

生成式人工智能用得越多，学得越好？

——基于 PISA 2025 中国四省(市)15 岁学生的调查

徐瑾劼¹ 安德烈亚斯·施莱歇尔²

- (1. 上海师范大学 国际教师教育中心/联合国教科文组织教师教育中心, 上海 200234;
2. 经济合作与发展组织教育与技能司, 法国巴黎 75775)

[摘要] 近期, 经济合作与发展组织发布了 PISA 2025 报告, 全球有 91 个国家和经济体 76 万余名 15 岁学生参加了测评。其中, 中国北京、上海、江苏和浙江四省(市)约 1.2 万名学生参测。PISA 2025 首次调查了以聊天机器人为代表的生成式人工智能在学校教学中的使用情况, 为回答人工智能何以促进学生学习提供了大规模、跨学科证据。本研究基于中国四省(市)PISA 2025 学生数据, 考察人工智能使用强度、认知功能及学生异质性与科学、数学和阅读成绩的关联。研究发现, 中国四省(市)15 岁学生尚未普遍高频依赖人工智能并伴随成绩明显下降, 但人工智能使用与学业表现的关联强度已呈现结构性差异, 主要体现在不同学科和不同能力水平的学生上。鉴于此, 本研究提出, 学校人工智能治理应从单一重视人工智能使用频率的管控转向动态监测与介入过程设计并重, 实施分学科、分群体差异化引导。

[关键词] OECD; PISA 2025; 生成式人工智能; 中国四省(市)

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2026)05-0042-09

一、问题提出

人工智能正加速进入学生的日常学习, 并日益渗透能力培养体系。2026 年, 教育部等五部门联合印发《“人工智能+教育”行动计划》, 明确提出推动人工智能融入教育教学全要素、全过程和全场景(教育部, 2026), 人工智能融入学校教育已成为大势所趋。然而, 人工智能的普及应用未必必然带来育人质量的提升。人工智能时代的教育面临

一个悖论: 学生完成学习任务越来越快, 并不意味着学得越来越好; 当越来越多本应由学生完成的认知活动交给人工智能, 其独立思考和问题解决能力反而可能因缺少必要的认知磨炼而受到削弱(Bastani, 2025)。人工智能在教育中真正值得期待的, 不只是提高任务完成效率, 更在于创造更具互动性的学习体验, 释放大规模因材施教和个性化辅助学习的潜能。但与促进学习的效果相比, 其提高任务完成效率的优势得到了更充分的实证支持。

[收稿日期] 2026-09-11

[修回日期] 2026-09-14

[DOI编码] 10.13966/j.cnki.kfjyyj.2026.05.005

[基金项目] 2024 年度国家社会科学基金教育学一般项目“基础性转向背景下中职生核心技能的内涵框架与创新培养研究”(BJA240174)。

[作者简介] 徐瑾劼, 教授, 博士生导师, 上海师范大学国际教师教育中心/联合国教科文组织教师教育中心, 教育部教育大数据与教育决策实验室, 研究方向: 比较教育、国际大规模教育测评(xujinjie@shnu.edu.cn); 安德烈亚斯·施莱歇尔, 博士, 经济合作与发展组织教育与技能司司长, 研究方向: 国际教育测量、教育政策(andreas.schleicher@oecd.org)。

[引用信息] 徐瑾劼, 安德烈亚斯·施莱歇尔(2026). 生成式人工智能用得越多, 学得越好? ——基于 PISA 2025 中国四省(市)15 岁学生的调查[J]. 开放教育研究, 32(5): 42-50.

已有研究发现,生成式人工智能能显著缩短专业写作和知识密集型任务的完成时间,并提高成果质量(Noy & Zhang, 2023; Brynjolfsson, 2025),但其能否促进学生的知识掌握和能力发展,现有研究尚未形成一致结论。教育的目的并非让学生以最少的时间获得答案、完成任务,而是使其在完成任务过程中获得知识和发展能力。生成式人工智能能够生成文本、答案和任务成果,但能否促进学习者建构意义和知识结构,即生成性学习(generative learning),仍是需要回答的关键问题。生成性学习指学习者主动理解学习材料,包括选择相关信息,把它组织成连贯的认知结构,并将新信息与已有知识相整合(Fiorella & Mayer, 2015)。可见,人工智能生成的答案、摘要和文本等外部内容,并不意味着生成性学习必然发生。外部学习成果的生成不能替代学习者认知结构的塑造,关键在于人工智能究竟促进还是替代了学习者必要的认知加工。

