A Study of Modern Detectors of Examinees with Pre-Knowledge Using Real, Marked Data

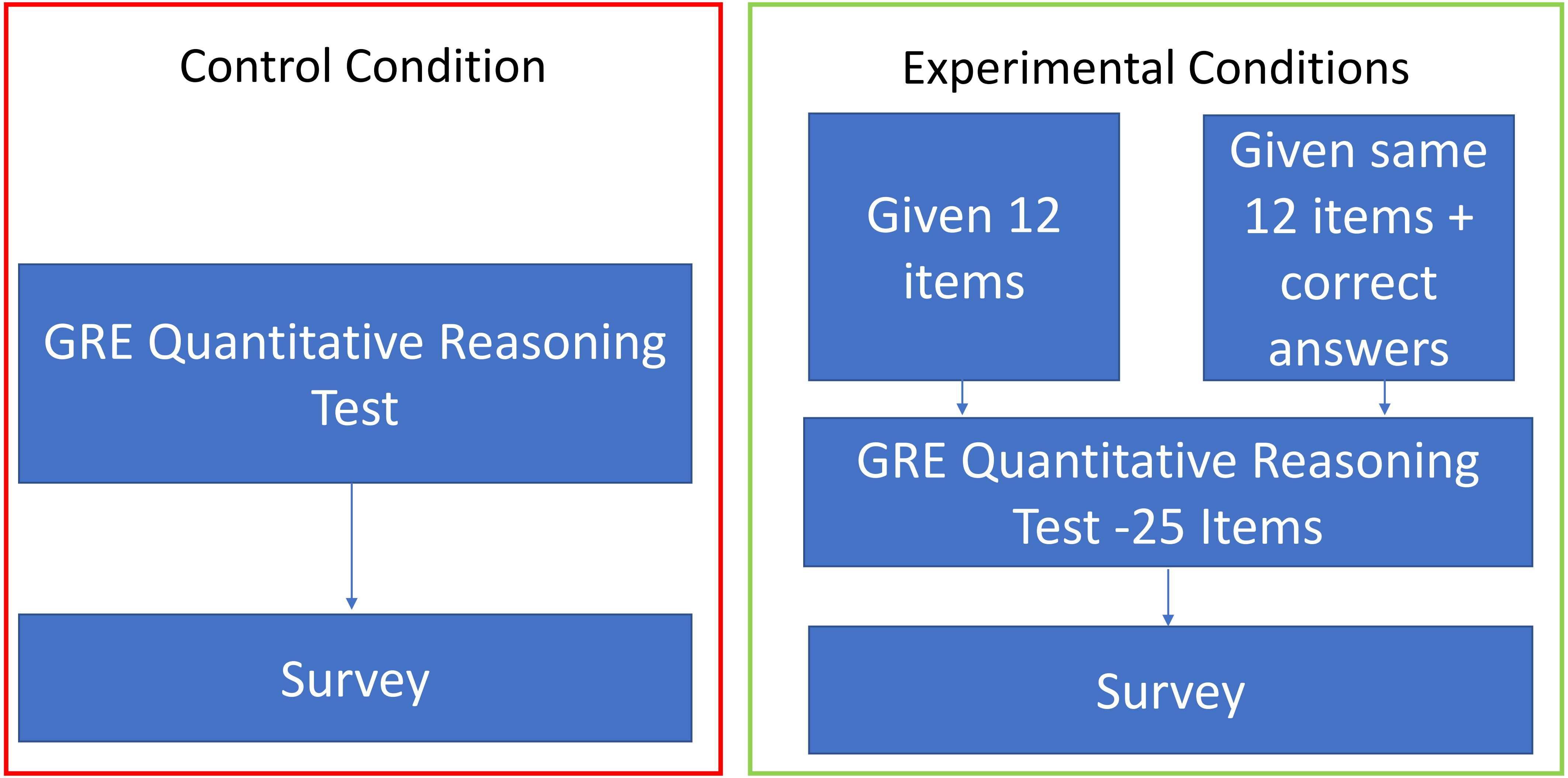
Sarah Toton
Caveon

Dmitry Belov
Law School Admission Council

Pre-Knowledge

- Very challenging to detect
- Involves unknown subset(s) of compromised items
- Involves unknown subset(s) of examinees using compromised items
- Occurring in addition to a variety of normal test-taking behaviors and potentially other types of cheating

