

CHEATING DETECTION IN LOW STAKES TESTING

Dipti Lal

Educational Initiatives (India)

dipti.lal@ei-india.com

An underlying assumption of a low-stakes test has been that mass cheating is not prevalent. Because the test gives feedback to stakeholders, such cheating could adulterate the data collected and feedback given to all stakeholders involved. Also, such instances of mass cheating could hint that the test is no longer perceived as a low-stakes test. The purpose of this study is to develop an algorithm-based system to detect mass cheating cases and investigate possible reasons, if detected, in ASSET.

This study will develop an algorithm and use it to detect possible cases of mass cheating using response data collected from ASSET. The outcome of this study, within our organization (Educational Initiatives Pvt. Ltd.), will be to analyse such cases, using a robust algorithm and supporting them with qualitative insights.

INTRODUCTION

What is cheating?

Cheating is any deceitful or fraudulent attempt to evade rules, standards, practices, customs, mores, and norms to gain an unfair advantage or to protect someone who has done so. Cheating includes, but is not limited to (Jones, 2001/ 2011):

- Giving or receiving information during an exam (“exam” includes tests, assessments, and quizzes, whether delivered in a classroom setting or on line.)
- Using unauthorized material (like notes) during any exam; unauthorized dissemination or receipt of exams, exam materials, contents, or answer keys in written or digital form.
- Taking an exam or writing a paper for another student—or asking someone to take an exam or write a paper for you (this includes sharing work and/or writing group-produced answers on take-home and on-line exams unless explicitly permitted by the instructor). This is also called “impersonation.”
- Submitting the same paper—or different versions of what is substantially the same paper—in other courses or in subsequent attempts to pass a course.
- Sabotaging, misrepresenting or fabricating written work, sources, research, or results as well as helping another student commit an act of academic dishonesty or lying to protect a student who has committed one.

For this research our findings mainly focus on cheating happening through a teacher where he/she reads out the answers in class or teaches through the test or asks a ‘bright’ student to solve the paper and read out the answers in class during the test.

WHAT IS ASSET?

ASSET (Assessment of Scholastic Skills through Educational Testing), developed by Educational Initiatives (EI), is a low-stakes, multiple choice-single response type and benchmarking test. It gives critical feedback to students, teachers and school management about the learning levels of students and also entire classes.

Rather than testing rote learning, through multiple-choice questioning, it focuses on measuring how well skills and concepts underlying the school syllabus have been learnt by the student (ASSET Website). Students of classes 3-10 are tested in English, Maths and Science, Social Studies and Hindi.

Basic Assumptions and Caveats

1. Cheating happens to secure a gain in performance. W. C. Fields, acting in the movie *You Can't Cheat an Honest Man*, opines, "If a thing is worth winning, it's worth cheating for." (Cook, in Cook & Sacerdote, 2003)
2. Only mass cheating is under the purview of this paper. The main objective of this paper is to identify cases of mass cheating where students receive external help from a teacher or an evaluator or students mark answers that the 'best' student reads out, etc.
3. Sections with at least 15 students are part of this study. Although, theoretically, cheating among a smaller number of students can be caught, we have restricted ourselves to sections with at least 15 students which helps strikes a balance between large sections and smaller sections.

Based on the performance and patterns in responses received via ASSET, the algorithm developed will help us identify sections of classes where cheating is suspected. The algorithm uses three criteria. To identify cheating cases one or more criteria should be met.

This algorithm was tested on four rounds of ASSET performance data. In each round, approximately 300 schools' performance data was looked at to ascertain if cheating had occurred.

The Three Criteria

Criteria 1 - Number of questions in each class where the performance of the class is greater than the national average by 30%. In other words, more the number of questions with 'high' performance, more likely cheating.

Criteria 2 - Number of questions in each class where the performance of the class in low performing questions (questions with performance less than 50%) shows a 90% match in responses chosen. Meaning, the more similar the responses in low performing questions, more likely cheating.

Criteria 3 - Average score compared to previous round shows a significant jump of 20 percentile points. Drastic improvements compared to previous performance is considered to be an indicator of cheating in the current round.

3. Unusually high scores in the round that cheating happened.

Unlike gains associated with true learning, however, one expects no persistence in the artificial test score gains due to cheating. Thus if the children in cheating classrooms this year are not in cheating classes next year, one expects the full magnitude of the cheating-related gain to evaporate (Jacob & Levitt, 2002)

The algorithm: The responses from students across the country in the ASSET tests administered are collected and scored. Post this, a log file can be created on demand in an internal system developed by EI's IT personnel. Thereafter, checks are made to see if above said criteria are met. Such cases are those where cheating is suspected and are filtered.

Checks:

Criteria 1 (where the performance is significantly greater than national performance): If the number of questions with high performance exceeds 30% of the total questions in the paper, it is flagged.

Criteria 2 (similar answering patterns in low performing questions): Number of questions with a 90% match should be at least 40% of all the low performing questions in that section¹.

Criteria 3 (drastic and sudden improvement in performance): The jump in the average percentile points in the current round compared to the previous round if greater than or equal to 20 points is flagged as suspicious.

The final output from this exercise is a list of schools and specific classes and sections within that school that are flagged as suspicious. There are two levels of suspicious cases – level 1 (greater levels of suspicion) and level 2 (moderate levels of suspicion).

