

Evidence of the Proliferation of Braindump Content in Real Data

Marcus Scott, Caveon Test Security

Dennis Maynes, Caveon Test Security

Measuring the Effects of Braindumps

- Three data sets with the following characteristics:
 - A braindump was found by the testing program
 - The braindump answer key contained errors
- Because the braindump contained errors, it was possible to detect likely users because they had similar incorrect responses
- Braindump usage was tracked, along with its effects on the pass rate (data sets 2 and 3), the mean score, and item p-values (data set 3 only)

Detection Method

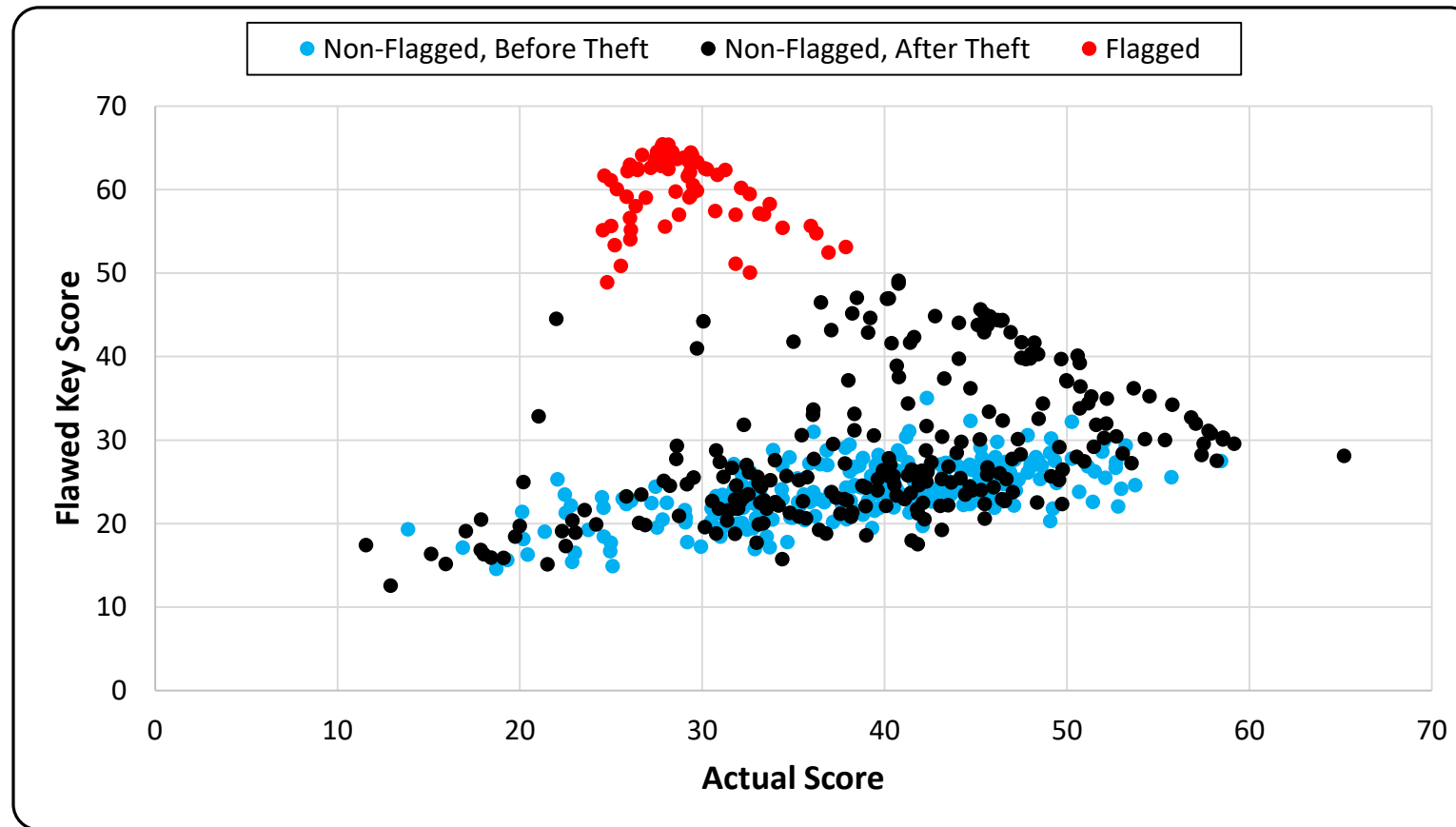
1. Score tests using the actual key and the flawed key
2. Transform the (actual score, flawed score) ordered pairs with the Gram-Schmidt Process. Use the most extreme ordered pair to determine the rotation angle
3. Build histogram of horizontal components of rotated data
4. Search histogram for optimal point of separation between likely users (right side) and likely non-users (left side)
5. Create a distribution for each group and apply to a Bayesian classifier
6. Flag at 1,000 to 1 odds in favor of braindump use

