# Edo State GPT-4 after-school English tutoring trial

IMPLEMENTED 2024 · EDUCATION · NIGERIA · BAND D · PARTLY PROVED
VERSION 1.0 · PUBLISHED 15 NOVEMBER 2026 · UPDATED 15 NOVEMBER 2026 · FACTS CHECKED 15 NOVEMBER 2026

World Bank researchers ran a randomised trial of a six-week after-school programme in which Nigerian secondary students used Microsoft Copilot (GPT-4) as an English tutor. Students chosen for the programme scored higher than those who were not. The result is a working paper from a single pilot in nine schools, and no scale-up has been reported.

**ABOUT THIS SYNOPSIS**

This is MarLuc's summary and grading of publicly available sources, for information only. It is not investment, legal, financial or professional advice, and should not be relied on for any decision. Readers must validate it against the original sources and, where appropriate, challenge its conclusions.

Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date. Check the original source before relying on any figure.

The impact band, proof marker and SEAM view are MarLuc's opinion. The organisation is not a MarLuc client and has not reviewed or endorsed this synopsis. Names and trademarks belong to their owners.

NOTICE WORDING LAST UPDATED 16 NOVEMBER 2026

AT A GLANCE

AI used: Generative AI tutor: Microsoft Copilot (GPT-4) used by pairs of students under teacher supervision, with lesson prompts written by the programme team.

Implemented: 2024.  Grade: band D, Task support, Partly proved.

Headline result, as reported: In a randomised trial in Benin City, Nigeria, a six-week after-school GPT-4 tutoring programme raised English scores by 0.24 standard deviations (0.31 on the full assessment), per World Bank researchers; the paper is a working paper and endline data were missing for about a third of students

## 1 · The case

The facts, as reported by the 3 sources listed at the end. Where a source is silent, the field says "Not reported".

### Organisation

World Bank, with the Edo State education authorities. World Bank Education Global Department research team evaluating a pilot in nine public secondary schools in Benin City, Edo State, Nigeria, for first-year senior secondary students aged about 15. Canvas block: 3. Customers and proposition. Also 5. People and skills.

## Business challenge

The challenge. The authors cite learning poverty of about 70% of 10-year-olds in low- and middle-income countries, a shortage of teachers, and the cost of one-to-one tutoring, which they say is unaffordable for most education systems.

How it was addressed. Students attended twelve 90-minute after-school sessions in school computer labs. They worked in pairs with Microsoft Copilot, powered by GPT-4, on topics from the national English curriculum. Teachers, trained for three days, started each session with a prompt and supervised but did not teach. Students were randomly chosen for the programme from those who volunteered. [1] [2] [3]

## Who implemented it

World Bank (University or research lab): Designed the programme and ran the randomised evaluation.
Microsoft (Model or cloud provider): Copilot, powered by GPT-4, was the tool students used.
The Edo State authorities are named in the working paper only as collaborators on implementation; their formal role is not described. The paper acknowledges financial support from the Mastercard Foundation.

## Technology (SEAM pillar E)

SEAM types: T1 Chat assistants. Deployment: Vendor SaaS. Vendors and products: Microsoft, OpenAI, Microsoft Copilot (GPT-4).

Where it was built in: Students used the free Copilot chat tool in school computer labs, in pairs, following lesson guides with prompts written by the programme team. There was no custom build or question bank.

## What changed

Before: students had regular classes only. In the pilot: programme students had extra supervised sessions with a generative AI tutor. Measured effects are in the figures. No change to the school system's way of working is reported. [1] [2] [3]

## What it cost

Per-pupil cost of the six-week pilot about $48 for 657 students, with marginal cost about $9, per the paper; a four-quarter programme is estimated at $124 per pupil.

## What it saved

Not a saving. The authors estimate gains equal to 1.5 to 2 years of typical schooling and 3.2 equivalent years of schooling per $100 invested, by their own calculation. [1] [2] [3]

## Governance and control

Guardian consent, teacher training on hallucinations and bias, teacher supervision and attendance monitoring, all per the authors. Accountable owner, data protection and vendor terms: Not reported.

**S2 · Accountable owner · Claimed**
The paper thanks named Edo State collaborators for implementation; a single accountable owner for the programme is not named. [2]

**A1 · Safe use · Claimed**
Guardians signed consent forms, and teachers and students were told about hallucinations, bias and over-reliance, per the paper. [2]

**A2 · Human in the loop · Claimed**
Teachers circulated during sessions to keep interactions on task, per the paper; this is the authors' own account. [2]

**A3 · Monitoring and assurance · Claimed**
Monitors tracked attendance and session conditions using a monitoring tool, per the paper. [2]

**A4 · Vendor AI · Not reported**
Not reported beyond the use of free Copilot accounts. [2]

**A5 · Agent controls · Not reported**
Not reported.

**A6 · Standards and regulation · Not reported**
Not reported.

## What worked, what did not, what is unproven

WORKED

– Programme students scored 0.31 standard deviations higher on the full assessment (weighted) and 0.238 higher in English, the main outcome.

