8th Annual Conference on Test Security

Collusion Detection Using an Extension of Yen's Q_3 Statistic

Brett P. Foley, Ph.D.
Director, Professional Credentialing &
Senior Psychometrician

Development of the Statistic: Logical and Theoretical Basis

Background: IRT is Awesome!

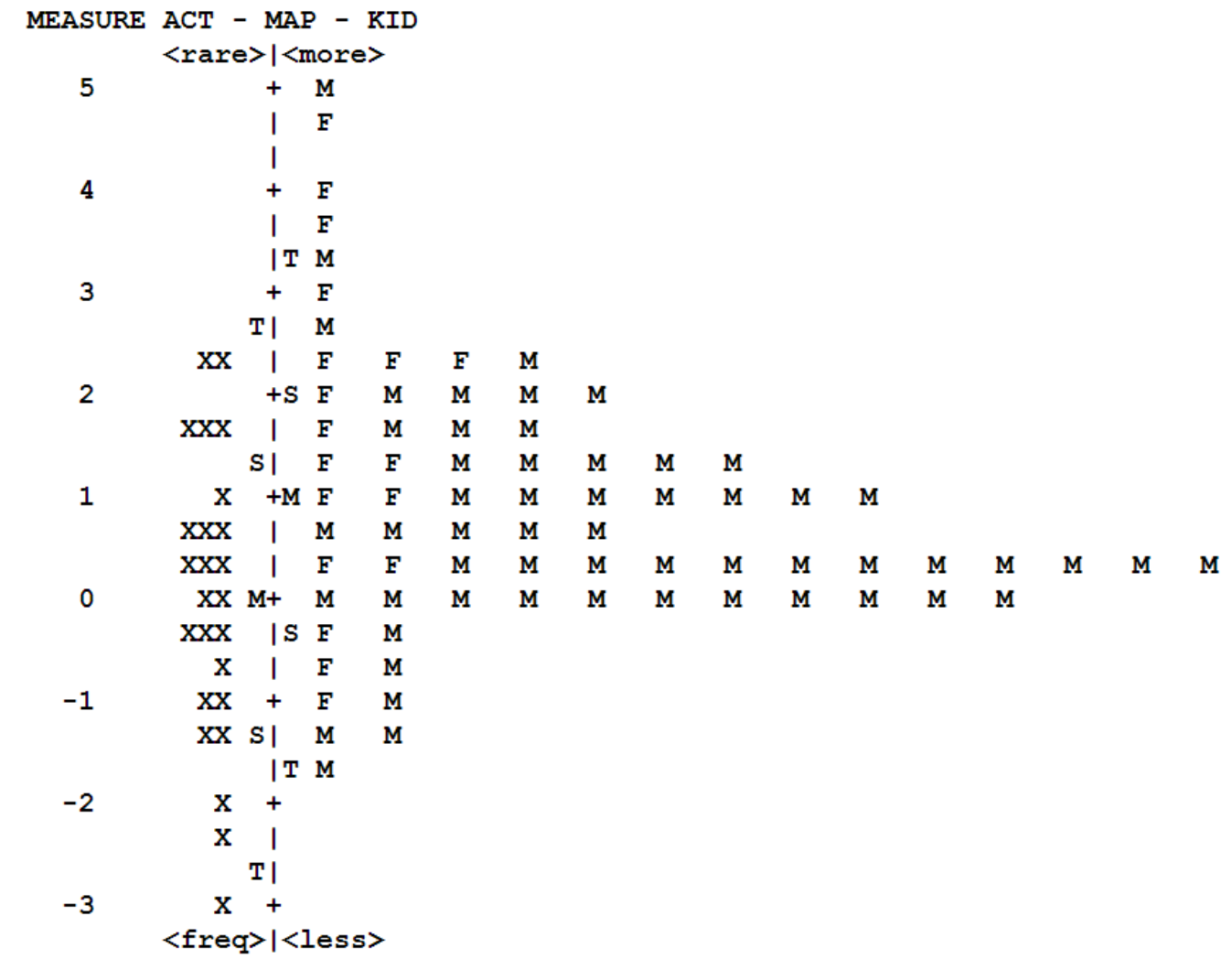


Image Source: winsteps.com

Example of Person/Item Symmetry

- Differential item functioning
  - Detection of potential bias
  - Integrates candidate group information into the analysis in order to identify potential bias issues with items
- Differential person functioning
  - Detection of potential pre-knowledge
  - Integrates item group information into the analysis in order to identify potential bias issues (e.g., item pre-knowledge) with candidates

Yen's Q_3 (1984)

- Widely used to test the assumption of local item independence in IRT analyses
- To calculate…
  - Estimated item and examinee parameters
  - Calculate the probability that examinee would answer an item correctly given his/her ability and the difficulty of the item
  - Calculate a residual by comparing this probability with the examinee's actual correct/incorrect response (or points achieved for polytomous items).
  - Repeat for all candidate/item combinations
  - Calculate correlation for each item pair

Yen's Q_3 (Continued)

- Residual

  $d_{ik} = u_{ik} - \hat{P}_i(\hat{\theta}_k)$

  - u_{ik} is the score of the *k*th examine on the *i*th item
  - $\hat{P}_i(\hat{\theta}_k)$ is the probability that the candidate will answer item *i* correctly given their estimated ability, $\hat{\theta}_k$

- Q_3
  - $q_{ij} = r_{d_i d_j}$
  - q_{ij} is the correlation across examinees of the residuals for items *i* and *j*

Yen's Q_3 (Continued)

- Expected value is approximately 0
- Indicates the similarity among a pair of items after accounting for item difficulty and examinee ability.
- High correlation -> items may be demonstrating local dependence.

- The two items are performing more similarly than one would expect.

Thought Experiment

- Consider the matrix of residuals used for calculating Q_3 statistics
  - Items across the columns and examinees down rows
- Q_3 statistics are estimated by calculating all pairwise correlations between the columns (i.e., items)
- What if we were to, instead, calculate the residual correlations between rows of this matrix (i.e., examinees)? Refer to as B_3 for convenience.

- These new results are an indication of the similarity of two examinees after accounting for item difficulty and examinee ability.
- What would a high B_3 value indicate?

Calculating B_3

- Residual (same as for Q_3)

  $d_{ik} = u_{ik} - \hat{P}_i(\hat{\theta}_k)$

  - u_{ik} is the score of the *k*th examine on the *i*th item
  - $\hat{P}_i(\hat{\theta}_k)$ is the probability that the candidate will answer item *i* correctly given their estimated ability, $\widehat{\theta}_k$

- B_3
  - $B_{3kl} = r_{d_k d_l}$
  - B_{3kl} is the correlation across items of the residuals for candidates *k* and *l*

Hypothesized Properties of B_3

- Expected value near zero
- High value would indicate that two examinees are performing more similarly than one would expect
- Exam security analysis context:
  - examinees performing more similarly than one would expect = collusion?