Data Set 1

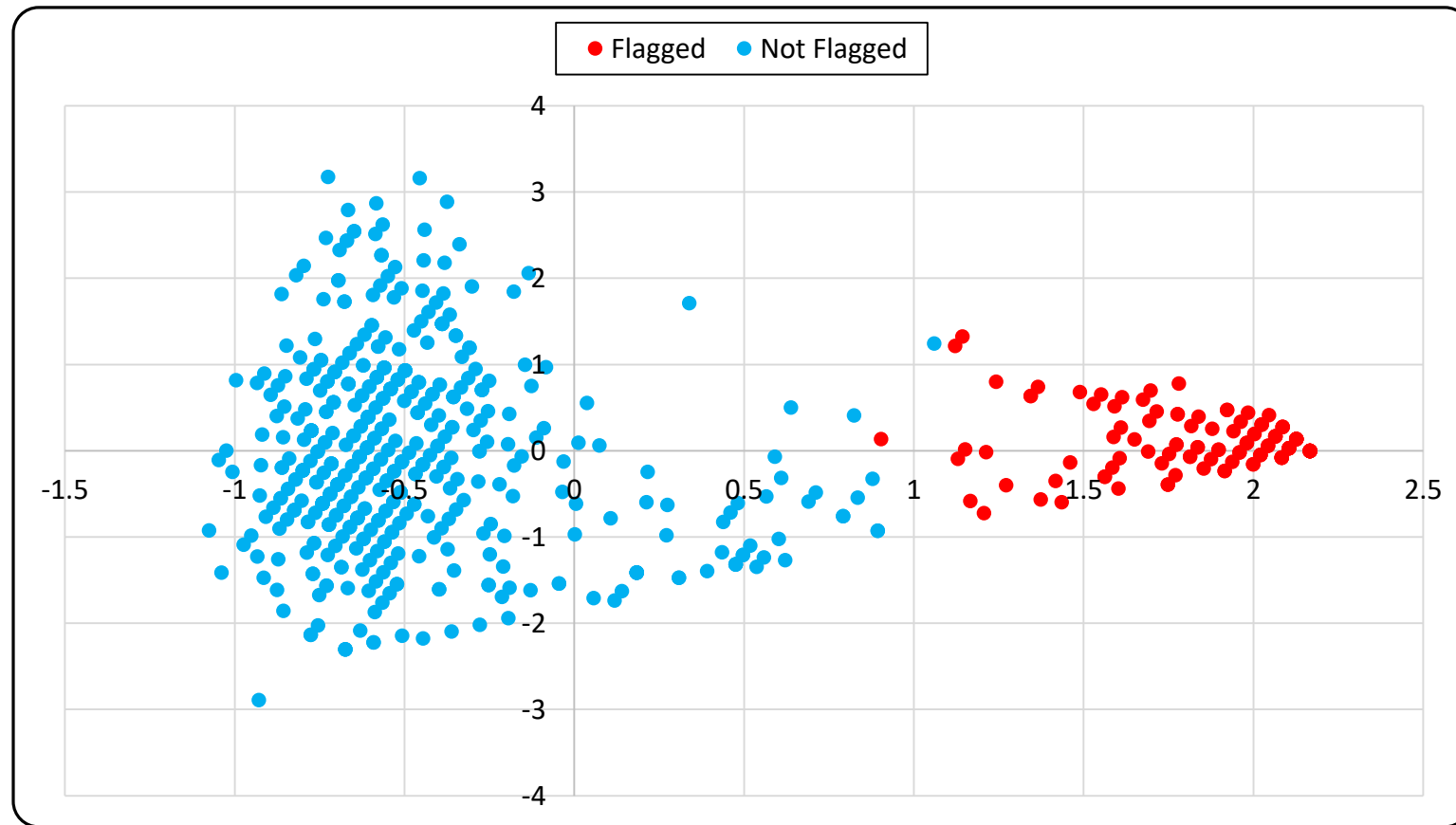
Test and Braindump Characteristics

- 65-item test administered to 599 examinees (January 5, 2014 to December 15, 2014)
- Items were stolen on July 7 (item ordering in braindump matched ordering in one test on this date)
- 236 tests taken before July 7; 363 taken on or after that date
- Braindump had correct answers to 28 of the 65 items (43% accurate)
- Only test date, actual score, flawed score, and number of incorrect matches with braindump were provided

