

人工智能能打破课堂沉默吗？

——以大学英语课堂会话为例

张欢¹, 李家齐²

¹上海外国语大学贤达经济人文学院外语学院, 上海

²上海外国语大学贤达经济人文学院数据科学院, 上海

收稿日期: 2025年3月11日; 录用日期: 2025年4月10日; 发布日期: 2025年4月18日

摘要

研究目的: 在大学英语课堂上, 课堂沉默现象较为普遍, 这通常是由教师提出的抽象问题以及学生知识储备不足所导致的。然而, 国内相关研究只关注课堂活跃氛围, 而忽略了沉默打破是否有效促进了学生能力的提高。为了有效打破沉默, 提升课堂互动效果以及学生的语言和认知能力, 本研究将GPT引入到两个英语班级中, 以辅助学生理解、回答教师提问。**研究方法:** 为了探究GPT是否能够、如何打破消极的课堂沉默(NCS), 本研究从中国一所民办高校中选取了两个班级, 分别作为实验组和控制组。为了衡量GPT对NCS的影响, 本研究从言语互动、语言能力和认知能力这三个角度进行了前测和两次后测, 采用语料库和Python算法进行测量。**研究结果:** 本研究发现, 打破沉默的效果应通过言语互动水平、技能、认知方面的提升来衡量。使用GPT的课堂和传统课堂都能够打破沉默, 但使用GPT的班级在言语表达的规范性、结构性、概括性、具体化方面表现更优, 语言质量提升明显, 而传统课堂在言语输出量方面表现更出色。当回答者存在对智能手机的行为成瘾问题、对问题准备不足、无法有效使用GPT的提示功能时, GPT的使用效果会大打折扣。

关键词

课堂沉默, 人工智能, 英语课, 大学生, 会话

Can GPT Break Classroom Silence?

—Case Study of a College English Course

Huan Zhang¹, Jiaqi Li²

¹Foreign Language Department, Xianda College of Economics & Humanities, Shanghai International Studies University, Shanghai

²School of Data Science, Xianda College of Economics & Humanities, Shanghai International Studies University, Shanghai

Received: Mar. 11th, 2025; accepted: Apr. 10th, 2025; published: Apr. 18th, 2025

文章引用: 张欢, 李家齐. 人工智能能打破课堂沉默吗? [J]. 教育进展, 2025, 15(4): 553-567.

DOI: 10.12677/ae.2025.154585

Abstract

Object: Classroom silence frequently occurs in college English courses. This phenomenon is often attributed to the teachers' abstract questions and the students' inadequate knowledge reserves. In order to effectively shatter this silence and enhance interaction as well as students' language and cognitive capabilities, this research incorporated GPT into two English classes. The aim was to help students better comprehend and respond to the teachers' questions. **Method:** With the objective of investigating whether GPT can break negative classroom silence (NCS) and the way it does so, this research selected two classes from a private university in China, where NCS is prevalent, to serve as the experimental group and the control group respectively. To assess the impact of GPT on NCS, this research carried out a pretest and two post-tests. These tests were conducted from three aspects: speech interaction, language ability, and cognitive ability, and were evaluated using corpus data and Python algorithms. **Result:** This research revealed that the effectiveness of breaking silence should be gauged not merely by increased speech interaction but also by the enhancement of skills and cognitive abilities. Both classes with the assistance of GPT and traditional classes are capable of breaking the silence. However, the GPT-assisted group demonstrated superior performance in terms of speech specification and drawing conclusions, while the traditional class had an advantage in the quantity of speech. GPT will prove to be ineffective when a respondent suffers from behavioral addiction to smartphones, lacks sufficient preparation for the questions, or is unable to use GPT prompts properly.

Keywords

Classroom Silence, Artificial Intelligence, English Class, College Students, Conversation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

英语教学中的一个典型问题是消极课堂沉默(NCS), 对学生、教师、课堂都会产生负面影响。由于多方面的原因, 这种沉默会对学生的学习积极性和教师的教学热情产生负面影响。言语交流和社交互动是课堂学习的两个重要元素[1] [2]。对于学生来说, 他们在“健谈程度”上的差异会明显影响其在课堂上的表现[3]。教师往往更倾向于回应那些健谈的学生, 并且认为沉默寡言的学生不如健谈的学生聪明, 学业表现也更差, 除非教师本身具有一定程度的内敛性格, 或者沉默的学生是女生[4]。此外, 对话者的沉默会引发学生对于自身知识储备不足的担忧, 担心自己不知道如何应对对话, 这会进一步加剧他们的焦虑和沉默。此外, 课堂沉默会让日本和英国的学生都感到不适, 这反映出课堂沉默在东西方文化中都具有普遍的负面影响[5]。

尽管许多研究已经探讨了课堂沉默的多种影响因素、教学方面的启示, 但关于通过有效打破沉默来提升学生认知能力和语言能力的实证研究, 还是少之又少。然而, 大多数同类研究都聚焦于课堂互动的改善、物质环境、教师的角色以及学生的学习态度, 但鲜有研究能够证明用于打破 NCS 的方法能够提升学生的数字能力、互动能力、认知能力和语言能力, 而这些能力在数字时代的语言教育中是至关重要的, 也是打破沉默的根本目的。

本文探讨的是如何通过打破课堂消极沉默, 来提升学生的认知能力、语言能力。文章首先通过对比实验组和对照组在互动、认知能力以及语言知识方面的情况, 探究成功打破课堂消极沉默的标准。接着, 文章重点关注生成式预训练变换器(GPT)技术是否能够弥补学生在回答策略方面的不足, 以及在应对抽象问题时知识储备方面的欠缺, 并且研究它如何提升学生的数字能力、认知能力、语言能力。随后, 通过对那些未能提升这些能力的学生展开研究, 来探讨 GPT 有效打破课堂消极沉默所需的条件。

2. 文献综述

焦虑与沉默被认为存在双向关系, 且有着认知行为方面的原因。沉默行为会加剧语言焦虑, 而焦虑的说话者则会保持沉默, 或者使用母语来避免使用第二语言[6]。语言焦虑的认知和行为原因是多方面的, 例如负面假设、令人担忧的预测、以自我为中心的形象、安全行为以及躯体和认知方面的问题[7]。当说话者因负面的言语和课堂预期而陷入沉默与焦虑的恶性循环时, 一种策略是进行反思活动, 帮助学习者重新构建对情境因素的负面认知, 并将其转化为建立自信的方式[8]。