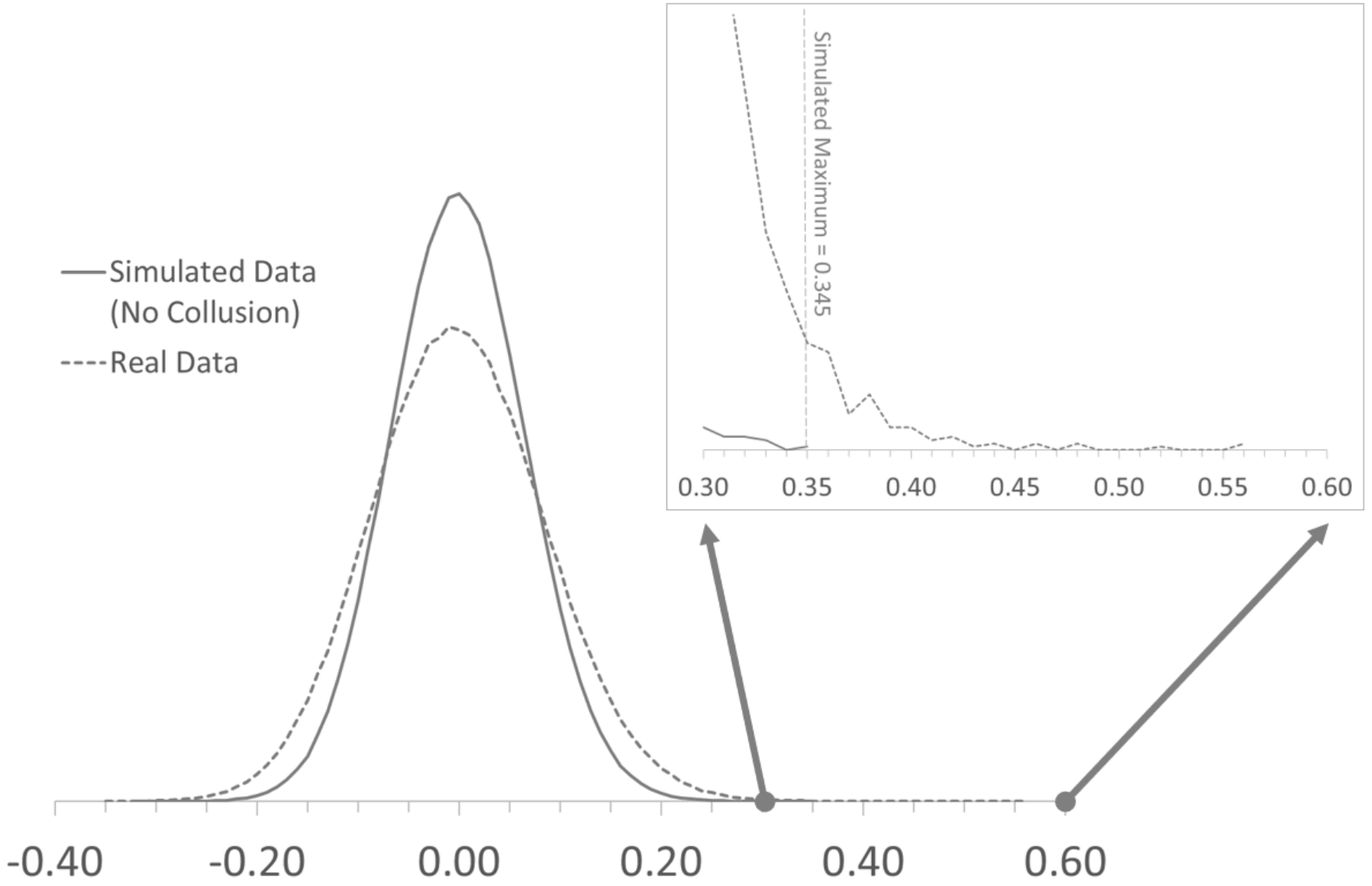
- Real data study:
  - Can B_3 identify cheating candidate pairs?
  - Identify a potential cutoff value

Research Study

Methodology

- Real Data
  - Credentialing data set (from Cizek and Wollack, 2017)
    - Form 1
    - 170 scored items
    - 1,636 candidates
    - 46 pre-flagged for suspected cheating (unspecified type)
    - Cheating may be more widespread
- Analysis
    - Calculate B_3 on real data (WINSTEPS, PCORFIL option)
    - Simulate collusion-free data
      - Use ability/difficulty estimates and reverse engineer candidate responses in WINSTEPS
      - Estimate B_3 on simulated data
      - Choose flagging threshold based on max B_3 value from simulated data
    - Compare B_3 flagging results with pre-flagged candidate list

Results: B_3 Distribution



Results: B_3 Summary Statistics

Candidates Pre-Flagged	Candidate Pairs	Average B_3		Maximum B_3		Std. Dev. B_3	
		Real	Simulated	Real	Simulated	Real	Simulated
None	1,263,255	0.00	0.00	0.44	0.35	0.09	0.07
One	73,140	0.00	0.00	0.48	0.29	0.09	0.07
Both	1,035	0.02	0.00	0.56	0.21	0.11	0.07
Overall	1,337,430	0.00	0.00	0.56	0.35	0.09	0.07

Results: B_3 Flagging Summary

B_3 Cut-off	0.345
# of flagged Pairs	121
# of flagged candidates	188
Total Candidates	1,636
# Pre-flagged Candidates	46
# also flagged by B_3	17

Probability of randomly selecting a pre-flagged candidate	2.81%
Expected # hits out of 188 random flags	5
Chances of randomly hitting 17+ correct out of 188 flags	~1 in 37,000
10 of the 12 candidates flagged in 3 or more pairs were pre-flagged	

Results: B_3 Flagging Evaluation

# of candidates…	
total	1,636
without pre-flag	1,590
with pre-flag	46

# of candidate pairs…	
total	1,337,430
without pre-flag	1,263,255
with at least 1 pre-flag	74,175
with 2 pre-flags	1,035

Probability of randomly sampling a pair with…	
at least 1 pre-flag	5.55%
2 pre-flags	0.08%

In 100 randomly sampled pairs, about how many would have…	
at least 1 pre-flag?	6
2 pre-flags?	0

Conclusions & Limitations

Conclusions

- B_3 has several appealing features:
  - Based on a well-known and widely-recognized statistic.
  - Accounts for both examinee ability and item difficulty in its calculation.
  - Easily interpretable.
  - Works with both dichotomous and polytomous items (or any mix thereof)
  - Already available in commercially available IRT software (e.g., WINSTEPS)
- Preliminary results are promising

Limitations

- Lacks a clear flagging threshold/underlying distribution
- Results show better-than-chance performance, but not yet compared to competing stats
- Need more information about false positives/false negative rates before operational use
- Preliminary analysis based on single exam
  - Cheating may be more widespread
  - Type of cheating (e.g., pre-knowledge) may not be as well suited for B_3

Questions?