Flawed Score Against Actual Score



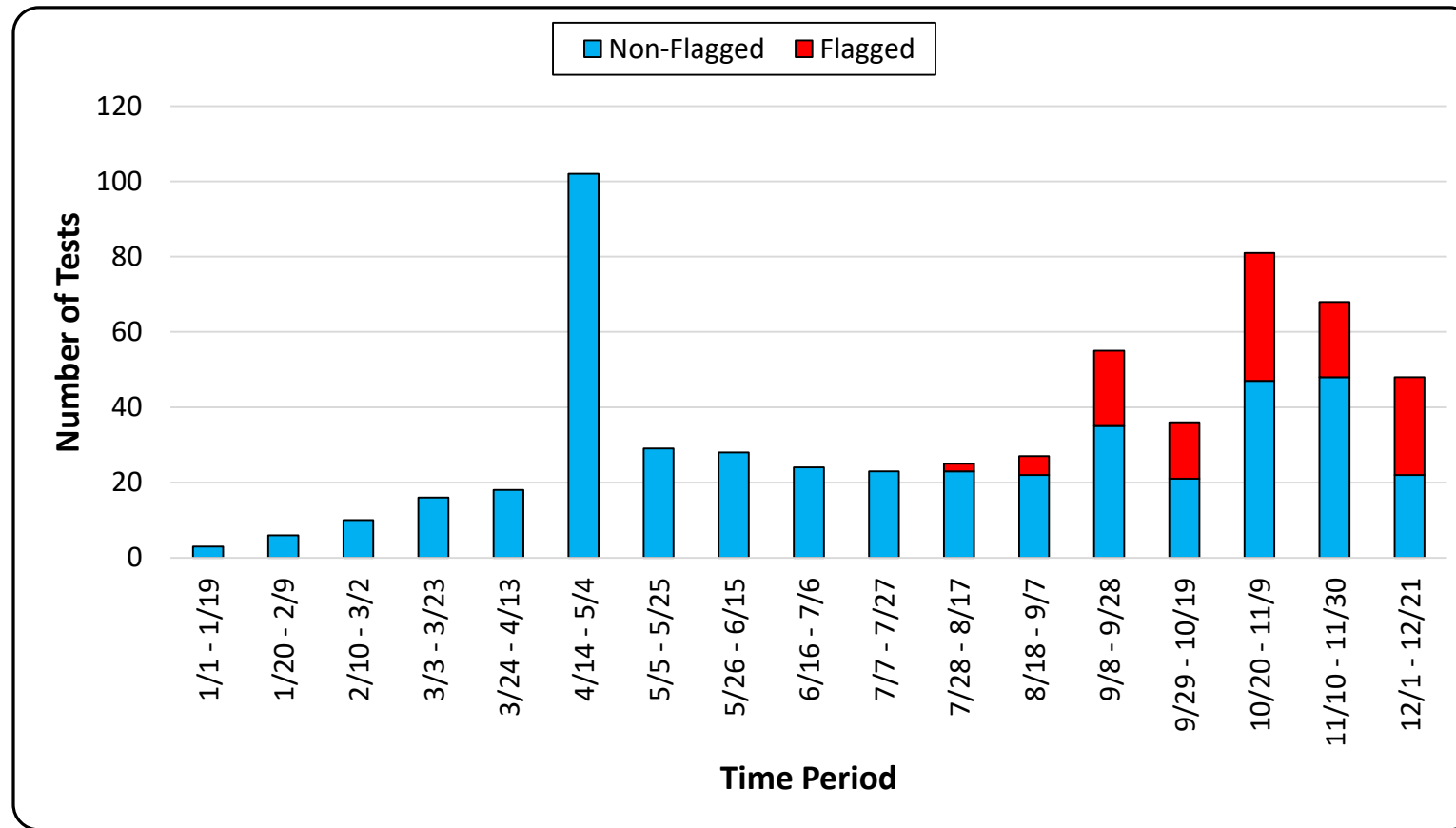
Rotated Data



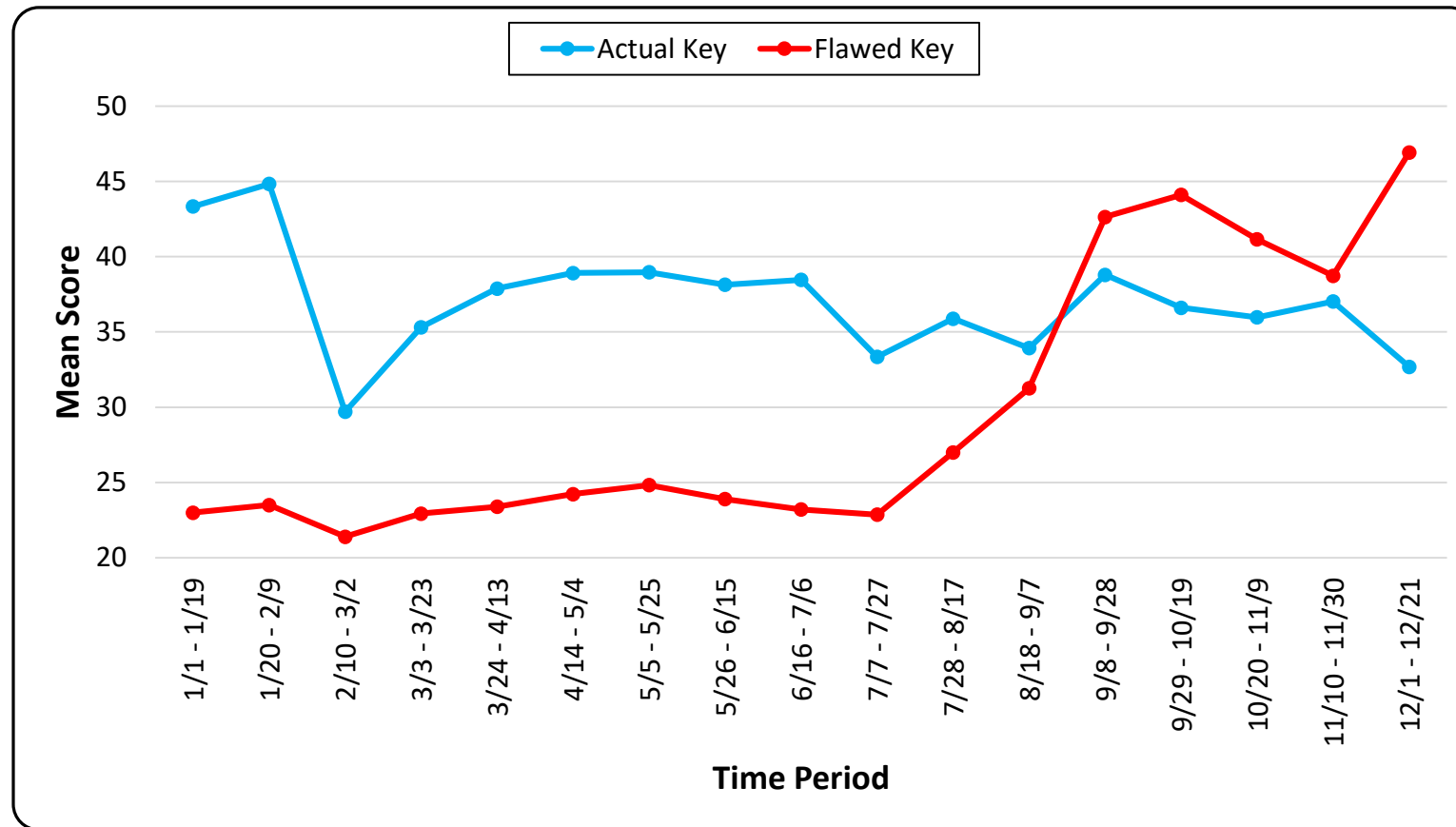
Flawed Key Analysis Results

- 122 examinees flagged as likely braindump users
- These 122 examinees had score increases between 14 and 37 points when the flawed key was used
- All flagged tests were taken after July 7
- Earliest flagged test was August 15
 - 39 days after theft
 - Response vector exactly matched the flawed key
- Test thief was not flagged (actual score – 16, flawed score – 15)

