An Initial Exploration into the Use of Erasure Analysis Results to Target Monitoring and Investigations

Steven G. Viger, Shiqi Hao and Dong Seo

Michigan Department of Education

MICHIGAN Department of **Education**

About the Michigan Merit Examination (MME)

- MME—Michigan statewide high school assessment (for grade 11 and eligible grade 12 students)

- Timed/speed tests

- Three days—Day 1 ACT Plus Writing (Reading, Writing, Mathematics and Science), Day 2 WorkKeys (WK Reading, WK Math & WK Locating Information) and Day 3 Michigan-developed (augmented) component (Mathematics, Science & Social Studies)

- Our students take full tests of ACT Plus Writing & WorkKeys, but only selected items are used for scoring and reporting

Components of the (MME)

MME Day	MME Component	Sections	Reading	Writing	Mathematics	Science	Social Studies
Day 1	ACT Plus Writing	English		S			
		Mathematics			S		
		Reading	S				
		Science				S	
		Writing		A			
Day 2	WorkKeys	Reading for Information	S				
		Applied Mathematics			S		
		Locating Information			S		S
Day 3	Michigan Component	Mathematics			A		
		Science				A	
		Social Studies					A

Note: The shaded area shows the sections in each component that contribute to a student's MME score in each subject area. An "A" means all operational items in that section contribute to the student's MME score, and an "S" means select items in that section contribute to the MME score.

Two Flagging Criteria

- Step 1, 2SD or more vs. 3SD or more above the state mean for the total erasures
- Step 2, 75% vs. 100% of total erasures that are W2R erasures
- Step 3, aggregate/summarize the results at school level and rank-order by # of flagged students in the building
- Data are from MME Spring 2013 all 3 day erasures data files

Day 1 ACT Plus Writing

Table 1. Comparison of Two Flagging Thresholds for MME Spring 2013 Day 1 ACT Plus Writing Erasure Analysis

Day 1 Subjects	State Mean	State SD	Min	Max	Flag Criteria	# of Students Flagged_Total Erasures	# of Students flagged*	# of Schools Involved	# of Schools (Short list)**	% of Schools Flagged	Top 10 Schools (overlapping)***
ACT English	0.72	1.24	0	42	3.20/4.45	4187/2045	1299/114	493/101	279/10	24.7/0.9	3
ACT Math	0.62	1.19	0	43	2.99/4.18	6874/1460	1024/40	434/33	240/3	21.2/0.3	2
ACT Reading	0.66	1.10	0	22	2.85/3.95	7628/3162	1531/217	520/168	327/35	46/3.1	8
ACT Science	0.52	1.00	0	27	2.51/3.50	5040/1997	1111/176	444/135	257/29	22.7/2.6	3

Notes: Total Ncounts=115122. *Flag criteria for the total erasures is 2 vs.3 times standard deviations or more above the state mean and the ratio of wrong-to-right erasure count and total erasure count is equal or greater than .75 vs. 1. *Students flagged for both total erasures and the ratio of W2R to the total erasures. **After removing school flagged n count =1. Total # of schools=1130. ***This refers to the number of schools flagged by the 3 SD and ratio of W2R and total erasure =1 that also appeared in Top 10 flagged schools list by the current more stringent flagging thresholds.

Day 2 WorkKeys

Table 2. Comparison of Two Flagging Thresholds for MME Spring 2013 Day 2 WorkKeys Erasure Analysis

Day 2 Subjects	State Mean	State SD	Min	Max	Flag Criteria	# of Students Flagged_Total Erasures	# of Students flagged*	# of Schools Involved	# of Schools (Short list)**	% of Schools Flagged	Top 10 Schools (overlapping)***
WK Reading	0.58	1.03	0	33	2.64/3.67	5667/2233	1393/214	532/167	314/35	27.8/3.1	7
WK Math	0.61	1.39	0	30	3.38/4.77	2314/1323	728/136	393/119	184/15	16.3/1.3	5
WK LI	0.86	1.20	0	36	3.27/4.47	3486/1428	888/89	430/78	208/8	18.4/0.7	2

Notes: Total Ncounts=114694. *Flag criteria for the total erasures is 2 vs.3 times standard deviations or more above the state mean and the ratio of wrong-to-right erasure count and total erasure count is equal or greater than .75 vs. 1. *Students flagged for both total erasures and the ratio of W2R to the total erasures. **After removing school flagged n count =1. Total # of schools=1129. ***This refers to the number of schools flagged by the 3 SD and ratio of W2R and total erasure =1 that also appeared in Top 10 flagged schools list by the recent more stringent flagging thresholds.