语言沉默也会因内容和情境因素而出现, 并且学生在会话课程中比在阅读课程中更容易感到焦虑[9]。课堂上的消极沉默(NCS)常常是由学生害羞的个性以及充满压力的学校环境所导致的, 在这样的环境中, 他们面临着社会评价和学业要求[10]。为了使情境引发更少的焦虑, 教师应该根据学生的情感需求设计活动, 比如在语言课程中安排小组活动和游戏, 以避免公开表现和即兴发挥所带来的尴尬[11]。至于不熟悉的内容, 学生应该有机会在小组活动中发言之前对口头任务进行排练, 并且任务主题应该对学生有意义, 因为他们的学习目标往往具有功利性[9]。此外, 外语类课程中的一些语言学理论课, 容易因课上提到的晦涩概念, 而引起学生的沉默现象, 如微语篇、概念整合机制、翻译的历史等, 师生之间的反馈机制显得尤为重要[12]-[15]。

教师在课堂消极沉默(NCS)方面也起着一定作用。教师往往会更多地与健谈的学生互动, 并且认为沉默的学生不够聪明, 在学业上的表现会比健谈的同学差[4]。由于沉默的原因是多维度的, 教师不应仅仅将学生不尽如人意的表现归因于他们在能力、天赋和动机方面的不足[9]。此外, 教师的教学方法也很重要。教师可以通过教授焦虑管理策略来帮助焦虑的学生适应不熟悉的情境, 并逐步引入活动[9]。而且, 为了打破沉默, 教师往往会对健谈的学生采用社会学习策略, 对害羞的学生采用以同伴为中心的策略[2], 但学生能力、课堂表现、回应质量的提升以及教学效果并未得到检验。

学生对课堂沉默的态度也可能导致 NCS, 尤其是在高等教育中, 学生将沉默视为一种权利, 而非问题。对于在澳大利亚的“自我一代”的中国学生来说, 沉默被视为一种选择、权利和抵制[16]。根据他们的自我报告, 他们清楚自己选择沉默, 并且没有人将课堂沉默视为一个问题, 也不认为发言是批判性思维的必要因素。此外, 沉默可以是学生抵制教师的一种权利。许多中国学生敢于以尊重的方式对教师表达批判性判断, 比如通过沉默抗议[17], 部分原因在于互联网的存在, 在互联网环境下, 教师可能失去了知识来源的角色以及学生的敬仰。

基于上述引用的文献, 本研究认为, 打破 NCS 的价值和目的应该是提升学生的认知和能力, 而不仅仅是实现生动的课堂对话, 因为生动的课堂对话可能不能伴随学生能力的提升。对于比儿童更成熟的大学生来说, NCS 常常是由教师提出的抽象问题、学生对该问题的知识储备以及他们对课堂沉默的态度所导致的, 而非焦虑。目前, 同行研究的方法和启示可能对营造对话氛围以及解决学生在个性和态度方面的问题是可行的, 但对提升他们的认知和语言能力方面作用甚微。因此, 有必要探索一种方法, 通过打破课堂消极沉默来提升学生分析教师抽象问题的能力、学生的知识储备以及语言能力, 而这可以借助 GPT 技术的辅助。

3. 研究方法

3.1. 研究问题

上述文献综述表明,有必要探索一种既能有效打破课堂沉默,又能提升学生认知能力和语言能力的方法。为实现此目的,本研究子问题如下:

1. 评估打破课堂沉默效果的标准和指标应该是什么?
2. GPT 技术能否在提升学生认知能力和语言能力的同时,有效地打破课堂上的消极沉默(NCS)? 如何做到? 人工智能辅助教学的课程与传统课程之间有何区别?
3. 人工智能辅助教学的课程对哪类学生无效? 人工智能辅助教学课程取得成效的前提条件是什么?

3.2. 数据

本研究的参与者是中国一所私立高校的 66 名英语专业本科生,在这所高校中,课堂上的消极沉默(NCS)现象非常普遍,其中有 13 名学生在课堂对话中较为突出。参与者大多为女性,年龄在 18 岁至 21 岁之间。所有参与者被分为两个班级,组成了一个实验组和一个对照组。他们的课程内容涉及有关研究方法的英语会话,这些内容学生能够理解,但要在课堂上回答相关问题仍具有挑战性。在保证参与者匿名以及数据保密的情况下,四个自然班级的所有参与者都自愿同意参与本研究。

本研究的数据可以按照判断打破 NCS 有效性的标准进行分类,即:互动情况、认知能力和语言能力。表明课堂互动情况的数据包括沉默的频率、学生的回应以及教师的反馈,这些可以通过语料库中的标志性词语、冒号以及教师的名字来统计。反映认知能力的数据包括课堂对话中的词类、词汇密度以及代词,其多样性可以通过 Python 算法进行计算。词汇密度可以体现讲话内容的多样性,而代词则可以表明讲话者对语境的参照以及讲话记忆,并且内容密度和讲话记忆与认知能力高度相关。说明语言能力的的数据包括单词数量、句子数量以及平均句子长度,这些数据能够体现学生句法的规模和多样性以及表达能力。所有数据均通过一款名为微信 8.0.52 的中文应用程序的语音转文字功能进行记录。课堂上师生对话的语音随后以文本形式记录下来,并保存在微信中,以便研究人员课后下载并建立语料库。借助这些数据,本研究能够检测课堂沉默情况的变化以及学生认知能力和语言能力的提升。用于沉默评估的标准、指标和数据如表 1 所示:

Table 1. Criteria for classroom silence
表 1. 课堂沉默的评价标准

标准	指标	数据
互动情况	回应者数量 师生问答频率 教师发言频率 学生回答频率 学生问答占比 沉默频率	回应者的姓名; 冒号; 教师的姓名; why are you silent?这类标志性话语语料
语言能力	单词数量 句子数量 平均句子长度	单词数量; 句号; 句子数量与单词数量的比例
认知能力	词类 词汇密度 代词	词类; 词汇密度 = 词类数量/单词数量; “it、they、he、she、this、those、these、that”等这类标志性代词的出现频率

