01	145	72.500%	424	53.000%	146	73.000%	379	47.380%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	15	7.500%	0	0.000%	12	6.000%	0	0.000%
	0.01	104	52.000%	66	8.250%	101	50.500%	39	4.880%

50-item Test, Normal Score Distribution, 80% of items

	Critical value	True SSI				Approximation SSI			
		Pre-knowledge		At least one without Pre-Knolwedge		Pre-knowledge		At least one without Pre-Knolwedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
	0.00001	1	0.010%	1	0.000%	0	0.000%	1	0.000%
	0.0001	5	0.030%	5	0.001%	1	0.005%	1	0.000%
	0.001	35	0.180%	57	0.012%	13	0.065%	19	0.004%
	0.01	350	1.760%	780	0.163%	320	1.608%	513	0.107%
Individuals flagged at least once	0.000001	0	0.000%	2	0.250%	0	0.000%	0	0.000%
	0.00001	2	1.000%	2	0.250%	0	0.000%	2	0.250%
	0.0001	7	3.500%	8	1.000%	2	1.000%	2	0.250%
	0.001	41	20.500%	85	10.630%	15	7.500%	31	3.880%
	0.01	122	61.000%	442	55.250%	108	54.000%	376	47.000%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	2	1.000%	1	0.130%	0	0.000%	0	0.000%
	0.01	45	22.500%	71	8.880%	42	21.000%	35	4.380%

50-item Test, Normal Score Distribution, 90% of items

	Critical value	True SSI				Approximation SSI			
		Pre-knowledge		At least one without Pre-Knolwedge		Pre-knowledge		At least one without Pre-Knolwedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	1	0.000%	0	0.000%	1	0.000%
	0.0001	0	0.000%	5	0.001%	0	0.000%	1	0.000%
	0.001	4	0.020%	62	0.013%	0	0.000%	19	0.004%
	0.01	45	0.230%	872	0.182%	25	0.126%	486	0.101%
Individuals flagged at least once	0.000001	0	0.000%	2	0.250%	0	0.000%	0	0.000%
	0.00001	0	0.000%	2	0.250%	0	0.000%	2	0.250%
	0.0001	0	0.000%	8	1.000%	0	0.000%	2	0.250%
	0.001	7	3.500%	92	11.500%	1	0.500%	31	3.880%
	0.01	51	25.500%	468	58.500%	37	18.500%	369	46.130%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	1	0.130%	0	0.000%	0	0.000%
	0.01	4	2.000%	91	11.380%	1	0.500%	35	4.380%

50-item Test, Normal Score Distribution, 100% of items

Critical value	Pairs with Pre-knowledge				Pairs with Pre-knowledge			
	Pre-knowledge		At least one without Pre-Knolwedge		Pre-knowledge		At least one without Pre-Knolwedge	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
0.000001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
0.00001	0	0.000%	1	0.000%	0	0.000%	1	0.000%
0.0001	0	0.000%	5	0.001%	0	0.000%	1	0.000%
0.001	0	0.000%	54	0.011%	0	0.000%	18	0.004%
0.01	1	0.010%	772	0.161%	0	0.000%	483	0.101%
0.000001	0	0.000%	2	0.250%	0	0.000%	0	0.000%
0.00001	0	0.000%	2	0.250%	0	0.000%	2	0.250%
0.0001	0	0.000%	8	1.000%	0	0.000%	2	0.250%
0.001	1	0.500%	84	10.500%	0	0.000%	30	3.750%
0.01	9	4.500%	441	55.130%	7	3.500%	368	46.000%
0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
0.001	0	0.000%	1	0.130%	0	0.000%	0	0.000%
0.01	0	0	71	0.0888	0	0	34	0.0425

Simulation Results, Uniform Score Distribution, 50 items

50-item Test, Uniform Score Distribution, 60% of items

		True SSI				Approximation SSI			
	Critical value	Pre-knowledge		At least one without Pre-Knowledge		Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	1	0.005%	0	0.000%	0	0.000%	0	0.000%
	0.0001	6	0.030%	0	0.000%	2	0.010%	0	0.000%
	0.001	62	0.312%	12	0.003%	34	0.171%	5	0.001%
	0.01	501	2.518%	222	0.046%	467	2.347%	156	0.033%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	2	1.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	11	5.500%	0	0.000%	4	2.000%	0	0.000%
	0.001	55	27.500%	18	2.250%	41	20.500%	8	1.000%
	0.01	109	54.500%	170	21.250%	106	53.000%	138	17.250%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	3	1.500%	0	0.000%	0	0.000%	0	0.000%
	0.01	67	33.500%	12	1.500%	63	31.500%	2	0.250%