Contact:
brett.foley@alpinetesting.com

Comparing B3 to Answer Similarity Index for Detecting Collusion

Russell Smith
VP Assessment Service & Senior Psychometrician
Alpine Testing Solutions

ASI vs. B3 (Person Residual Correlations)

Answer Similarity Index (ASI)

- Generalized binomial probability of matching scores/responses
- Has an underlying distribution; probabilities are straightforward
- IRT models with 1/0 work well, when they fit. Polys are challenging.
- Recursive calculations can be labor intensive, functions can shorten that

Person Residual Correlations (B3)

- Calculate expected vs. actual for each person on each item, then correlate those residuals
- Does NOT have an underlying distribution; probabilities are not straightforward
- Handles polytomous items with ease
- Readily available out of standard software

		Perfect Model Fit		Reasonable		Unexpected	
		PersonA	PersonB	PersonC	PersonD	PersonE	PersonF
Item	Measure	1	0.5	1	0.5	1	0.5
Item001	-1.00	1	1	1	1	0	0
Item002	-0.75	1	1	1	1	0	0
Item003	-0.50	1	1	1	1	1	0
Item004	-0.25	1	1	1	1	1	0
Item005	-0.10	1	1	0	0	1	1
Item006	0.10	1	1	1	1	1	1
Item007	0.25	1	0	1	0	1	1
Item008	0.50	1	0	1	1	1	1
Item009	0.75	0	0	0	0	1	1
Item010	1.00	0	0	1	0	1	1
Total Score		8	6	8	6	8	6
Number matching		8		8		8	
Person Residual Correlation/B3		0.2453		0.5274		0.7462	
Binomial probability		0.1389		0.1389		0.1389	

Three sets of analyses

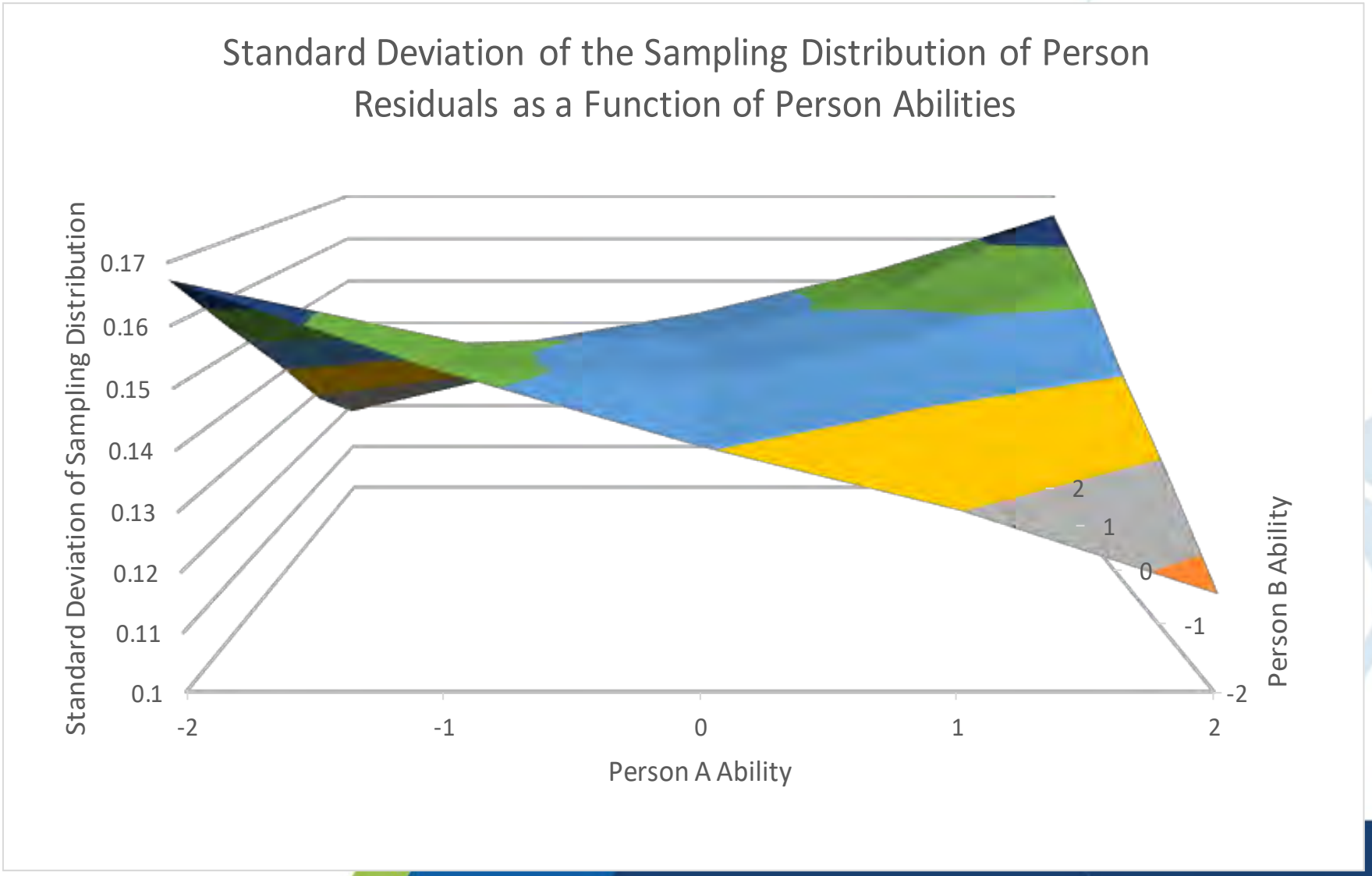
- Simulation 1
  - 50-dichotomous items, 3,000 simulees, no collusion
  - Calibrate, create fitting data, use fitting data to establish probabilities for B3
  - Compare flagging for ASI and B3
- Simulation 2
  - 8 conditions: 2 exam lengths (50 and 100 items) by 4 sizes of item difficulty variance (none, small, medium, and large), 25 pairs of candidates (abilities -2 to 2 in whole numbers), 10,000 samples for each pair, no collusion
  - Run ASI and two variations of B3: one with simulated fitting data and one with adjustments for standard deviations
  - Compare flagging for ASI and both B3s
- Real data
  - Use real exam data with known exposure problems
  - Run ASI and two variations of B3 (same as above, adjusted and not adjusted)
  - Compare flagging