3.3. 研究流程

本研究的过程可分为四个步骤, 如图 1 所示: 1. 实验前的准备工作包括明确问题和进行文献综述。为了实现总体目标, 将研究题目细分为三个子问题, 如 3.1 中所述。此外, 本研究回顾了同行学者针对 NCS 的解决方案和启示, 并探究了本研究的独特必要性。2. 预测试阶段包括对两个班级在无 GPT 干预的传统教学方法指导下的课堂对话进行记录和分析。他们所有的对话都由教师提出的相同问题引导。为了模拟一般的课堂讨论, 教师挑选了 3 名积极回应者和 3 名在之前课程中倾向于保持沉默的消极回应者, 因为教师通常会选择优秀的学生回答问题, 并挑选少数沉默的学生以提醒他们专注于课堂。3. 在预测试之后, 本研究从三个角度, 通过分析学生的言语行为和对话语料库来观察沉默问题。在进行反思并收集学生通过调查给出的反馈后, 本研究设计了人工智能辅助教学法来干预 NCS。4. 在干预之后, 后测阶段通过语音转文字功能记录对话数据, 并将其存储在语料库中, 以便与预测试数据进行比较。当结果不理想时, 本研究将重新调整干预方法, 并重复进行后测, 以确保有效地打破 NCS, 其标志为活跃的课堂互动以及学生语言能力和认知能力的提升。研究流程概括如图 1 所示:

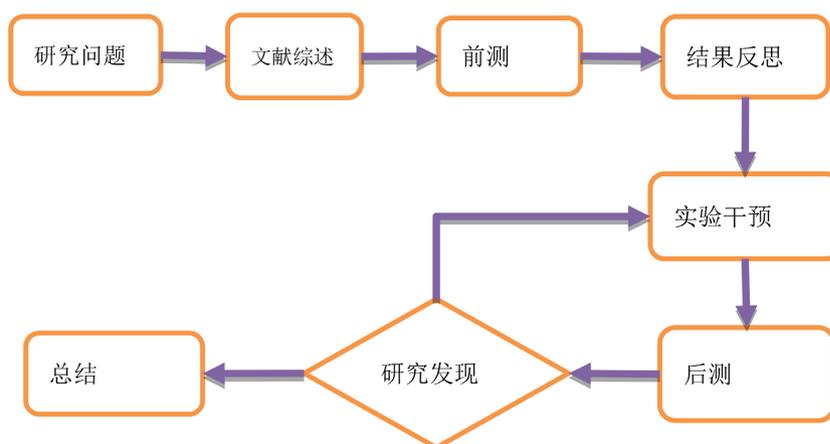


Figure 1. Research roadmap
图 1. 研究路线图

4. 研究发现

4.1. 前测：传统英语课堂的沉默

预测试中用于互动的教师提问:

“What’s your research topic, general question and sub-questions?”

In your questionnaire, which questions will your mention?

What’s the relation between these questions and your research topic?”

本研究选择了二班和三班作为实验组和对照组, 不同变量的对比情况如表 2 所示。

在表 2 中, 二班和三班的学生人数相同, 在课堂对话的互动情况、语言能力和认知能力方面存在细微差异。就课堂互动而言, 尽管两个班级在总体发言频率、教师和学生的发言频率上略有不同, 但学生发言频率在课堂发言频率中的占比仅相差 0.0007。其次, 两个班级在语言能力方面也略有不同。三班在课堂发言长度(单词数量)、词类数量和词汇密度方面总体上高于二班。尽管它们在发言长度上的差异可能较为明显, 但词汇密度仅相差 0.01。第三, 由词类数量、词汇密度和代词所体现的两个班级的认知能力差异更为显著, 且三班更具优势。

Table 2. Pretest data of Classes 2 and 3**表 2.** 二班和三班的预测试数据

标准	变量	二班	三班
	测试	前测	后测
	班级人数	32	32
互动性	回应者数量	6	7
	师生问答频率	92	107
	教师发言频率	67	78
	学生回答频率	25	29
	学生问答占比	0.2717	0.2710
	沉默频率	5	4
语言能力	课堂发言长度(单词数)	1526	1932
	句子数量	191	226
	平均句子长度	7.99	8.55
认知能力	词类数量	439	575
	词汇密度	0.288	0.298
	代词数量	932	1092

Table 3. Silence reasons and methods in pretest**表 3.** 预测试中的沉默原因及应对方法

沉默原因	教师采用的方法	说明
1. 翻译能力不足	允许使用母语	当一些回答者因翻译能力不足而无法回答问题, 但并非缺乏想法时, 教师允许他们使用母语来打破沉默。这种沉默类型的典型标志性用语是 “you can speak Chinese”。
2. 对问题理解不足	简化问题; 拆分问题; 使用幽默	当一些回答者因问题难度或自身知识储备不足而对问题理解不够时, 教师会用更简单的词汇和情景示例来简化问题, 或者将问题拆分成更小的问题, 引导学生逐步回答。这种情况的标志性用语有 “difficult” 和 “let me ask you simple questions” 之类的词。当使用这两种方法后, 一些回答者仍然无法回答时, 教师会使用幽默的问题来打破沉默。
3. 准备不充分	分析沉默原因	当一些回答者在课前或课中没有做好准备时, 在被提问后他们会保持沉默。在课堂上, 教师将问题准备和反应时间限制为 15 秒, 这对他们来说是不够的, 但更长的时间也会使问答节奏变慢。这种沉默原因是通过小组访谈诊断出来的, 典型的用语有 “you could have prepared better” 和 “why didn't you prepare before?”。
4. 注意力分散	更换回答者	当一些回答者被智能手机娱乐内容或观察前一位回答者的发言所分散注意力时, 他们会对教师的问题保持沉默。教师通常会直接将沉默的回答者更换为下一位来继续问答, 而不是重复问题的介绍和解释。这种沉默原因的标志性用语是 “你没有集中注意力”。
5. 性格内向	提出幽默且无关的问题; 更换回答者	一些性格内向的回答者坐在最后一排, 只回答几个字后就保持沉默。为了继续对话, 教师会提出一些简单且幽默的问题, 比如 “are you happy?”, “what do you like most?”, “what's your ideal career?” 等等。然而, 他们对幽默问题的回答往往是否定的, 比如 “no”, “I don't know”, “nothing”, “I'm not sure” 等等。他们带着微笑的回应表明他们理解了幽默。性格内向导致沉默的标志性用语是同时包含 “your silence” 和 “happy” 的语境, 因 why are you silent? 可以索引所有沉默的情况, 而 “are you happy?” 可以进一步筛选出教师提出的幽默问题, 这类问题通常用于引导性格内向的回答者继续作答。