50-item Test, Uniform Score Distribution, 70% of items

		True SSI				Approx SSI			
	Critical value	Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	4	0.020%	0	0.000%	0	0.000%	0	0.000%
	0.001	54	0.271%	11	0.002%	36	0.181%	6	0.001%
	0.01	436	2.191%	204	0.043%	424	2.131%	151	0.031%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	7	3.500%	0	0.000%	0	0.000%	0	0.000%
	0.001	43	21.500%	17	2.125%	34	17.000%	9	1.125%
	0.01	113	56.500%	158	19.750%	105	52.500%	134	16.750%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	4	2.000%	0	0.000%	2	1.000%	0	0.000%
	0.01	57	28.500%	9	1.125%	58	29.000%	2	0.250%

50-item Test, Uniform Score Distribution, 80% of items

		True SSI				Approx SSI			
	Critical value	Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
	0.001	12	0.060%	14	0.003%	6	0.030%	7	0.001%
	0.01	181	0.910%	209	0.044%	147	0.739%	137	0.029%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	1	0.500%	1	0.125%	0	0.000%	0	0.000%
	0.001	24	12.000%	20	2.500%	16	8.000%	10	1.250%
	0.01	91	45.500%	162	20.250%	79	39.500%	125	15.625%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	27	13.500%	9	1.125%	20	10.000%	3	0.375%

50-item Test, Uniform Score Distribution, 90% of items

		True SSI				Approx SSI			
	Critical value	Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	1	0.005%	13	0.003%	1	0.005%	3	0.001%
	0.01	18	0.090%	202	0.042%	13	0.065%	115	0.024%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	4	2.000%	19	2.375%	2	1.000%	6	0.750%
	0.01	43	21.500%	171	21.375%	33	16.500%	123	15.375%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	1	0.500%	10	1.250%	0	0.000%	1	0.125%

50-item Test, Uniform Score Distribution, 100% of items

Pairs		True SSI				Approx SSI			
	Crit value	Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	1	0.005%	12	0.003%	0	0.000%	4	0.001%
	0.01	7	0.035%	209	0.044%	3	0.015%	133	0.028%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	2	1.000%	19	2.375%	1	0.500%	7	0.875%
	0.01	26	13.000%	173	21.625%	22	11.000%	133	16.625%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	2	1.000%	9	1.125%	1	0.500%	1	0.125%

Simulation Results, Skewed Score Distribution, 50 items

50-item Test, Skewed Score Distribution, 60% of items

	Critical value	True SSI				Approximation SSI			
		Pre-knowledge		At least one without Pre-Knowledge		Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	7	0.035%	1	0.000%	1	0.005%	0	0.000%
	0.001	66	0.332%	14	0.003%	45	0.226%	9	0.002%
	0.01	506	2.543%	290	0.060%	512	2.573%	262	0.055%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	10	5.000%	2	0.250%	2	1.000%	0	0.000%
	0.001	53	26.500%	25	3.125%	39	19.500%	15	1.875%
	0.01	111	55.500%	254	31.750%	107	53.500%	229	28.625%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	5	2.500%	0	0.000%	1	0.500%	0	0.000%
	0.01	65	32.500%	8	1.000%	64	32.000%	8	1.000%

50-item Test, Skewed Score Distribution, 80% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	2	0.010%	1	0.000%	0	0.000%	0	0.000%
	0.001	20	0.101%	13	0.003%	9	0.045%	6	0.001%
	0.01	159	0.799%	261	0.054%	143	0.719%	215	0.045%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	4	2.000%	2	0.250%	0	0.000%	0	0.000%
	0.001	23	11.500%	23	2.875%	9	4.500%	12	1.500%
	0.01	69	34.500%	254	31.750%	63	31.500%	220	27.500%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	1	0.500%	0	0.000%	0	0.000%	0	0.000%
	0.01	23	11.500%	5	0.625%	21	10.500%	5	0.625%

