INNOVATION IN LANGUAGE LEARNING 2026· FLORENCE

From Perceptions to Practice

Exploring EFL university teachers' prompt literacy in generative AI-assisted language teaching

Hessah Saleh Aldayel

Assistant Professor of Applied Linguistics · English Language Skills Department
Common First Year Deanship, King Saud University, Riyadh, Saudi Arabia



BEFORE WE BEGIN

Two questions, and I need your chat box

1

Type **YES** in the chat if you have used a generative AI tool for your teaching in the last seven days.

2

Now type **YES** again if you would say you are good at writing prompts.

I will read your answers out as they come in.



THE SAME TWO QUESTIONS

Twenty EFL teachers answered them almost exactly as you just did

19 of 20

use generative AI tools in their teaching

3.87 / 5

their average rating of their own prompt writing

Then I asked them to write a prompt.



THE GAP

What the research tells us, and what it does not

What we know

- EFL teachers hold positive views of GenAI for materials, planning and feedback

Alzubi & Alelaiwi (2025)

- They ask for training in writing effective prompts

Bhatti (2026)

- Prompt quality predicts output quality

Knoth et al. (2024)

- Most GenAI studies in language classrooms rely on self-reported data

Lee et al. (2025)

What we do not know

- What teachers actually write when they prompt
- Whether their self-rating matches the quality of those prompts
- Which elements of a prompt they leave out
- Which teachers most need prompt-focused training

Studies of real prompts have focused almost entirely on students, not teachers.



Theoretical framework

Two ideas this study is built on

IDEA ONE

Prompt literacy

A new literacy skill: writing precise prompts for generative AI, reading what comes back, and revising the prompt until it does the job.

Prompting is a linguistic and cognitive activity — give the input, define the task, specify the output — not a list of tricks.

Hwang et al. (2023); Tour & Zadorozhnyy (2025)

IDEA TWO

Teacher cognition

What teachers say they believe, know and can do does not always match what they actually do.

In digital competence specifically, teachers' self-ratings have been shown not to measure what they can actually do.

Borg (2003); Maderick et al. (2015); González-Mujico (2023)



RESEARCH QUESTIONS

Four questions, in one instrument

Q1

How far do these teachers use GenAI, and how do they perceive their own prompt literacy?

Q2

What is the quality of the prompts they write to get GenAI feedback on student paragraphs?

Q3

How closely does self-perceived prompt literacy match the quality of the prompts they write?

Q4

What challenges do they meet when writing prompts for GenAI?



DESIGN

One instrument, three kinds of evidence

PART 1

What they say

Questionnaire: background, how often and why they use GenAI, and twelve Likert items on how they write prompts.

PART 2

What they write

Two tasks: write, word for word, the prompt you would use to get GenAI feedback on this student paragraph.

PART 3

How they explain it

Open questions: the challenges they meet, how they judge GenAI feedback, and what support would help.

Mixed-methods: all three collected at the same time, from the same teachers, and interpreted together.

PARTICIPANTS AND SETTING

One department, one common syllabus

200
teachers invited

20
completed the survey

A response rate of about 10 per cent. Teachers follow a common syllabus, teach similar proficiency levels and give written feedback on paragraphs regularly— so the tasks are close to their daily work.

Participation was voluntary and anonymous, so the sample is self-selected.

Characteristic	n = 20
Women / men	18 / 2
11 or more years' experience	11
Doctorate / Master's / Bachelor's	10 / 9 / 1
Trained on using GenAI tools	10
Trained on writing prompts	6

MEASUREMENT

Nine things you can see in a written prompt

Input

I1 the student paragraph

I2 the learners' level

I3 the course or task context

Task

T1 a clear action

T2 the focus of the feedback

T3 the teaching purpose

Output

O1 the format of the answer

O2 the tone or style

O3 the length or scope

Each item scored **0** absent · **1** vague · **2** specific. Maximum **18** per task.

Two raters scored all 37 prompts independently. Weighted κ = **0.87**; exact agreement 84.1%.