Simulation 1 results

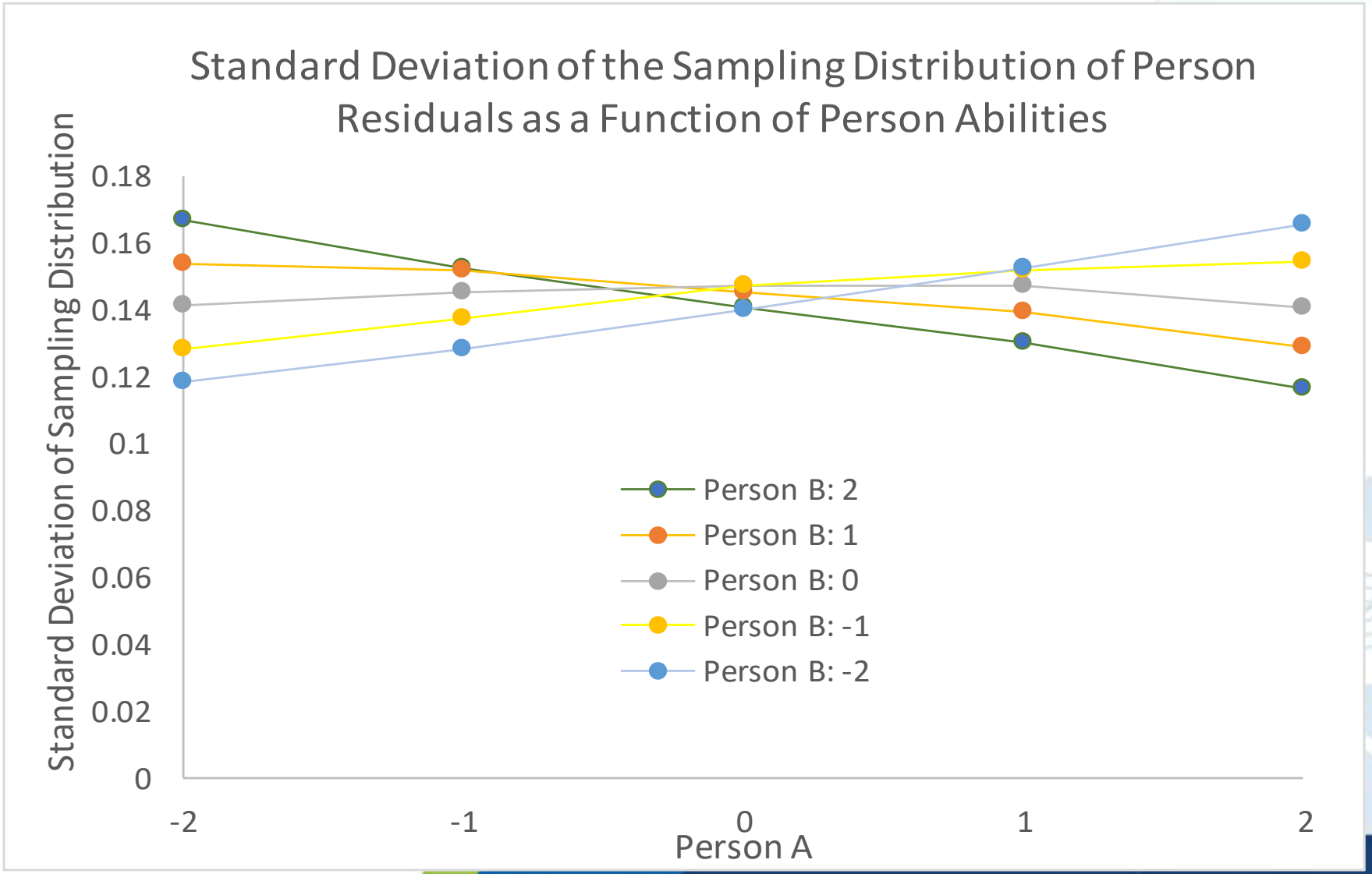
- ASI flagged 0.24% using α of 0.01
- B3 flagged 1.01% using α of 0.01
- Decision consistency: 99.1%
- Besides ASI flagging less than expected….

Boring

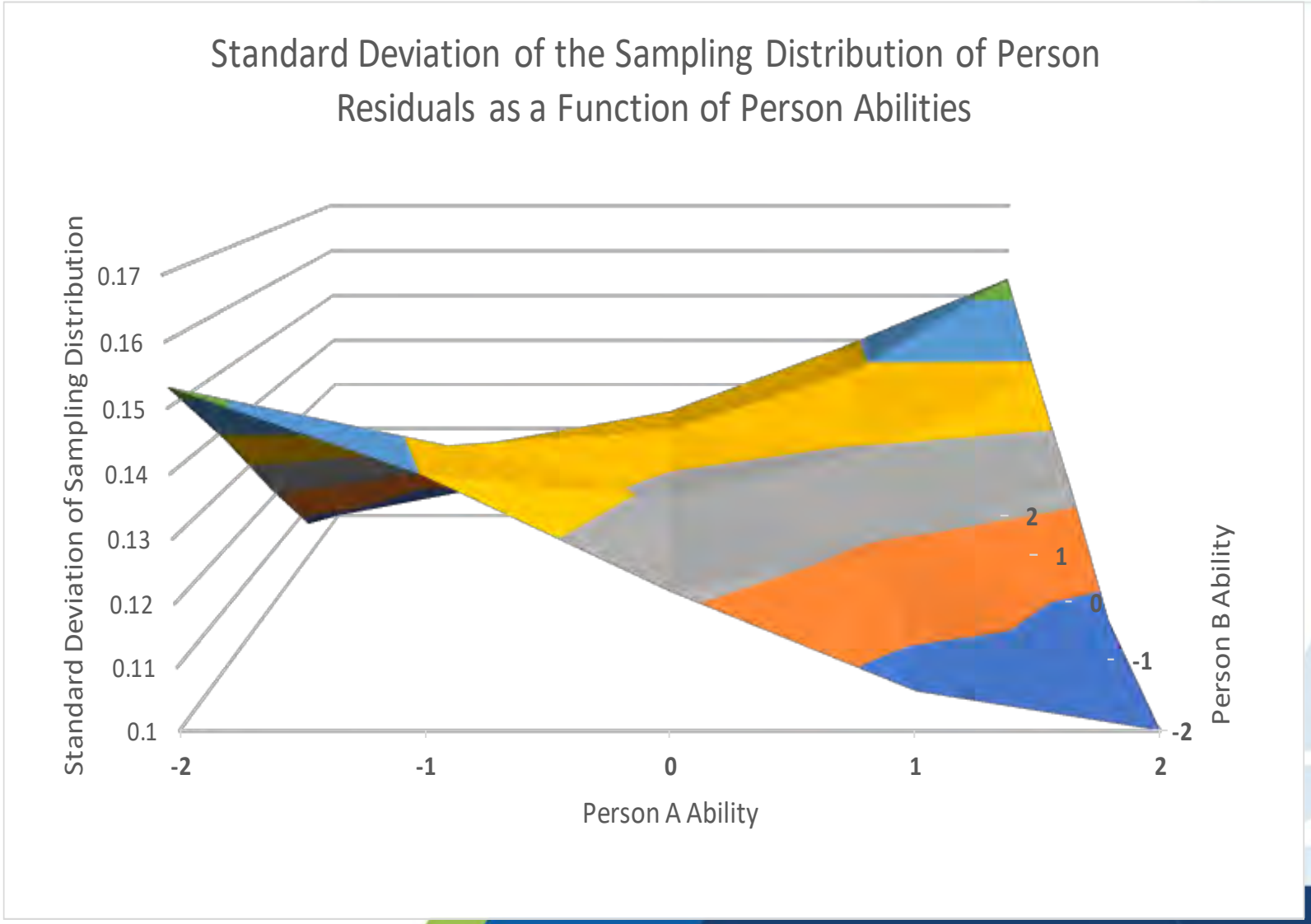
Why adjust B3?
(simulated data, 50 items, medium variance)



