Uniwersytet Komisji Edukacji Narodowej w Krakowie

Krakowskie Forum AI w Edukacji

Od badań do praktyki szkolnej

Kraków

23–24 November 2026 / Kraków

AI in Marking Language Exams: The Potential and Limitations of Using ChatGPT

AI w ocenianiu egzaminów językowych: możliwości i ograniczenia wykorzystania umiejętności w ChacieGPT



Anna Pieczka, PhD

Uniwersytet Komisji Edukacji Narodowej w Krakowie

With support from the City of Kraków and the Erasmus+ SmartStart Teacher Academy

Kuratorium Oświaty w Krakowie

Koło Naukowe Nauczyciel z Pasją

MCDN Instytucja Województwa Małopolskiego www.mcdn.edu.pl

Małopolska

smart start ERASMUS+ Teacher Academy

Co-funded by the European Union

Uniwersytet Komisji Edukacji Narodowej w Krakowie

Krakowskie Forum AI w Edukacji

Od badań do praktyki szkolnej

23–24 November 2026 / Kraków

Can AI support teachers in assessing exams?



Kraków

What about…

DYI
with no/little technical knowledge

Ethics

Individualised yet universal

ASSESSING EXAMS WITH AI

Accuracy of the assessment

Broader context

Effective or time consuming?

What about…

DYI
with no/little technical knowledge

Ethics

Individualised yet universal

ASSESSING EXAMS WITH AI

Accuracy of the assessment

Broader context

Effective or time consuming?