Researcher-developed checklist adapted from Li's (2025) input–task–output framework. Disagreements resolved by codebook rules and discussion.



FINDING 1 · WHAT TEACHERS SAID

Regular users, and confident ones

How they use GenAI

11 of 20 use it daily or almost daily; 5 weekly; 3 monthly; 1 never.

ChatGPT 17 · Gemini 12 · Copilot 5 · Claude 3

Most common purposes: making materials (16), quizzes and tests (15), lesson planning (13).

Generating feedback on student writing — the task in this study — was chosen by only **9 of 20**.

How they rated themselves

3.80
Input

3.90
Task

3.82
Output

3.95
Overall confidence

Self-perception score 3.87 (SD 0.75) on a 1–5 scale. Cronbach's α = .86.

FINDING 2 · WHAT THEY WROTE

The prompts told a different story

6.6 / 18

mean consensus score across 37 scored prompts

SD 2.62 · teacher means ranged from 1 to 11.5

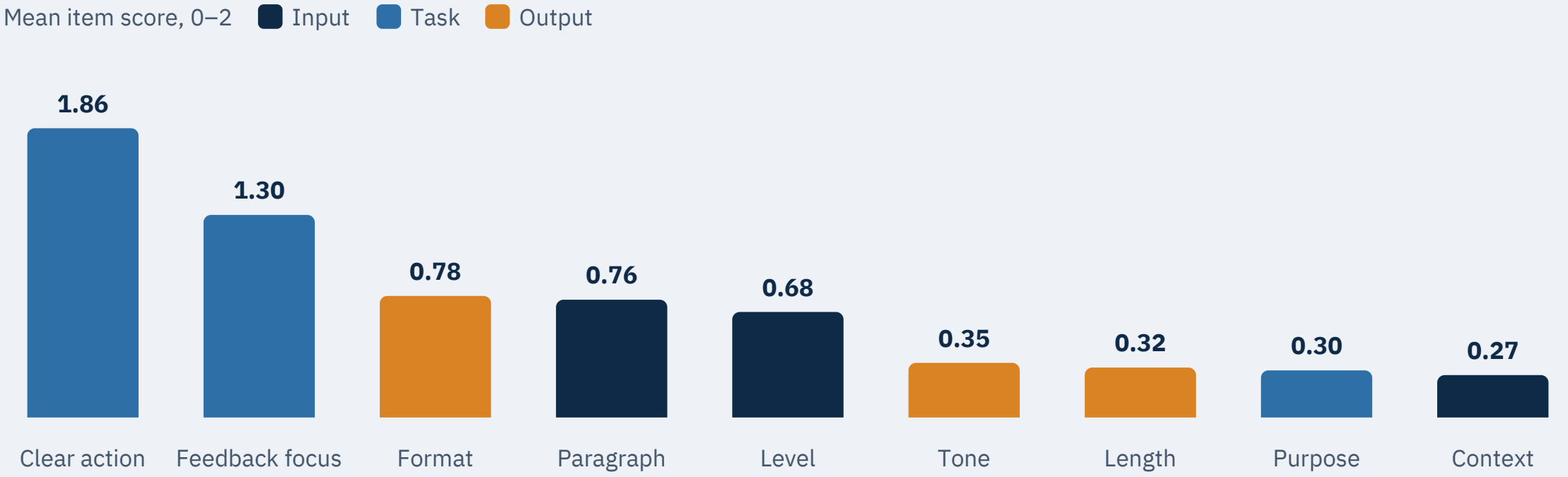
Task 1 mean **6.8** · Task 2 mean **6.4**

8 teachers submitted the **same prompt text** for both tasks

Roughly one third of the available points

FINDING 2 · WHERE THE POINTS WENT

They said what to do, and almost nothing else



Consensus scores, 37 prompts. Maximum 2 per item.

FINDING 3 · THE CENTRAL COMPARISON

ρ = −.30

Teachers who rated themselves higher did not write better prompts

Spearman's rank correlation between the self-perception score and the consensus prompt score, n = 19.