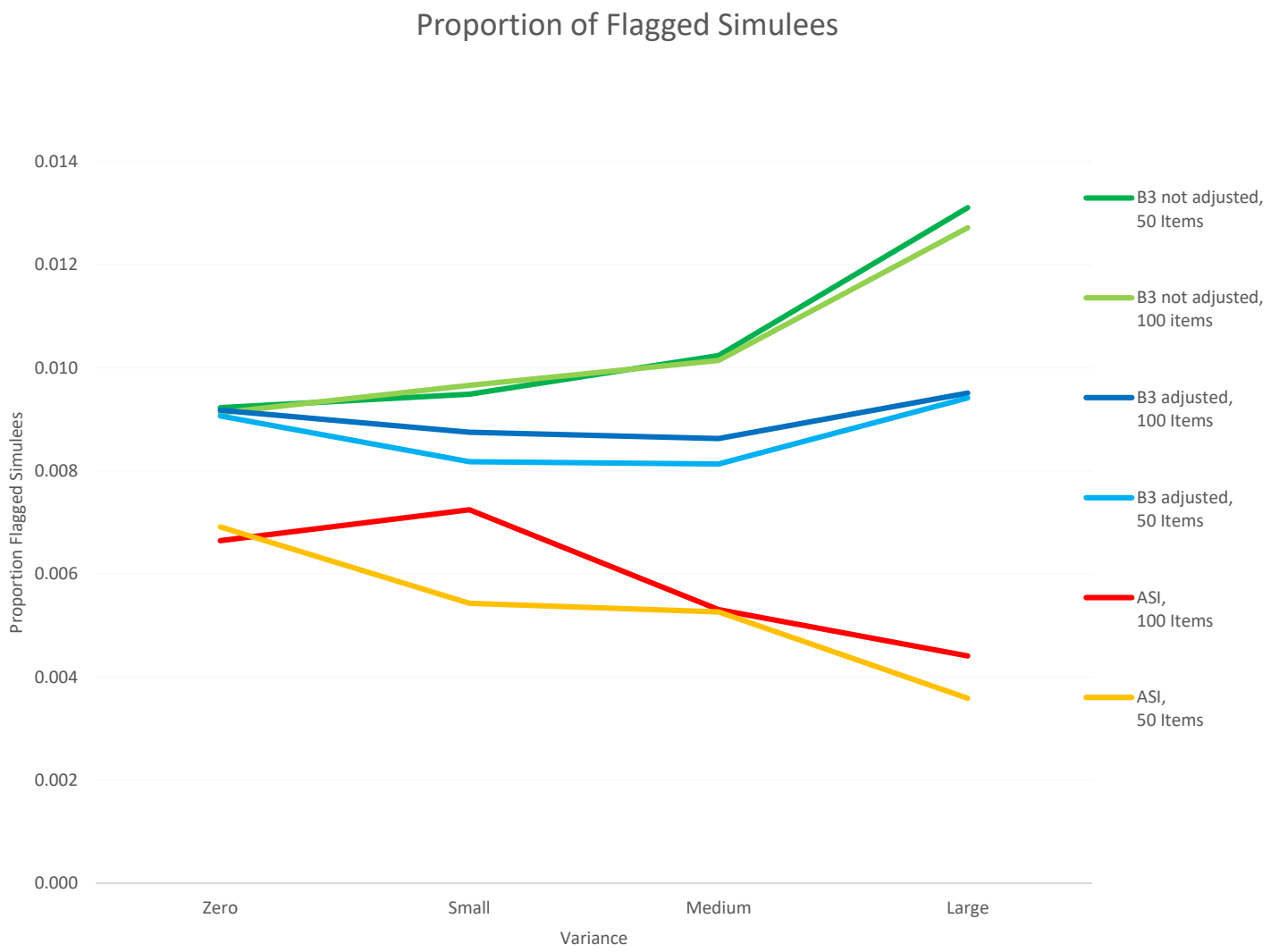
Why adjust B3?
(simulated data, 50 items, medium variance)



Why adjust B3? (real data)



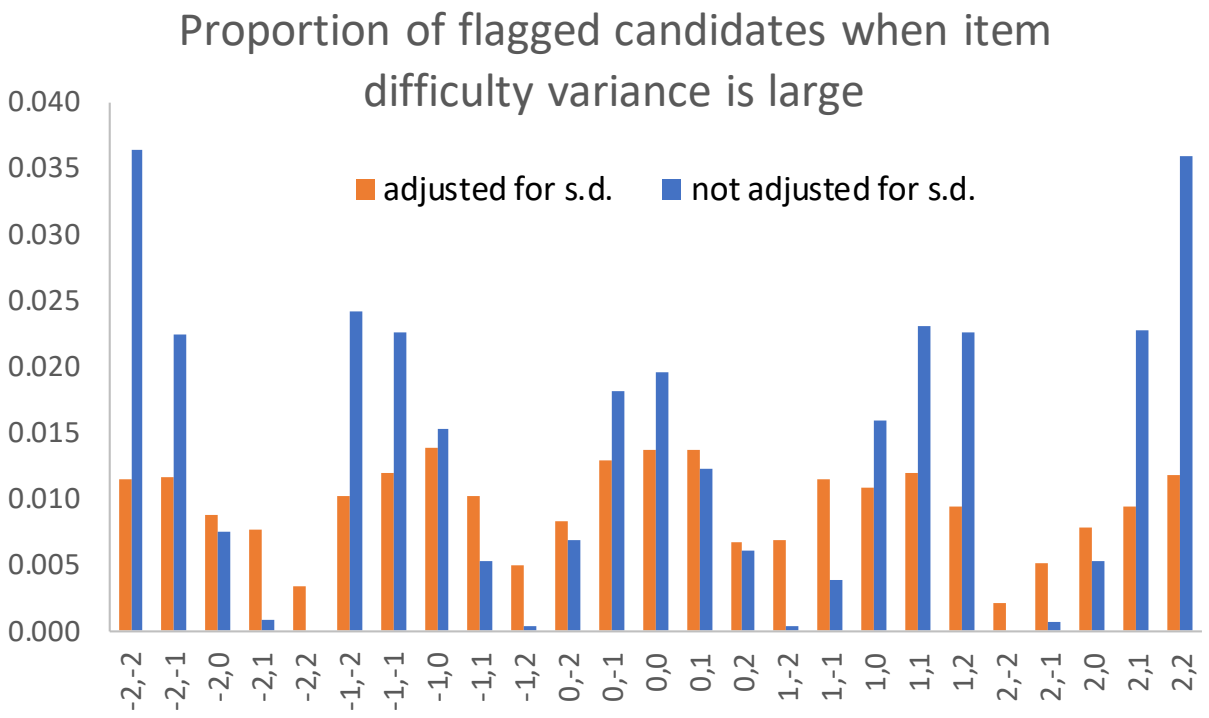
Simulation 2 Results (using α of 0.01)