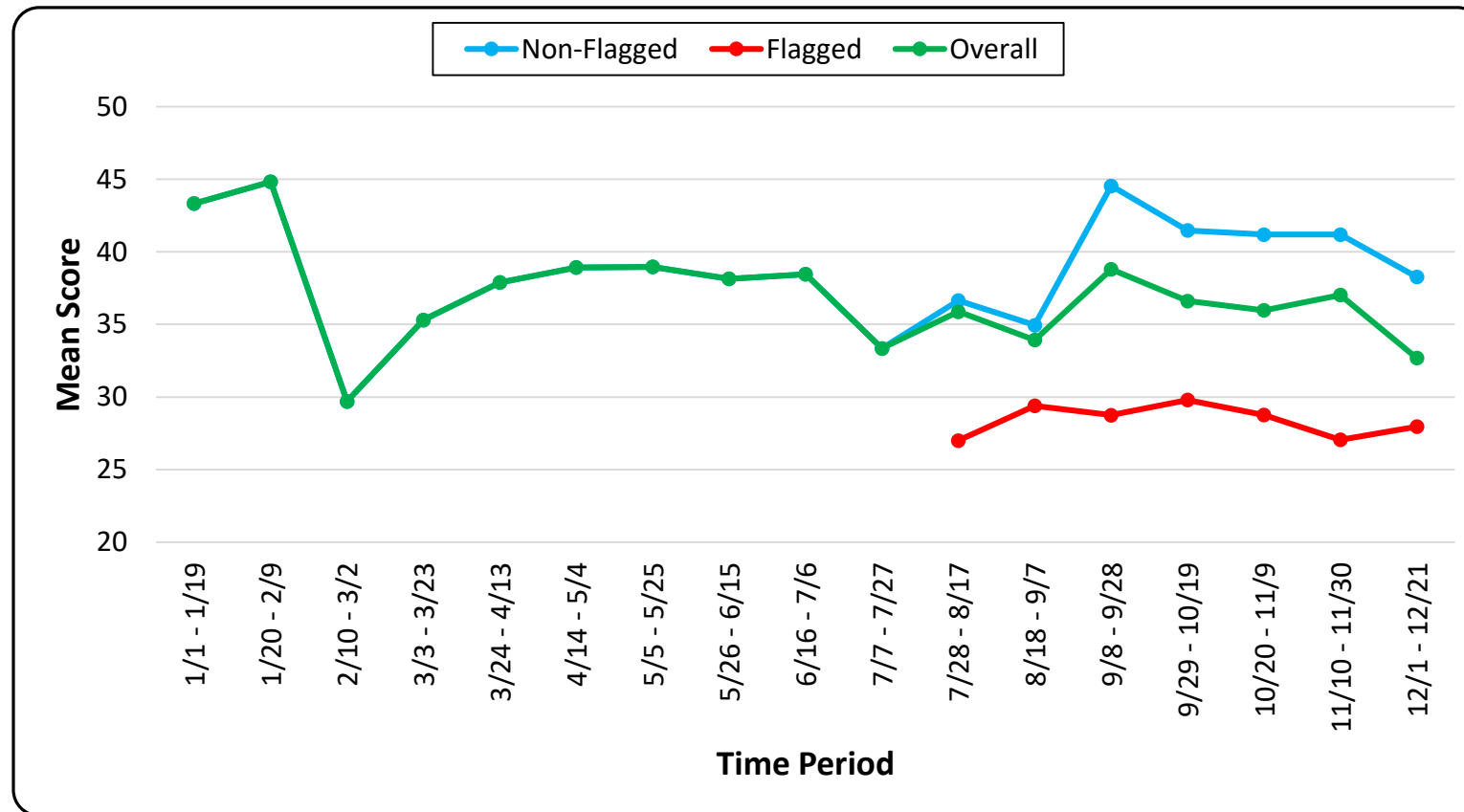
Braindump Use Over Time



Effect on Scores



Effect on Scores

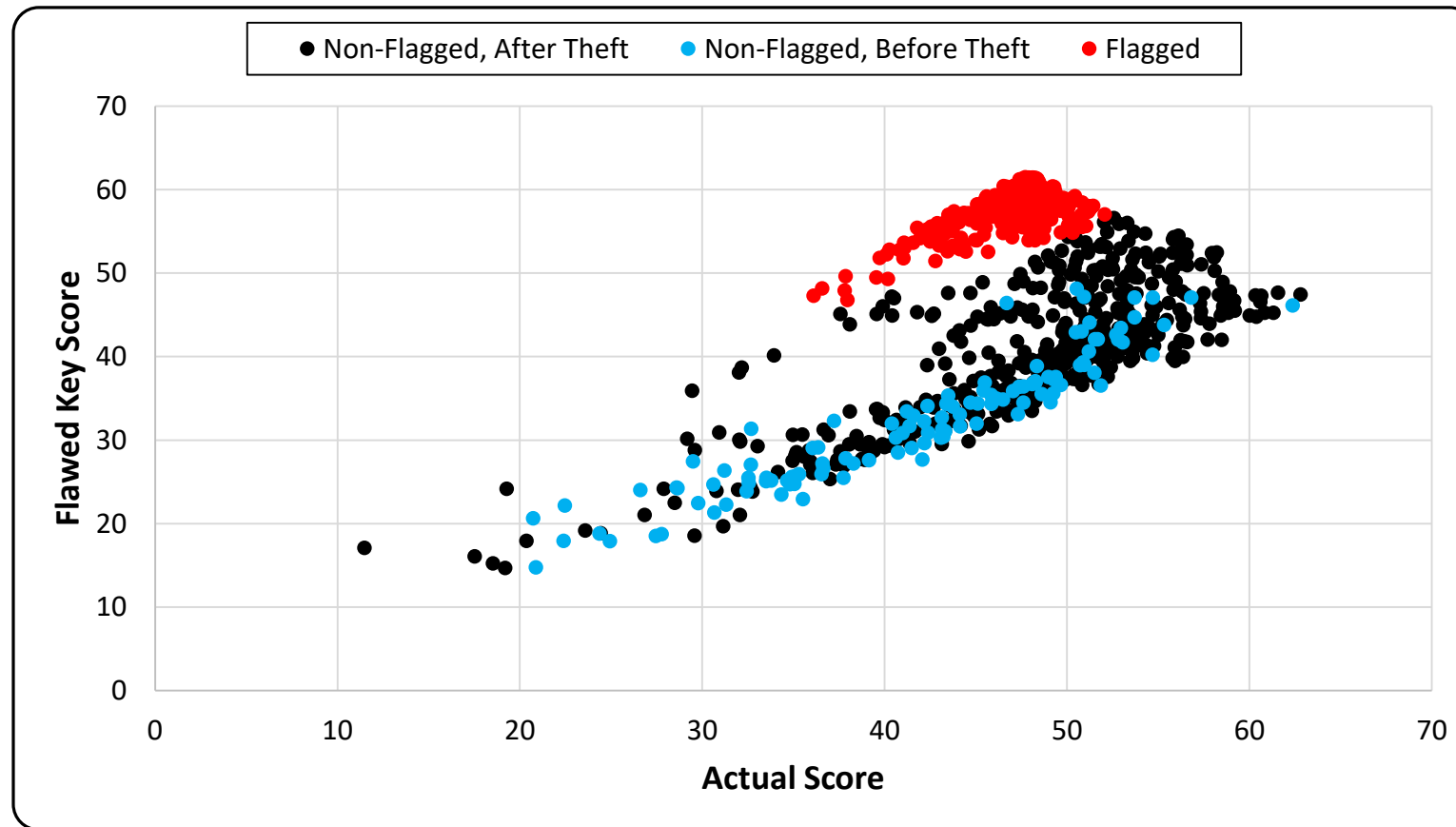


Data Set 2

Test and Braindump Characteristics

- 63-item test administered to 884 examinees (October 6, 2014 to June 8, 2015)
- Items were stolen on January 16, 2015
- 117 tests taken before January 16; 767 taken on or after that date
- Braindump had correct answers to 48 of the 63 items (76% accurate)
- Exam cut score was 38
- Only test date, actual score, flawed score, pass/fail status, and number of incorrect matches with braindump were provided