Day 3 Components

Table 3. Comparison of Two Flagging Thresholds for MME Spring 2013 Day 3 Michigan Components Erasure Analysis

Day 3 Subjects	State Mean	State SD	Min	Max	Flag Criteria	# of Students Flagged_Total Erasures	# of Students flagged*	# of Schools Involved	# of Schools (Short list)**	% of Schools Flagged	Top 10 Schools (overlapping)***
Math	0.92	1.40	0	21	3.72/5.13	5795/1618	1332/56	478/43	286/5	25.4/0.4	3
Science	1.44	1.93	0	45	5.30/7.23	4473/1706	567/18	324/18	133/0	11.8/0	0
Social Studies	1.22	1.72	0	40	4.64/6.38	5072/1813	847/38	409/4	209/2	18.5/0.2	1

Notes: Total Ncounts=114357. *Flag criteria for the total erasures is 2 vs. 3 times standard deviations or more above the state mean and the ratio of wrong-to-right erasure count and total erasure count is equal or greater than .75 vs. 1. *Students flagged for both total erasures and the ratio of W2R to the total erasures. **After removing school flagged n count =1. Total # of schools=1128. ***This refers to the number of schools flagged by the 3 SD and ratio of W2R and total erasure =1 that also appeared in Top 10 flagged schools list by the current more stringent flagging thresholds.

Summary and Conclusion

- In spite of some overlapping on the top 10 flagged schools list, the two thresholds may capture or focus on different schools.

- If time and resources allowed, start from more rigorous flag criterion and then narrow down.

- Erasures alone not enough—other evidence (multiple indicators), such as seating chart, answer document images can help identify possible anomalies and possible test irregularity cases.

Michigan Educational Assessment Program (MEAP)

- Erasure Analysis
  - If a students' total erasures count is greater than 2SD of the grade level total erasure mean in the state
  - and the ratio of the number of WtoR to the number of total erasures is larger than .75, the student would be flagged as aberrant.
- Person-fit analysis
  - To identify abnormal test behavior like cheating and test anxiety (Wright & Stone 1979)
  - If l_z statistics is less than -2SD of the grade level l_z mean, then the student would be flagged as a misfit student

Erasure Analysis: Math

The Summary of Erasure Analyses of Fall 2012 Mathematics Assessment

Grade	# of Erasures				WtoR				# of Flagged (ratio)
	Mean	SD	Min	Max	Mean	SD	Min	Max	
03	1.978	2.145	0	37	1.204	1.490	0	33	1,008 (.009)
04	2.858	3.128	0	46	1.739	2.146	0	34	1,192 (.011)
05	2.348	2.724	0	49	1.289	1.730	0	30	1,168 (.011)
06	2.532	2.774	0	41	1.329	1.741	0	28	720 (.006)
07	2.237	2.585	0	59	1.185	1.633	0	43	932 (.008)
08	1.978	2.369	0	47	0.976	1.430	0	31	817 (.007)

