Comparison of Proctored and Self-Proctored COMAT Clinical Subject Examination Scores

Hotaka Maeda, PhD, Stuart Barnum, MA

Corresponding author:

Hotaka Maeda, PhD

National Board of Osteopathic Medical Examiners

hmaeda@nbome.org

Presentation type: 10 minute express talk

Learning objectives:

1. Recognize the nuanced differences in scores between self-proctored COMAT Clinical Subjects exams and traditionally proctored exams
2. Understand how to interpret the differences in scores between self-proctored COMAT Clinical Subjects exams and traditionally proctored exams

SHORT DESCRIPTION (200 words)

Self-proctored (SP) examinations were introduced during the COVID-19 pandemic to let students take the web-delivered COMAT Clinical Subject examinations from any location without a proctor. We compared the score differences between 46,893 proctored examinations from 2018 to 2019 and 22,162 SP examinations in 2020. The raw mean (SD) of the scores was 100.9 (10.7) for proctored and 102.5 (10.3) for SP. Average internal consistency test reliability for proctored (.79) and SP (.80) was not significantly different (p = .12). Multiple regression was used to model proctored scores from seven explanatory variables serving as control variables. The multiple regression model explained 10.8% of the variance of the proctored scores. Using the proctored data-based model as the baseline, SP scores were higher than proctored exams under otherwise equivalent circumstances by an average of 1.3 points (SE = 0.07, p < .0001). Although the prevalence of widespread cheating is unlikely and the score difference may be negligible in most applications, careful interpretation is recommended when comparing SP scores to a close threshold. Changes to the school curriculum and student daily lifestyles as a result of COVID-19 were important study limitations that could have affected the scores.

ABSTRACT (700 words)

CONTEXT

COMAT Clinical is a series of eight nationally standardized subject exams designed to assess osteopathic medical students' knowledge and ability in core osteopathic medical and foundational biomedical sciences principles. Every exam contains 125 items, typically administered on campus or in a rotation site under the supervision of a proctor. With the onset of COVID-19, on-campus examinations became impossible for nearly all schools. To maintain the accessibility of COMAT, self-proctored (SP) examinations were introduced on March 30, 2020. This allowed students to take the same COMAT exams from at any location without a proctor. Although the convenience of SP exams enabled continued administration of COMAT during the pandemic, test security and changes to score interpretation became concerns.

OBJECTIVE

We examined the score differences between the proctored COMAT Clinical Subject examinations in 2018-2019 with the SP examinations in 2020.

METHODS

This was an observational study, where two independent groups under different conditions were compared.

Data

The proctored data included 46,893 COMAT Clinical Subject examinations from July 16, 2018, to June 28, 2019, which was one full year of data without changes to the forms or score calculation algorithm. The SP data included 22,162 examinations from March 30, 2020, to October 20, 2020. Students with missing data were excluded. Examinations with both 5 or more missing responses and 5 or more remaining minutes were removed because these were potentially associated with low-effort or technical difficulties.

Analysis

Cronbach's alpha internal consistency reliability was compared between 32 common test forms between proctored and SP using a paired t-test. Forms required at least 100 examinations records to be included.

Examination score differences were also compared. In order to minimize the effects of external factors, seven variables were controlled in the analysis, including the 32-category student's school affiliation (school), an 8-category examination subject (subject), student age in years at the time of the examination (age), 3-category gender (gender), total number of COMAT Clinical examinations taken prior to the examination (seq_test), and total number of exams of the same subject taken prior to the examination (retaken). The final control variable was years-in-school, which was the number of years that the student had been in osteopathic medical school when the examination took place.

The following multiple regression model was estimated using the proctored data:

Score = school + (age + gender + subject + years-in-school + seq_test + retaken)^2,

where the "^2" shows that all main and 2-way interaction effects were included between all explanatory variables within the parentheses. The estimated model was used to calculate expected scores for the SP examinations. Finally, the differences between expected and observed SP scores were compared using bias and percentiles. This statistical approach allowed flexible post-hoc comparisons.

RESULTS

The raw mean (SD) of the exam scores was 100.9 (10.7) for proctored and 102.5 (10.3) for SP. Average exam reliability for proctored (.79) and SP (.80) was not significantly different (p = .12).

The multiple regression model explained 10.8% of the variance of the proctored scores (R^2 = .108, Adj R^2 = .106, RMSE = 10.1, p < .0001). The model explained nearly the same amount of the variance in SP exam scores (R^2 = .104, RMSE = 9.7).

Using the proctored data-based model as the baseline, SP scores were higher than proctored exams under otherwise equivalent circumstances by an average of 1.3 points (SE = 0.07, p < .0001). This bias was greater for the lower scores than the higher scores, which were 1.8, 1.5, 1.3, 1.0, and 0.9 for the 10th, 30th, 50th, 70th, and 90th percentiles, respectively.