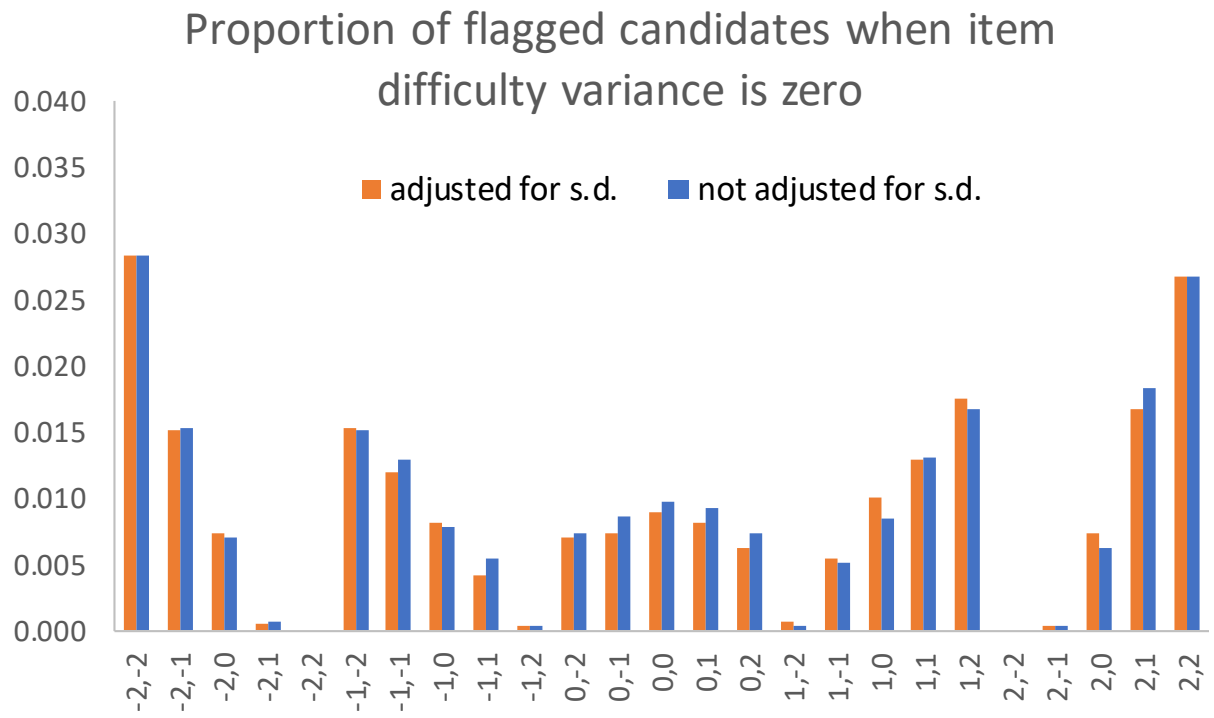


Simulation 2 Results (using α of 0.01): proportion flagged

		ASI Flagging								B3 Flagging - adjusted for standard deviation								B3 Flagging - no standard deviation adjustment							
	Items	50				100				50				100				50				100			
	Variance	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large
PersonA	PersonB																								
-2	-2	0.0029	0.0053	0.0040	0.0016	0.0078	0.0061	0.0038	0.0044	0.0283	0.0204	0.0173	0.0114	0.0203	0.0161	0.0140	0.0117	0.0283	0.0299	0.0317	0.0364	0.0199	0.0225	0.0255	0.0313
-2	-1	0.0067	0.0058	0.0025	0.0025	0.0068	0.0068	0.0065	0.0031	0.0152	0.0137	0.0135	0.0116	0.0129	0.0116	0.0122	0.0095	0.0153	0.0165	0.0182	0.0224	0.0148	0.0176	0.0165	0.0216
-2	0	0.0076	0.0058	0.0033	0.0027	0.0051	0.0062	0.0046	0.0049	0.0073	0.0063	0.0059	0.0088	0.0095	0.0101	0.0081	0.0083	0.0071	0.0069	0.0068	0.0074	0.0100	0.0084	0.0081	0.0081
-2	1	0.0064	0.0057	0.0081	0.0048	0.0071	0.0064	0.0062	0.0043	0.0005	0.0011	0.0021	0.0076	0.0038	0.0036	0.0039	0.0069	0.0007	0.0001	0.0004	0.0008	0.0027	0.0021	0.0008	0.0008
-2	2	0.0056	0.0086	0.0070	0.0048	0.0056	0.0064	0.0050	0.0043	0.0000	0.0001	0.0002	0.0034	0.0000	0.0001	0.0001	0.0041	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-1	-2	0.0067	0.0048	0.0020	0.0019	0.0074	0.0057	0.0041	0.0025	0.0153	0.0130	0.0123	0.0101	0.0143	0.0137	0.0124	0.0107	0.0152	0.0172	0.0180	0.0242	0.0121	0.0145	0.0169	0.0228
-1	-1	0.0064	0.0041	0.0039	0.0031	0.0062	0.0069	0.0042	0.0070	0.0119	0.0122	0.0104	0.0119	0.0112	0.0112	0.0107	0.0129	0.0129	0.0144	0.0141	0.0226	0.0113	0.0123	0.0135	0.0233
-1	0	0.0079	0.0045	0.0042	0.0048	0.0048	0.0092	0.0050	0.0049	0.0082	0.0085	0.0095	0.0138	0.0090	0.0096	0.0094	0.0123	0.0078	0.0106	0.0107	0.0153	0.0097	0.0102	0.0089	0.0154
-1	1	0.0068	0.0044	0.0085	0.0069	0.0062	0.0097	0.0070	0.0048	0.0041	0.0046	0.0069	0.0101	0.0073	0.0075	0.0076	0.0105	0.0054	0.0036	0.0044	0.0052	0.0083	0.0059	0.0064	0.0052
-1	2	0.0068	0.0054	0.0082	0.0043	0.0062	0.0052	0.0058	0.0035	0.0003	0.0013	0.0021	0.0050	0.0027	0.0022	0.0037	0.0051	0.0003	0.0001	0.0004	0.0004	0.0029	0.0017	0.0018	0.0007
0	-2	0.0075	0.0058	0.0035	0.0024	0.0067	0.0061	0.0042	0.0044	0.0071	0.0070	0.0075	0.0082	0.0097	0.0079	0.0086	0.0093	0.0073	0.0063	0.0049	0.0068	0.0096	0.0100	0.0078	0.0066
0	-1	0.0069	0.0038	0.0050	0.0040	0.0067	0.0093	0.0045	0.0048	0.0074	0.0106	0.0101	0.0129	0.0093	0.0101	0.0081	0.0122	0.0087	0.0085	0.0099	0.0182	0.0089	0.0093	0.0100	0.0164
0	0	0.0081	0.0062	0.0065	0.0048	0.0065	0.0079	0.0063	0.0045	0.0089	0.0106	0.0103	0.0137	0.0101	0.0103	0.0119	0.0136	0.0097	0.0106	0.0120	0.0196	0.0100	0.0104	0.0141	0.0191
0	1	0.0080	0.0041	0.0040	0.0039	0.0076	0.0091	0.0054	0.0049	0.0081	0.0088	0.0099	0.0136	0.0101	0.0088	0.0099	0.0111	0.0093	0.0101	0.0131	0.0123	0.0095	0.0092	0.0089	0.0151
0	2	0.0079	0.0059	0.0041	0.0024	0.0067	0.0085	0.0051	0.0042	0.0062	0.0062	0.0060	0.0067	0.0073	0.0105	0.0092	0.0081	0.0073	0.0074	0.0067	0.0060	0.0083	0.0079	0.0086	0.0059
1	-2	0.0078	0.0053	0.0088	0.0045	0.0058	0.0050	0.0058	0.0033	0.0007	0.0011	0.0015	0.0068	0.0027	0.0030	0.0039	0.0068	0.0004	0.0003	0.0010	0.0004	0.0039	0.0020	0.0015	0.0005
1	-1	0.0055	0.0040	0.0075	0.0056	0.0070	0.0102	0.0065	0.0062	0.0055	0.0053	0.0065	0.0114	0.0084	0.0073	0.0078	0.0109	0.0051	0.0037	0.0050	0.0038	0.0073	0.0061	0.0057	0.0052
1	0	0.0079	0.0052	0.0047	0.0028	0.0057	0.0079	0.0040	0.0038	0.0100	0.0096	0.0107	0.0108	0.0095	0.0096	0.0087	0.0109	0.0085	0.0096	0.0119	0.0159	0.0099	0.0089	0.0114	0.0135
1	1	0.0080	0.0041	0.0049	0.0040	0.0072	0.0076	0.0059	0.0060	0.0129	0.0113	0.0118	0.0120	0.0127	0.0124	0.0121	0.0113	0.0131	0.0136	0.0160	0.0230	0.0129	0.0161	0.0167	0.0210
1	2	0.0081	0.0056	0.0025	0.0019	0.0066	0.0057	0.0054	0.0029	0.0176	0.0127	0.0112	0.0094	0.0137	0.0115	0.0123	0.0115	0.0167	0.0170	0.0178	0.0225	0.0157	0.0167	0.0182	0.0224
2	-2	0.0046	0.0076	0.0070	0.0035	0.0080	0.0073	0.0053	0.0058	0.0000	0.0000	0.0000	0.0021	0.0000	0.0000	0.0006	0.0043	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-1	0.0073	0.0041	0.0087	0.0055	0.0059	0.0067	0.0063	0.0050	0.0003	0.0010	0.0019	0.0051	0.0025	0.0045	0.0046	0.0079	0.0003	0.0004	0.0003	0.0006	0.0030	0.0015	0.0015	0.0003
2	0	0.0088	0.0052	0.0044	0.0027	0.0069	0.0079	0.0050	0.0040	0.0073	0.0074	0.0082	0.0078	0.0087	0.0085	0.0093	0.0075	0.0062	0.0058	0.0051	0.0053	0.0073	0.0108	0.0089	0.0068
2	1	0.0079	0.0061	0.0036	0.0027	0.0084	0.0063	0.0070	0.0026	0.0167	0.0141	0.0125	0.0094	0.0155	0.0121	0.0126	0.0103	0.0183	0.0165	0.0177	0.0227	0.0126	0.0133	0.0164	0.0237
2	2	0.0047	0.0084	0.0048	0.0016	0.0073	0.0070	0.0038	0.0042	0.0268	0.0176	0.0150	0.0118	0.0182	0.0165	0.0140	0.0101	0.0268	0.0281	0.0299	0.0359	0.0176	0.0241	0.0255	0.0322