生成式人工智能可能通过两条相反的路径影响生成性学习。第一,当人工智能作为辅助学习的“教练”时,它通过解释概念、提供提示、反馈错误和回应追问,为学生搭建认知脚手架,激发学生的好奇心和探索意愿,帮助其建立知识联系、修正和加深理解,并完成超出原有能力的学习任务。已有实验研究发现,嵌入学习者保护机制并依据教学原则设计的人工智能辅助工具,能提高学生的数学和物理学习效果(Kestin et al., 2025)。第二,当人工智能作为学生依赖的认知“拐杖”时,情况则完全不同。学生可以直接让人工智能筛选信息、总结阅读材料、完成推理或生成文本。基于中国学生30个月追踪数据的研究表明,人工智能使用提高了学生的作业成绩并缩短了完成作业的时间,但大多数学生表现出“作业外包”的特征,其闭卷考试成绩反而随着人工智能使用时间的增加而下降(Strömberg, 2026)。可见,与生成式人工智能进行互动,人工智能既可以通过解释、提示和反馈支持学生思考,也可以通过直接提供答案替代学生思考。现有研究多聚焦特定人工智能工具、特定课程和短期实验情境,对于普通学生日常学习使用人工智能的频率、目的及其与不同学科学业表现之间的关系,仍缺乏大规模、系统性的证据。

基于上述生成式人工智能在学习中可能发挥双重认知功能的分析,以及现有研究的证据缺口,PISA 2025 学生问卷纳入人工智能聊天机器人学习使用的相关问题,来自 91 个国家和经济体的 76 万余名 15 岁学生参加了测试,这为考察学生日常学习中人工智能使用与认知能力表现之间的关系提供了依据。本研究基于中国四省(市)数据,试图回答一个核心问题:当以人工智能聊天机器人为代表的生成式人工智能进入学生日常学习后,它究竟是帮助学生形成认知能力,还是在替代形成这些能力所必需的认知活动?

二、分析框架与研究设计

(一)分析维度与研究问题

生成式人工智能只有在特定的条件下才能促进生成性学习,进而推动学生的知识建构与能力发展。其学习效应并非由技术使用本身决定,而是取决于人工智能介入学习的程度、在学习过程中承担的认知功能,以及与学习者特征之间的相互作用。PISA 2025 综合考虑了上述关联因素,围绕人工智能使用与学业成绩的关联,对学生使用人工智能聊天机器人开展学习活动进行了调查,分别从使用强度、认知功能和学生异质性维度考察并回答三个研究问题(见图 1):

第一,使用强度。人工智能使用与成绩是线性关系,还是存在适度区间和高频使用风险?适度使用人工智能可能为学生提供及时的解释与反馈,但随着使用频率的提高,人工智能在缩短学习任务完成时间的同时,也可能更多地替代学生独立阅读、练习和问题解决,挤占能力形成所必需的认知投入。经济合作与发展组织基于 PISA 2022 测试数据的研究发现,适度使用数字资源工具与较高的数学成绩相关,但随着每日使用时间的延长,这种正向关

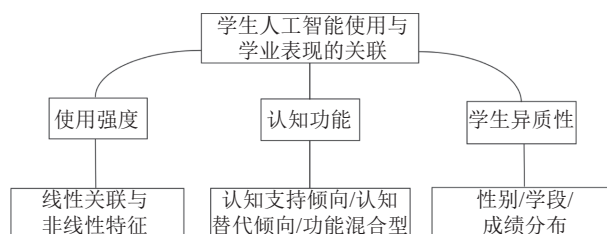


图 1 PISA 2025 学生人工智能使用分析框架

系会逐渐减弱并转为负向(OECD, 2023a)。因而, 人工智能使用强度与学业成绩是否同样呈非线性关系, 适度使用的潜在收益是否会在高频使用下转化为风险, 仍有待探讨。

第二, 认知功能。人工智能是促进学生加工信息, 还是在替代学生加工信息? 认知卸载理论认为, 个体可以借助外部工具改变任务的信息加工要求, 降低自身承担的认知负荷。在教育情境中, 其作用取决于被卸载的认知要求相对于学习目标是否必要, 以及节省的认知资源能否被重新投入有意义的学习活动(Risko & Gilbert, 2016)。据此, 判断人工智能引发认知卸载的教育影响, 关键不在于它是否降低了学生的认知负担, 而在于它卸载了何种认知活动, 以及这种活动是否构成学习目标本身。不同学习活动对信息选择、组织、整合与表达要求不同。PISA 2025 分别调查了学生使用人工智能聊天机器人帮助学习、初步研究新主题、总结阅读材料和起草写作类作业的频率。结合这些活动涉及的核心认知过程, 四类人工智能使用目的可以被区分为认知支持、认知替代和功能混合型。其中, “帮助学习”更具认知支持倾向, “总结阅读材料”更具认知替代倾向, “初步研究新主题”和“起草写作类作业”兼具支持与替代的可能。基此, 本研究比较不同认知功能取向的人工智能使用与学生学业成绩之间是否呈现差异化关联。