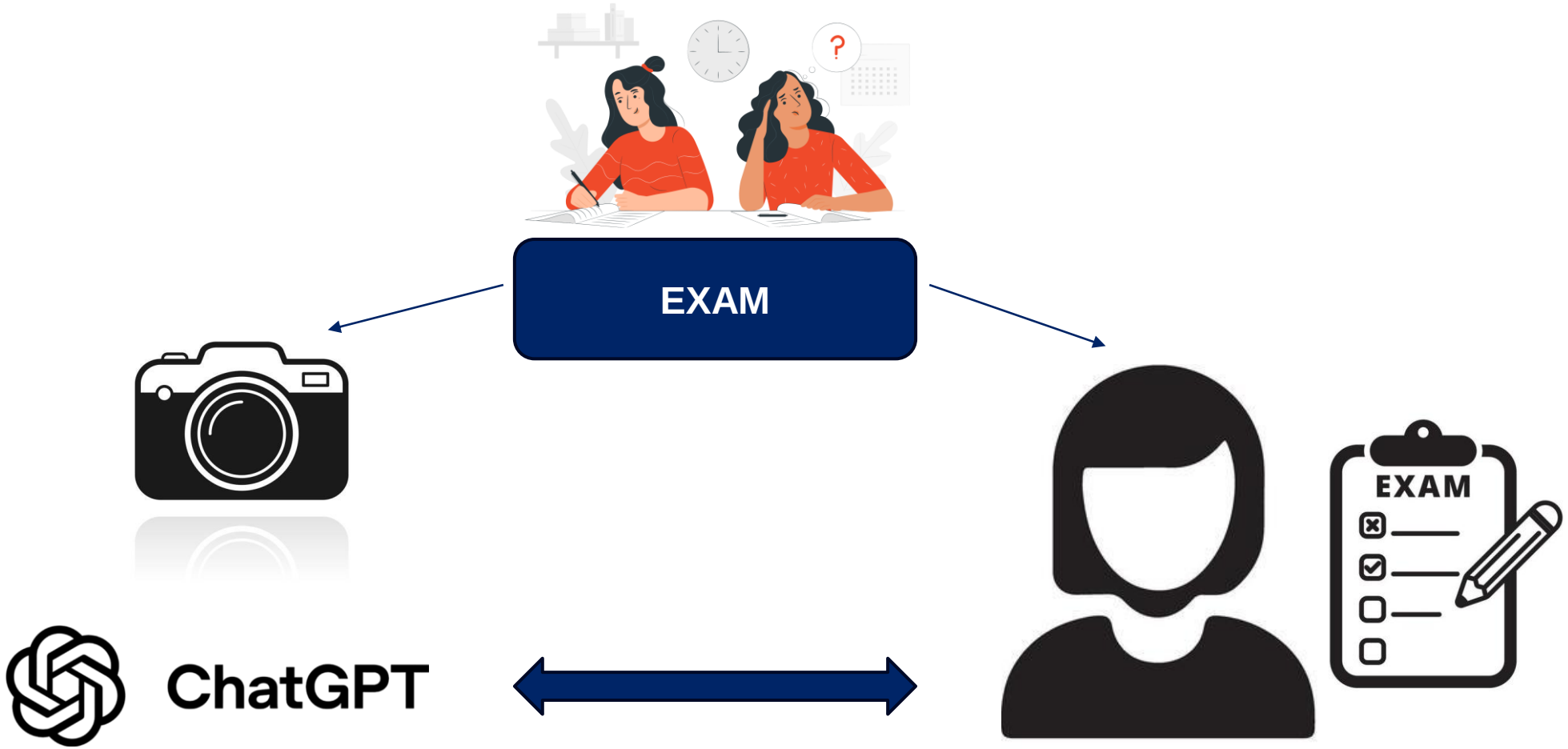
ACCURACY

- Can AI assess exams as accurately as human examiners?
- What margin of error would we consider acceptable?
- What are the main reasons for differences between scores awarded by human examiners and AI?
- How could we reduce these differences?

Exploratory study

- **Tool:** ChatGPT Plus (paid version): creating a skill (an organised instruction)
- **Exam:** Italian language exam for first-year university students
  - *Reading comprehension, vocabulary and grammar*
  - *11 tasks across 6 pages*
  - *Open-ended and closed-ended tasks*
  - *Maximum score: 75 points*
- **Sample:** 5 exam papers × 11 tasks = 55 tasks analysed

Exploratory study



Testing the skill

Exam 1

Task [max]	GPT	HUM
1 [6]	4,5	4,5
2 [8]	8	8
3 [10]	1	1
4 [9]	4,75	4
5 [8]	5	6
6 [5]	4	4
7 [6] open-ended	3,75	1,5
8 [9] open-ended	7,25	7,75
9 [6]	3	3
10 [4]	2	2
11 [4]	4	4
	47,25	**45,75**

(1,5)

Difference range
0
0,25-0,5
0,75-1
>1

Causes of scoring discrepancies

- Inaccurate answer key (0,75)
- Handwriting recognition error (1,0)
- Inaccurate answer key (2,25)
- Handwriting recognition error (0,5)

Comparison of results

Task [max]	Exam 1	
	GPT	HUM
1 [6]	4,50	4,5
2 [8]	8	8
3 [10]	1	1
4 [9]	4	4
5 [8]	6	6
6 [5]	4	4
7 [6] open-ended	1	1,5
8 [9] open-ended	7,5	7,75
9 [6]	3	3
10 [4]	2	2
11 [4]	4	4
	45,00	45,75

Difference range
0
0,25-0,5
0,75-1
>1

Comparison of results

Task [max]	Exam 1		Exam 2		Exam 3		Exam 4		Exam 5	
	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM
1 [6]	4,50	4,5	3	3	6	6	4	4	5	5
2 [8]	8	8	4	4	6	6	7	7	7	7
3 [10]	1	1	3	2,5	5	5	8	9	1	1
4 [9]	4	4	1	1	8	8	8	8	7	6
5 [8]	6	6	4,75	5	7	7	6	6	5,25	5
6 [5]	4	4	3	3	4	4	2	2	1	1
7 [6] open-ended	1	1,5	2,5	2,5	5,5	5,5	4,5	4,5	2	2
8 [9] open-ended	7,5	7,75	6,25	6,25	8	8,5	8,75	8,5	5,5	4,75
9 [6]	3	3	2	2	5	5	5	5	3	3
10 [4]	2	2	1,5	1,5	4	4	3	3	0	0
11 [4]	4	4	1	1	3	3	2	1	2	2
	45,00	45,75	32	31,75	61,5	62	58,25	58	38,75	36,75

Difference range
0
0,25-0,5
0,75-1
>1

Comparison of results

INTERPRETATION

Task [max]	Exam 1		Exam 2		Exam 3		Exam 4		Exam 5	
	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM
1 [6]	4,50	4,5	3	3	6	6	4	4	5	5
2 [8]	8	8	4	4	6	6	7	7	7	7
3 [10]	1	1	3	2,5	5	5	8	9	1	1
4 [9]	4	4	1	1	8	8	8	8	7	6
5 [8]	6	6	4,75	5	7	7	6	6	5,25	5
6 [5]	4	4	3	3	4	4	2	2	1	1
7 [6] open-ended	1	1,5	2,5	2,5	5,5	5,5	4,5	4,5	2	2
8 [9] open-ended	7,5	7,75	6,25	6,25	8	8,5	8,75	8,5	5,5	4,75
9 [6]	3	3	2	2	5	5	5	5	3	3
10 [4]	2	2	1,5	1,5	4	4	3	3	0	0
11 [4]	4	4	1	1	3	3	2	1	2	2
	45,00	45,75	32	31,75	61,5	62	58,25	58	38,75	36,75