显然, 二班比三班更加沉默。在 45 分钟内, 三班有 3 名学生主动举手回答教师的问题, 而有 4 名学生是被动点名回答的。二班有 6 名学生被点名回答问题, 且都是被动点名的, 其中 3 名学生回答流畅, 另外 3 名学生回答结结巴巴。为了避免学生资质和课堂随机点名对人工智能辅助教学课程效果的影响, 本研究选择二班作为实验组, 三班作为对照组, 并进一步与在预测试中回答问题的学生进行互动。如果一个资质相对较差的小组在 GPT 工具的帮助下能够取得比资质较好的小组更好的效果, 那么人工智能的效果就能得到更明显的证明。

根据课堂观察和学生的反馈, 本研究发现 NCS 是由五个原因导致的, 例如翻译能力不足、对问题的理解不够、课前准备不充分、课程开始时注意力分散以及性格内向。一些学生保持沉默是因为他们无法将自己的想法转化为英语; 尽管大多数同学都能理解问题, 但仍有少数学生由于对问题的理解不够和课前准备不充分而觉得难以回答; 一些学生无法回答问题是因为在教师介绍课程目标为提问做铺垫时, 他们在课程开始就被智能手机分散了注意力; 只有少数学生性格过于内向, 在被突然点名时无法回答问题。

在采用传统教学方式的两个班级中, 当出现 NCS 时, 教师采用了六种提问方法: 原因分析、问题引导、问题简化、幽默和无关问题、母语以及更换回答者。当出现 15 秒的沉默, 且回答者沉默并非是因为思考, 而是对问题理解不足时, 教师通常会询问回答者沉默的原因; 第二种方法是将一个问题分解成小问题, 以此引导学生逐步回答; 第三种方法是用非正式的语言简化问题的描述; 当使用了前三种方法后, 判断出回答者无法回答问题时, 教师会使用幽默和无关问题来缓解尴尬的气氛; 当回答者在思维上有了准备, 但在英语技能上遇到困难时, 教师通常会允许他用母语继续回答; 除了使用幽默的问题之外, 有时在多次提问后回答者仍无法回答时, 教师也会直接更换回答者。预测试中不同沉默情况的原因、方法、解释和标志性词语如表 3 所示。

4.2. 后测: GPT 对课堂沉默的影响

4.2.1. 后测 1: GPT 有效性的影响因素

后测 1 中用于互动的教师提问

“What might be the reasons for classroom silence?”

What might be the typical words in the context of classroom silence?”

Table 4. Pre and post test data of Classes 2 and 3

表 4. 2 和 3 班的前测和后测数据

评价标准	变量	Class 2	Class 2	Class 3	Class 3
		Pretest	Post1	Pretest	Post1
	测试				
	回答者数量	6	6	7	6
互动性	师生问答频率	92	140	107	154
	教师问答频率	67	99	78	88
	学生回答频率	25	41	29	66
	学生问答占比	0.2717	0.2929	0.2710	0.4286
	沉默频率	5	0	4	2
语言能力	班级会话篇幅(字数)	1526	2545	1932	2236
	会话句子数量	191	297	226	351
	平均句子长度	7.99	8.57	8.55	6.37
认知能力	词种类数	439	642	575	655
	词密度	0.288	0.252	0.298	0.293
	代词数量	932	1423	1092	1210

为确保其他变量不会影响 GPT 在实验组中的作用, 本研究控制了点名名单、样本的能力差异以及技术工具的使用。如果点名是随机的, 课堂表现可能更多地取决于学生的个人状况。换句话说, 如果教师挑选了 6 名准备不足或性格内向的学生, 那么原本活跃的班级在后测中可能会变得更加沉默。此外, 选择更沉默的班级作为实验组, 是因为如果在 GPT 辅助后, 这个班级的表现超过了原本更活跃的班级, 那么 GPT 的效果会更具说服力。最后, 对照组的学生被禁止使用手机应用程序来翻译或概括教师的问题内容, 以模拟传统课堂。前测和后测 1 的测试结果如表 4 所示。

表 4 显示, 二班(实验组)和三班(对照组)在课堂互动、语言技能和认知能力方面的表现均明显优于前测。除了共同的进步之外, 两个班级在这三个方面的差距也缩小了。

然而, 两个小组的进步很难体现出 GPT 工具对 NCS 的影响。因此, 为了进一步探究是哪些变量导致了这种共同的进步, 并阻碍了人工智能对课堂对话的作用, 本研究进一步组织了对两个小组特定点名名单上的学生进行小组访谈。小组访谈中的问题如表 5 所示:

Table 5. Post test 1 questions for group interview

表 5. 后测 1 的小组访谈提问

Experiment group (Class 2):
In this lesson, which benefits of GPT are obvious?
In this lesson, what limited the effect of GPT tool on classroom silence?
Control group (Class 3):
Compared with your performance in pretest lesson, is your performance in this lesson better, the same, or worse? Why?

