Detecting Test Tampering at the Group Level

James A. Wollack
University of Wisconsin

Carol Eckerly
University of Wisconsin

- Few approaches to detection at group-level
- Unusually large score gains
  - Likely to lose power as group expands to include non-tampered individuals/classes/schools
- Empirical estimates of number of erasures
  - No clear understanding of error rates
  - No accurate probabilistic statement of the likelihood of results
- Very little is known about how well these approaches actually work
- Current study focused on a model-based approach to detect tampering at the group-level

Erasure Detection Index (EDI)

- EDI (Wollack, Cohen, & Eckerly, 2013) compares individual’s WTR score with that person’s expected WTR score
  - Expected number is estimated as the expected number correct score across all erased items
  - Appropriate IRT model is used to estimate $P(x_{ij} = 1)$
  - Estimate θ_j across non-erased items only: $\theta \downarrow j[i \notin I \downarrow E,j\ ]$
- $EDI = X\downarrow j, I\downarrow E,j\ - \sum i \in I\downarrow E,j \uparrow \text{▒} P(x\downarrow ij = 1)\ - 1/2\ /\text{■}\surd \sum i \in I\downarrow E,j \uparrow \text{▒} P(x\downarrow ij = 1)[1 - P(x\downarrow ij = 1)\ ]\ \ = WTR - E(WTR|I\downarrow E,j\ ) - 1/2\ /SE(WTR)$

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

EDI Properties

- Properties were examined in simulation study
  - Multiple types of tampering and benign erasures
  - Manipulated the ability-level of tampered student
  - 5 – 15 tampered items per student
- EDI had strong Type I error control and power

Power of EDI for Individuals

5 Tampered items

Quintile	.00001	.0001	.0005	.001	.005	.01	.05
1	.140	.258	.385	.458	.676	.765	.961
2	.005	.018	.046	.075	.287	.420	.794
3	.000	.001	.007	.014	.081	.162	.605
4	.000	.000	.000	.000	.011	.035	.304
5	.000	.000	.000	.000	.000	.000	.086

10 Tampered items

1	.587	.779	.888	.927	.980	.991	.999
2	.077	.250	.473	.584	.834	.904	.990

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Extension of EDI to the Group Level

- Computation of EDI at student-level involves three components: WTR, E(WTR)*, and SE(WTR)*
  - * denotes that $\theta_{j[i \notin I_{E,j}]}$ is used in place of θ_{j}.
- $EDI_g = \sum_{j \in g} \left[X_{j_g, I_{E,j}} - \sum_{i \in I_{E,j}} P(x_{\;g} = 1) \right] - 1/2 \Big/ \sqrt{\sum_{j \in g} \left[\sum_{i \in I_{E,j}} P(x_{ij} = 1)[1 - P(x_{ij} = 1)] \right]}$
- Compute EDI components for each student in group
- Essentially treats the class as a single student taking one really long test, except that each student's $\theta_{j[i \notin I_{E,j}]}$, erased items, and WTR data are used for summary statistic.

Simulating Erasures

- Data simulated under the nominal response model
    - 50-item test
- Included both fraudulent and benign erasures
- Within each level of fraudulent erasures studied, benign erasures were simulated for all examinees.
    - Misalignment Erasures for random 2% of examinees
        - # Misaligned ~ Bin(50, .25)
    - Random Erasures remaining 98% examinees
        - # Random erasures ~ Bin (50, .02)
        - Approximately 1/3 students had no benign erasures

Simulating Fraudulent Erasures

- Simulated on top of benign erasures
  - 1,000 replications (Schools) per condition
  - School-Level Variables
    - School Selection: Random or Mean Ability-Weighted
    - Classes/School (1, 3, 6) × % Tampered Classes (0%, 33%, 67%, 100%)
      - 0% provided null data for Type I error study
      - 33% and 67% conditions not possible with 1 Class—7 power conditions
  - Class-Level Variables
    - # Erasure Victims per class: 1, 3, 5, 10
    - Victim Selection: Random or Ability-Weighted
    - # Tampered Items per victim: 3, 5, 10
    - Class size: 15, 25, 35
  - Tampered questions were simulated to be answered correctly
  - α (7 levels): .05, .01, .005, .001, .0005, .0001, .00001