Experimental Design



Statistics to Detect Pre-Knowledge

- lz (Drasgow, Levine, & Williams, 1985)
  - Measures person-fit, quantifying the number of unexpected responses to items
- Neyman-Pearson lemma-based statistic (NPL; Drasgow et al., 1996)
  - Likelihood ratio between probability of a response vector assuming item pre-knowledge is present and the probability of a response vector assuming no item pre-knowledge
- Posterior shift statistic (PSS; Belov, 2017)
  - Compares posteriors of ability between subsets of items (e.g., known compromised and others or known uncompromised and others)
- Alternating Minimization statistic (AM; Belov, 2016b)
  - Injects a statistic (similar to statistics measuring score gain from one item subset to another) into a specially organized Markov Chain Monte Carlo
  - The Monte Carlo Posterior Shift (MCPS) provides the initial estimate of examinees with pre-knowledge

Alternating Minimization Statistic

Minimizes the distance between 1) subsets of examinees and 2) subsets of items in an alternating cycle.

C is the subset of items currently detected as compromised.

PK is the subset of examinees who are currently detected as having pre-knowledge.

- Step 1: Draw random subsets of items and compute the posterior shift. The average posterior shift becomes the statistic by which the first estimate of *PK* is obtained.
- Step 2: For each item on the test, the uncompromised item that has the most similar item difficulty value is identified. The posterior shifts are computed using the responses of *PK* examinees to the item and then to the similar uncompromised item. *C* is formed by selecting items with a posterior shift equal to or greater than a critical value.

Alternating Minimization Statistic

- Step 3: *PK* is reformed by selecting examinees who performed better on *C* than on uncompromised items.
- The alternating cycle stops when the minimum distance stabilizes.