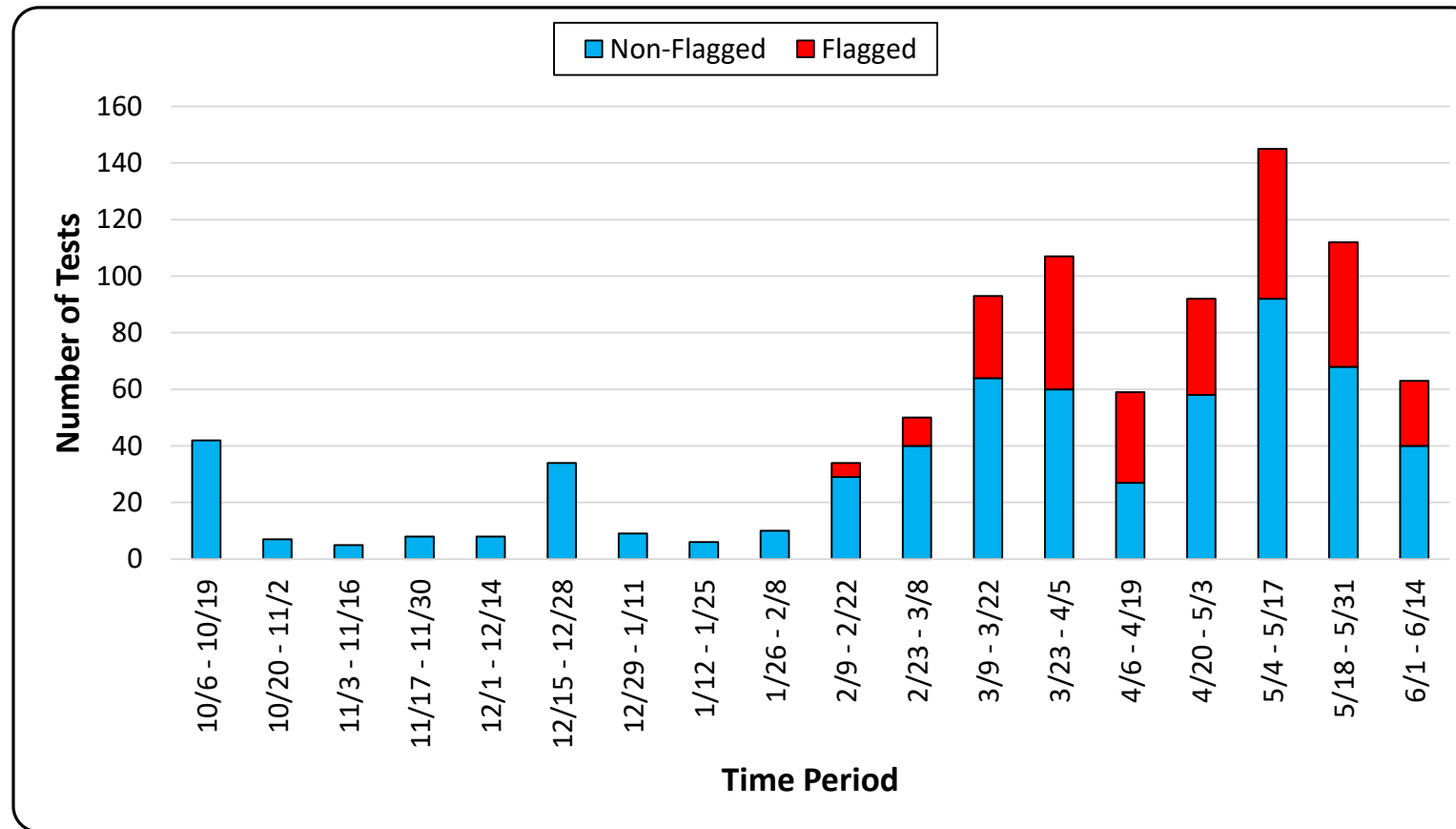
Flawed Score Against Actual Score



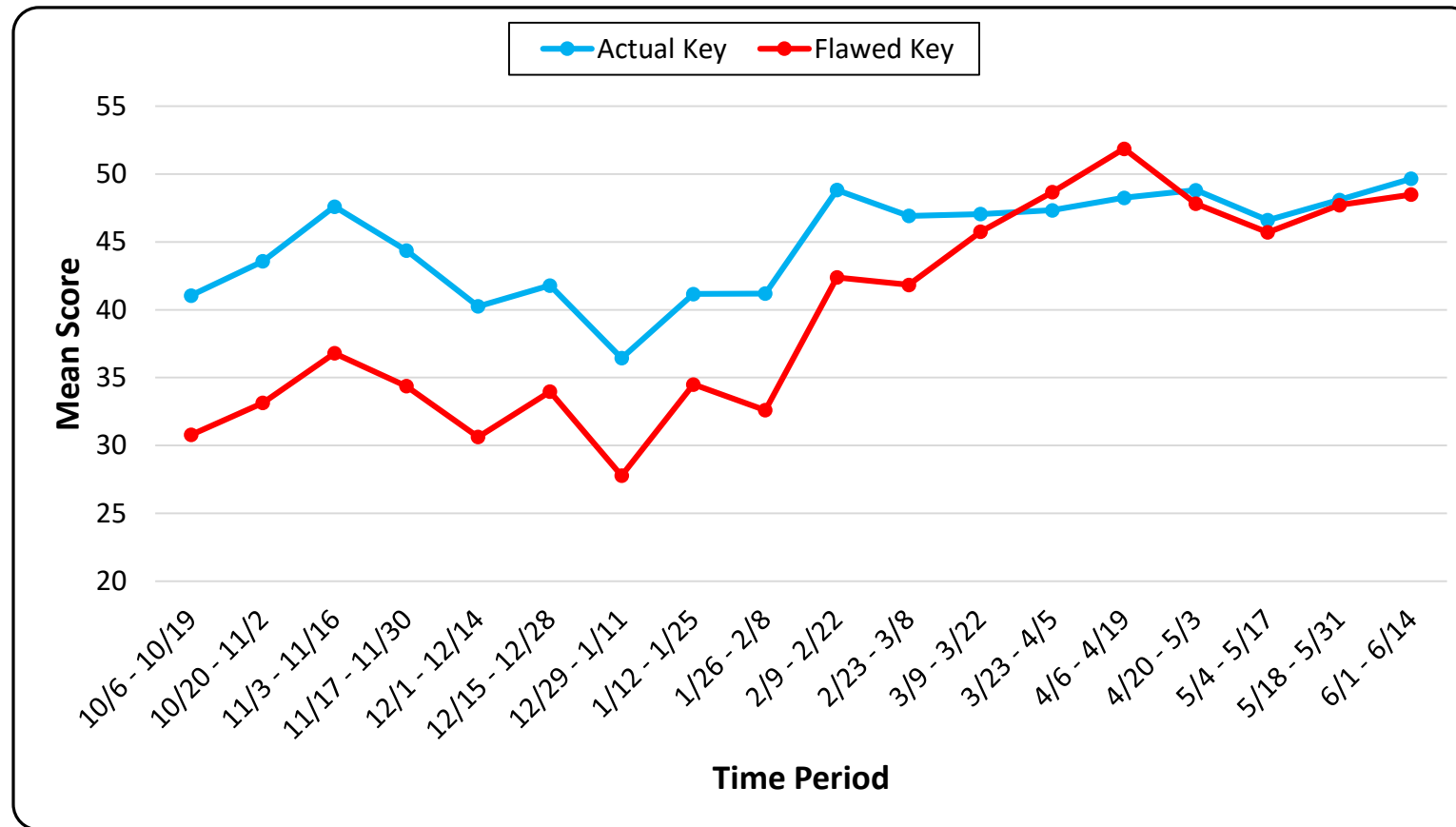
Flawed Key Analysis Results

- 277 examinees flagged as likely braindump users
- These 277 examinees had score increases between 5 and 14 points when the flawed key was used
- All flagged tests were taken after January 16
- Earliest flagged test was February 13
 - 28 days after theft
 - Response vector matched the flawed key for 56 of the 63 items
- Test thief was not flagged (actual score – 11, flawed score – 17)