50-item Test, Skewed Score Distribution, 90% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	1	0.005%	16	0.003%	0	0.000%	6	0.001%
	0.01	17	0.085%	298	0.062%	13	0.065%	210	0.044%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	4	0.500%	0	0.000%	0	0.000%
	0.001	3	1.500%	28	3.500%	0	0.000%	12	1.500%
	0.01	31	15.500%	275	34.375%	25	12.500%	218	27.250%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	2	1.000%	13	1.625%	1	0.500%	7	0.875%

50-item Test, Skewed Score Distribution, 100% of items

	Crit value	True SSI				Approx SSI			
Pairs		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	1	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	15	0.003%	0	0.000%	7	0.001%
	0.01	0	0.000%	258	0.054%	0	0.000%	202	0.042%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	2	0.250%	0	0.000%	0	0.000%
	0.001	2	1.000%	24	3.000%	1	0.500%	13	1.625%
	0.01	11	5.500%	259	32.375%	9	4.500%	216	27.000%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	0	0.000%	7	0.875%	0	0.000%	5	0.625%

50-item Test, Skewed Score Distribution, 70% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	1	0.005%	0	0.000%	0	0.000%	0	0.000%
	0.0001	5	0.025%	1	0.000%	2	0.010%	0	0.000%
	0.001	63	0.317%	16	0.003%	38	0.191%	7	0.001%
	0.01	424	2.131%	274	0.057%	429	2.156%	244	0.051%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	2	1.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	9	4.500%	2	0.250%	4	2.000%	0	0.000%
	0.001	48	24.000%	26	3.250%	32	16.000%	13	1.625%
	0.01	110	55.000%	253	31.625%	99	49.500%	225	28.125%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	5	2.500%	0	0.000%	3	1.500%	0	0.000%
	0.01	58	29.000%	5	0.625%	58	29.000%	7	0.875%

Simulation Results, Normal Score Distribution, 100 items

100-item Test, Normal Score Distribution, 60% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knolwedge		Pairs with Pre-knowledge		At least one without Pre-Knolwedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	4	0.020%	0	0.000%	0	0.000%	0	0.000%
	0.0001	50	0.251%	2	0.000%	7	0.035%	0	0.000%
	0.001	301	1.513%	44	0.009%	134	0.673%	10	0.002%
	0.01	1578	7.930%	974	0.204%	981	4.930%	389	0.081%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	6	3.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	47	23.500%	4	0.500%	11	5.500%	0	0.000%
	0.001	104	52.000%	70	8.750%	76	38.000%	20	2.500%
	0.01	163	81.500%	494	61.750%	145	72.500%	340	42.500%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	1	0.500%	0	0.000%	0	0.000%	0	0.000%
	0.001	46	23.000%	0	0.000%	11	5.500%	0	0.000%
	0.01	124	62.000%	108	13.500%	88	44.000%	25	3.125%

100-item Test, Normal Score Distribution, 70% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	22	0.111%	2	0.000%	2	0.010%	0	0.000%
	0.001	204	1.025%	45	0.009%	66	0.332%	12	0.003%
	0.01	1241	6.236%	928	0.194%	658	3.307%	392	0.082%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	33	16.500%	4	0.500%	4	2.000%	0	0.000%
	0.001	99	49.500%	70	8.750%	55	27.500%	22	2.750%
	0.01	157	78.500%	490	61.250%	128	64.000%	344	43.000%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	27	13.500%	0	0.000%	4	2.000%	0	0.000%
	0.01	116	58.000%	102	12.750%	77	38.500%	24	3.000%

100-item Test, Normal Score Distribution, 80% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	5	0.025%	2	0.000%	0	0.000%	0	0.000%
	0.001	51	0.256%	47	0.010%	10	0.050%	10	0.002%
	0.01	536	2.693%	904	0.188%	263	1.322%	386	0.080%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	10	5.000%	4	0.500%	0	0.000%	0	0.000%
	0.001	50	25.000%	74	9.250%	14	7.000%	20	2.500%
	0.01	131	65.500%	486	60.750%	102	51.000%	342	42.750%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	74	37.000%	101	12.625%	37	18.500%	24	3.000%