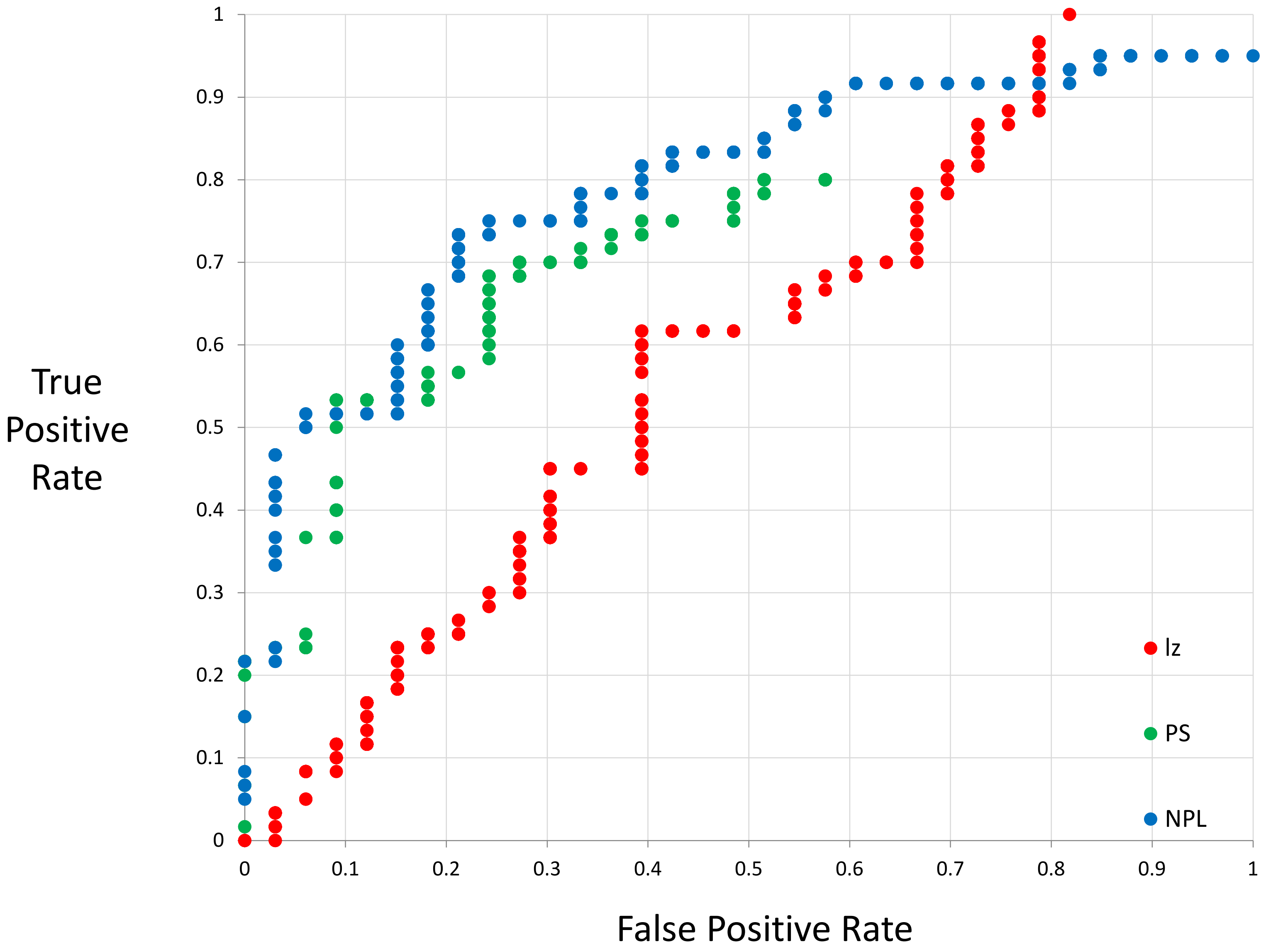
Analysis Design

- Study 1
  - Compromised items are known
  - Compute lz, posterior shift statistic (PSS), and Neyman-Pearson lemma (NPL)
  - Hypotheses based on Belov (2016a)
- Study 2
  - Compromised items are unknown
  - A subset of uncompromised items is known
  - Compute PSS, MCPS, and AM
- Analysis groups:
  - participants from all conditions,
  - participants from only the control and Item conditions, and
  - participants from only the control and Item+Answer condition.

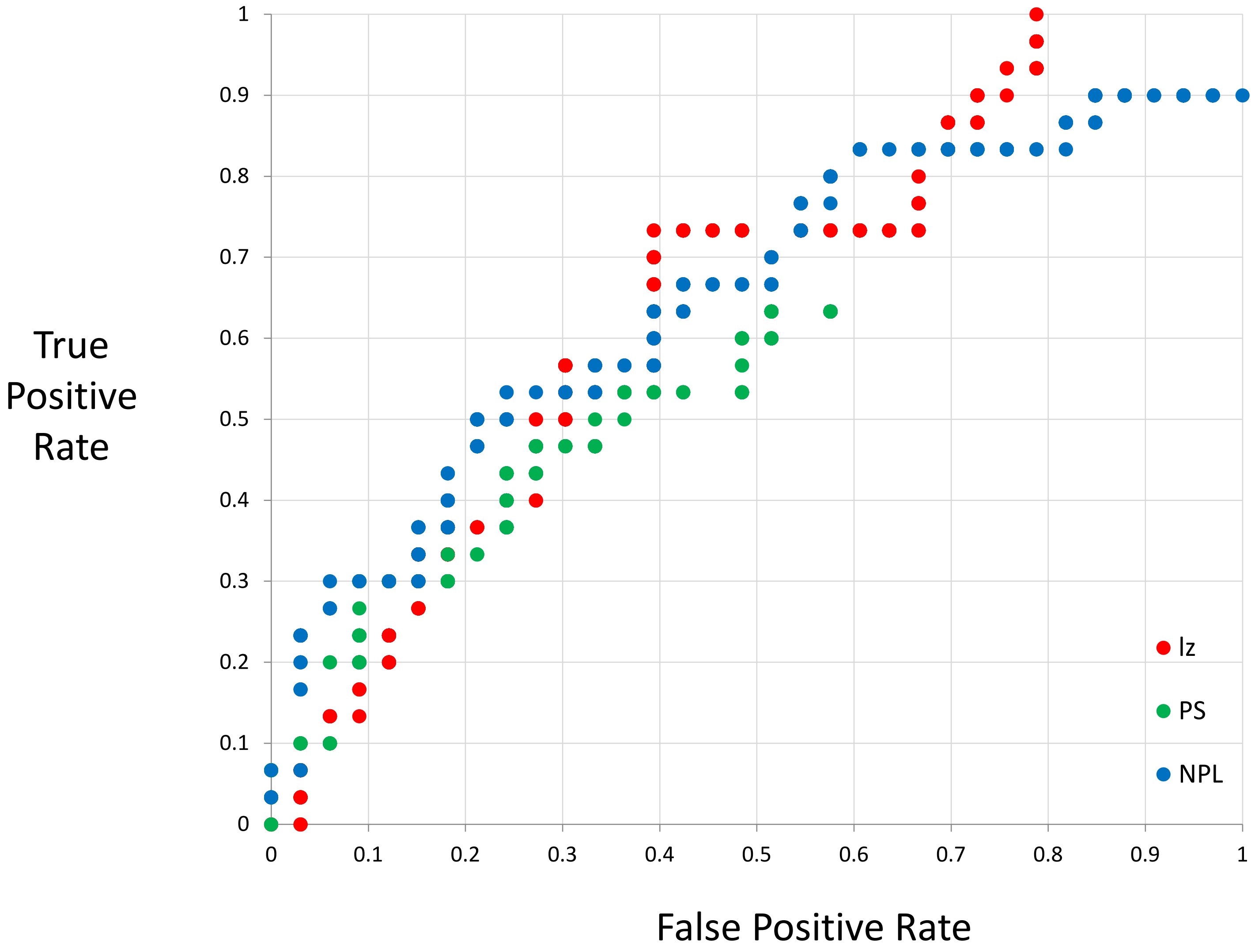
Study 1 Results- Known Compromised Subset

