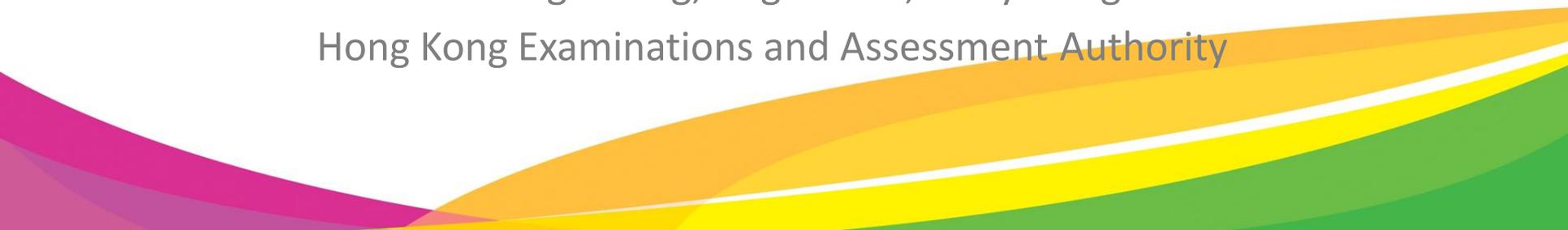# An Empirical Study on the Effectiveness of Double Marking in Hong Kong Diploma of Secondary Education Examination

Xiaoting Huang, Angela Lee, Tracy Hung

Hong Kong Examinations and Assessment Authority

Outline

- Background and Research Questions

- Data and Methods

- Results

- Conclusion and Discussions

Double marking in HKDSE

- Double marking in HKDSE
- The double marking scheme in HKDSE Chinese
  - Mark difference ≤ 11 points
  - Mark difference > 11 points
  - Final mark = the closest highest pair of marks

Example 1
M1:50   M2:70   M3:60
Final mark: 60 and 70

Example 2
M1:40   M2:55   M3:70
M4 (Assistant Examiner):50
Final mark: 50 and 55

Check Marking

- Who are the check markers
  - Chief Examiners, Assistant Examiners, subject managers, very experienced markers

- What kind of answers scripts will be checked
  - 3 scripts from different mark ranges (H, M, L) from each marker in each marking session
  - Large mark discrepancy

Research Questions

1. Is double marking more reliable than single marking?

2. How can we improve marking efficiency?
   – Can we improve marking efficiency by selecting marker with more years of marking experience and/or good performance ratings?

Data

- 7195 case from 2018 HKDSE Chinese writing
  - The minimum score is 0, and the maximum score is 103

N	Marker	Mean	Std. Deviation	Minimum	Maximum
	M1	55.45	10.91	0	90
7195	FM	55.97	9.98	0	100
	C1	55.76	9.96	0	89

Methods

- **Assumption**: check mark is considered to be our best estimate of the students' true score

- Based on $X = T + E$, compare whether double marking results in significantly smaller E than single marking (*paired sample t-test*)

  $M1 = C1 + E_1$ $FM = C1 + E_2$

Methods (*cont.*)

- Can double marking better explain the variation of students' true proficiency than single marking?

$$C1 = a_1 + b_1 \, M1 + e_1 \qquad C1 = a_2 + b_2 \, FM + e_2$$

- Are marking experience and performance rating good indicators of marking proficiency?

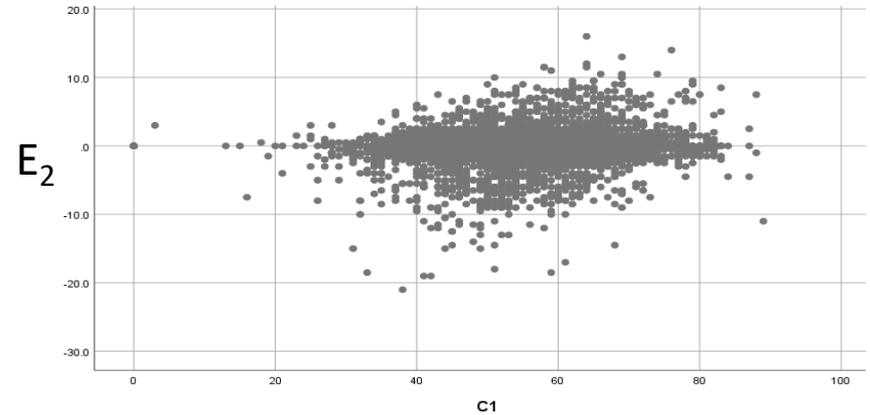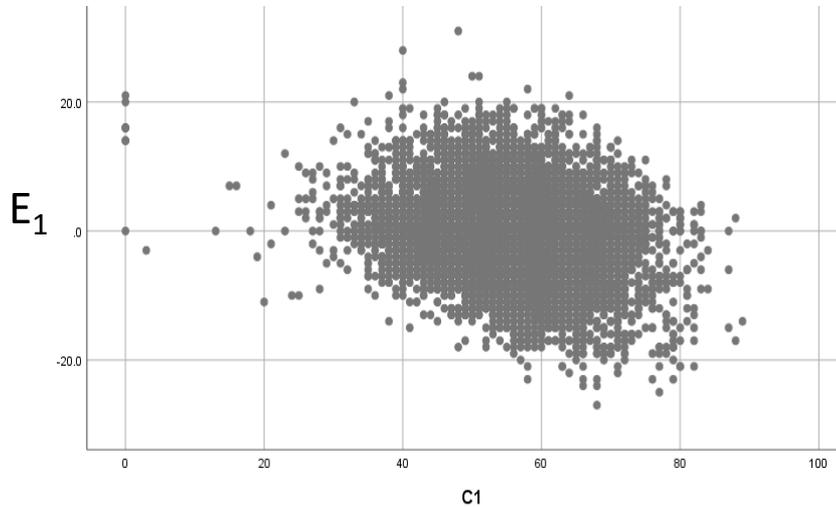$$C1M1_{ABSDIFF} = a + b_1 EXP + b_2 PER + e$$