Simulation 2: 50 items **large** compared to **zero** variance





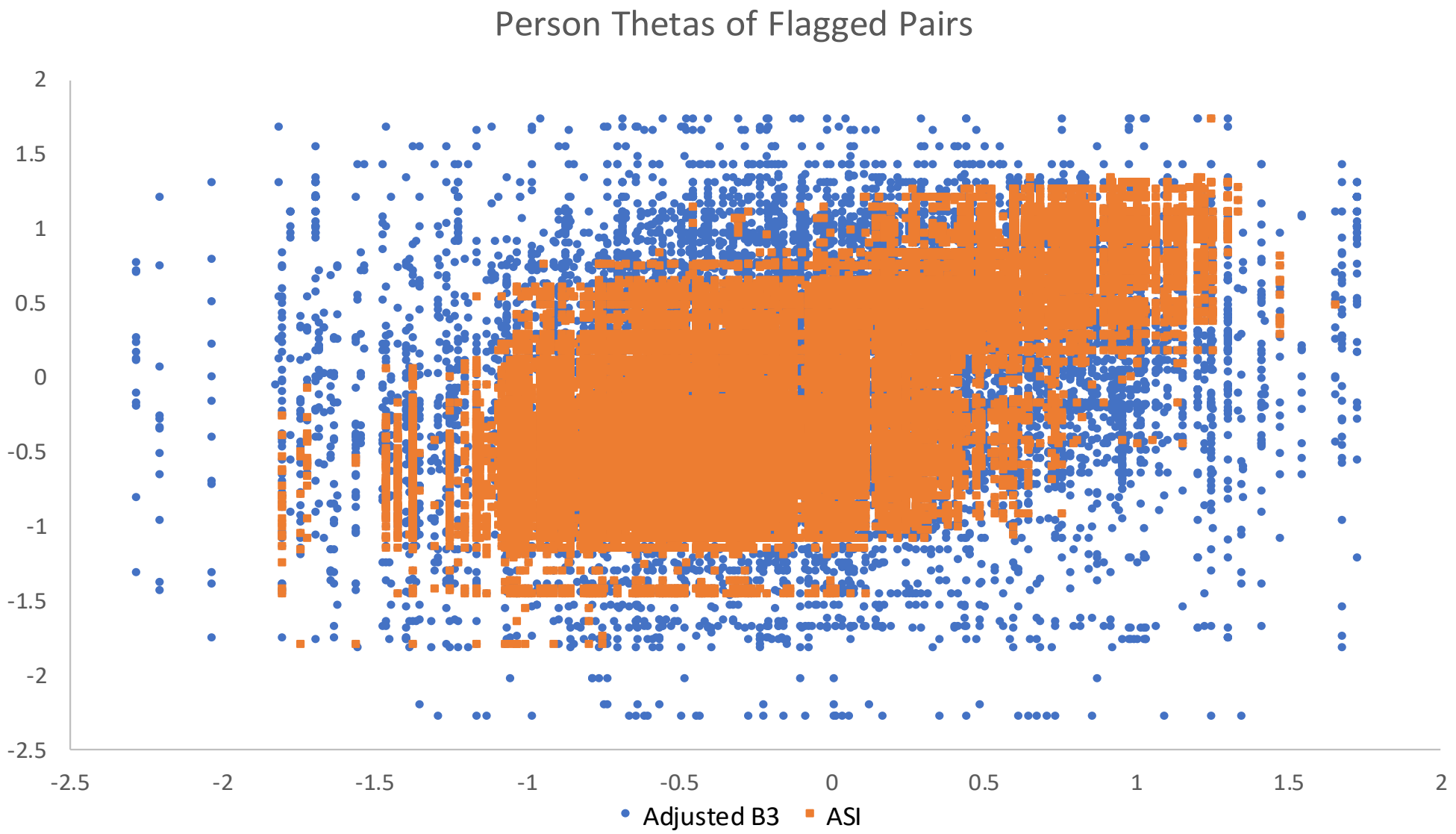
Simulation 2 Results (using α of 0.01): decision consistency