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Implementation and Evaluation

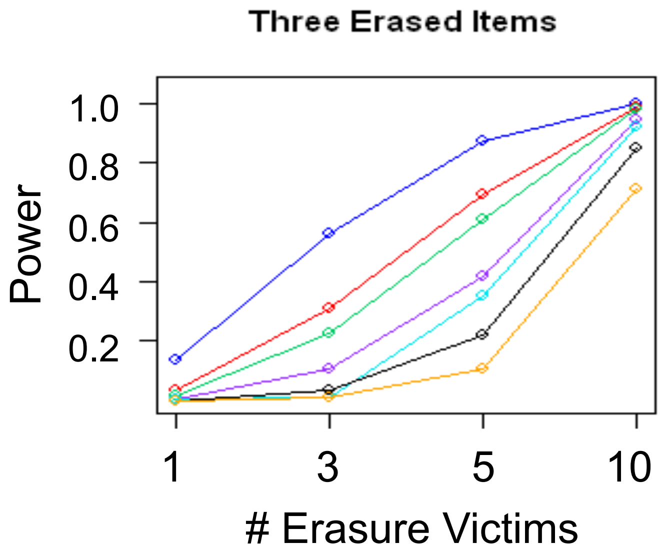
- Nominal response model used to estimate $P(x_{ij} = 1)$
  - Could have also used a dichotomous model
- Item parameters treated as known
  - No attempt was made to mirror reality with respect to amounts and magnitudes of tampering
- EDI computed
  - At Individual Student Level
  - At Class Level
  - At School Level
- Evaluative Measures
  - Type I Error rate and Power at each of the three levels
  - Only results from Random School Selection are presented
    - Class and School-Level only

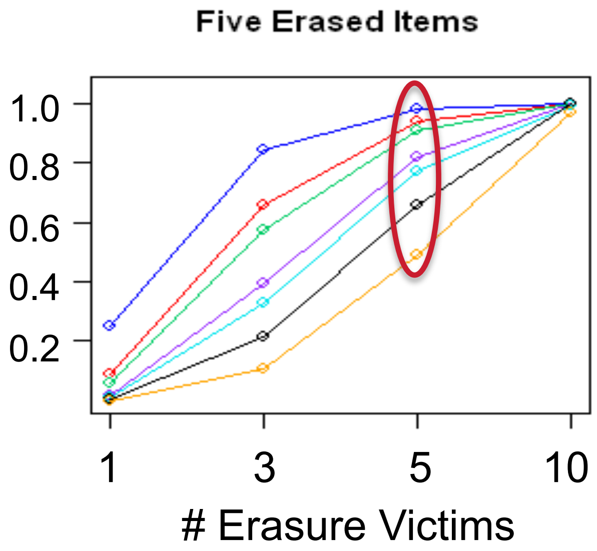
Type I error results

Over all null conditions

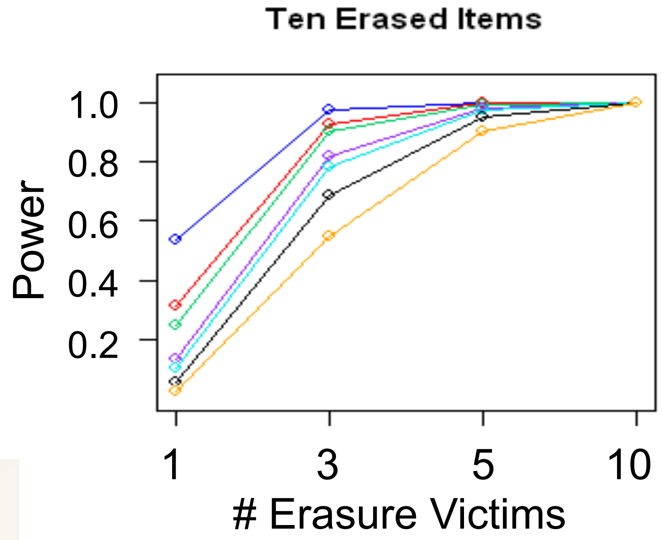
Level	.00001	.0001	.0005	.001	.005	.01	.05
Class	0.00000	0.0000	0.0002	0.0004	0.0022	0.005	0.029
School	0.00000	0.0001	0.0003	0.0006	0.0035	0.007	0.037