100-item Test, Normal Score Distribution, 90% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	1	0.005%	1	0.000%	0	0.000%	0	0.000%
	0.001	12	0.060%	39	0.008%	0	0.000%	11	0.002%
	0.01	166	0.834%	864	0.181%	48	0.241%	393	0.082%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	2	1.000%	2	0.250%	0	0.000%	0	0.000%
	0.001	14	7.000%	62	7.750%	1	0.500%	21	2.625%
	0.01	96	48.000%	475	59.375%	48	24.000%	342	42.750%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	15	7.500%	90	11.250%	4	2.000%	24	3.000%

100-item Test, Normal Score Distribution, 100% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knolwedge		Pairs with Pre-knowledge		At least one without Pre-Knolwedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	2	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	39	0.008%	0	0.000%	10	0.002%
	0.01	1	0.005%	824	0.172%	1	0.005%	402	0.084%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	4	0.500%	0	0.000%	0	0.000%
	0.001	4	2.000%	61	7.625%	0	0.000%	20	2.500%
	0.01	15	7.500%	464	58.000%	14	7.000%	347	43.375%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	2	1.000%	86	10.750%	0	0.000%	24	3.000%

Simulation Results, Uniform Score Distribution, 100 items

100-item Test, Uniform Score Distribution, 60% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	10	0.050%	0	0.000%	0	0.000%	0	0.000%
	0.0001	66	0.332%	2	0.000%	14	0.070%	0	0.000%
	0.001	304	1.528%	17	0.004%	104	0.523%	2	0.000%
	0.01	1225	6.156%	196	0.041%	678	3.407%	58	0.012%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	15	7.500%	0	0.000%	0	0.000%	0	0.000%
	0.0001	44	22.000%	4	0.500%	21	10.500%	0	0.000%
	0.001	95	47.500%	24	3.000%	68	34.000%	4	0.500%
	0.01	122	61.000%	190	23.750%	109	54.500%	76	9.500%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	2	1.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	42	21.000%	0	0.000%	9	4.500%	0	0.000%
	0.01	97	48.500%	8	1.000%	78	39.000%	0	0.000%

100-item Test, Uniform Score Distribution, 70% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	1	0.005%	0	0.000%	0	0.000%	0	0.000%
	0.00001	8	0.040%	0	0.000%	1	0.005%	0	0.000%
	0.0001	38	0.191%	2	0.000%	12	0.060%	0	0.000%
	0.001	222	1.116%	10	0.002%	77	0.387%	2	0.000%
	0.01	1074	5.397%	201	0.042%	566	2.844%	58	0.012%
Individuals flagged at least once	0.000001	2	1.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	13	6.500%	0	0.000%	2	1.000%	0	0.000%
	0.0001	42	21.000%	4	0.500%	17	8.500%	0	0.000%
	0.001	89	44.500%	16	2.000%	60	30.000%	4	0.500%
	0.01	121	60.500%	197	24.625%	108	54.000%	75	9.375%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	30	15.000%	0	0.000%	8	4.000%	0	0.000%
	0.01	97	48.500%	8	1.000%	68	34.000%	0	0.000%

100-item Test, Uniform Score Distribution, 80% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	1	0.005%	0	0.000%	0	0.000%	0	0.000%
	0.0001	3	0.015%	2	0.000%	1	0.005%	0	0.000%
	0.001	55	0.276%	13	0.003%	12	0.060%	2	0.000%
	0.01	392	1.970%	195	0.041%	190	0.955%	58	0.012%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	2	1.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	6	3.000%	4	0.500%	2	1.000%	0	0.000%
	0.001	54	27.000%	20	2.500%	18	9.000%	4	0.500%
	0.01	102	51.000%	186	23.250%	81	40.500%	75	9.375%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	2	1.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	57	28.500%	7	0.875%	25	12.500%	0	0.000%

100-item Test, Uniform Score Distribution, 90% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	1	0.005%	2	0.000%	0	0.000%	0	0.000%
	0.001	4	0.020%	14	0.003%	1	0.005%	3	0.001%
	0.01	65	0.327%	201	0.042%	18	0.090%	60	0.013%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	2	1.000%	4	0.500%	0	0.000%	0	0.000%
	0.001	10	5.000%	20	2.500%	3	1.500%	5	0.625%
	0.01	74	37.000%	192	24.000%	27	13.500%	76	9.500%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	4	2.000%	6	0.750%	0	0.000%	0	0.000%