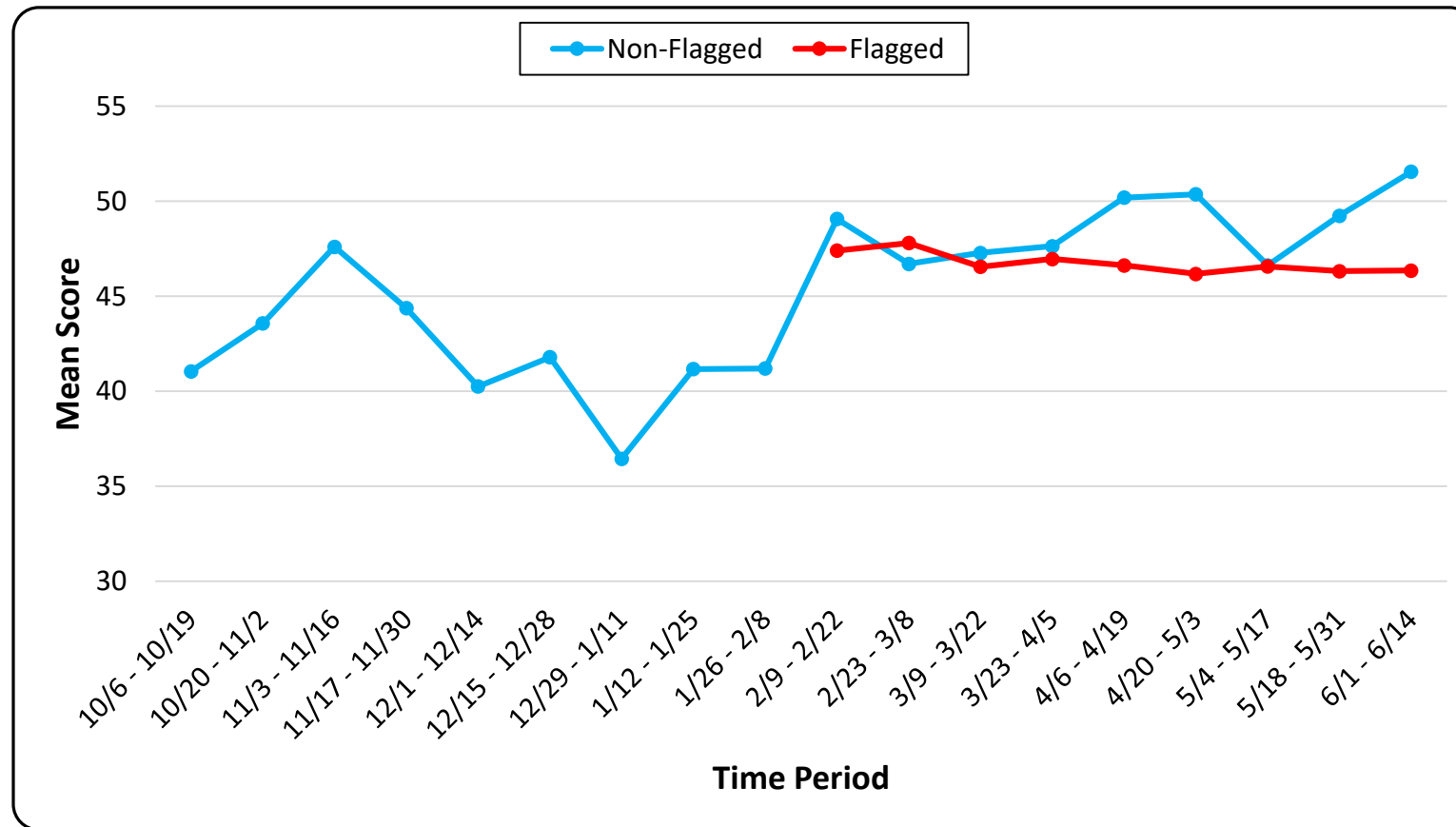
Brindump Use Over Time



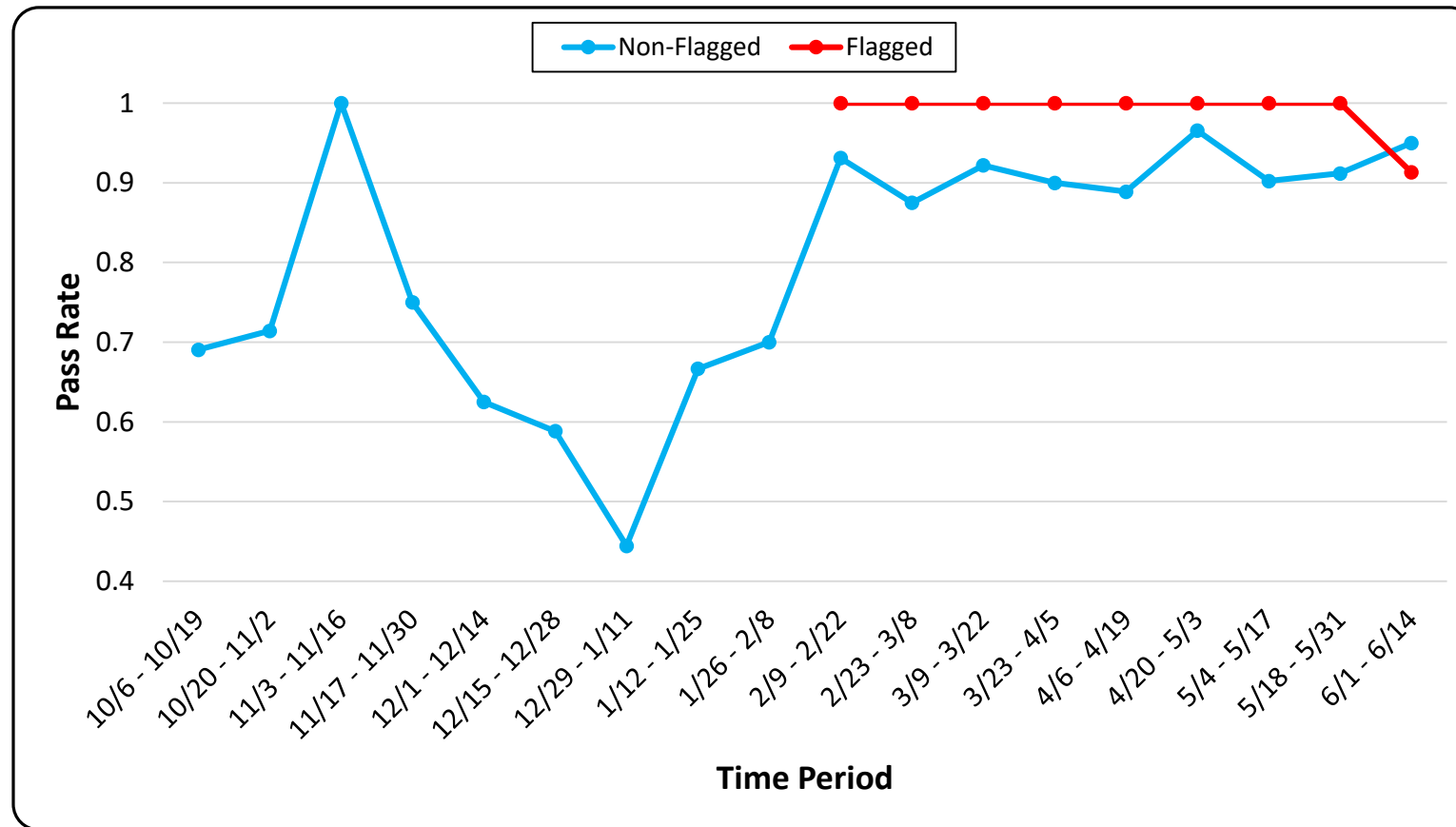
Effect on Scores



Effect on Scores



Effect on Pass Rate

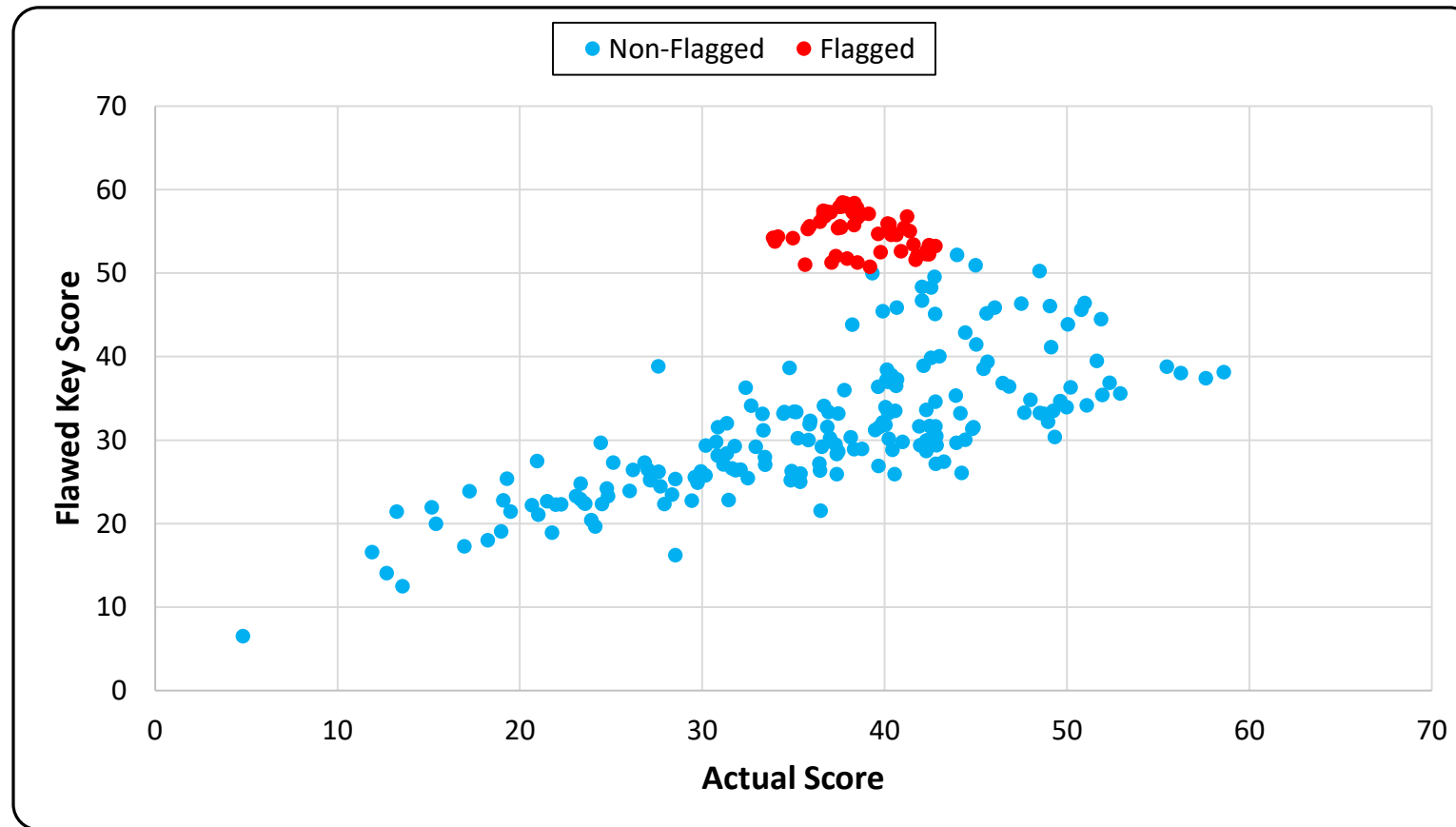


Data Set 3

Test and Braindump Characteristics

- 60-item test administered to 248 examinees (November 6, 2015 to June 1, 2016)
- Unknown theft date; braindump purchased on May 3, 2016
- Braindump had correct answers to 38 of the 60 items (63% accurate)
- Exam cut score was 39
- Item responses were provided, along with the actual key and the flawed key