第三, 学生异质性。相同程度的人工智能使用是否与不同性别、学段和成绩水平学生的学业表现呈现不同关联? 同一教学干预不会对所有学习者产生同一效果, 其作用取决于教学方式与学习者能力倾向、知识基础及其他个体特征之间的匹配(Cronbach & Snow, 1977)。人工智能提供的外部支持也可能因学生的能力基础和发展阶段不同而发挥不同作用: 既可能弥补学生的知识和能力不足, 也可能替代学生原本能够独立完成的认知活动。已有研究发现, 生成式人工智能带来的学业惩罚在男生、低年级学生和成绩较高的学生中更明显(Strömberg, 2026)。鉴于此, 本研究比较人工智能使用与成绩的关系是否因性别和学段而异, 以及这种关系是否随学生处于不同成绩水平而发生变化, 从而揭示人工智能使用与成绩关联的群体差异和

分布特征。

(二)数据、变量与分析方法

1. 数据来源与变量选取

本研究数据源于 PISA 2025 中国四省(市)15 岁在校学生调查数据。PISA 采用分层两阶段抽样设计, 先抽取学校, 再从样本学校中抽取符合年龄要求的学生。样本涵盖北京、上海、江苏和浙江四省(市)的 374 所学校和 12,385 名学生。由于不同变量的有效作答数存在差异, 有效样本介于 12,241—12,360 人之间。根据上述研究问题, 本研究纳入分析的变量如下:

第一, 学业表现。本研究将学生的 PISA 2025 科学、数学和阅读素养成绩作为因变量。每个测评领域均采用 10 个似真值估计学生成绩, 每个似真值都分别估计模型, 再合并 10 次估计结果与相应方差。

第二, 人工智能使用。依据前述分析框架, 本研究从“使用强度”和“认知功能取向”维度测量和反映学生的人工智能使用情况, 并将相关指标作为核心自变量纳入分析。首先是人工智能使用强度。问卷调查了学生使用人工智能聊天机器人完成四类学校学习活动的频率, 分别为: 1) 总结一篇所阅读的文章, 2) 初步研究一个新主题, 3) 起草写作类作业的电子文档, 4) 帮助学习; 回答选项为“从来没有或几乎没有”“每年约一或两次”“每月约一或两次”“每周约一或两次”和“每天或几乎每天”。PISA 根据上述四个题项构建人工智能使用指数(变量名: AIUSESCH)以衡量学生使用人工智能完成学习活动的总体强度。该指数以经济合作与发展组织成员国学生为参照进行标准化, 平均值为 0、标准差为 1; 指数值越高, 表示学生使用人工智能越频繁。同时, 为便于分析, 本研究依据四类使用目的中的最高使用频率形成四类分析性分组: 未使用组为四项均选择“从来没有或几乎没有”的学生; 低频使用组为最高频率仅达到“每年一或两次”的学生; 适度使用组为最高频率达到“每月一或两次”或“每周一或两次”的学生; 高频使用组为至少一项达到“每天或几乎每天”的学生。其次是人工智能使用的认知功能取向。本研究将上述四类使用目的每周使用情况分别作为自变量

纳入模型,并结合相应学习活动的认知加工要求加以区分:“帮助学习”具有较强的认知支持倾向,“总结阅读材料”具有较强的认知替代倾向,“初步研究新主题”和“起草写作类作业”则兼具认知支持与认知替代的可能。

第三,控制变量与异质性变量。基准模型控制学生的家庭经济、社会和文化地位指数(ESCS)、性别和学段。学生的家庭经济、社会和文化地位指数综合反映父母职业地位、父母受教育程度和家庭财力等信息;根据抽样资料编码,女生被编码为1、男生为0;高中为1、初中为0。在学生异质性分析中,性别和学段还分别与人工智能使用变量构成交互项,用于检验人工智能使用与学业成绩的关联是否存在群体差异。

2. 分析方法

本研究数据分析过程包括:使用学生最终权重 W_FSTUWT 计算人工智能使用指数均值、不同用途占比和使用频率分组比例,并将中国四省(市)与其他参测经济体进行比较;建立回归模型,依次估计人工智能使用指数的一次项模型和加入平方项的非线性模型;将四个用途的“每周及以上使用”指标同时纳入模型,估计在其他用途、父母职业地位和性别、学段相同条件下,各用途与成绩的关联;计算四类使用者在加权平均父母职业地位和样本性别构成、不同学段下的调整后成绩,直观呈现群体差异。由于四种用途可能同时发生,所有参数均使用学生最终权重进行估计。标准误采用 80 个 Fay 修正的平衡重复抽样权重进行估计,Fay 系数为 0.5;成绩模型对 10 个似真值分别运行,再依据多重估计规则合并抽样方差与可能值方差。PISA 复杂抽样下,普通回归标准误会低估抽样不确定性,因此不能采用常规独立同分布假设替代重复权重法(OECD, 2024)。