	N	Truncated ROC			ROC Area		
		lz	PSS	NPL	lz	PSS	NPL
All Conditions	93	0.05	0.30	0.42	0.59	0.69	0.79
Control vs Item	63	0.08	0.14	0.21	0.65	0.51	0.65
Control vs Item+Answer	63	0.01	0.45	0.62	0.52	0.88	0.94

Study 1 Results- All Conditions, ROCs



Study 1 Results- Control vs Item, ROCs



Study 1 Results-Control vs Item+Answer, ROCs

Study 2

- The performance of the detectors was studied under a variety of conditions. The uncompromised subsets provided were:
  - all uncompromised items (*N* = 13),
  - approximately 75% of the uncompromised items (*N* = 10),
  - approximately 50% of the uncompromised items (*N* = 7), and
  - approximately 25% of the uncompromised items (*N* = 3).
- We expected the best detection rate would be achieved when all of the uncompromised items were known and when the control and Item+Answer conditions were compared.

Study 2 Results- All Uncompromised as Subset

	N	Truncated ROC			ROC Area		
		PSS	MCPS	AM	PSS	MCPS	AM
All Conditions	93	0.17	0.08	0.34	0.72	0.74	0.78
Control vs Item	63	0.03	0.07	0.18	0.60	0.61	0.66
Control vs Item+Answer	63	0.37	0.37	0.63	0.85	0.91	0.90

Study 2 Results- All Uncompromised as Subset

	N	Accurately Detected Items/ Total Detected Items	Accurately Detected Examinees/ Total Detected Examinees
All Conditions	93	8/8	29/33
Control vs Item	63	8/8	9/12
Control vs Item+Answer	63	7/7	26/36

Ideally 11/11

Ideally 60/60 30/30 30/30

Study 2 Results- All Uncompromised as Subset

	N	Detection Rate for Items	Precision for Items	Detection Rate for Examinees	Precision for Examinees
All Conditions	93	.73	1.00	.48	0.88
Control vs Item	63	.73	1.00	.30	0.75
Control vs Item+Answer	63	.64	1.00	.87	0.72

Conclusions

- Study 1 showed similar results to those presented in Belov 2016b, except the performance was lower (similar to when more uncertainty/noise was introduced)
- Study 2 showed that MCPS and AM methods provide very useful starting points to identify groups of compromised items and examinees with pre-knowledge
- Both studies showed that stronger pre-knowledge is more easily detectable. For example, pre-knowledge is stronger in examinees who receive correct answers