人工智能在辅助课堂会话上的优缺点

在对实验组进行访谈后, 被点名的学生描述了 GPT 工具在解决课堂沉默问题上的优点和缺点。人工智能的优点包括: 提供思考方向、节省思考时间、辅助翻译过程、自动组织语言。然而, 人工智能工具也存在四个缺点: GPT 的回答复杂且难以理解、GPT 的回复不准确、回答准备时间不足、GPT 的回答标准化且重复、打字速度慢。具体来说, 当人工智能的回答过长或过于深奥时, 学生很难在短时间内理解其内容和主旨, 从而无法回答教师的问题。由于学生提问的方式未经训练, 人工智能的回答有时与学生的意图不符。回答准备时间不足, 因为使用 GPT 工具实际上延长了向人工智能输入问题的等待时间, 以及阅读、总结和记忆人工智能给出的答案的时间。此外, 当学生搜索相同的问题时, GPT 给出的答案是标准化的, 并且经常重复, 这使得后来的回答者由于超限效应, 在他们的想法被先回答的人说出来后, 不愿意再回答。在智能手机屏幕上打字速度慢也使得学生难以与人工智能进行互动。

打破沉默的原因

在对对照组进行访谈后, 被点名的学生描述了他们中一些人在后测课程中表现更好的原因。这些原因包括课前的英语口语练习、回答顺序、在公开演讲时的焦虑感降低、对话题的兴趣、良好的英语基础、学习态度、技术准备。具体来说, 一些回答者会定期进行诸如课前英语口语练习之类的衔接性活动, 所以他们额外的课后努力提高了他们的课堂表现。回答顺序和固定的点名名单让排在中间和后面顺序的学生有更多时间准备回答。此外, 有了之前的表现经验和充足的准备时间, 大多数回答者对公开演讲的敏感度和焦虑感降低, 能够更自然地表达自己的想法。第四, 一些积极的回答者因为对问题话题感兴趣而发言更多, 这与他们的个人经历有关。一些回答者有良好的英语基础和学习态度, 因为他们能从课程中感受到收获和兴趣。最后一个原因是更好的技术准备, 因为所有回答者都比前测时更熟悉之前安装的语音转文字应用程序。

变量的控制

在研究人员对两个回答者小组的反馈进行分析后, 发现两个小组表现更好的四个主要原因包括固定的点名名单、回答顺序、在公开演讲时的焦虑感降低以及对技术的熟悉程度。另一方面, 在后测 1 中限制 GPT 工具效果的四个主要因素包括 GPT 使用者的指令技巧、回答准备时间不足、人工智能回答的重复导致的超限效应、打字速度慢。下一部分将进一步研究如何控制这四个制约因素, 以便让 GPT 在后测 2 中更好地解决课堂沉默问题。

4.2.2. 后测 2: GPT 有效性的证据

后测 2 中用于互动的教师提问:

“Is there any difference in part of speech among different social groups’ daily conversation?”

Are girls better at using adjectives in daily conversation than boys?”

Are professors better at using nouns in daily conversation than farmers?”

制约性变量

根据后测 1 的数据和小组访谈, 本研究发现了四个制约人工智能在课堂会话发挥作用的, 如: GPT 使用者的指令技巧、回答准备时间不足、人工智能回答的重复性、打字速度慢。首先, 后测 1 中的大多数回答者没有专业的指令技巧, 这导致 GPT 给出的答案冗长、复杂且不准确。其次, 15 秒的等待时间对于回答者与 GPT 互动并组织答案来说是不够的, 因为输入指令问题、阅读 GPT 的答案以及记住人工智能答案的主旨, 使得一些回答者的等待时间从 15 秒延长到了 60 秒。第三, 当实验组中所有被点名的回答者向 GPT 提出相同的问题时, 他们从 GPT 得到的所有答案都是标准化的, 这使得后来的回答者不愿意重复先回答者的答案。第四, 所有回答者都只是通过在狭窄的智能手机屏幕上缓慢打字来与 GPT 互动, 这延长了等待时间。

控制变量

为了使人工智能的有效性不受这四个制约因素的影响, 本研究调整了课堂规则。为了避免 GPT 给出冗长、复杂和不准确的答案, 研究人员在后测 2 课程开始前, 对固定点名名单上的学生进行了 GPT 指令技巧培训, 通过教授指令模型和特定的指令语言, 来调整人工智能给出的冗长、复杂和重复的答案。第二, 为了减少等待时间并为回答者提供足够的准备时间, 本研究在课程开始时, 给所有回答者 5 分钟时间, 让他们同时利用 GPT 共同准备答案, 以避免突然点名和回答者的连续等待。第三, 为了避免超限效应, 所有回答者被允许并鼓励以积极的反馈回应先回答者重复过的观点。他们还接受了进一步向 GPT 获取创意观点的培训, 以避免回答的重复。第四, 通过鼓励学生使用语音转文字应用程序向 GPT 输入指令, 解决了打字速度慢的问题。在控制了这四个干扰人工智能发挥作用的, 因素后, 后测 2 的结果如下表 6 所示。

表 6 显示, 两个小组在后测 2 中的表现呈起伏变化模式, 并且差距缩小了。他们的起伏变化类似于儿童的语言发展, 儿童通常在童年时期话多, 而在青春期话少。根据研究人员对语料库的观察和小组访谈, 两个小组的回答者在后测 1 中随意说出的单词更多, 句子结构零碎, 但在后测 2 中, 他们更理性地说出较少的单词, 且结构清晰, 这体现在“first”, “second”, “third”等表述上。此外, 两个小组在词汇密度、学生发言量以及学生发言量与全班发言量的比率方面都有所提升, 这意味着回答频率降低了, 但回答质量提高了。

表 6 中两个小组在回答者数量、互动频率、全班发言长度和词汇密度方面差距的缩小, 能够反映出他们在对话技巧方面的同步提升, 但 GPT 组在发言规范性上表现更优。首先, 尽管两个小组在后测 2 中的互动频率都有所下降, 但回答者数量的减少和发言长度的增加表明, 每个回答者的发言比前两次测试

时更频繁、篇幅更长。其次, 两个小组在词类数量上都有明显进步, 尤其是 GPT 组, 在后测 2 中词类数量达到了 824。第三, GPT 组在发言结构、规范性和全面性方面也有所提升, 从表 7 中的标志性词语可以看出, 后两个方面优于对照组。

Table 6. Result of post-test 2

表 6. 后测 2 的结果

标准	变量	Class 2			Class 3		
	测试	preT	post1	post2	preT	post1	post2
互动性	回答者数量	6	6	4	7	6	4
	师生问答频率	92	140	83	107	154	92
	教师问答频率	67	99	58	78	88	51
	学生回答频率	25	41	25	29	66	41
	学生问答占比	0.2717	0.2929	0.3012	0.2710	0.4286	0.4457
	沉默频率	5	0	1	4	2	0
语言能力	班级会话篇幅(字数)	1526	2545	2576	1932	2236	2421
	会话句子数量	191	297	222	226	351	279
	平均句子长度	7.99	8.57	11.60	8.55	6.37	8.68
认知能力	词种类数	439	642	824	575	655	738
	词密度	0.288	0.252	0.320	0.298	0.293	0.305
	代词数量	932	1423	1501	1092	1210	1382