– Programme students also scored 0.206 standard deviations higher on the school's third-term English exam, which covered more than the programme.

– Gains rose with attendance: about 0.031 standard deviations per extra day attended.

– Effects were positive across baseline ability, and larger for girls, per the working paper.

DID NOT

– Endline data were missing for many students: 422 of 657 programme students and 337 of 671 control students completed the final assessment. The authors say the difference in attrition was significant and report bounds and weighting checks.

– Randomisation was among volunteers, so results may not apply to all students.

– Some control students gained access to sessions and there were early account and connection problems, per the authors.

UNPROVEN

– Whether the gains last; the authors list long-term effects as an open question.

– Whether the chatbot itself caused the gain. The authors say they cannot separate it from structured prompts, curriculum alignment, teacher support and peer work.

– Results beyond English, beyond after-school sessions, in rural schools, or at scale.

## Timeline

Jun 2024 Six-week after-school programme starts in nine public schools in Benin City, Edo State [2]

Jul 2024 Programme ends; students sit a pen-and-paper endline assessment [1]

Jan 2025 World Bank blog reports the first results of the randomised evaluation [1]

May 2025 Policy Research Working Paper 11125 published [2]

Jun 2025 World Bank blog sets out results, cost-effectiveness and open questions [3]

## 2 · Grade and SEAM assessment

MarLuc's opinion, made on the evidence above. How cases are graded: marluc.co.uk/case-studies/how-we-rate

### Level of change and impact

Claimed: A.  Grade: D, Task support.  Proof: Partly proved.  Level of change: Task support.

Task support (Level 1, band D), claimed as transformational (band A). The programme added a supervised tutoring session; no process, role or budget change is reported.

Graded D, claimed A. The World Bank team calls the result's potential transformative. Sources show learning gains from a six-week pilot with teacher support, but no change to how schools operate and no scale-up. The potential is unproven.

### The SEAM view and lessons learnt

The design is a useful template: structured prompts tied to the curriculum, teacher supervision, a control group and cost recorded from the start. The effect sizes come from the evaluators' own working paper with large missing endline data, so treat the headline as promising evidence rather than a settled result. For a larger organisation the transferable point is to test an AI tool against a comparison group and publish the limits alongside the gain.

1. Wrap the AI tool in structured prompts and human supervision.
2. Use a control group and record cost from the start.
3. Report attrition and limits next to the headline effect.

### Evidence

Score 2 of 4. Independent design, own evaluation: randomised trial reported by the programme's World Bank team in a working paper; not independently replicated.

Evaluation method. Randomised controlled trial among volunteering students, with intent-to-treat estimates, attrition bounds and a cost-effectiveness calculation; reported as a World Bank working paper.

Sources disagree. The paper's abstract gives 0.23 standard deviations for English and its introduction gives 0.24; the results table gives 0.238. The introduction describes 0.31 as the score on the final assessment while the item-response-theory score is 0.26. The January 2025 blog says about 0.3 standard deviations and nearly two years of typical learning in six weeks; the paper says 1.5 to 2 years.

Status. Pilot complete. No scale-up or follow-up trial is reported in the sources opened.

## 3 · Original sources

Read these in full. To challenge a fact or a grade, use the contact form at marluc.co.uk or the link on https://marluc.co.uk/case-studies/edo-state-gpt4-english-tutoring-trial


[1] World Bank: From chalkboards to chatbots: Transforming learning in Nigeria, one prompt at a time (blog, January 2025)
https://blogs.worldbank.org/en/education/From-chalkboards-to-chatbots-Transforming-learning-in-Nigeria

[2] World Bank: From Chalkboards to Chatbots: Evaluating the Impact of Generative AI on Learning Outcomes in Nigeria (Policy Research Working Paper 11125, May 2025)
https://documents1.worldbank.org/curated/en/099548105192529324/pdf/IDU-c09f40d8-9ff8-42dc-b315-591157499be7.pdf

[3] World Bank: Addressing the learning crisis with generative AI: lessons from Edo State in Nigeria (blog, June 2025)
https://blogs.worldbank.org/en/developmenttalk/addressing-the-learning-crisis-with-generative-ai--lessons-from-

World Bank, with the Edo State education authorities website: https://www.worldbank.org



Version 1.0 · Published 15 November 2026 · Updated 15 November 2026 · Facts checked against sources 15 November 2026 · Downloaded 26 November 2026. This PDF may be out of date: the latest version, and a way to report an error or a better source, is at https://marluc.co.uk/case-studies/edo-state-gpt4-english-tutoring-trial


VERSION HISTORY

v1.0 15 November 2026 First published