1,25 / 7,25

→ **17%**

Difference range
0
0,25-0,5
0,75-1
>1

Comparison of results

ANSWER KEY

1,75 / 7,25

→ **24%**

Task [max]	Exam 1		Exam 2		Exam 3		Exam 4		Exam 5	
	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM
1 [6]	4,50	4,5	3	3	6	6	4	4	5	5
2 [8]	8	8	4	4	6	6	7	7	7	7
3 [10]	1	1	3	2,5	5	5	8	9	1	1
4 [9]	4	4	1	1	8	8	8	8	7	6
5 [8]	6	6	4,75	5	7	7	6	6	5,25	5
6 [5]	4	4	3	3	4	4	2	2	1	1
7 [6] open-ended	1	1,5	2,5	2,5	5,5	5,5	4,5	4,5	2	2
8 [9] open-ended	7,5	7,75	6,25	6,25	8	8,5	8,75	8,5	5,5	4,75
9 [6]	3	3	2	2	5	5	5	5	3	3
10 [4]	2	2	1,5	1,5	4	4	3	3	0	0
11 [4]	4	4	1	1	3	3	2	1	2	2
	45,00	**45,75**	**32**	**31,75**	**61,5**	**62**	**58,25**	**58**	**38,75**	**36,75**

Difference range
0
0,25-0,5
0,75-1
>1

Comparison of results

HANDWRITING

Task [max]	Exam 1		Exam 2		Exam 3		Exam 4		Exam 5	
	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM
1 [6]	4,50	4,5	3	3	6	6	4	4	5	5
2 [8]	8	8	4	4	6	6	7	7	7	7
3 [10]	1	1	3	2,5	5	5	8	9	1	1
4 [9]	4	4	1	1	8	8	8	8	7	6
5 [8]	6	6	4,75	5	7	7	6	6	5,25	5
6 [5]	4	4	3	3	4	4	2	2	1	1
7 [6] open-ended	1	1,5	2,5	2,5	5,5	5,5	4,5	4,5	2	2
8 [9] open-ended	7,5	7,75	6,25	6,25	8	8,5	8,75	8,5	5,5	4,75
9 [6]	3	3	2	2	5	5	5	5	3	3
10 [4]	2	2	1,5	1,5	4	4	3	3	0	0
11 [4]	4	4	1	1	3	3	2	1	2	2
	45,00	**45,75**	**32**	**31,75**	**61,5**	**62**	**58,25**	**58**	**38,75**	**36,75**

4,25 / 7,25

→ **59%**

Difference range
0
0,25-0,5
0,75-1
>1

Comparison of results

Task [max]	Exam 1		Exam 2		Exam 3		Exam 4		Exam 5	
	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM	GPT	HUM
1 [6]	4,50	4,5	3	3	6	6	4	4	5	5
2 [8]	8	8	4	4	6	6	7	7	7	7
3 [10]	1	1	3	2,5	5	5	8	9	1	1
4 [9]	4	4	1	1	8	8	8	8	7	6
5 [8]	6	6	4,75	5	7	7	6	6	5,25	5
6 [5]	4	4	3	3	4	4	2	2	1	1
7 [6] open-ended	1	1,5	2,5	2,5	5,5	5,5	4,5	4,5	2	2
8 [9] open-ended	7,5	7,75	6,25	6,25	8	8,5	8,75	8,5	5,5	4,75
9 [6]	3	3	2	2	5	5	5	5	3	3
10 [4]	2	2	1,5	1,5	4	4	3	3	0	0
11 [4]	4	4	1	1	3	3	2	1	2	2
	45,00	45,75	32	31,75	61,5	62	58,25	58	38,75	36,75

59%
HANDWRITING

24%
ANSWER KEY

17%
INTERPRETATION

Difference range
0
0,25-0,5
0,75-1
>1

Conclusions

Challenges to address

- ❑ **Accurate handwriting recognition**
- ❑ Broader verification process
- ❑ Broader sample

Basic rules

- ✓ **Provide an accurate answer key**
- ✓ Review exam papers with scores close to grade boundaries
- ✓ Have a sample to compare

Uniwersytet Komisji Edukacji Narodowej w Krakowie

Krakowskie Forum AI w Edukacji

Od badań do praktyki szkolnej

Kraków

23–24 November 2026 / Kraków

Thank you

Questions and discussion



Anna Pieczka, PhD

anna.pieczka1@uken.krakow.pl

Kuratorium Oświaty w Krakowie

Koło Naukowe Nauczyciel z Pasją

MCDN Instytucja Województwa Małopolskiego

Małopolska

smart start ERASMUS+ Teacher Academy

Co-funded by the European Union