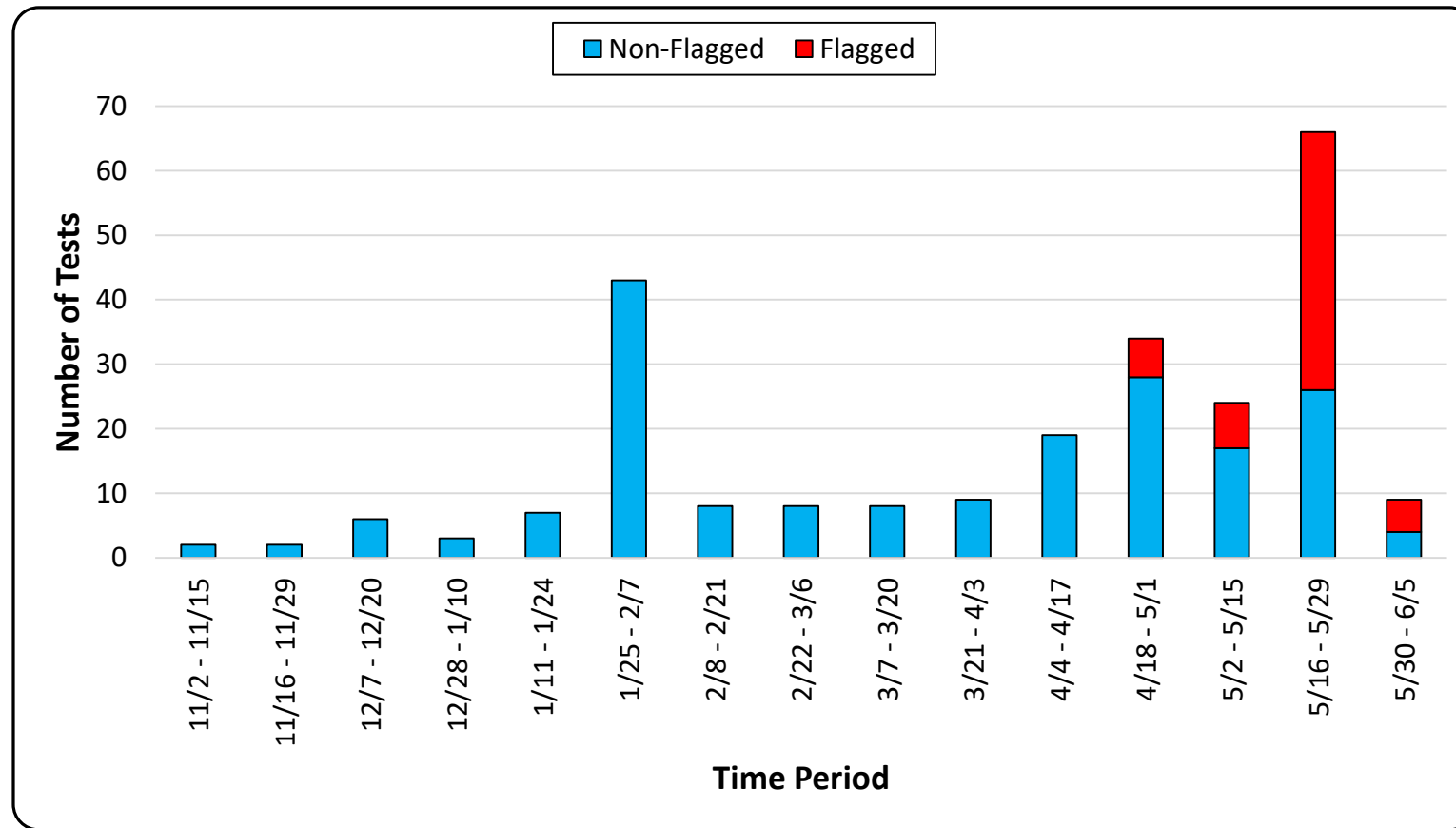
Flawed Score Against Actual Score



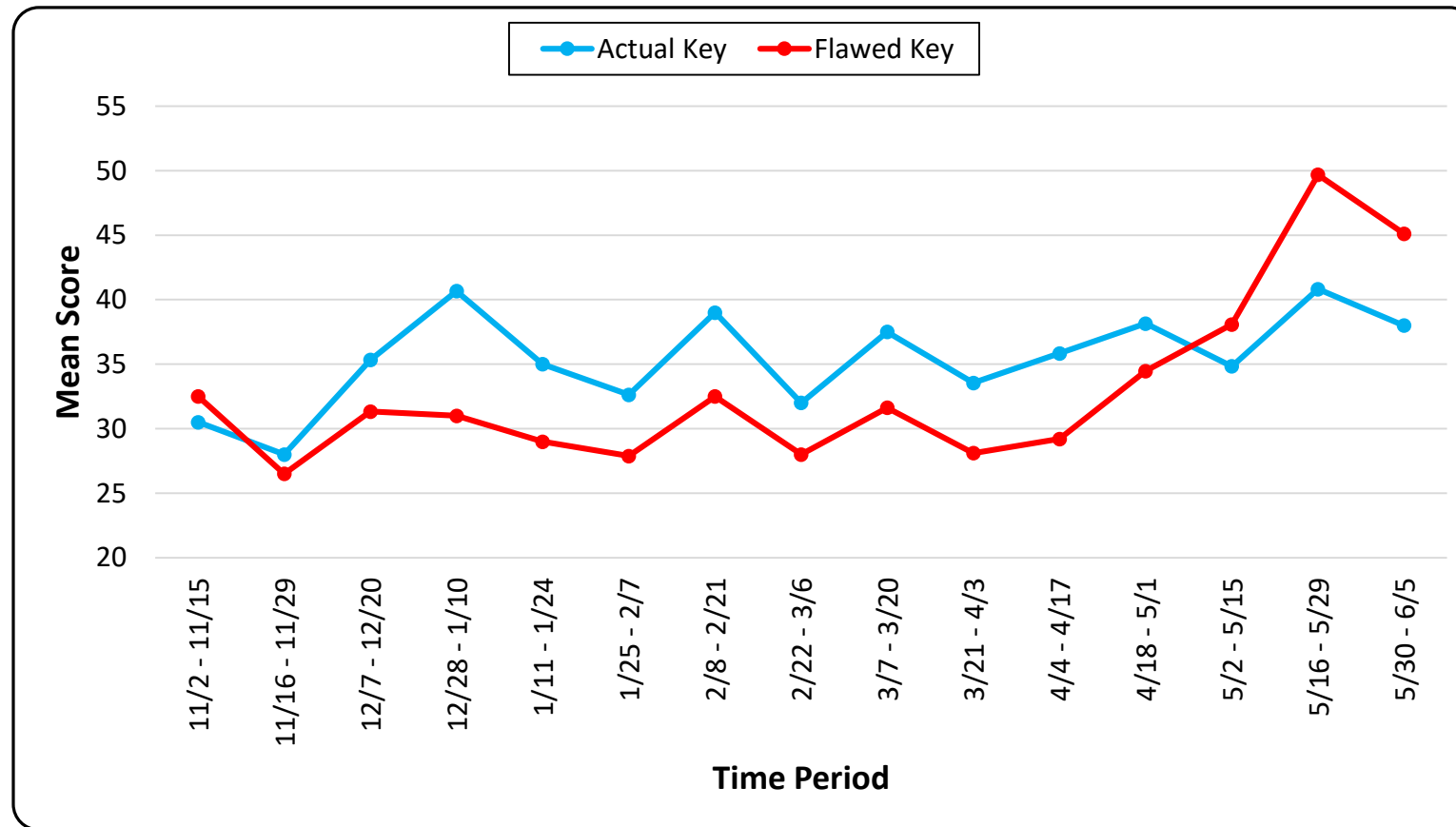
Flawed Key Analysis Results

- 58 examinees flagged as likely braindump users
- These 58 examinees had score increases between 10 and 20 points when the flawed key was used
- Earliest flagged test was April 27, 2016
 - 6 days before the braindump was purchased
 - Response vector matched the flawed key for 56 of the 60 items
- No exact matches with the flawed key

