Presented at the 2022 Conference on Test Security
October 26-28 Princeton, NJ

Alpine Testing Solutions

Approximation Score Similarity Index (aSSI) Analysis: An analysis of its effectiveness compared to true score similarity index

Russell W. Smith
Chief Innovation Officer

Purpose

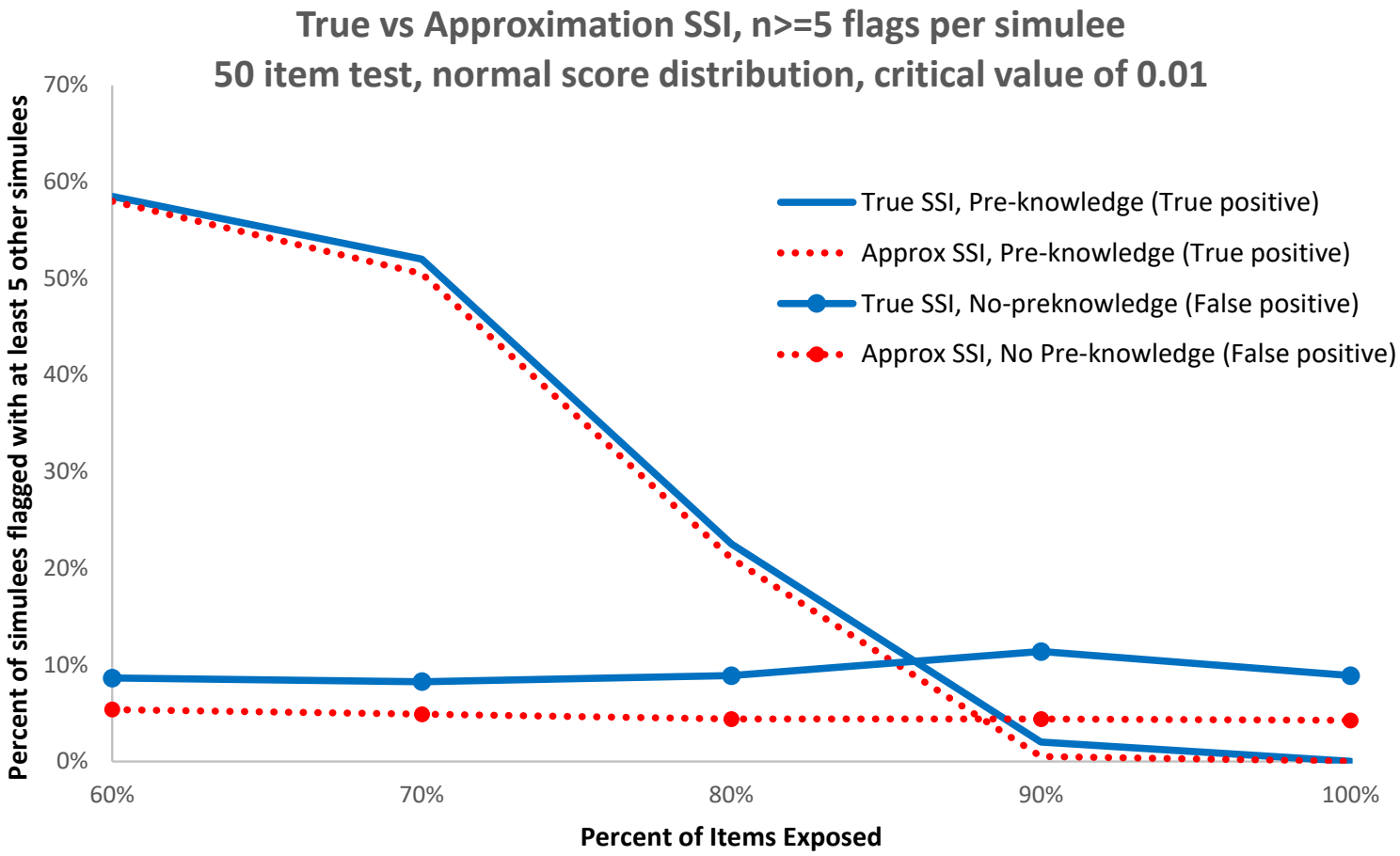
- Approximate score similarity index
  - Without Item Response Theory (IRT)/specialized software
  - Less computationally intense
  - Can be run in real time or near real time
- Does the percent of people with pre-knowledge matter?
- Does the percent of items exposed matter?
- Does exam length matter?