三、研究结论与主要发现

(一)中国四省(市)15 岁学生在校学习广泛接触人工智能,但总体使用强度偏低,用于认知支持取向相对突出

调查显示,82.3% 的中国学生至少使用人工智能完成了一种学习活动,这说明人工智能已较广泛

地进入学生学习过程,人工智能使用指数加权均值为-0.28(低于 OECD 平均值),在 85 个具有有效数据的参测经济体中由高到低居第 79 位,表明总体使用频率相对较低。从使用强度看,56.8% 的中国学生属于适度使用者,即在四类使用人工智能的学习活动中,至少一种使用达到“每月”或“每周使用”,但没有达到“每天使用”。以 85 个具有可比数据的参测经济体为参照,四种用途均为“从不或几乎从不使用”的学生占 17.7%,位居第 8 位;至少一种用途达到每周使用的学生占 45.0%,居第 78 位;至少一种用途达到“每天或几乎每天”使用的学生占 10.8%,居第 83 位。上述结果说明,人工智能已进入中国相当一部分学生的学习,但尚未成为多数学生每天依赖的常规工具。

不同用途的使用频率存在明显差异。帮助学习是最常见的,每周至少使用一次的学生占 40.7%,明显高于起草写作类作业(20.6%)、总结阅读材料(19.0%)和初步研究新主题(17.1%);其每日使用比例为 9.8%,其余三类用途均不足 4%(见表 1)。从用途结构看,学生使用人工智能总体偏向认知支持,即倾向于借助人工智能获取解释、提示和反馈,而非持续利用人工智能总结材料或生成文本。需注意的是,使用目的不完全等同于促进认知,帮助学习也可能包括直接索取答案;同时,有 10.8% 的学生至少在某方面每天使用人工智能。因此,人工智能究竟发挥认知支持还是认知替代作用,需结合不同使用强度、使用目的与学业成绩的关联加以判断。

(二)人工智能使用强度与学业表现呈先升后降的非线性关联,且存在学科差异

在线性模型中,人工智能使用指数每提高 1 个单位,科学、数学和阅读成绩分别降低 3.98 分、5.98 分和 5.58 分(见表 2)。在加入平方项后,三个领域的平方项系数均显著为负,表明人工智能使用强度与成绩之间不是简单的线性关系,而呈向下弯曲的形态:在较低使用频率区间,使用强度提高与预测成绩上升相关;超过曲线最高点后,继续提高使用强度则与预测成绩下降相关,且负向关联逐渐增强。科学、数学和阅读拟合曲线的最高点分别位于 $AIUSESCH=-0.64$ 、 -0.98 和 -1.19 处(见图 2)。这表明,人工智能使用强度增强后,其与成绩的关

表 1 人工智能使用用途频率分布(加权占比均值)

人工智能使用用途	从不/ 几乎从不(%)	每年 1—2 次 (%)	每月 1—2 次 (%)	每周 1—2 次 (%)	每天/ 几乎每天(%)	每周至少 一次(%)
总结一篇所阅读的文章	37.6	20.8	22.6	15.7	3.3	19.0
初步研究一个新主题	40.7	20.6	21.5	14.0	3.1	17.1
起草写作类作业的电子文档	35.3	19.5	24.6	17.1	3.5	20.6
帮助学习	23.0	14.5	21.8	30.9	9.8	40.7

注:各用途有效样本数为 12,337—12,360,“每周至少一次”为“每周 1—2 次”与“每天或几乎每天”之和。

表 2 人工智能使用指数与学业表现的回归结果

变量	科学线性	科学非线性	数学线性	数学非线性	阅读线性	阅读非线性
AIUSESCH	-3.98**	-5.03***	-5.98***	-6.92***	-5.58***	-6.28***
AIUSESCH 平方	—	-3.91***	—	-3.55**	—	-2.63**
拟合曲线的最高点		-0.64		-0.98		-1.19

注:控制学生家庭社会经济地位背景、性别和学段;* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

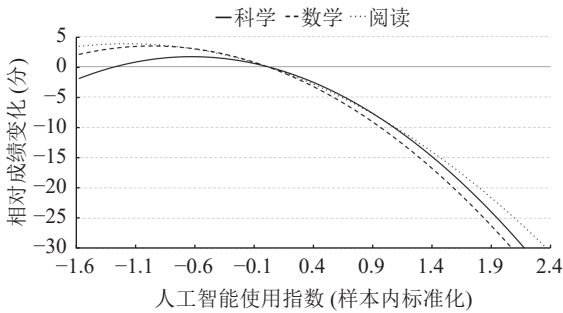


图 2 人工智能使用强度与学业成绩的非线性关联

联在阅读领域最早由正转负,数学次之,科学最晚。与此同时,三科平方项系数分别为-3.91、-3.55和-2.63,表明科学曲线向下弯曲的程度最大,数学次之,阅读相对较小。换言之,科学领域虽然较晚越过最高点,但越过之后,其负向关联随使用强度的提高而增加得更快。上述结果表明,人工智能使用强度与三科素养表现的关联并非一致,学科差异体现在负向关联开始出现的使用强度,以及越过最高点后负向关联随使用强度提高而增强的速度上。