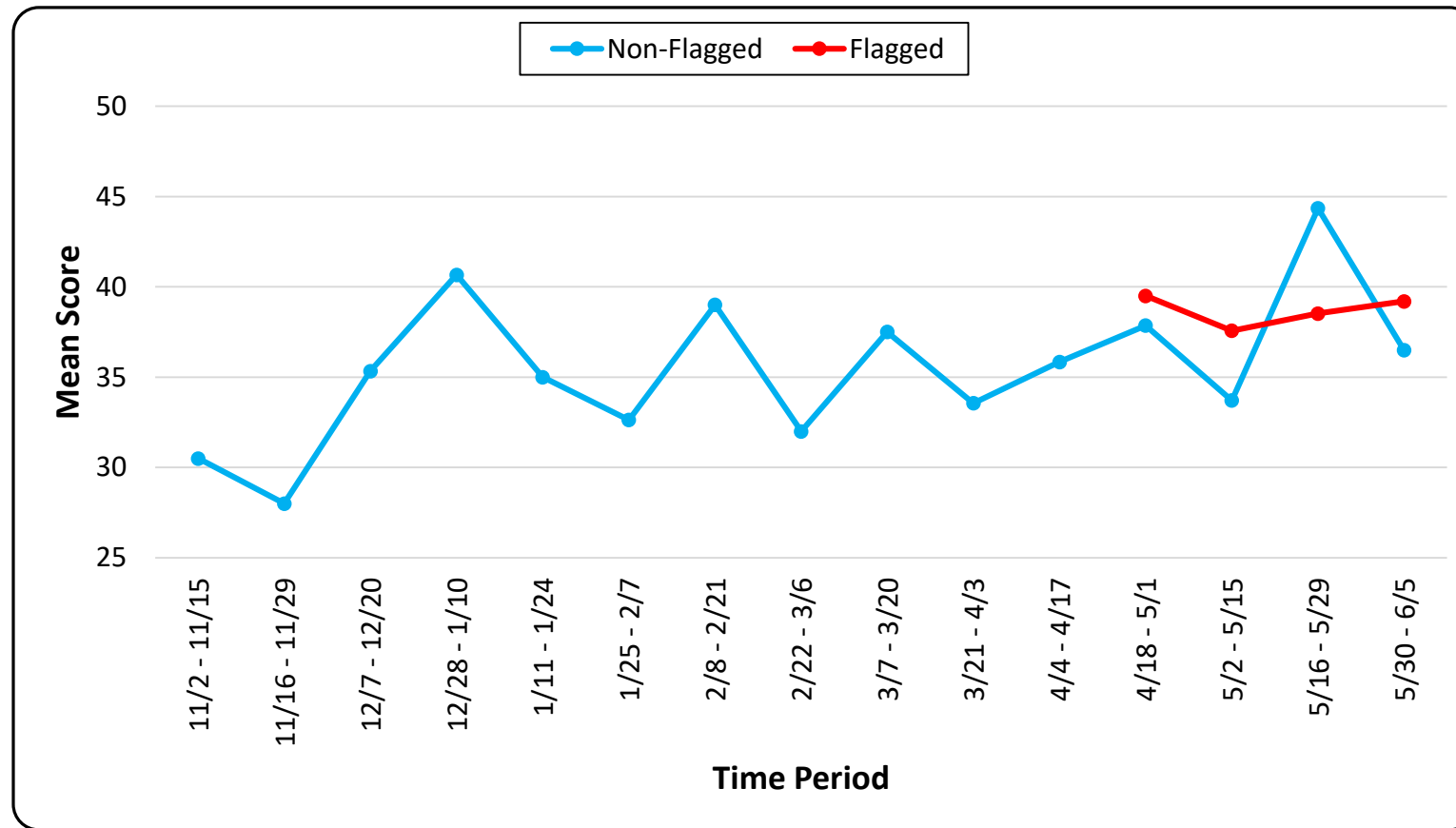
Brindump Use Over Time



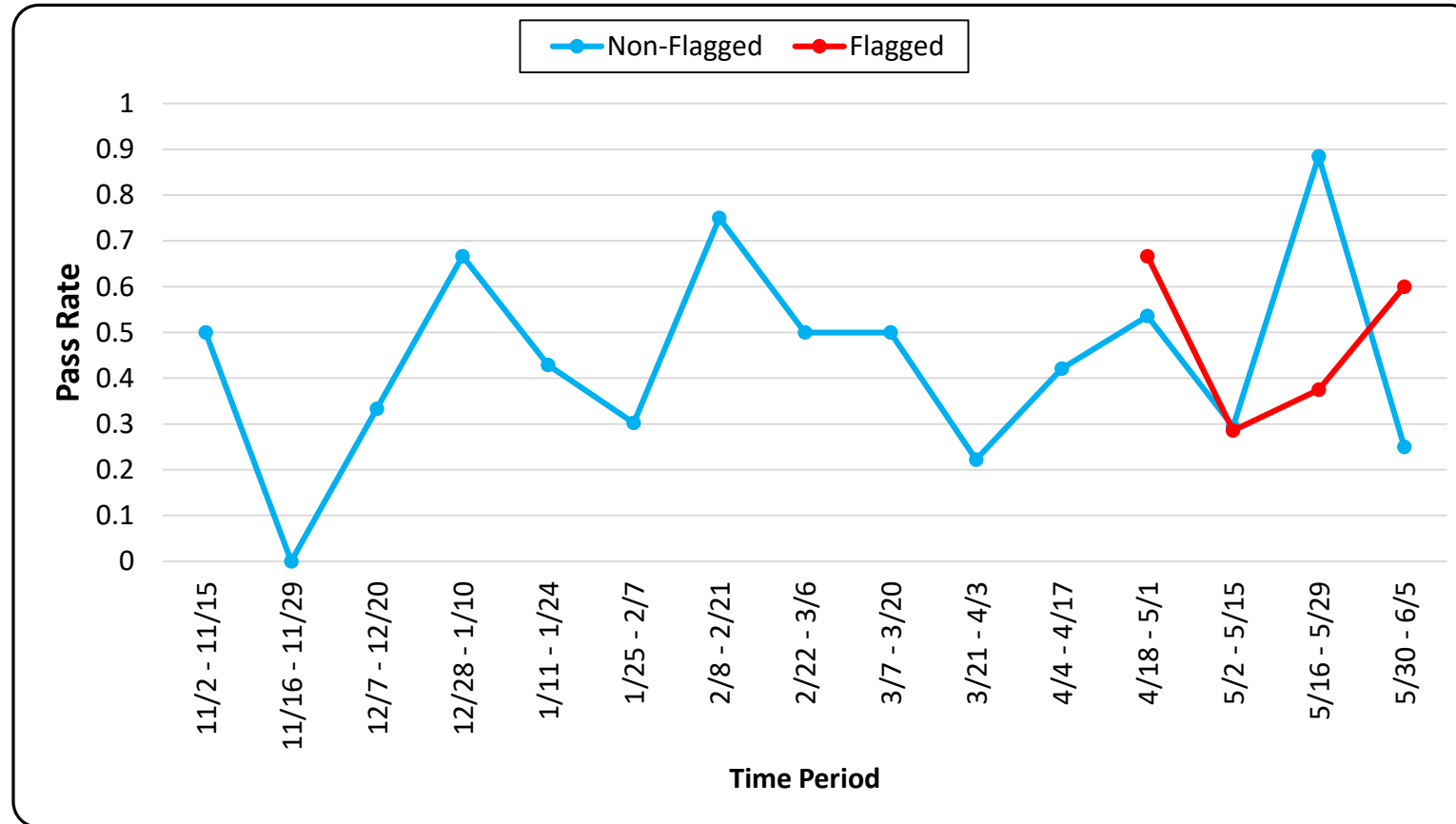
Effect on Scores



Effect on Scores



Effect on Pass Rate



Effect on P-Values

- A logit shift model and dynamic programming were used to analyze item p-value changes (weekly basis)
- 28 items had p-value changes
 - 21 had p-value decreases ranging between 0.10 and 0.42 (all were from the group of 22 with incorrect responses in the flawed key)
 - 7 had p-value increases ranging between 0.34 and 0.57
 - For 26 items, the change occurred during the week of April 25 (first flagged test was April 27)

Findings

1. Braindump quality appears to be an important factor in how the braindump affects mean score and pass rate
2. Flawed braindumps continued to be used, and there did not appear to be a concerted effort to correct them
3. Monitoring p-values can detect items that were disclosed with incorrect answer keys
4. Mean scores and pass rates for non-flagged examinees can increase after a braindump appears

Thank You!

Marcus Scott

marcus.scott@caveon.com

Dennis Maynes

dennis.maynes@caveon.com