Purpose

- Extension of an NCME presentation

  Approximation answer and response similarity analyses: A practical approach

  - 60% or more of the content exposed, 20% of examinees with pre-knowledge



True SSI / GBT

- Calculation involves the use of a z-score (i.e., assumes a normal distribution)
- IRT-based model

$$GBT = \frac{O - E}{\sqrt{npq}}$$, where

E = sum of joint probabilities of matching scores (0,1) between two examinees given the ability of each examinee and the item's IRT parameters

O = observed agreement between two examinees

aSSI

- aSSI = z-score (for persons 1 and 2)

$$z_{12} = \frac{(M_{12} - E^*_{12})}{\sqrt{npq}}$$, where

M is count of observed matches

n is the number of items

p is E^*_{12}/n and ***q*** is $(1 - p)$

E^*_{12} is the adjusted expected value:

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

where, s_i is proportion correct score for person *i*, ***b*** is an adjustment to the magnitude of the correction set at 12.5%

Study

- **Simulations**
  - True & Approx SSI, 18 conditions
  - Simulated stochastically (+3.0 logits)
  - Multiple critical values

	NCME 2022	CoTS 2022
Number of Test Items	50 & 100	100
Person score distribution	skewed, uniform, normal	normal
Percent of exposed content	60, 70, 80, 90, 100	0, 5, 10, 15, 20, 40, 60, 80, 100
Percent of examinees with pre-knowledge	20%	1% & 5%

True Positive **1%** of people with pre-knowledge

Approximation Score Similarity Index

1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0%	0%	0%	15.6%	17.8%	22.2%	37.8%	17.8%	0%
0.005	0%	0%	0%	0%	6.7%	13.3%	17.8%	4.4%	0%
0.0005	0%	0%	0%	0%	0%	2.2%	4.4%	0%	0%
0.00005	0%	0%	0%	0%	0%	0%	0%	0%	0%
0.000005	0%	0%	0%	0%	0%	0%	0%	0%	0%

True Score Similarity Index

1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0%	2.2%	15.6%	26.7%	24.4%	28.9%	46.7%	26.7%	0%
0.005	0%	0%	0%	4.4%	8.9%	17.8%	26.7%	6.7%	0%
0.0005	0%	0%	0%	0%	0%	6.7%	4.4%	0%	0%
0.00005	0%	0%	0%	0%	0%	2.2%	0%	0%	0%
0.000005	0%	0%	0%	0%	0%	0%	0%	0%	0%