(三)人工智能介入学习的认知功能分化明显:认知支持倾向与较高成绩相关,认知替代倾向与较低成绩相关,且这种功能分化在数学和科学领域表现明显

为识别人工智能介入学习所承担的认知功能,本研究重点关注学生每周至少一次使用人工智能完成不同学习活动的情况。首先,每周使用表明具

有一定的持续性,较能反映人工智能能否稳定介入学生的信息加工过程;月度及以下使用可能表明只是偶尔求助,难以形成相对稳定的认知支持或认知替代模式。其次,四类用途每周至少使用一次的学生占比为 17.1%—40.7%,具有分析意义。

第一,认知支持倾向与较高成绩稳定相关。控制其他三类用途的人工智能使用及学生家庭社会经济背景、性别和学段后,每周至少一次使用人工智能“帮助学习”的学生,科学、数学和阅读成绩比每周低于一次的学生分别高 22.06 分、22.47 分和 18.85 分。这表明,若人工智能用于解释概念、提示步骤和反馈过程等保留学生主动加工的认知活动,可能有助于发挥认知支架的作用。第二,认知替代倾向与较低成绩稳定相关。每周至少一次使用人工智能“总结阅读材料”的学生,三科成绩比每周低于一次的学生分别低 19.07 分、20.30 分和 15.92 分。总结是学生选择、组织并整合信息获得认知发展的重要环节,若由人工智能完成,则可能造成认知卸载。第三,关联程度因学科而异,认知支持或替代倾向不明确的人工智能用途与成绩的关联不一致。“帮助学习”的正向关联和“总结材料”的负向关联在数学、科学领域均大于阅读,表明人工智能的支持与替代作用可能在多步推理和证据整合要求较高的学习中同时被放大。“初步研究新主题”仅与阅读成绩显著负相关,“起

草写作类作业”与三科成绩均无显著关联,反映两者可能兼具支持和替代功能(见表3)。

表3 人工智能使用用途与学业成绩的关联

变量	科学成绩	数学成绩	阅读成绩
每周至少一次:总结阅读材料	-19.07***	-20.30***	-15.92***
每周至少一次:初步研究新主题	-4.09	-6.96	-6.62*
每周至少一次:起草写作类作业	1.44	0.05	0.67
每周至少一次:帮助学习	22.06***	22.47***	18.85***

注:用途虚拟变量回归模型四个用途的“每周至少一次”指标同时进入模型,参照组为该用途低于每周一次;模型控制学生家庭社会经济地位背景、性别和学段;* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$ 。

(四)人工智能使用与学生成绩的负向关联在中高成绩学生中更明显,关联程度因学科而异。

本研究从学生性别、学段和成绩分布位置分析人工智能使用与学业表现关联的异质性。交互项分析未发现该关联因性别或学段而存在显著差异。相比之下,分位数回归结果显示,不同成绩分布位置的关联强度及其非线性特征存在差异。

为分析人工智能使用与学业表现的关联是否因学生成绩位置而异,本研究采用分位数回归,分别估计成绩分布第25、50和75百分位点(P25、P50和P75)的结果。其中,P25、P50和P75分别代表在控制家庭社会经济背景、性别和学段等因素后,学生成绩分布在较低、中间和较高的位置。

第一,人工智能使用强度与中高成绩位置的负向关联明显,较高成绩位置的学生不意味着能够免于潜在人工智能使用的学业惩罚。总体来看,如表4所示,在不含平方项的分位数回归模型中,人工智能使用指数与三科成绩的回归系数均为负,且系数绝对值由P25至P50、P75依次增大。这表明,人工智能使用强度与成绩的负向关联并非仅集中于成绩分布较低位置的学生,而处于中高成绩位置的学生同样未表现对潜在人工智能学业惩罚的抵御能力,其原有认知优势甚至可能受到明显削弱。

第二,在科学学习中,高强度的人工智能使用与成绩之间加速增强的负向关联体现在各个成绩水平;在数学和阅读中,这一特征分别体现在成绩低和高的学生群体中(见表4)。科学成绩呈明显的非线性特征,P25、P50和P75的平方项均显著为负,表明超过各自拟合曲线最高点后,人工智能使用强

表4 人工智能使用强度与学业成绩百分位分布的关联

学习领域	变量	P25	P50	P75
科学	AIUSESCH	-3.19	-7.03***	-7.62***
	AIUSESCH 平方	-4.32**	-2.42*	-3.35**
	拟合曲线的最高点	-0.37	-1.45	-1.14
数学	AIUSESCH	-5.09*	-9.09***	-9.15***
	AIUSESCH 平方	-3.86*	-1.82	-2.53
	拟合曲线的最高点	-0.66	-2.49	-1.81
阅读	AIUSESCH	-5.20**	-7.33***	-7.97***
	AIUSESCH 平方	-2.42	-1.76	-2.33*
	拟合曲线的最高点	-1.08	-2.08	-1.71

注:分位数回归模型控制学生家庭社会经济地位背景(ESCS)、性别和学段;* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$ 。

度越高,与成绩的负向关联越强,且贯穿科学成绩分布。相比之下,数学的非线性特征主要集中于较低成绩位置,阅读则主要集中于较高成绩位置,分别呈“低位成绩集中型”和“高位成绩集中型”特征。