Descriptive Statistics

- Percentage of M1 not included in the FM: 59%

- Descriptive Statistics of E1 and E2

	Minimum	Maximum	Mean	Std. Deviation
E1（M1-C1）	-27	31	-0.270	7.073
E2（FM-C1）	-21	16	-0.208	2.362
Absolute E1	0	31	5.503	4.450
Absolute E2	0	21	1.393	1.918

Results

- T-tests between E_1 and E_2, and the absolute values of E_1 and E_2 are significant ($p<0.01$)
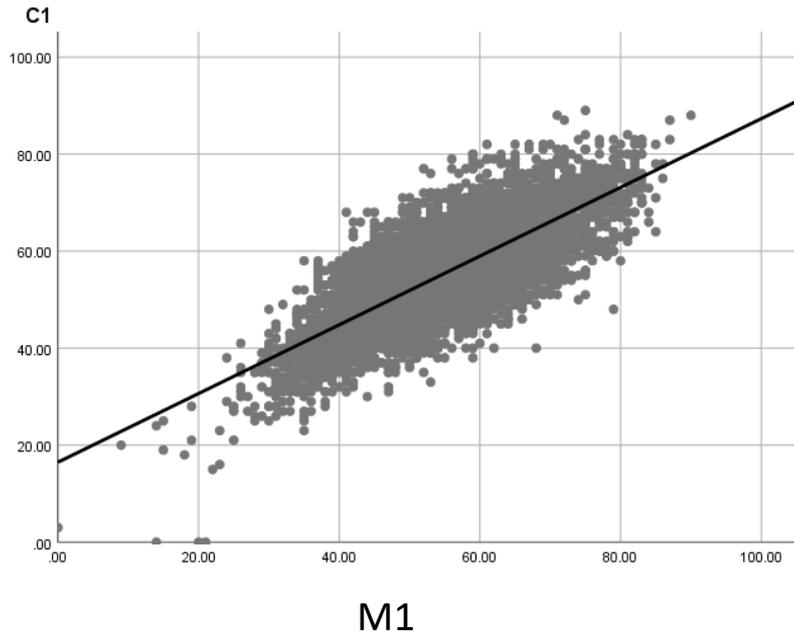
Results (*cont.*)

- Correlations

	M1 and C1	M1 and FM	FM and C1
R	0.776	0.771	0.976

- Summary of regression analyses

	N	R^2	a	b
$C1 = a_1 + b_1 M1$	7195	0.602	16.48	0.708
$C1 = a_2 + b_2 FM$	7195	0.944	1.511	0.969

C1 = $a_1 + b_1$ M1

M1

C1 = $a_2 + b_2$ FM

FM

12
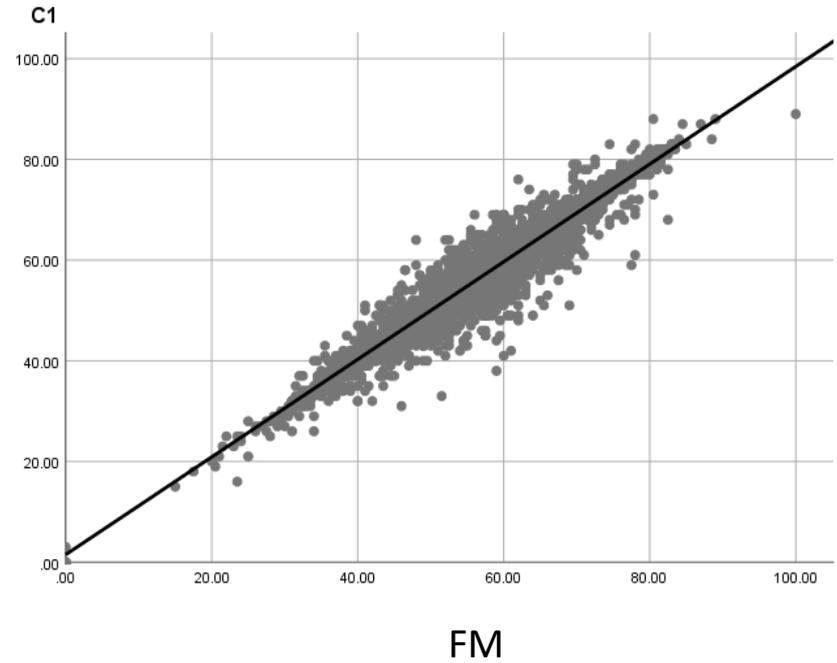
Results (*cont.*)

- Marker experience and performance rating

		N	Min Abs Diff	Max Abs Diff	Mean Abs Diff	Std
Experience	More	4455 (62%)	0	31	5.57	4.47
	Less	2740 (38%)	0	24	5.33	4.36
Perf Rating	High	3535 (49%)	0	31	5.31	4.42
	Average	3660 (51%)	0	27	5.64	4.44

Results (*cont.*)

- Regression analysis result

  $$C1M1_{ABSDIFF} = 5.47 + 0.35\text{EXP} - 0.42\text{PER}$$

  – more experienced markers tended to give marks farther away from the reference mark

  – markers with good performance ratings deviated less from the reference mark

Conclusions and Discussions

- Using C1 as the reference, double marking did help improve marking accuracy
  - Future directions: similar analysis on the subjects/item formats

- Markers' previous performance record was a good indicator to select more reliable markers
  - Look for other potential indicators of markers marking accuracy

Thank you!