100-item Test, Uniform Score Distribution, 100% of items

	Critical value	True SSI				Approx SSI			
		Pairs with Pre-knowledge		At least one without Pre-Knowledge		Pairs with Pre-knowledge		At least one without Pre-Knowledge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	2	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	12	0.003%	0	0.000%	2	0.000%
	0.01	3	0.015%	198	0.041%	1	0.005%	60	0.013%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	4	0.500%	0	0.000%	0	0.000%
	0.001	3	1.500%	18	2.250%	0	0.000%	4	0.500%
	0.01	25	12.500%	192	24.000%	12	6.000%	77	9.625%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	2	1.000%	7	0.875%	0	0.000%	0	0.000%

Simulation Results, Skewed Score Distribution, 100 items

100-item Test, Skewed Score Distribution, 60% of items

	Critical value	True SSI — Pairs with Pre-knowledge		True SSI — At least one without Pre-Knowedge		Approx SSI — Pairs with Pre-knowledge		Approx SSI — At least one without Pre-Knowedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	3	0.015%	0	0.000%	0	0.000%	0	0.000%
	0.00001	7	0.035%	0	0.000%	3	0.015%	0	0.000%
	0.0001	41	0.206%	0	0.000%	15	0.075%	0	0.000%
	0.001	168	0.844%	22	0.005%	99	0.497%	7	0.001%
	0.01	623	3.131%	313	0.065%	456	2.291%	143	0.030%
Individuals flagged at least once	0.000001	4	2.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	9	4.500%	0	0.000%	4	2.000%	0	0.000%
	0.0001	29	14.500%	0	0.000%	18	9.000%	0	0.000%
	0.001	64	32.000%	33	4.125%	47	23.500%	14	1.750%
	0.01	109	54.500%	275	34.375%	88	44.000%	174	21.750%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	3	1.500%	0	0.000%	0	0.000%	0	0.000%
	0.001	21	10.500%	0	0.000%	10	5.000%	0	0.000%
	0.01	59	29.500%	16	2.000%	50	25.000%	2	0.250%

100-item Test, Skewed Score Distribution, 70% of items

	Critical value	True SSI — Pairs with Pre-knowledge		True SSI — At least one without Pre-Knowedge		Approx SSI — Pairs with Pre-knowledge		Approx SSI — At least one without Pre-Knowedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	10	0.050%	0	0.000%	0	0.000%	0	0.000%
	0.001	66	0.332%	22	0.005%	23	0.116%	7	0.001%
	0.01	389	1.955%	294	0.061%	240	1.206%	145	0.030%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	15	7.500%	0	0.000%	0	0.000%	0	0.000%
	0.001	47	23.500%	32	4.000%	25	12.500%	14	1.750%
	0.01	97	48.500%	269	33.625%	81	40.500%	173	21.625%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	5	2.500%	0	0.000%	1	0.500%	0	0.000%
	0.01	45	22.500%	16	2.000%	36	18.000%	2	0.250%

100-item Test, Skewed Score Distribution, 80% of items

	Critical value	True SSI — Pairs with Pre-knowledge		True SSI — At least one without Pre-Knowedge		Approx SSI — Pairs with Pre-knowledge		Approx SSI — At least one without Pre-Knowedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	11	0.055%	24	0.005%	2	0.010%	7	0.001%
	0.01	147	0.739%	306	0.064%	70	0.352%	148	0.031%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	18	9.000%	35	4.375%	4	2.000%	14	1.750%
	0.01	75	37.500%	274	34.250%	55	27.500%	173	21.625%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	1	0.125%	0	0.000%	0	0.000%
	0.01	20	10.000%	16	2.000%	5	2.500%	2	0.250%

100-item Test, Skewed Score Distribution, 90% of items

	Critical value	True SSI — Pairs with Pre-knowledge		True SSI — At least one without Pre-Knowedge		Approx SSI — Pairs with Pre-knowledge		Approx SSI — At least one without Pre-Knowedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	1	0.005%	0	0.000%	0	0.000%	0	0.000%
	0.001	3	0.015%	21	0.004%	1	0.005%	7	0.001%
	0.01	44	0.221%	309	0.064%	19	0.095%	155	0.032%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	2	1.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	6	3.000%	31	3.875%	2	1.000%	14	1.750%
	0.01	52	26.000%	277	34.625%	31	15.500%	176	22.000%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	1	0.500%	14	1.750%	0	0.000%	2	0.250%