第三,数学学习的潜在人工智能学业惩罚主要集中于较低成绩位置的学生,而中高成绩位置呈现较稳定的线性负相关。如表4所示,仅P25的平方项显著为负,表明数学成绩低的学生对增强使用人工智能的强度更敏感,超过拟合曲线最高点后,其负向关联呈加速趋势。因而,数学教学应重点关注成绩低学生的人工智能使用强度,避免因高强度使用人工智能增加其学习劣势。

第四,阅读学习中潜在人工智能学业惩罚的加速风险集中于较高成绩位置,较低和中间成绩位置呈较稳定的线性负向关联(见表4)。仅P75的平方项显著为负,表明处于阅读成绩分布较高位置的学生对人工智能使用强度提高更敏感,超过拟合曲线最高点后,其负向关联呈加速趋势。因而,在规范学生阅读学习中人工智能使用的同时,应特别关注阅读成绩较好学生的使用频率,避免频繁借助人工智能替代深度理解、跨文本整合和批判性评价,削弱其原有的阅读优势。

四、结论、反思与建议

上述研究表明,中国四省(市)15岁学生尚未普遍高频使用以聊天机器人为代表的生成式人工智能(如DeepSeek、豆包等)。超过半数的学生属

于适度使用者,高频使用者仅占 10.8%;未使用者和适度使用者的科学、数学和阅读成绩未呈现显著差异。总体来看,调查尚未发现学生普遍高频依赖人工智能并伴随成绩明显下降现象。但要警惕的是,基于 PISA 2025 数据,从各国及地区的总体情况看,人工智能学业惩罚的潜在风险多集中于高频使用者。随着人工智能日益深入学生学习,对使用强度的持续监测仍是防范认知依赖的基本防线。

然而,值得关注的是,中国四省(市)学生人工智能使用与学业表现的关联已呈现结构性差异:人工智能介入学习所承担的认知功能、与成绩关联的方向及其程度,均因学科学习特点和成绩分布位置而异。这意味着,判断人工智能能否促进学生学习,不能仅依据学生使用人工智能的频率,更要具体分析人工智能作为提升学习力的工具对学习过程的影响,尤其是与学习目标一致的认知功能设定,依据学科核心认知过程设计介入顺序,以及基于学生能力水平差异调整支持方式和强度等。概言之,学校人工智能治理应由单一的使用频率管控转向动态监测与介入过程设计并重,通过分学科、分群体的差异化引导,推动人工智能规范而有效地赋能学生学习。

(一)坚守人工智能应用的教育属性,以学习目标引领其使用功能定位、介入顺序和使用边界,以评价与监测强化过程调控,防止陷入“效率至上”的使用逻辑

本研究发现,人工智能使用强度与科学、数学和阅读成绩均呈先升后降的非线性关联。从用途看,“帮助学习”与较高成绩的关联贯穿三科,使用人工智能直接“总结阅读材料”均与较低成绩相关。这表明,评价人工智能的学习价值应以能否促进真实学习为标准,而不能仅看任务完成效率或产出质量,即借助人工智能提高任务表现不等于学生真正学会(OECD, 2026a)。防止这一价值定位发生偏移的措施包括:首先,应依据课程标准、学段特点和学生发展水平,明确学生应发展的核心能力,划分“学生独立完成—人工智能提供支架—人工智能适度代理”的任务边界,并将是否具有明确的教学增值作为人工智能介入的基本标准(日本文部科学省, 2024; 法国教育部, 2025);其次,应建立以

过程性评价为基础的持续监测与调控机制,在学习任务和评价中明确人工智能可以介入和不得替代的核心环节(澳大利亚教育部, 2023),将学生的初稿、对话记录、提示词、修改依据和核验过程作为评价证据,并设置必要的无人工智能独立迁移任务,以区分“借助人工智能完成得好”与“学生真正学会”,确保人工智能使用最终服务于学习力发展。

(二)人工智能介入学习越深入,越应坚持基础能力优先发展:重新定义人工智能时代的阅读能力,筑牢跨学科学习的认知底座

PISA 2025 调查显示,在可监测长期表现的 23 个国家中,15 岁学生的平均阅读成绩较 2022 年下降约 16 分;此前十年, PISA 阅读成绩已呈下降趋势(OECD, 2023b)。这一趋势早于生成式人工智能的大规模普及,因此不能简单归因于使用人工智能工具。但生成式人工智能能直接完成信息提取、文本概括和内容整合,可能进一步减少学生自主加工文本的机会,成为放大既有阅读危机的新变量。本研究为认识阅读这一跨学科基础能力面临的潜在冲击提供了新证据。中国四省(市)学生使用人工智能“总结阅读材料”与数学、科学成绩负向关联程度反而大于与阅读成绩的关联程度,表明信息选择、结构识别和意义整合不仅是阅读活动,也是理解数学情境、提取问题条件和开展科学证据推理的重要基础。经济合作与发展组织早就将阅读界定为“建构和验证知识”的过程,强调辨别事实与观点、识别偏见、整合多源信息和处理复杂性,凸显了阅读的跨学科属性(OECD, 2021)。如果学生在形成这些基础能力前借助人工智能绕过必要的“认知挣扎”,就可能损害深度阅读、持续注意和独立推理能力的发展,相关风险还可能累积并延伸至数学和科学学习。为此,新加坡、瑞典等国家降低低龄学生人工智能使用或屏幕暴露时间、加强成人监督,以及增加学校阅读时间,倡导阅读纸质教材和图书馆资源,优先保障读写、计算与推理等基础能力得到发展。