Erasure Analysis: Math

Math Grade 3
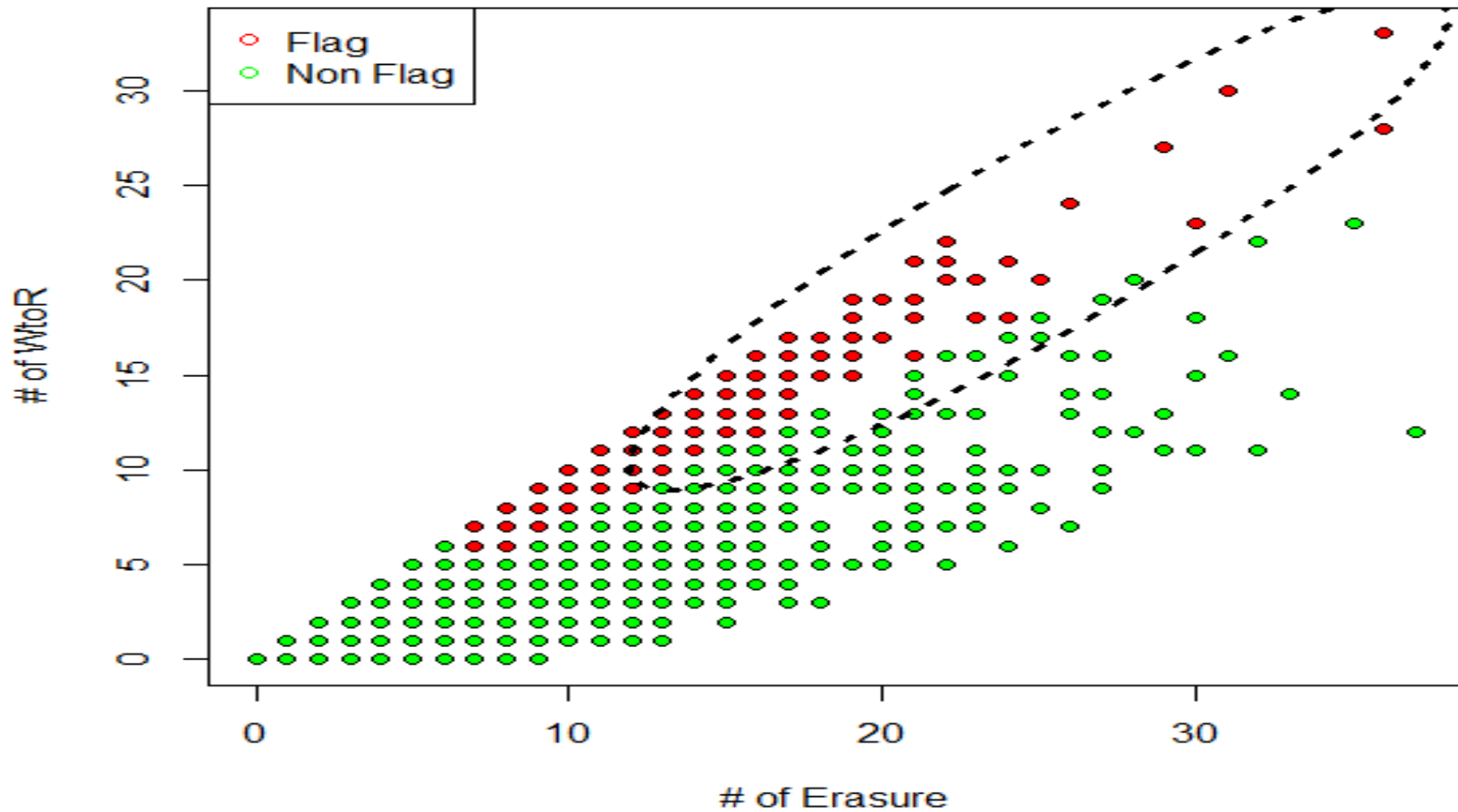
The l_z Person-fit

- $lo = \sum_{i=1}^{n} [u_{ij} \ln P_i(\hat{\theta}_j) + (1 - u_{ij}) \ln Q_i(\hat{\theta}_j)],$

  $P_i(\theta_j)$ = The probability of person j correctly answering item i

  $Q_i(\theta_j)$ = 1- $P_i(\theta_j)$, the probability of person of person j incorrectly answering the item i,

  u_{ij} = the scored (1 or 0) response to items i and person j

- $E(lo) = \sum_{i=1}^{n} \{P_i(\hat{\theta}_j) \ln P_i(\hat{\theta}_j) + [1 - P_i(\hat{\theta}_j)] \ln[1 - P_i(\hat{\theta}_j)]\},$

- VAR $(lo) = \sum_{i=1}^{n} P_i(\hat{\theta}_j)[1 - P_i(\hat{\theta}_j)] \left\{ \ln \left[\frac{P_i(\hat{\theta}_j)}{1 - P_i(\hat{\theta}_j)} \right] \right\}^2 .$