Class-Level Power


Three Erased Items


Five Erased Items

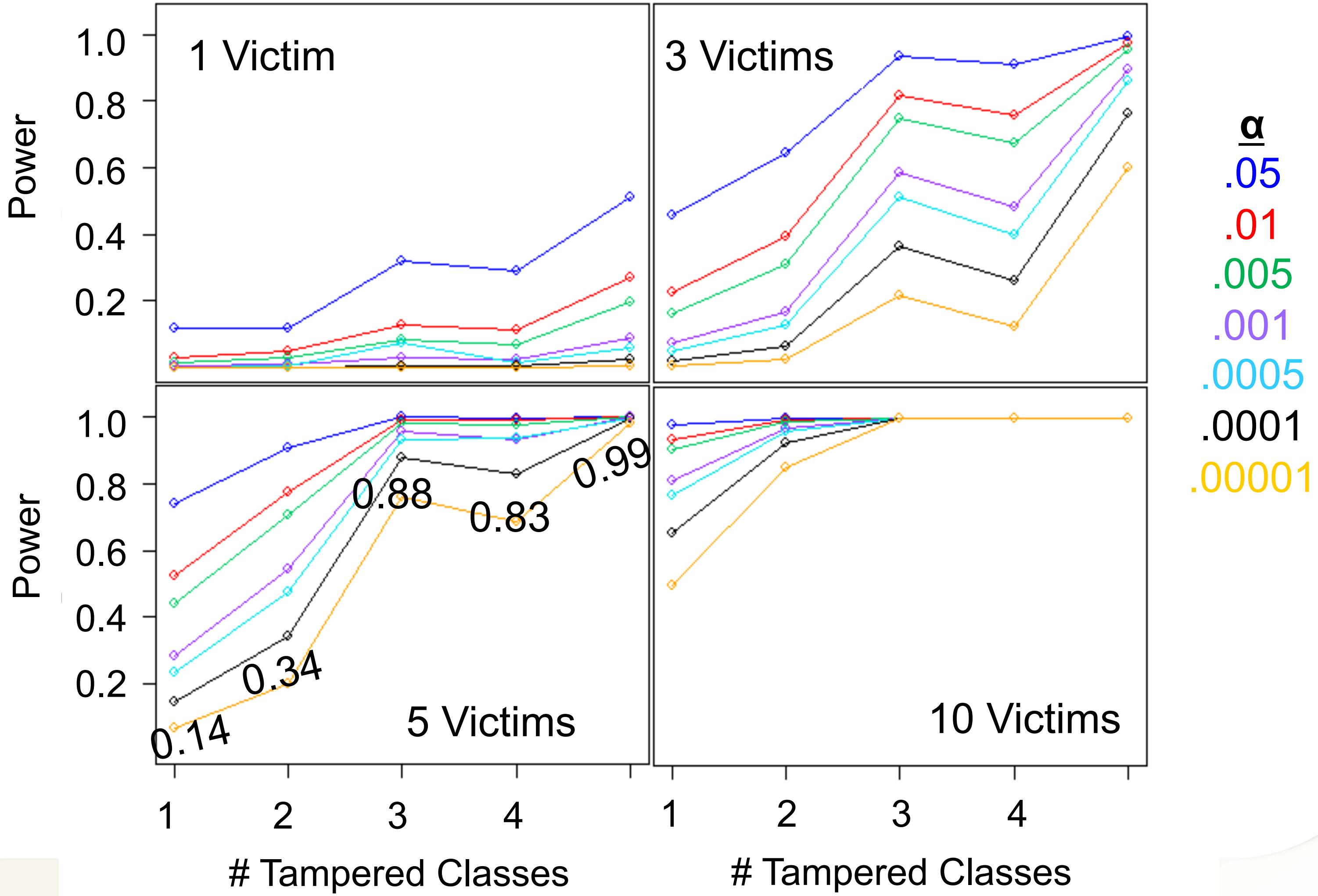
α
.05
.01
.005
.001
.0005
.0001
.00001


Ten Erased Items

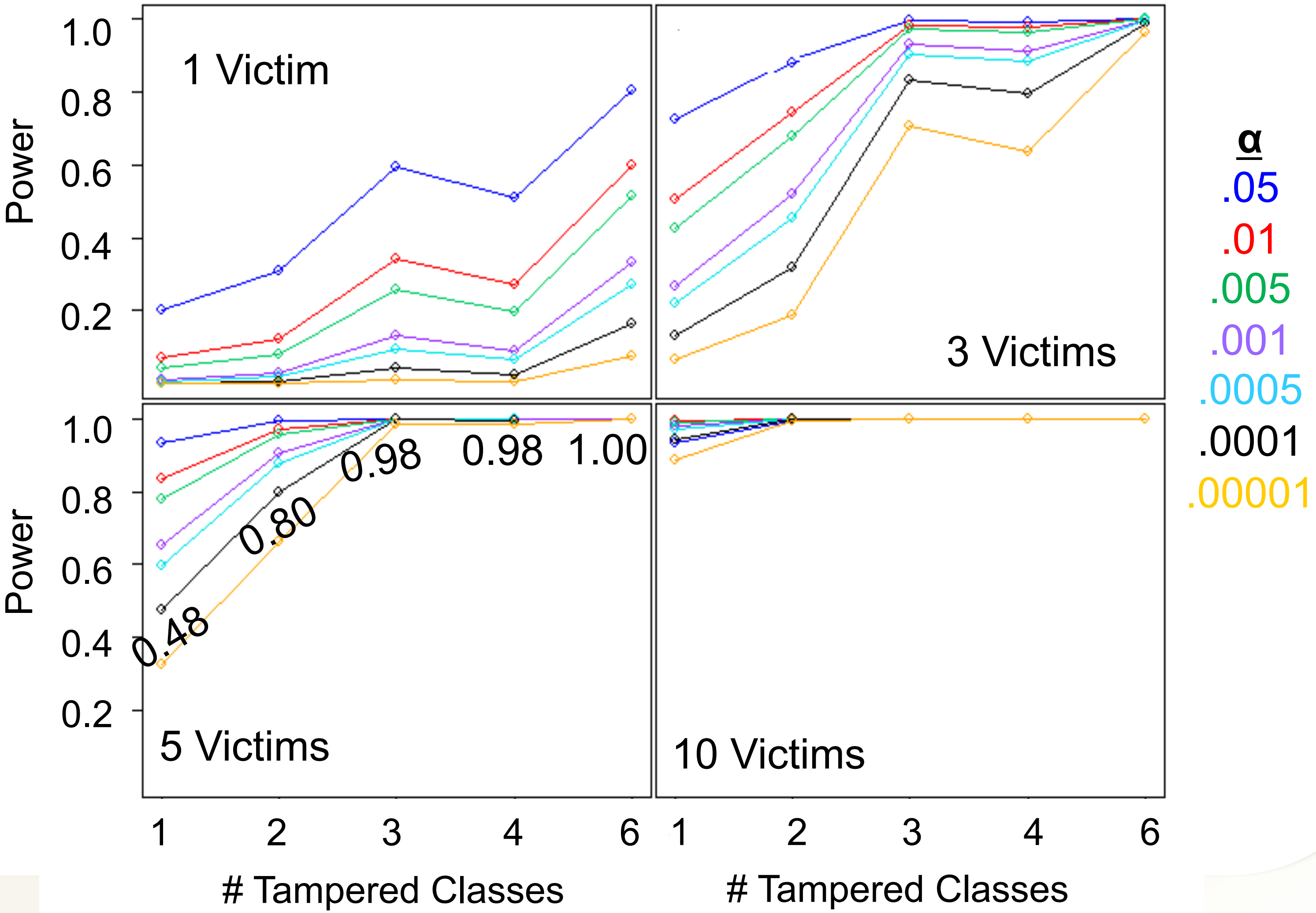
5 Erased, 5 Victims

α	Power
.05	0.99
.01	0.94
.005	0.91
.001	0.82
.0005	0.77
.0001	0.66
.00001	0.49