True Positive **5%** of people with pre-knowledge

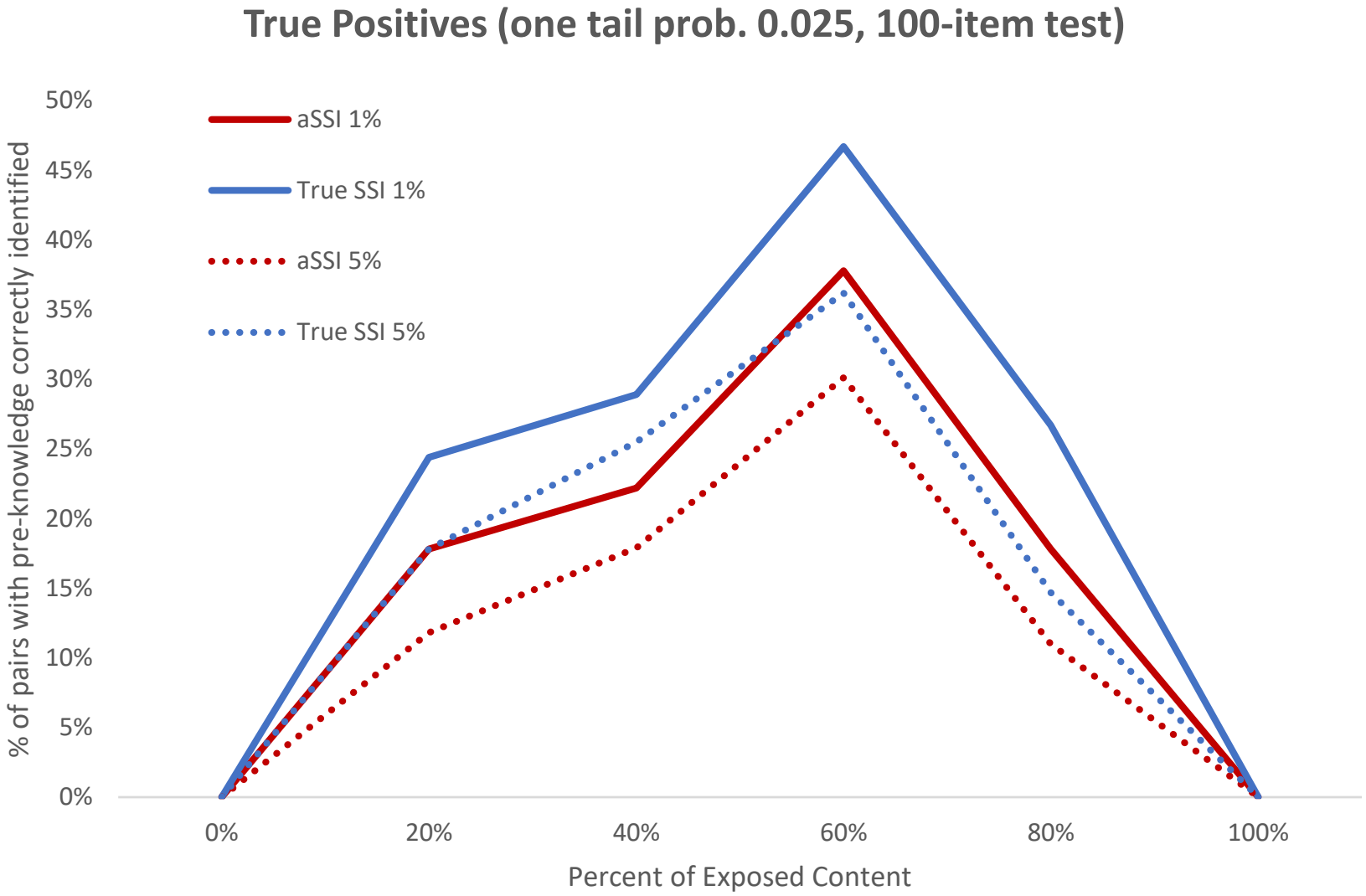
Approximation Score Similarity Index

1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0%	2.8%	5.4%	9.3%	11.8%	17.9%	30.1%	11.0%	0%
0.005	0%	0.5%	1.2%	1.8%	2.9%	4.5%	11.1%	1.8%	0%
0.0005	0%	0.1%	0.2%	0.5%	0.3%	0.7%	2.4%	0.2%	0%
0.00005	0%	0%	0.1%	0.1%	0.1%	0.2%	0.3%	0%	0%
0.000005	0%	0%	0%	0%	0%	0%	0%	0%	0%

True Score Similarity Index

1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0%	4.1%	7.7%	13.8%	17.8%	25.5%	36.2%	14.7%	0%
0.005	0%	0.8%	1.9%	3.4%	5.3%	8.8%	16.7%	4.2%	0%
0.0005	0%	0.1%	0.3%	0.7%	0.7%	2.0%	5.2%	0.7%	0%
0.00005	0%	0%	0.1%	0.2%	0.2%	0.5%	0.7%	0.2%	0%
0.000005	0%	0%	0.1%	0.1%	0.1%	0.2%	0.2%	0%	0%

Does the percent of people with pre-knowledge matter?
Does the percent of items exposed matter?