- $l_z = \dfrac{l_0 - E(l_0)}{[Var(l_0)]^{1/2}},$

Non-fitting Response Cases

- Spuriously high(SH) condition: the responses from low-ability persons are manipulated correctly in the difficult items (WtoR)

b-parameter	-3 -2 -1 0 1 2 3

responses	1 1 0 0 0 0 0

→

-3 -2 -1 0 1 2 3

1 1 0 0 1 1 1

- Spuriously low(SL) condition: the responses from high-ability persons are manipulated incorrectly in the easy items (RtoW)

b-parameter	-3 -2 -1 0 1 2 3

responses	1 1 1 1 1 1 0

→

| -3 -2 -1   0   1   2   3 |
|---|

0 0 0 1 1 1 0

13

Person-fit Analysis: Math

The Summary of Person-fit analysis of Fall 2012 Mathematics Assessment
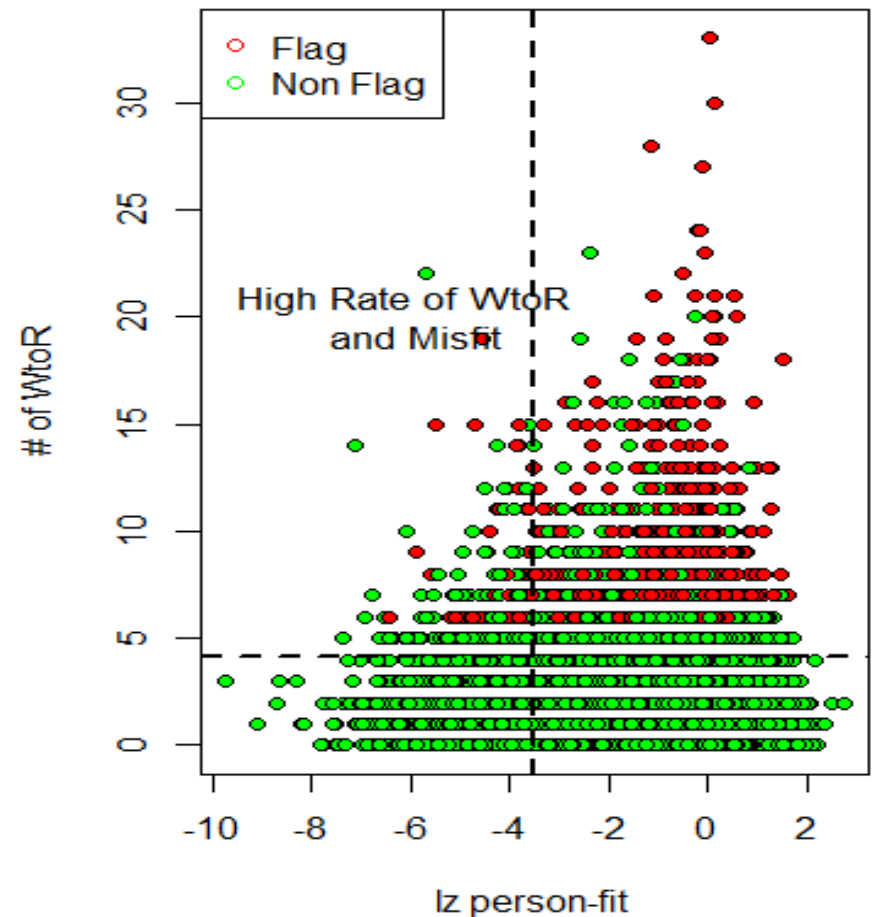
Grade	l_z statistics				Hybrid
	Mean	SD	Cutoff	# of Misfit (ratio)	# of Flagged (ratio)
03	$-.986$	1.316	-3.618	5,282 (.048)	238 (.002)
04	-1.455	1.536	-4.527	4,526 (.042)	343 (.003)
05	-1.517	1.444	-4.405	4,293 (.040)	363 (.003)
06	-1.667	1.603	-4.873	4,290 (.038)	277 (.002)
07	-1.549	1.610	-4.769	4,708 (.041)	294 (.003)
08	-1.016	1.233	-3.482	4,266 (.038)	139 (.001)