Table 7. Word frequency of structure, specification and comprehensiveness

表 7. 结构化、具体化、综合性内容的标签词频率

Tag words	Class 2	Class 3
third	4	4
In conclusion	3	0
For example	5	0
comprehensive	7	3

在表 7 中, GPT 组在发言结构和规范性方面的提升很明显。third, for example, comprehensive 和 in conclusion 分别可以表明回答者的发言结构、规范性、全面性和概括性。third 这个词反映出回答者能够使用第三个理由, 通过有逻辑、有条理的表达来支持或分析某个观点。这样的回答也得到了教师 thanks for your comprehensive answer 这样的积极反馈, 其中包含了 comprehensive 这个词。二班的学生使用 for example 这个词, 通过举例来说明他们观点的规范性。此外, 为了证实 GPT 组在发言规范性上表现更优这一假设, 研究人员对二班组织了一次小组访谈。所有 4 名回答者都证实, GPT 辅助了他们的发言组织、翻译、规范性和全面性。人工智能提供和总结的丰富案例、例子和情景, 增强了他们发言的规范性。

除了发言规范性和概括性之外, GPT 组在解决沉默问题方面也有提升。表 7 显示, 两个小组在沉默问题上都有改善, 经过三次测试后, 沉默出现的频率逐渐降低, 并且沉默的出现被教师的标志性话语 why are you silent? 标记和索引。二班在后测 1 中沉默问题的消失比三班更早。然而, 在后测 2 中, 二班出现了一次沉默, 而三班没有出现沉默。在对保持沉默的回答者进行访谈后, 研究发现, 5 分钟的共同回答准

备时间这一因素受到了 GPT 使用者因智能手机分心的影响。具体来说, 在教师指导的共同准备过程中, 沉默的回答者被微信朋友圈分散了注意力, 所以他无法立即回答教师的问题, 而是不得不再次向 GPT 搜索答案, 这使他的回答延迟了 50 秒。在进一步访谈后, 这位沉默的回答者表示, 他是因为看手机的习惯而不自觉地分心了, 并非是对微信朋友圈感兴趣或有特定目的, 并且明确表示愿意改变这个习惯, 这是一种行为成瘾。此外, 本研究还观察到, 一个即兴提出的有建设性、新颖且相关的问题, 虽然没有导致沉默, 但延迟了 GPT 组的回应。因此, 人工智能在打破沉默方面的有效性, 部分受到对智能手机的行为成瘾和即兴提出的新颖问题的影响。三次测试中的沉默频率如表 8 所示:

Table 8. Silence frequency of 3 tests
表 8. 3 次测试的课堂沉默频率

test	Class 2	Class 3
pretest	5	4
post-test1	0	2
post-test2	1	0

综上所述, 根据沉默的原因以及 GPT 组的激活效果, 本研究发现了人工智能在打破 NCS 方面发挥有效性的三条原则: 1. 足够的对话前准备时间; 2. 回答者具备 GPT 指令技巧; 3. 回答者的行为成瘾受到监督。尽管人工智能辅助教学的班级和传统教学的班级在沉默问题上都有改善, 且差距较小, 并且对对照组在课堂上的发言量更多, 但 GPT 组在学生的语言结构、规范性和结论方面表现更优, 且有明显的标志性词语体现。

4.3. GPT 与传统课堂如何打破沉默

4.3.1. 人工智能(AI)组

AI 课堂沉默的原因

根据后测 1 之后对 GPT 组的访谈和观察, 本研究发现人工智能辅助英语课堂出现沉默的 7 个原因: GPT 答案难以理解、使用者的 GPT 指令能力不足、准备时间有限、人工智能的标准化答案、使用者的打字速度、对智能手机的行为成瘾、即兴提出的新颖问题。首先, 当一些回答者无法阅读、理解、记住并复述冗长、复杂且深奥的 GPT 答案时, GPT 组会保持沉默, 这延迟了他们在课堂上的回应。其次, GPT 使用者的指令能力不足导致 GPT 给出难以理解、不准确且不相关的反馈, 因为 GPT 组的一些使用者没有向人工智能提供足够清晰的结构元素信息, 比如使用者的身份、背景、问题意图、答案要求以及修改答案的指令。因此, GPT 有时无法为他们提供相关的答案。第三, 紧张的答案准备时间使得一些回答者难以与人工智能互动。人机互动比传统课堂需要更多的时间来准备答案。GPT 组的回应因输入问题、阅读和记忆人工智能的反馈而被延迟。第四, 由于超限效应, 人工智能的标准化答案使得一些回答者不愿意回答并重复之前回答者说过的观点。第五, 大多数回答者因打字速度慢而无法高效地与 GPT 互动, 而这一问题可以通过语音转文字功能轻松解决。第六, 对于一些长期养成了不良使用智能手机习惯的回答者来说, 行为成瘾是一个难以控制的因素, 这种习惯短期内很难改变, 并且在教师授课过程中会导致注意力分散和理解偏差。最后, 当教师用即兴提出的新颖问题进一步挑战回答者时, 由于没有足够的准备时间就突然使用 GPT 工具, 他们的回应通常会被延迟。

AI 的积极作用

根据表 8 和小组访谈, GPT 组的沉默因以下 5 个原因得到缓解: 自动生成观点、自动翻译、回应规

范化、自动总结材料、从多个角度拓宽回应的广度。首先, 自动生成观点节省了 GPT 组组织观点所需的脑力劳动, 大多数学生只需要模仿人工智能的答案来回应和学习, 因为模仿也是一种学习方式。其次, 人工智能的翻译功能帮助一些回答者用丰富的词汇表达深刻的观点, 这从表 7 中更高的词类数量和词汇密度可以体现出来。

由于人工智能的数据库以及带有示例的反馈, GPT 组的回应更加具体, 而对照组在没有足够的工作和生活经验时很难回想起这些示例。有了更多的例子, GPT 组能够说更多的话, 且发言时间更长。如前测和后测 2 所记录的那样, 标志性词语 “for example” 在后测 2 中 GPT 组的对话中出现了 5 次, 而在对照组中没有出现。然而, 在前测中, 两组都没有出现这个词, 这意味着 GPT 组在发言规范性方面有了明显提高, 其语料库证据如表 9 所示:

Table 9. The occurrence of “For example” in post-test 2 conversation

表 9. “For example” 在后测 2 对话中的出现和语境

Left context	hit	Right context
won't be slowed, sold by me and if the bicycle is really, really important,	for example,	It's a gift that a friend gave me or my parents gave
types of words they use, influenced by their educational background, professional context, and cultural environment.	For example,	professionals in technical fields tend to use more specialized nouns and jargon, while
no longer need to use it and throw it away? TangWenQi 09:00 I think it depends.	For example,	the bicycle, if I truly ride it for six years but it's
a cute pet or a beautiful dress. Boys also use adjectives in their own ways.	For example,	when talking about sports, they may use “amazing”, “powerful” to describe players' performances
basketball and so on and some, some, some topic that girls may not familiar with.	For example.	Uh, something that only used by men, not by women and if they

第三, GPT 组的回答也更好, 因为其总结能力更强, 这体现在频繁使用 “in conclusion” 一词上, 这可以帮助听众更好地记住主要观点。这样的总结性短语是模仿人工智能的反馈风格, 人工智能通常会以清晰的逻辑顺序总结并向用户呈现信息, 同时给出结论。in conclusion 一词在后测 2 中在 GPT 组出现了 3 次, 在对照组中出现了 0 次, 而在前测中两组都没有出现。其语料库证据如表 10 所示:

Table 10. The context of “In conclusion” in Post-test 2 conversation

表 10. “In conclusion” 在后测 2 对话中的语境

Left context	hit	Right context
think professors are better than farmers in using nouns in daily conversation? TangWenQi 08:54 So the	conclusion	is that professors and farmers may have different noun usage based on their professional context,
language can influence other adjective use, but this virus among individuals regardless of gender. In	conclusion,	it cannot be definitely stated that girls are more are using more adjectives than
in their work and talk. These nouns help them with farming and sharing info. In	conclusion,	they have different needs, and we can't say one is better at using

最后, GPT 组的回应从多个角度拓宽了广度, 体现在 first, second, third 等标志性词语上, 这些词的使用频率与对照组相当。这种结构词的高度相似性可能表明, 人工智能在角度多样性方面的影响并不显著。

至于教师为打破沉默而改变提问方法, 教师只使用了原因分析的方法来引导学生反思他们的沉默, 因为在人工智能的帮助下, 预计他们的沉默不是由问题难度、翻译能力不足或回答者的知识储备等原因造成的。因此, 没有使用其他提问方法, 如简化问题、幽默以及允许使用母语。

总之, GPT 工具确实帮助学生解决了由 7 个原因导致的人工智能辅助课堂的沉默问题。在后测 2 中, 通过语料库、研究人员的观察以及对参与者的访谈证实, 沉默问题通过 5 种方式得到了缓解。然而, 尽管沉默问题的出现频率降低了, 但由于某个回答者对智能手机成瘾, 点击无关内容, 沉默问题仍然偶尔出现, 而这种成瘾是长期形成的, 短期内很难改变。这种手机成瘾是无意识发生的, 且使用者表示愿意改变。

4.3.2. 对照组

根据后测 1 后对对照组的访谈和观察, 回答者保持沉默且表现比前测更差有四个原因: 对问题不熟悉、缺乏翻译能力、性格内向以及注意力分散。首先, 一些回答者发现老师提出的问题比他们的知识储备更具挑战性和新颖性。前测的问题基于回答者之前的作业, 这让他们对回答问题准备得更充分。第二个原因是缺乏翻译能力, 因为一些回答者不知道该用哪些词汇和句法来描述他们对新问题的想法。因此, 在没有充分准备的情况下, 新问题可能引发的不仅是对问题的不熟悉, 在外语课程中还可能导致翻译能力不足。第三, 一些沉默的回答者称性格内向是一个影响因素, 但研究人员发现性格并非主要原因, 因为性格内向的回答者在后测 2 中表现要好得多, 这是由于他们有了更多的准备时间和之前被点名的经历, 这些让他们在公开演讲时不那么焦虑。第四, 一些学业基础好且自律的回答者在回答问题时表现较差, 原因是他们在观察前面回答者的回答过程中分散了注意力, 而没有专注于自身的准备。学生作为回答者和观察者这两个角色是相互矛盾的。

基于后测 2 中的小组访谈和观察, 对照组回答得更好有三个原因: 课前练习、回答方式以及充分的准备。两名积极的回答者表示, 他们参加过一些与英语演讲技能相关的课前活动, 比如以观众或选手的身份参加英语比赛, 以及经常在课后与老师交流。第二, 一些回答者在改变回答方式后回答得更好。他们将一次性的回答转变为连贯的回答, 有条不紊地梳理自己的想法, 并在每个间隔处停顿, 以便以稳定的语速清晰地表达每一部分内容。第三, 由于有固定的点名名单, 并且在对话前有 5 分钟的集体准备时间, 所有回答者在内容和情绪上都准备得更充分了。

总之, 传统课堂也能够有效地打破沉默。GPT 组和对照组在沉默原因和激发效果方面存在相似之处。经过三次测试后, 他们在对话中所展现出的互动、语言能力和认知方面的差距缩小了。至于他们的差异, GPT 组的沉默原因和激发原因更加多样化且更难控制。尽管对照组的公开演讲积极性和发言量高于 GPT 组, 但 GPT 组的发言质量更好, 在规范性和总结性方面表现更佳, 且有明显的标志性词语, 便于回答者模仿具有清晰模式的人工智能反馈风格, 而传统课堂中的回答者只能模仿老师的发言, 老师发言的模式和结构则更加随意。

5. 结论

对 3.1 中 3 个研究问题的回答如下:

根据三次测试后的研究结果, 评判英语课堂对话表现的标准应考虑回答者在言语互动方面的提升, 同时兼顾语言能力和认知能力的提高, 因为课堂对话的本质目标是教育性的, 而不仅仅是互动性的。至于标准指标, 互动情况可以通过问答频率和回答者的发言频率来量化, 而语言能力可以用词数、词类和词汇密度来体现, 认知能力则可以通过使用 Python 中的困惑度和连贯性算法计算子话题数量来量化。