CONCLUSION

Reliability was no different between SP and proctored exams, but SP scores were slightly higher. Although the difference may be negligible in most applications, careful interpretation is recommended when comparing SP scores to a close threshold. Given that the reliability is consistent and the gap is equivalent to only about 1 to 2 additional correct items per exam, the indication of serious widespread cheating is unlikely. Potentially, items may have become easier simply from repeated exposure through item parameter drift. The changes to the school curriculum and student daily lifestyles as a result of COVID-19 were important study limitations that could have affected the scores.

Comparison of Proctored and Self-Proctored COMAT Clinical Subject Examination Scores

Hotaka Maeda, PhD
Stuart Barnum, MA
hmaeda@nbome.org

NBOME

NATIONAL BOARD OF OSTEOPATHIC MEDICAL EXAMINERS

BACKGROUND

- COMAT Clinical is a series of eight nationally standardized subject examinations designed to assess osteopathic medical students' knowledge and ability in core osteopathic medical and foundational biomedical sciences principles.

- These exams are currently used by over 80% of COMs.

- Traditionally, they have been proctored on campus or at clinical rotation sites.

- After COVID-19 caused COM campus shutdowns, <u>self-proctored (SP) examinations were introduced on March 30, 2020</u> to allow students to take the same COMAT exams from at any location without a proctor.

- Although convenient, test security and questions about score interpretation became concerning.

NBOME

OBJECTIVE

- We examined the score differences between the proctored COMAT Clinical Subject examinations in 2018-2019 with the SP examinations in 2020.

- We aim to provide examples on <u>how to interpret the SP scores</u>.

NBOME

DATA

- Proctored data included 46,893 exam scores from July 16, 2018 to June 28, 2019

- SP data included 22,162 exams from March 30, 2020 to October 20, 2020

- Examinations with both 5 or more missing responses and 5 or more remaining minutes were removed because these were potentially associated with low-effort or technical difficulties.

NBOME

ANALYSIS

- Least squares multiple regression to <u>predict proctored scores</u> from

  - School affiliation, exam subject, age in years, 3-category gender

  - Test sequence - number of COMAT Clinical examinations taken prior to the examination

  - Retakes - total number of exams of the same subject taken prior to the examination

  - Years-in-school - number of years that the student had been in osteopathic medical school when the examination took place

  - All 2-way interactions (e.g., the effect of retakes on scores when test sequence is low vs high)

- Fitted the proctored model to the SP data to examine the <u>differences between expected and observed SP scores</u>

NBOME

MODEL FIT

- Multiple regression model explained 10.8% of the variance of the proctored scores (R^2 = .108, Adj R^2 = .106, RMSE = 10.1, $p < .0001$).

- Model explained nearly the same amount of the variance in SP exam scores (R^2 = .104, RMSE = 9.7).

- Overall, the factors related to score performance were similar between proctored and SP.
  - E.g., if School 1 scored higher than School 2 for proctored Family Medicine exams, this pattern usually remained consistent if they had taken SP instead

NBOME

SCORE BIAS

- Raw mean standard scores was 100.9 for proctored and 102.5 for SP.

- **SP scores were higher than proctored exams under otherwise equivalent circumstances, especially for lower performers** ($p < .0001$).
  - +0.9 points for high performers (90th percentile)
  - +1.3 points for median and average performers (50th percentile)
  - +1.8 points for lower performers (10th percentile)

- Careful interpretation is recommended when comparing low SP scores to a close threshold.
  - E.g., if a student wanted to obtain a preset score of 90 (16th percentile) and scored exactly 90 on an SP exam, the student may not have achieved this goal if the proctored exam was taken instead.

NBOME

RELIABILITY

- Cronbach's alpha internal consistency reliability was compared between 32 common test forms between proctored and SP using a paired t-test.

- Average exam reliability for proctored (.79) and SP (.80) was not significantly different (p = .12).

- This is evidence <u>against</u> widespread cases of cheating

- It shows that <u>SP exams</u> can be compared to other <u>SP exams</u> without a problem
  - E.g., if Student A's score on a Family Medicine SP exam was 10 points higher than Student B's score, Student A had a higher knowledge level than Student B

NBOME

DISCUSSIONS

- <u>Reliability was not different </u>between SP and proctored exams.

- SP scores were slightly higher than Proctored scores, by about 1 or 2 additional correct items in the 125 item exams, which <u>may be negligible in most applications</u>.

- Why scores may have increased:

  1. Minor or periodic cases of cheating (widespread serious cases are unlikely and have not been observed)

  2. Repeated use of items caused them to become easier over time

  3. Changes to school curriculum and student lifestyles as a result of COVID-19

  4. Students may be less anxious or distracted when taking exams at home

NBOME

THE END

HOTAKA MAEDA, PHD
STUART BARNUM, MA
HMAEDA@NBOME.ORG

#DOProud | #NBOME | #RoadToDOLicensure

NBOME