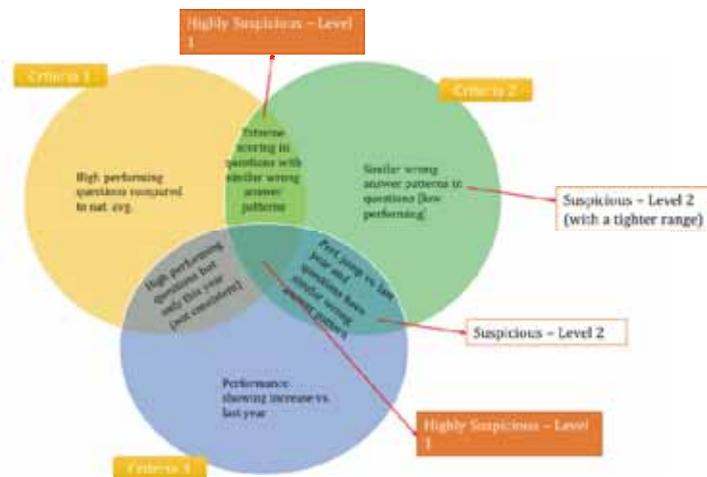


Figure 2: The figure above shows which cases fall under level 1 and 2 respectively.

¹ The number of questions that satisfy these criteria should make up at least 10% of the questions in the paper.

RESULTS

Performance data from the December, 2014 round of ASSET was collected. A total of 9890 sections data was fetched. These sections belonged to classes 3 to 10 from 260 schools.

Out of these 9890 sections, 20 sections have been deemed to cheating cases – 14 in level 1 and 6 in level 2.

These 20 sections belong to a total of 11 unique schools.

Section	Student Count	Question Count	Benchmark	Criteria 1	Criteria 2 checked	Criteria 2 matched	Criteria 3 student check	Criteria 3 percentile difference
Section 1	33	30	147	15	4	3	Null	Null
Section 2	32	40	138	18	9	4	30	48
Section 3	34	40	152	17	7	4	33	-11
Section 4	34	45	118	14	19	9	31	-5
Section 5	27	40	127	18	11	10	22	34
Section 6	27	40	111	12	18	12	24	19
Section 7	32	60	131	30	14	10	30	33
Section 8	30	60	118	25	21	13	25	20
Section 9	30	45	115	14	19	8	25	34
Section 10	31	45	117	15	17	7	29	33
Section 11	43	50	132	24	6	5	43	42
Section 12	30	45	127	19	16	10	29	-8
Section 13	24	35	140	18	10	10	17	62
Section 14	31	40	123	14	15	6	30	33
Section 15	17	30	115	0	5	3	Null	Null
Section 16	40	40	103	2	7	4	37	-30
Section 17	43	40	129	11	7	4	43	-2
Section 18	26	40	73	5	25	22	Null	Null
Section 19	26	45	97	10	26	18	Null	Null
Section 20	27	40	123	9	8	5	20	25

Table 1: Showing the cases of cheating filtered and how they fare of each criterion

DISCUSSION

Individually, the criteria used do not fully substantiate cheating. But when corroborated, they substantiate cheating. Criteria 2 (similar wrong answer patterns) is believed to be the most compelling of the three. Criteria 1 and 3 help in supplementing and affirming the suspicion. Identifying similar answering patterns using strings of responses suffers from a drawback – the process is complicated. Simply identifying number of questions where such similarities are seen are not limited to blocks of questions showing similar responses patterns. Even if the cheating pattern in the test is sporadic, it can be identified here. This research would be incomplete without conducting a retest on a sample set of students suspected to have cheated. For the retest, the same paper will be administered again under supervised conditions (based on a forthcoming publication).

The overall perception of low stakes testing is challenged in this research.

Another kind of standardized testing, in which the results only matter somewhat for teachers and students, is commonly referred to as low-stakes. On the scale of testing, this is

better. It holds teachers and students collaboratively rather than competitively accountable for student outcomes, and uses test scores to guide things like professional development and budget allocation (Save our Schools blog page, 2012).

A low-stakes, benchmarking test like ASSET focusses on testing concepts and skills and providing feedback on the learning gaps identified through this test. Students need not prepare for the test as it assess them on what they have learnt and understood so far, rather than recall facts. In such a test or any other low-stakes test, if cheating is brought to light, it is interesting to find out if the stakeholders of this test know the true meaning and purpose of a low-stakes test.

High-stakes testing is also not free from the obstructive issues of cheating.

As school systems across the country have raised the stakes associated with standardized testing, cheating on these tests has become a tempting option for some teachers and administrators. The investigation for the Chicago Public Schools by Brian Jacob and Steven Levitt (2003) has documented cheating by 5 percent or more of the teachers.

CONCLUSION

The basic results of the analysis of the data points to mass cheating being prevalent in a low-stakes test.

Our first goal in the coming months is to administer retest and compare data with the first round of tests. The process will then be finalized and ongoing after every round of ASSET (which is currently conducted twice a year; in August and December)

As a long term goal, this algorithm and understanding can be extended to other similar tests.

Acknowledgements

I would like to thank Sridhar Rajagopalan, Rahul Venuraj, Meghna Kumar and other colleagues at Educational Initiatives Private Limited.

References

- Cook, P. J., & Sacerdote, B. (2003). Comments. In *Brookings-Wharton Papers on Urban Affairs* 2003(1), 210-218. Brookings Institution Press. Retrieved November 10, 2015, from Project MUSE database.
- Jones, L. R. (2001 & 2011). *Academic integrity & academic dishonesty: A handbook about cheating & plagiarism*.
- Jacob, B. A., & Levitt, S. D. (2003). Catching cheating teachers: The results of an unusual experiment in implementing theory. In *Brookings-Wharton Papers on Urban Affairs: 2003* (pp. 185-220) US: Brookings Institution Press.
- Save our Schools Blog (2012). Retrieved from www.saveourschoolswisconsin.wordpress.com/2012/06/12/high-stakes-low-stakes-no-stakes-the-testing-dilemma/
- ASSET Website (n.d). Retrieved from www.ei-india.com/asset