False Positive

1% with pre-knowledge

Approximation Score Similarity Index

1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%
0.005	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.1%
0.0005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.00005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.000005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

True Score Similarity Index

1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%
0.005	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
0.0005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.00005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.000005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

5% with pre-knowledge

Approximation Score Similarity Index

1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0.5%	0.5%	0.5%	0.5%	0.5%	0.4%	0.4%	0.4%	0.4%
0.005	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
0.0005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.00005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.000005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

True Score Similarity Index

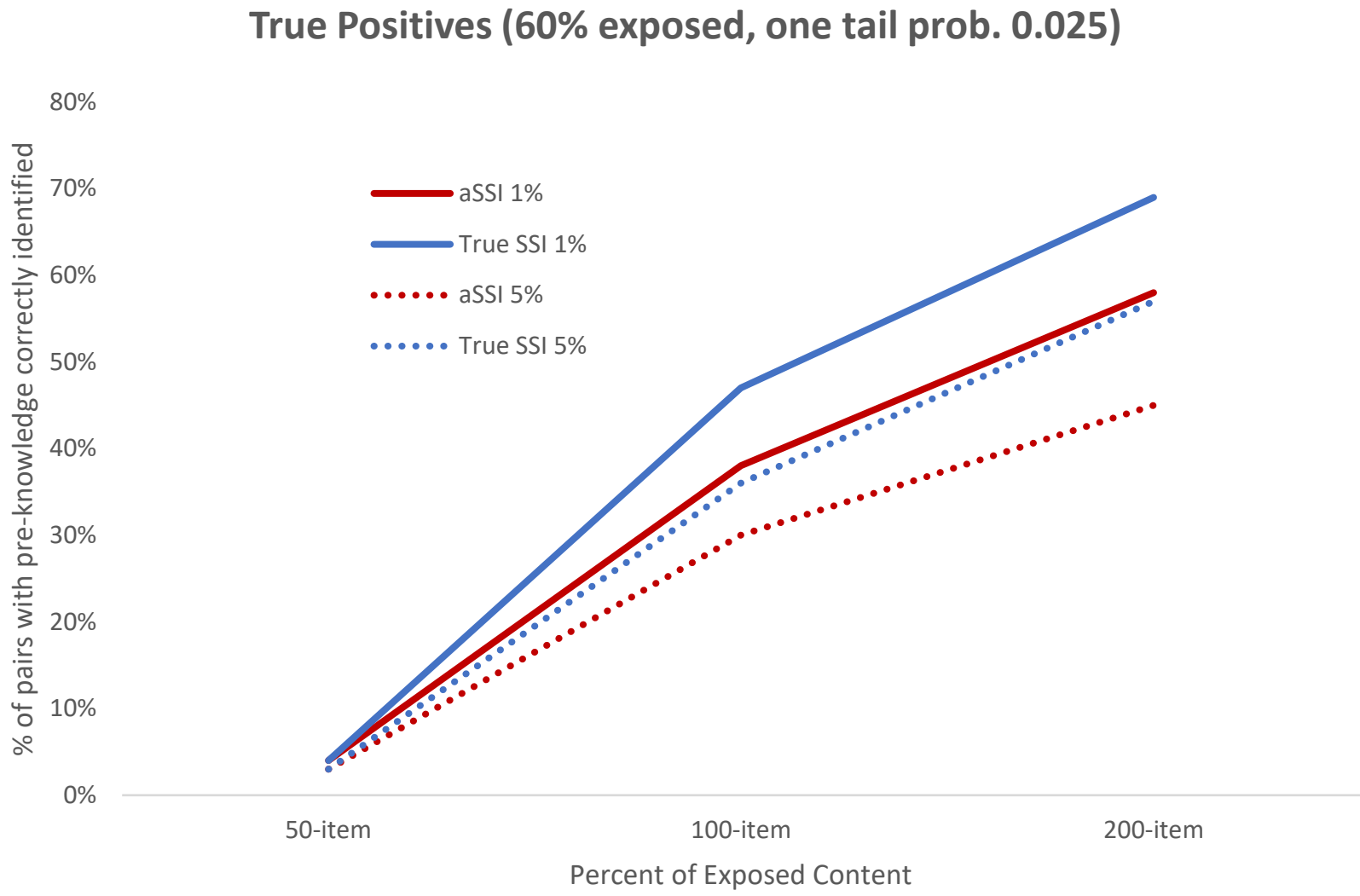
1-tail prob	% of exposed content								
	0	5	10	15	20	40	60	80	100
0.025	0%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	1%
0.005	0%	0%	0%	0.1%	0.1%	0.1%	0.1%	0.1%	0%
0.0005	0%	0%	0%	0%	0%	0.0%	0.0%	0%	0%
0.00005	0%	0%	0%	0%	0%	0.0%	0%	0%	0%
0.000005	0%	0%	0%	0%	0%	0%	0%	0%	0%