100-item Test, Skewed Score Distribution, 100% of items

	Critical value	True SSI — Pairs with Pre-knowledge		True SSI — At least one without Pre-Knolwedge		Approx SSI — Pairs with Pre-knowledge		Approx SSI — At least one without Pre-Knowedge	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
Pairwise	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	23	0.005%	0	0.000%	8	0.002%
	0.01	1	0.005%	279	0.058%	0	0.000%	156	0.033%
Individuals flagged at least once	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	3	1.500%	32	4.000%	1	0.500%	15	1.875%
	0.01	12	6.000%	260	32.500%	8	4.000%	180	22.500%
Individuals flagged at least 5 times	0.000001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.00001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.0001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.001	0	0.000%	0	0.000%	0	0.000%	0	0.000%
	0.01	1	0.500%	12	1.500%	1	0.500%	2	0.250%

Appendix B Simulations results using critical values of 0.01 and 0.001
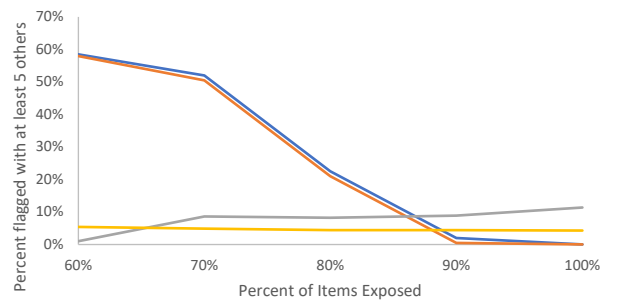
50-item test

Legend:
- True SSI (Power)
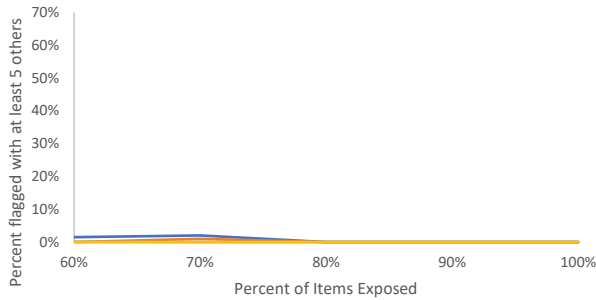- Approx SSI (Power)
- True SSI (Type I Error)
- Approx SSI (Type I Error)
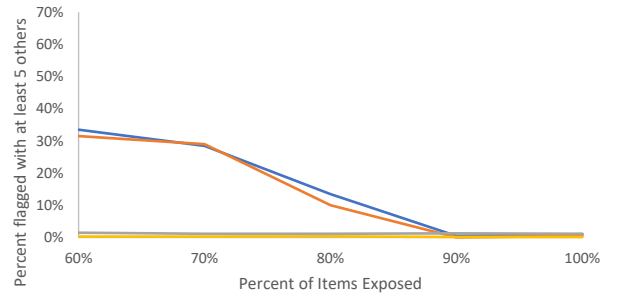
50-item, Normal Distribution, p < 0.001
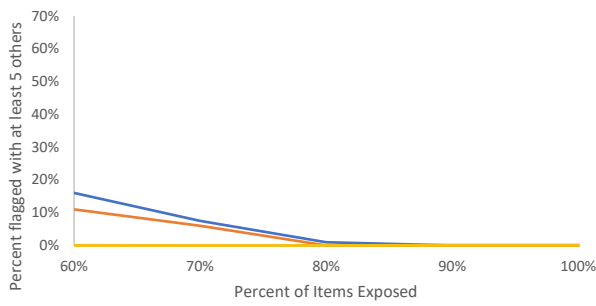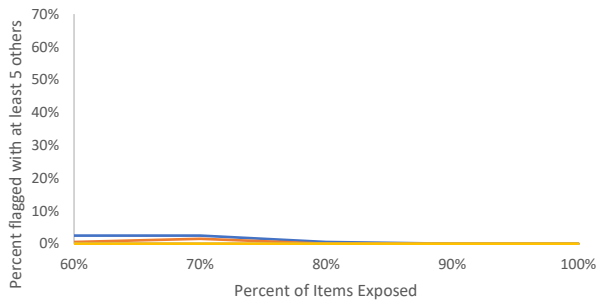
50-item, Normal Distribution, p < 0.01