Decision Consistency		ASI vs. Adjusted B3								ASI vs. Non-Adjusted B3								Adjusted B3 vs. Non-Adjusted B3							
		50 Items				100 Items				50 Items				100 Items				50 Items				100 Items			
	Variance	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large	None	Small	Medium	Large
PersonA	PersonB																								
-2	-2	0.9680	0.9790	0.9840	0.9950	0.9760	0.9760	0.9800	0.9880	0.9680	0.9680	0.9650	0.9700	0.9760	0.9730	0.9650	0.9770	1.0000	0.9905	0.9856	0.9750	0.9996	0.9936	0.9885	0.9804
-2	-1	0.9870	0.9860	0.9910	0.9930	0.9890	0.9810	0.9890	0.9940	0.9760	0.9830	0.9870	0.9740	0.9800	0.9740	0.9880	0.9820	1.0000	0.9965	0.9941	0.9877	0.9995	0.9961	0.9959	0.9891
-2	0	0.9880	0.9940	0.9930	0.9910	0.9920	0.9870	0.9930	0.9880	0.9930	0.9910	0.9960	0.9890	0.9880	0.9870	0.9920	0.9890	1.0000	0.9999	0.9993	0.9992	0.9997	0.9995	0.9995	0.9988
-2	1	0.9950	0.9950	0.9960	0.9890	0.9940	0.9920	0.9930	0.9880	0.9920	0.9940	0.9930	0.9980	0.9950	0.9940	0.9950	0.9970	1.0000	0.9990	0.9989	0.9940	1.0000	0.9991	0.9969	0.9940
-2	2	0.9970	0.9960	0.9950	0.9880	0.9940	0.9920	0.9950	0.9940	0.9960	0.9930	0.9960	0.9950	0.9860	0.9890	0.9970	0.9930	1.0000	1.0000	1.0000	0.9979	1.0000	1.0000	0.9994	0.9957
-1	-2	0.9760	0.9870	0.9920	0.9890	0.9800	0.9790	0.9900	0.9930	0.9870	0.9840	0.9890	0.9840	0.9900	0.9770	0.9840	0.9800	1.0000	0.9965	0.9955	0.9874	0.9992	0.9971	0.9953	0.9867
-1	-1	0.9940	0.9910	0.9950	0.9890	0.9830	0.9930	0.9870	0.9890	0.9920	0.9900	0.9890	0.9820	0.9830	0.9920	0.9870	0.9720	0.9990	0.9978	0.9963	0.9893	0.9999	0.9989	0.9972	0.9896
-1	0	0.9970	0.9930	0.9900	0.9880	0.9940	0.9910	0.9870	0.9930	0.9890	0.9880	0.9940	0.9870	0.9880	0.9920	0.9930	0.9890	0.9996	1.0000	0.9994	0.9976	0.9996	0.9999	0.9992	0.9968
-1	1	0.9920	0.9940	0.9920	0.9940	0.9930	0.9950	0.9930	0.9900	0.9920	0.9910	0.9930	0.9980	0.9880	0.9910	0.9900	0.9920	0.9999	0.9983	0.9979	0.9938	0.9999	0.9986	0.9986	0.9943
-1	2	0.9850	0.9930	0.9920	0.9890	0.9950	0.9900	0.9900	0.9920	0.9980	0.9990	0.9960	0.9920	0.9960	0.9940	0.9930	0.9970	1.0000	0.9991	0.9985	0.9953	0.9996	0.9972	0.9972	0.9928
0	-2	0.9930	0.9910	0.9940	0.9880	0.9880	0.9870	0.9930	0.9890	0.9880	0.9940	0.9950	0.9950	0.9920	0.9870	0.9930	0.9900	1.0000	1.0000	0.9990	0.9980	0.9999	0.9999	0.9997	0.9983
0	-1	0.9900	0.9880	0.9950	0.9880	0.9900	0.9920	0.9930	0.9910	0.9960	0.9930	0.9900	0.9880	0.9940	0.9920	0.9880	0.9890	0.9995	1.0000	0.9996	0.9956	0.9999	0.9997	0.9994	0.9959
0	0	1.0000	0.9900	0.9910	0.9880	0.9980	0.9930	0.9910	0.9870	0.9990	0.9900	0.9910	0.9840	0.9980	0.9930	0.9910	0.9840	0.9992	1.0000	0.9983	0.9941	0.9999	0.9999	0.9978	0.9945
0	1	0.9910	0.9930	0.9900	0.9860	0.9920	0.9920	0.9870	0.9910	0.9920	0.9880	0.9910	0.9870	0.9930	0.9920	0.9930	0.9850	0.9993	0.9995	0.9976	0.9985	1.0000	0.9996	0.9998	0.9958
0	2	0.9930	0.9950	0.9920	0.9900	0.9890	0.9880	0.9910	0.9920	0.9850	0.9910	0.9940	0.9930	0.9860	0.9850	0.9890	0.9920	1.0000	1.0000	0.9985	0.9982	0.9996	0.9994	0.9993	0.9984
1	-2	0.9920	0.9940	0.9940	0.9930	0.9950	0.9930	0.9960	0.9890	0.9950	0.9960	0.9950	0.9950	0.9930	0.9940	0.9920	0.9950	0.9999	0.9992	0.9989	0.9928	0.9999	0.9984	0.9976	0.9936
1	-1	0.9920	0.9890	0.9920	0.9950	0.9880	0.9910	0.9890	0.9870	0.9910	0.9940	0.9930	0.9960	0.9930	0.9950	0.9930	0.9930	0.9990	0.9991	0.9981	0.9937	1.0000	0.9986	0.9981	0.9947
1	0	0.9900	0.9890	0.9950	0.9910	0.9930	0.9920	0.9930	0.9910	0.9910	0.9920	0.9900	0.9840	0.9920	0.9920	0.9830	0.9900	0.9996	0.9992	0.9980	0.9977	0.9998	0.9999	0.9985	0.9976
1	1	0.9890	0.9890	0.9860	0.9940	0.9900	0.9920	0.9920	0.9880	0.9890	0.9850	0.9820	0.9840	0.9900	0.9880	0.9870	0.9790	0.9998	0.9977	0.9958	0.9890	0.9998	0.9963	0.9954	0.9903
1	2	0.9910	0.9840	0.9910	0.9900	0.9820	0.9830	0.9810	0.9870	0.9830	0.9880	0.9780	0.9860	0.9840	0.9810	0.9790	0.9770	1.0000	0.9971	0.9947	0.9869	0.9998	0.9954	0.9944	0.9879
2	-2	0.9960	0.9930	0.9960	0.9950	0.9860	0.9890	0.9960	0.9910	0.9970	0.9960	0.9950	0.9900	0.9940	0.9920	0.9950	0.9980	1.0000	0.9999	0.9998	0.9966	1.0000	0.9999	0.9999	0.9959
2	-1	0.9980	0.9980	0.9940	0.9870	0.9960	0.9920	0.9930	0.9940	0.9850	0.9930	0.9950	0.9910	0.9950	0.9910	0.9920	0.9970	1.0000	0.9991	0.9982	0.9956	0.9997	0.9993	0.9978	0.9952
2	0	0.9850	0.9910	0.9940	0.9930	0.9860	0.9860	0.9890	0.9900	0.9930	0.9950	0.9930	0.9910	0.9890	0.9870	0.9910	0.9920	1.0000	0.9996	0.9991	0.9986	1.0000	0.9997	0.9997	0.9987
2	1	0.9830	0.9900	0.9830	0.9930	0.9840	0.9860	0.9820	0.9860	0.9910	0.9810	0.9870	0.9800	0.9840	0.9800	0.9730	0.9750	0.9993	0.9962	0.9935	0.9867	0.9989	0.9982	0.9959	0.9878
2	2	0.9780	0.9790	0.9860	0.9880	0.9830	0.9880	0.9840	0.9950	0.9780	0.9740	0.9730	0.9680	0.9830	0.9840	0.9790	0.9790	1.0000	0.9895	0.9851	0.9759	0.9994	0.9924	0.9885	0.9779

Results: Real Data Decision Consistency

		B3 non-adjusted		
B3 Adjusted		Flagged	Not Flagged	
	Flagged	5.12%	0.16%	5.28%
	Not Flagged	0.03%	94.69%	94.72%
		5.15%	94.85%	
		ASI		
B3 Adjusted		Flagged	Not Flagged	
	Flagged	3.95%	1.33%	5.28%
	Not Flagged	1.83%	92.89%	94.72%
		5.79%	94.21%	
		ASI		
B3 non-adjusted		Flagged	Not Flagged	
	Flagged	3.94%	1.20%	5.15%
	Not Flagged	1.84%	93.01%	94.85%
		5.79%	94.21%	

Results: Real Data Decision Consistency



Conclusions

- B3 is intuitive
- Has basis in measurement
- Easy to calculate
- Adds something beyond ASI
- We need to figure out how to better estimate the underlying distribution (especially if used as additional evidence)
- Even without that, it can be used to identify systemic problem (especially if combined with cluster analyses)

Thank you!

For a copy of the paper and/or presentation:
russell.smith@alpinetesting.com