True Positives: Ad hoc analyses

Approx SSI	60% of items, 1% people, 50 item test	60% of items, 5% people, 50 item test	60% of items, 1% people, 100 item test	60% of items, 5% people, 100 item test	60% of items, 1% people, 200 items	60% of items, 5% people, 200 item test
1.96	4%	3%	38%	30%	58%	45%
2.58	0%	0%	18%	11%	20%	23%
3.29	0%	0%	4%	2%	7%	6%
3.89	0%	0%	0%	0%	0%	1%
4.42	0%	0%	0%	0%	0%	0%

True SSI						
0.025	4%	3%	47%	36%	69%	57%
0.005	0%	1%	27%	17%	31%	35%
0.0005	0%	0%	4%	5%	11%	12%
0.00005	0%	0%	0%	1%	7%	4%
0.000005	0%	0%	0%	0%	0%	1%

Does exam length matter?



False Positives: Ad hoc analyses

Alpine Testing Solutions

Approx SSI

	60% of items, 1% people, 50 item test	60% of items, 5% people, 50 item test	60% of items, 1% people, 100 item test	60% of items, 5% people, 100 item test	60% of items, 1% people, 200 items	60% of items, 5% people, 200 item test
1.96	0.83%	0.77%	0.48%	0.44%	0.32%	0.30%
2.58	0.10%	0.1%	0.05%	0.05%	0.0%	0.0%
3.29	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

True SSI

0.025	0.71%	0.71%	0.79%	0.78%	0.88%	0.87%
0.005	0.09%	0.10%	0.11%	0.11%	0.12%	0.13%
0.0005	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Contributions: Approximation SSI

- Does the percent of people with pre-knowledge matter?

  Yes. Fewer is easier to detect

- Does the percent of items exposed matter?

  Yes. There is a sweet spot at about 60%.

- Does exam length matter?

  Yes. Longer is better.

Contributions: Approximation SSI

- Approx SSI is good for “real-time”
- Solves a real-world problem
- No IRT, no calibration, computationally less intense
- Provides pairwise estimated probabilities
- All that’s required is scores, length, and a weight
- It does well under the right conditions
  - Normal distribution
  - “Sweet spot”
    - Some exposure, but not too much
  - Fewer cheaters, easier to detect

russell.smith@alpinetesting.com

References

Smith, R. W. (2021, October 6-7). *A Practical Approximation of Response Similarity*. Conference on Test Security, online.

Smith, R. W. (2022, April). *Approximation answer and response similarity analyses: A practical approach* [Paper presentation]. Annual meeting of the National Council on Measurement in Education (NCME), San Diego, CA.

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.

russell.smith@alpinetesting.com