Hybrid Method (Erasure and Person-fit)
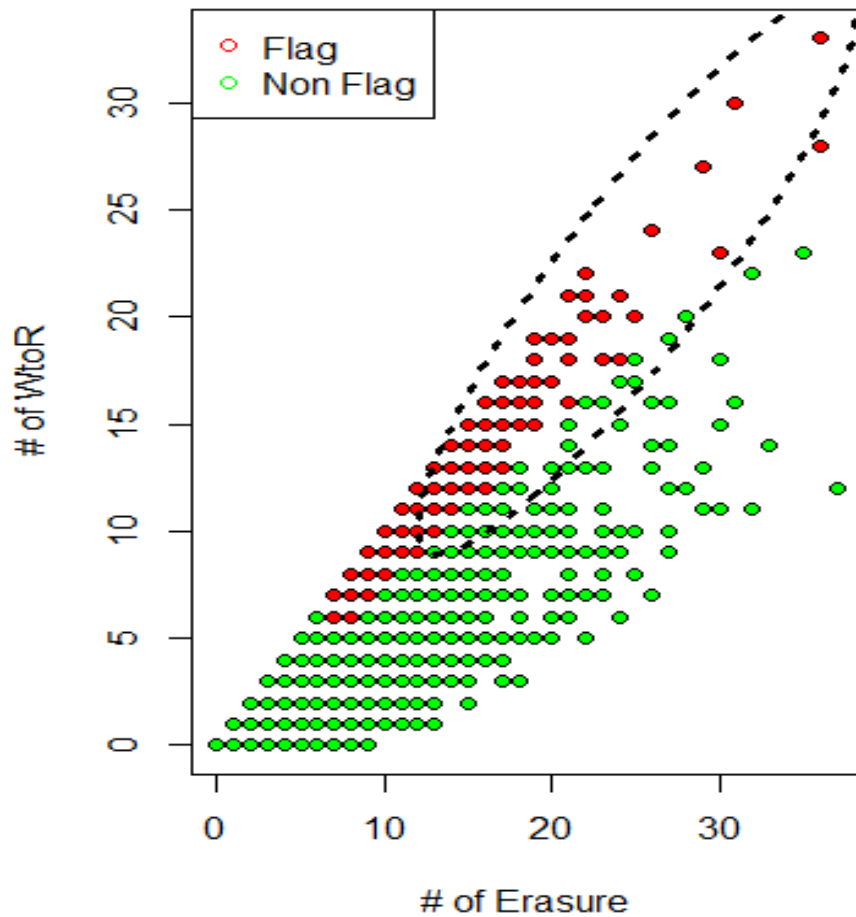
Interpretations

- High rate of WtoR and misfit:
  - Spuriously high(SH) from erasures
- High rate of WtoR and good fit:
  - Erasures were made by real ability or
  - Erasures were made by cheating
- Low rate of WtoR and misfit:
  - Spuriously low (SL) or high (SH)
- Low rate of WtoR and good fit:
  - Normal responses fitting to a model

School Analysis

Top 10 Schools Showing High WtoR Ratio and Misfit Ratio in Math All Grades

School Code	# of Students	Erasure		Person-fit	
		# of Flags	Proportion	# of Flags	Proportion
8800	69	29	42.0	29	42.0
4156	174	40	23.0	78	44.8
1600	155	35	22.6	46	29.7
9314	57	11	19.3	9	15.8
6811	338	49	14.5	158	46.7
8669	62	8	12.9	27	43.5
8572	178	17	9.6	82	46.9
1489	63	6	9.5	23	36.5
3181	33	3	9.1	3	9.1
3117	80	7	8.8	38	47.5

Limitations

- Before person-fit analyses are performed, a model should be fit to the data
  - Rasch model does not fit well to MEAP data
- High rate of WtoR and good fit case
  - Erasures are made by real ability or cheating?
- Minimize false positive cases but maximize false negative cases
- Other evidence
  - Answer booklet return rate

Contact information

Steve Viger

Chief Psychometrician/Psychometrics and Measurement Research Manager

vigers@michigan.gov

Dong Seo

MEAP Psychometrician

Seod@michigan.gov

Shiqi Hao

MME Psychometrician

Haos@michigan.gov