(三)人工智能介入学习不宜采用统一模式,应依据学科核心认知过程设计介入顺序,实施分科分层引导

本研究发现,潜在人工智能学业惩罚既非“学

科中性”,也非“群体均质”:科学领域的非线性风险呈现“全成绩位置覆盖”特征,数学的成绩加速下降主要集中于低成绩位置学生,阅读主要集中于较高成绩位置学生。这表明,相同强度的人工智能介入可能在不同学科、不同能力水平学生中引发不同风险,学校不能将一套使用规范用于不同学习任务和学生群体。PISA 历次评估框架对三个领域核心认知过程的界定,为差异化设计提供了依据:数学学习遵循“形成数学表征—运用数学知识—解释与评价结果”的问题解决过程,科学学习强调“解释现象—设计与评价探究—解释数据和证据—评价科学信息”,阅读包括“定位信息—理解文本—整合信息—评价与反思”(OECD, 2019、2023a、2026)。据此,本研究建议:在科学学习中,应覆盖不同成绩位置的学生,保护提出问题、形成解释和评价证据等核心环节,坚持“学生形成假设与初步解释—人工智能提供模拟、追问或反例—学生依据证据核验并作出结论”的路径;数学学习应重点保护低成绩位置学生的情境理解、数学表征和步骤推理,坚持“学生提取条件并尝试求解—人工智能提供分级提示或局部示例—学生独立完成并检验”的路径,避免人工智能直接生成解答;阅读学习应防止人工智能替代较高水平学生的深度理解、跨文本整合和批判性评价,将人工智能主要用于提供不同观点、质疑既有解释和辅助来源核验。学校由此应建立“学科—任务—能力水平”三维介入规则,重点明确人工智能在哪个环节介入、提供何种程度的支持及何时退出等规范。

(四)人工智能支持学习不宜止于提供认知支架,应依据学生既有水平动态调节支持强度与认知挑战,守护并拓展中高成绩水平学生的高阶思维和自主学习空间

本研究发现,在科学、数学和阅读三个领域,人工智能使用强度与成绩负向关联的程度随成绩分布位置提高而增大。这意味着,潜在人工智能学业惩罚并非仅集中于低成绩位置,中高成绩位置学生面临的潜在认知替代风险可能更突出。专长反转效应(expertise reversal effect)可为这一结果提供理论解释:对初学者有效的指导,可能随着学习者知识和经验增长而成为冗余信息,增加额外认知负

荷,进而妨碍自主加工与问题解决(Kalyuga, 2007)。因此,较高成绩水平的学生不能天然抵御人工智能潜在的学业惩罚。对这类学生,增加支持反而可能降低任务挑战,压缩自主探索和高阶思考的空间。人工智能应用个性化不等同于提供更多帮助,而应实现“低位增支架、高位增挑战”:对较低成绩位置的学生,教师应借助人工智能提供分步提示、基础解释和即时反馈,防止支架异化为答案代理;对中高成绩位置的学生,应减少直接解答和程序化提示,提高任务的开放性、复杂性与迁移要求,将人工智能用于提出反例、比较竞争性解释、质疑论证和检验观点,使其由答案提供者变为认知挑战者。

[参考文献]

- [1] 澳大利亚教育部(2023). Australian Framework for Generative Artificial Intelligence (AI) in Schools. [EB/OL]. [2026-09-03]. <https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools>.
- [2] Bastani, H., Bastani O., Sungu A., Ge H., Kabakcı Ö. & Mariman, R. (2025), “Generative AI without guardrails can harm learning: Evidence from high school mathematics”, *Proceedings of the National Academy of Sciences*, Vol. 122/26, e2422633122, <https://doi.org/10.1073/pnas.2422633122>.
- [3] Brynjolfsson, E., Li D. & Raymond L. (2025), “Generative AI at work”, *The Quarterly Journal of Economics*, Vol. 140/2, pp. 889-942, <https://doi.org/10.1093/qje/qjae044>.
- [4] Cronbach, L. J. & Snow R. E. (1977). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions* [M]. New York: Irvington Publishers.
- [5] 法国教育部(2025). Cadre d’usage de l’IA en éducation. [EB/OL]. [2026-09-03]. <https://www.education.gouv.fr/cadre-d-usage-de-l-ia-en-education-450647>.
- [6] Fiorella, L. & Mayer R. E. (2015). *Learning as a Generative Activity: Eight Learning Strategies That Promote Understanding* [M]. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781107707085>.
- [7] 教育部等五部门(2026). 教育部等五部门关于印发《“人工智能+教育”行动计划》的通知 [EB/OL]. [2026-09-03]. https://www.moe.gov.cn/srcsite/A16/s3342/202604/t20260410_1433240.html.
- [8] Kalyuga, S. (2007), “Expertise reversal effect and its implications for learner-tailored instruction”, *Educational Psychology Review*, Vol. 19, pp. 509-539, <https://doi.org/10.1007/s10648-007-9054-3>.
- [9] Kestin, G., Miller K., Klales A., Milbourne T. & Ponti G. (2025), “AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting”, *Scientific Reports*, Vol. 15, Article 17458, <https://doi.org/10.1038/s41598-025-97652-6>.
- [10] Noy, S. & W. Zhang (2023), “Experimental evidence on the