Repeating the analysis with the second rater's independent scores gave the same picture: ρ = −.37



FINDING 3 · THREE GROUPS OF TEACHERS

Confidence rose as prompt quality fell

Group	n (%)	Self-perception (1–5)	Prompt score (/18)
Rated themselves higher than they scored	6 (31.6%)	4.41	3.8
About the same	8 (42.1%)	3.89	6.9
Rated themselves lower than they scored	5 (26.3%)	3.20	9.2

Group boundaries were decided before any teacher was grouped. With n = 19, these comparisons are descriptive; no significance tests were run.

THREE TEACHERS

What the gap looks like up close

1 / 18

Answered *strongly agree* or *agree* to every self-perception item.

The whole prompt: “grammar mistakes”

The largest gap in the sample.

11.5 / 18

Self-ratings among the more modest in the sample.

Named the error types and gave an explicit learning purpose

The highest score in the sample.

4 / 18

Among the most experienced; trained in GenAI use and in prompt writing; a daily user.

Rated themselves *agree* or above throughout; did not attempt Task 1

Experience and training were not enough on their own.

Participant responses, anonymised.

FINDING 4 · WHAT THEY SAID WAS HARD

Three themes in their own words

1 **Making the prompt complete and specific enough**

“Ensuring everything I need is within the prompt” · “I guess I do not detail it enough”

2 **The tool misunderstanding, or answering unreliably**

“sometimes AI does not fully understand what you need exactly” · “it often gives me its own opinion”

3 **No challenges at all**

Four teachers answered *“nothing”*, *“none”*, or similar

Theme 1 matches the scored prompts exactly: the elements teachers found hard to include are the elements the checklist found missing.

Thematic content analysis of open-ended responses.

THE GROUP THAT WORRIES ME

Seeing no difficulty was not a sign of doing it well

The lowest-scoring teacher in the sample answered the three open questions:

“nothing” · “do not know” · “nothing”

One teacher declined both prompt tasks, objecting to using GenAI for student feedback, and wrote:

“None. Writing prompts is easy … None, I have a brain.”

Voluntary training reaches the teachers who already know they need it.

IMPLICATIONS

What prompt-focused training should actually teach

Not “using AI”

Teachers already state clear actions. General AI-awareness sessions teach them what they can already do.

Three missing habits

Context: who the learner is, what the course and task are. Purpose: what the feedback is for. Output: format, tone, length.

Hands-on, not theory

Teachers asked for practical examples and practice; three asked for ready-made prompts they could save and reuse.

Make room for refusal

Departments introducing GenAI feedback should expect principled disagreement and plan for it.

The six teachers with prompt-writing training averaged 7.5 of 18 against 6.1 for the thirteen without it. With groups this small the comparison is descriptive only — but its direction is consistent with the item profile.

LIMITATIONS

What this study cannot tell you

Twenty teachers, one department

About 10% of those invited, and self-selected. The findings describe this group, not the department and not the field.

Unsupervised online completion

Outside help cannot be ruled out, though the low scores make inflated results unlikely.

One occasion, two fixed tasks

A single written prompt per paragraph, not the conversation a teacher would normally have with the tool.

One checklist, small subgroups

Another instrument might weight the elements differently, and every subgroup comparison here is descriptive.

Next: larger samples, other departments and countries, and a test of whether the training suggested here actually closes the gap.

THREE THINGS TO TAKE AWAY

1. Confidence in prompting told us nothing useful about the quality of the prompts teachers wrote.
2. What was missing was specific and teachable: context, purpose, and output specification.
3. The teachers who thought they had no problem had the biggest one — and they are the least likely to come to training.

Thank you — questions welcome

Hessah Saleh Aldayel · King Saud University · haldayel@ksu.edu.sa

The checklist and codebook are in Appendix A of the paper and are free to adapt. Approved by the Standing Committee for Research Ethics, King Saud University (KSU-HE-26-1159).