School-Level Power: 3 Erased Items

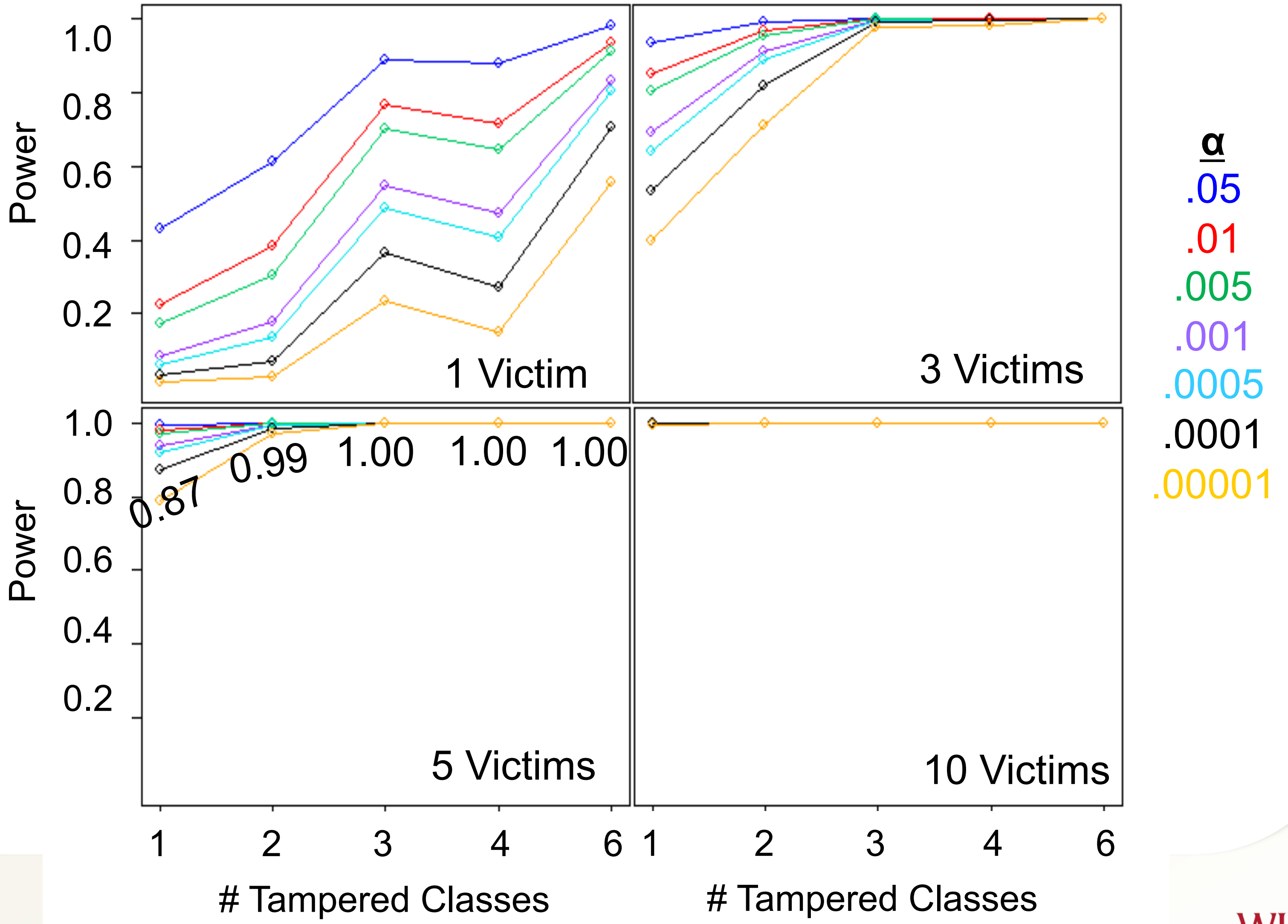


School-Level Power: 5 Erased Items



WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

School-Level Power: 10 Erased Items



Conclusion

- EDI appears to work very well for group-level tampering detection.
  - Type I error rate was well controlled at nearly all α levels
    - Small amounts of inflation evident within high-ability schools
  - Power was quite strong, even when few items were tampered for relatively small numbers of students, and at small α levels

Thank You

For more information, contact:

James Wollack

University of Wisconsin

jwollack@wisc.edu