productivity effects of generative artificial intelligence”, *Science*, Vol. 381/6654, pp. 187-192, <https://doi.org/10.1126/science.adh2586>.

[11] OECD (2019). *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing. [EB/OL]. [2026-09-03]. <https://doi.org/10.1787/b25efab8-en>.

[12] OECD (2021). *21st-Century Readers: Developing Literacy Skills in a Digital World*. Paris: OECD Publishing. [EB/OL]. [2026-09-03]. <https://doi.org/10.1787/a83d84cb-en>.

[13] OECD (2023a). *PISA 2022 Assessment and Analytical Framework*. Paris: OECD Publishing. [EB/OL]. [2026-09-03]. <https://doi.org/10.1787/dfe0bf9c-en>.

[14] OECD (2023b). *PISA 2022 Results (Volume II): Learning During and From Disruption*. Paris: OECD Publishing. [EB/OL]. [2026-09-03]. <https://doi.org/10.1787/a97db61c-en>.

[15] OECD (2024). *PISA 2022 Technical Report*. Paris: OECD Publishing. [EB/OL]. [2026-09-03]. <https://doi.org/10.1787/01820d6d-en>.

[16] OECD (2026a). *OECD Digital Education Outlook 2026: Ex-*

ploring Effective Uses of Generative AI in Education. Paris: OECD Publishing. [EB/OL]. [2026-09-03]. <https://doi.org/10.1787/062a7394-en>.

[17] OECD (2026b). *PISA 2025 Assessment and Analytical Framework*. Paris: OECD Publishing. [EB/OL]. [2026-09-03]. <https://doi.org/10.1787/86c36975-en>.

[18] 日本文部科学省(2024). *Guideline for the Use of Generative AI in Primary and Secondary Education (Ver. 2.0)*. [EB/OL]. [2026-09-03]. https://www.mext.go.jp/a_menu/other/mext_02412.html.

[19] Risko, E. F. & Gilbert S. J. (2016), “Cognitive offloading”, *Trends in Cognitive Sciences*, Vol. 20/9, pp. 676-688, <https://doi.org/10.1016/j.tics.2016.07.002>.

[20] Strömberg, D., Lei V. & Wu Y. (2026). *The Generative AI Learning Penalty: Evidence from Chinese secondary education*. CEPR Discussion Paper No. 21577. Paris and London: CEPR Press. [EB/OL]. [2026-09-03]. <https://cepr.org/publications/dp21577>.

(编辑: 赵晓丽)

More Generative AI Use, Better Learning? Evidence from 15-Year-Old Students in Four Chinese Provincial-Level Regions in PISA 2025

XU Jinjie¹ & Andreas Schleicher²

(1. *International Center for Teacher Education / UNESCO Teacher Education Centre, Shanghai Normal University, Shanghai 200234, China*; 2. *Directorate for Education and Skills, OECD, Paris 75775, France*)

Abstract: *The OECD recently released the PISA 2025 results, which cover more than 760,000 15-year-old students across 91 countries and economies, including approximately 12,000 students from Beijing, Shanghai, Jiangsu, and Zhejiang (B-S-J-Z) in China. For the first time, PISA 2025 surveyed students' use of generative artificial intelligence (GenAI), particularly AI chatbots, in school learning activities, providing large-scale, cross-disciplinary evidence for examining whether—and under what conditions—GenAI may support student learning. Drawing on PISA 2025 student data from B-S-J-Z (China), this study investigates how the intensity and cognitive functions of GenAI use, together with student heterogeneity, are associated with performance in science, mathematics, and reading. The findings do not reveal a general pattern in which Chinese 15-year-olds rely on AI frequently and extensively while experiencing marked declines in achievement. Nevertheless, the association between GenAI use and academic performance already exhibits clear structural variation across subject domains and students at different achievement levels. Accordingly, school governance of AI should move beyond a singular focus on regulating frequency of use toward an approach that combines dynamic monitoring with deliberate design of when and how AI enters the learning process, supported by differentiated guidance across subjects and student groups.*

Key words: *OECD; PISA 2025; generative artificial intelligence; B-S-J-Z (China)*