50-item, Uniform Distribution, p < 0.001

50-item, Uniform Distribution, p < 0.01
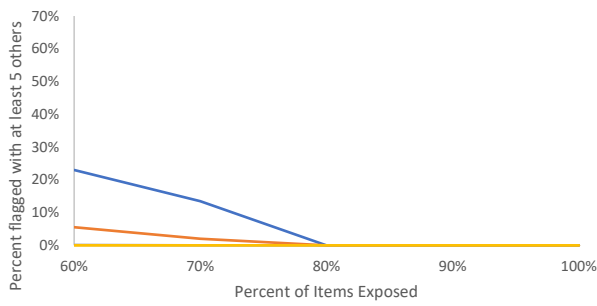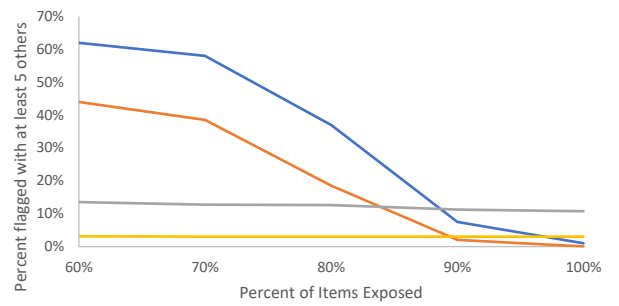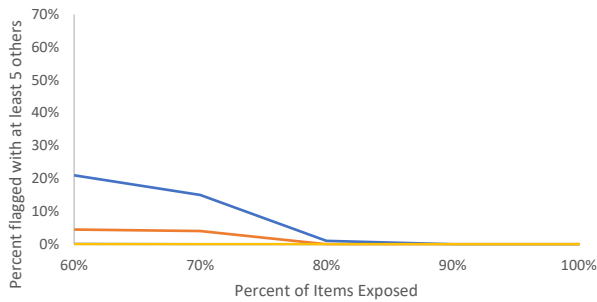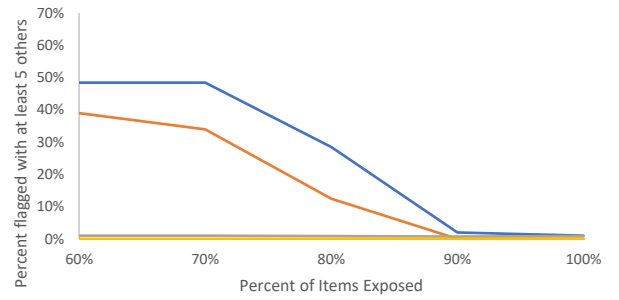
50-item, Skewed Distribution, p < 0.001

50-item, Skewed Distribution, p < 0.01
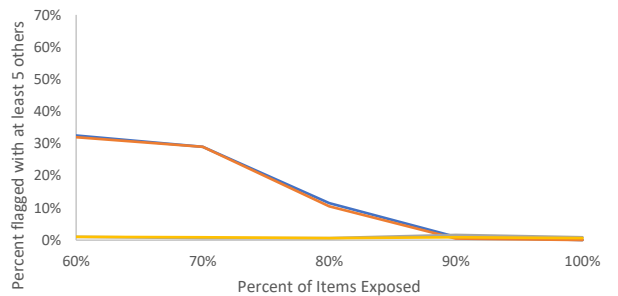
100-item test

True SSI (Power)

Approx SSI (Power)

True SSI (Type I Error)

Approx SSI (Type I Error)

100-item, Normal Distribution, p < 0.001
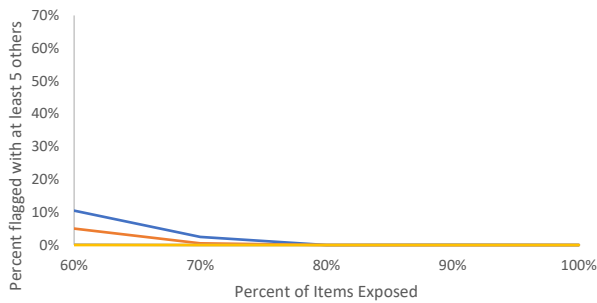

100-item, Normal Distribution, p < 0.01

100-item, Uniform Distribution, p < 0.001

100-item, Uniform Distribution, p < 0.01

100-item, Skewed Distribution, p < 0.001

100 item, Skewed Distribution, p < 0.01