至于打破沉默的效果, GPT 辅助教学的课堂和传统课堂都能够在提升语言能力和认知能力的同时打破沉默, 但 GPT 组的效果可能会受到对智能手机的行为成瘾的影响, 这需要教师进行监督。与发言量表

现更好的传统组相比, GPT 能够通过让回答者模仿人工智能的反馈风格, 以明显的标志性词语在规范性和总结性方面提高回答者的发言质量。

教师在面对沉默情况时的提问方法变得简化, 仅采用原因分析这一种方法, 因为在回答者在课前准备阶段接受了指令培训的情况下, 预计其他导致沉默的因素不会成立。经过访谈发现, 在 GPT 组中, 由无意识的对智能手机的行为成瘾所导致的注意力分散是造成沉默以及人工智能失效的主要原因。不过, 沉默的回答者在访谈中也承认有改变这种行为成瘾的意愿。研究发现, 人工智能在打破沉默方面发挥有效性的三个前提条件包括: 1. 足够的对话前准备时间; 2. 回答者具备 GPT 指令技巧; 3. 回答者的行为成瘾受到监督。本研究也期望未来的研究能够关注由对智能手机的行为成瘾所导致的 NCS 问题。

由此, AI 课堂的互动效果比较考验教师对学生手机行为上瘾的控制和 AI 互动课堂教案设计和组织的能力。本研究也期待未来能进一步探索这两个问题。

致 谢

本研究得到了“上海外国语大学贤达经济人文学院”的资助, 所属科研项目为“人工智能在英语会话教学中的应用”。

参考文献

- [1] Butler, K.G. (1999) From Oracy to Literacy. *Topics in Language Disorders*, **20**, 14-32. <https://doi.org/10.1097/00011363-199911000-00004>
- [2] Williams, C. (2006) Teacher Judgements of the Language Skills of Children in the Early Years of Schooling. *Child Language Teaching and Therapy*, **22**, 135-154. <https://doi.org/10.1191/0265659006ct3040a>
- [3] Evans, M.A. (1996) Reticent Primary Grade Children and Their More Talkative Peers: Verbal, Nonverbal, and Self-Concept Characteristics. *Journal of Educational Psychology*, **88**, 739-749. <https://doi.org/10.1037/0022-0663.88.4.739>
- [4] Coplan, R.J., Hughes, K., Bosacki, S. and Rose-Krasnor, L. (2011) Is Silence Golden? Elementary School Teachers' Strategies and Beliefs Regarding Hypothetical Shy/Quiet and Exuberant/Talkative Children. *Journal of Educational Psychology*, **103**, 939-951. <https://doi.org/10.1037/a0024551>
- [5] King, J. and Aono, A. (2017) Talk, Silence and Anxiety during One-to-One Tutorials: A Cross-Cultural Comparative Study of Japan and UK Undergraduates' Tolerance of Silence. *Asia Pacific Education Review*, **18**, 489-499. <https://doi.org/10.1007/s12564-017-9503-8>
- [6] Maher, K. and King, J. (2022) 'The Silence Kills Me.': 'Silence' as a Trigger of Speaking-Related Anxiety in the English-Medium Classroom. *English Teaching & Learning*, **46**, 213-234. <https://doi.org/10.1007/s42321-022-00119-4>
- [7] Gkonou, C., Daubney, M. and Dewaele, J. (2017) New Insights into Language Anxiety: Theory, Research and Educational Implications. In: *Multilingual Matters*, Channel View Publications.
- [8] Maher, K. and King, J. (2023) Language Anxiety and Learner Silence in the Classroom from a Cognitive-Behavioral Perspective. *Annual Review of Applied Linguistics*, **43**, 105-111. <https://doi.org/10.1017/s0267190523000077>
- [9] Kim, S. (2009) Questioning the Stability of Foreign Language Classroom Anxiety and Motivation across Different Classroom Contexts. *Foreign Language Annals*, **42**, 138-157. <https://doi.org/10.1111/j.1944-9720.2009.01012.x>
- [10] Coplan, R.J. and Arbeau, K.A. (2008) The Stresses of a "Brave New World": Shyness and School Adjustment in Kindergarten. *Journal of Research in Childhood Education*, **22**, 377-389. <https://doi.org/10.1080/02568540809594634>
- [11] Young, D.J. (1991) Creating a Low-Anxiety Classroom Environment: What Does Language Anxiety Research Suggest? *The Modern Language Journal*, **75**, 426-437. <https://doi.org/10.1111/j.1540-4781.1991.tb05378.x>
- [12] Feng, Q., Zhang, H. and Sun, H. (2022) The Translation of Chinese Culture: Past, Present and Future. *International Journal of Languages, Literature and Linguistics*, **8**, 249-253. <https://doi.org/10.18178/ijll.2022.8.4.357>
- [13] Zhang, H. (2017) A Study on the Micro-Discourse Patterns of Stance-Taking in English News. *English Square*, **5**, 30-32. <https://doi.org/10.16723/j.cnki.yygc.2017.05.015>
- [14] 张欢, 相玟, 卢婧. 概念整合在翻译教学中的意义运作机制[J]. 世纪之星(高中版), 2021(10): 56-58.
- [15] Zhang, H., Feng, Q., Lu, J., Guo, Z. and Wu, B. (2022) Research on the Feedback Mechanism of Mobile and Digitized English Teaching in College during Covid-19. *Proceedings of the 7th International Conference on Distance Education and Learning*, Beijing, 20-23 May 2022, 20-28. <https://doi.org/10.1145/3543321.3543325>

- [16] Ha, P.L. and Li, B. (2012) Silence as Right, Choice, Resistance and Strategy among Chinese 'Me Generation' Students: Implications for Pedagogy. *Discourse: Studies in the Cultural Politics of Education*, **35**, 233-248. <https://doi.org/10.1080/01596306.2012.745733>
- [17] Grimshaw, T. (2007) Problematizing the Construct of 'the Chinese Learner': Insights from Ethnographic Research. *Educational Studies*, **33**, 299-311. <https://doi.org/10.1080/03055690701425643>