arXiv:2603.02567v1 [cs.HC] 3 Mar 2026

“展示出来，别光说”：指令多模态对软件指导的互补效应

Emran Poh
emran.poh.2025@smu.edu.sg 计算与信息系统学院 新加坡管理大学 新加坡，新加坡

Yueyue Hou
yueyue.hou.2024@smu.edu.sg 计算与信息系统学院 新加坡管理大学 新加坡，新加坡

Tianyi Zhang
tianyizhang.2023@smu.edu.sg 计算与信息系统学院 新加坡管理大学 新加坡，新加坡

Jiannan Li
jiannanli@smu.edu.sg 新加坡管理大学计算与信息系统学院 新加坡，新加坡

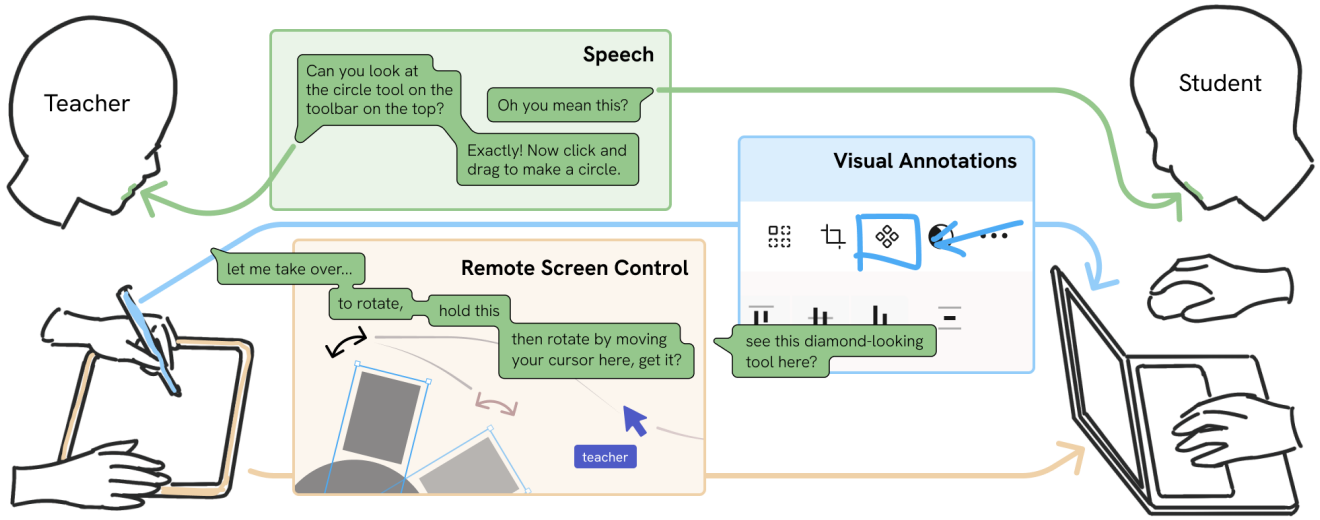


图1：在这种观察性研究场景中，人类教师协调运用多种沟通模态——语音（双向）、视觉标注（单向）和远程屏幕控制（单向）——以提供有效的软件教学，展现了可为人工智能辅导系统设计提供参考的成熟策略。

摘要

为软件学习设计自适应辅导系统，需要应对如何确定恰当教学模态这一挑战。为给此类系统的设计提供参考，我们对十组人类师生配对（N=10）开展了观察性研究。研究中，经验丰富的图形设计软件用户通过多步骤流程，向新手教授两项新的图形设计软件功能。为模拟人工智能辅导模态的可能形式，这些课程仅使用三种沟通渠道（语音、视觉标注和远程屏幕控制）。我们发现，批注能够与语音互补，空间精确性与远程控制以空间与时间精确性对其进行补充，但二者都会侵扰学习者能动性。教师会自适应地选择模态，以平衡指导推进的需要与学生的认知参与及数字领地所有权感。我们的结果进一步支持了邻接原则以及学习中能动性的价值，同时表明精确性—能动性权衡和数字领地性是自适应软件引导的新设计约束。

CCS 概念

• **以人为中心的计算 → 交互设计中的实证研究。**

*CHI '26，2026*年*4*月*13*日至*17*日，西班牙巴塞罗那© 2026 版权所有者/作者保留。出版权已授权给 ACM。此为作者版本，仅供个人使用。请勿再分发。正式版本发表于*2026*年第*17*届*2026*年*4*月*13*日至*17*日101145 37723183791200*CHI*计算机系统人因大会（*CHI '*），*4*月–西班牙巴塞罗那，https://doi.org/ 。/。.

关键词
多模态指导、软件学习、学习者能动性、视觉标注、认知负荷理论、自适应辅导系统、数字领地性

ACM 引用格式：
Emran Poh、Yueyue Hou、Tianyi Zhang 和 Jiannan Li。2026年。“展示出来，别光说”：指令多模态对软件指导的互补效应。载于*2026*年*CHI*计算机系统中的人为因素会议论文集（*CHI '26*），*2026*年*4*月*13*日至*17*日，西班牙巴塞罗那。ACM，纽约州纽约，18页。https://doi.org/10.1145/3772318.3791200

1 引言
随着重要产品和服务日益数字化，软件熟练度不仅对提高生产力至关重要，也是人们参与日常生活的基本条件 [39]。尽管用户体验设计不断进步，许多现代应用程序仍然复杂、零散且难以掌握 [38]。因此，用户经常依赖教程和帮助系统，但这些资源往往不完整、前后不一致，或难以大规模扩展 [32, 67]。生成式人工智能的最新进展提供了一种极具吸引力的替代方案：大规模提供个性化、循序渐进的软件指导 [27]。然而，大语言模型生成的纯文本指令往往难以理解和遵循，因为它们并未以实际软件界面为依据。用户可能难以理解陌生术语，也难以将抽象的文字描述与具体的用户界面操作联系起来 [37]。例如，“点击工具面板中的锁形按钮”这一指令，默认用户能够找到工具面板，并从众多选项中通过视觉识别出锁形图标。

多模态生成式人工智能有望在帮助他人使用软件时模仿人类教学策略，整合视觉标注、空间引导和演示，创造丰富而有效的学习体验。多媒体学习理论认为，在教学内容中结合多种模态能够实现时空邻近性，也就是让文字和图片在空间和时间上相互对应。这种邻近性能够促进学习 [49, 50]。人机交互研究表明，多模态引导通过视觉突出显示 [36] 和演示 [10, 30]，等额外提示补充文本指令，能够提升理解力和可学习性。此外，人类教师非常善于调整知识表征的选择，以适应其功能 [1]（例如，用图表呈现结构、用文本说明规则）、教学目标以及学生的接受程度 [47]。

然而，在如何有效建模并实现此类多模态适应性，以便在自动化软件指导系统中生成更具表现力且更具情境感知的指导方面，仍存在关键空白。因此，我们的研究由一个首要问题所引导：

在软件教学中，各模态如何相互补充，以回应教程内容和学生状态？

受此前分析人类实践以指导数字系统设计的研究启发 [26, 41, 46]，，我们对十组师生配对进行了观察性研究（N=10）。在研究中，经验丰富的图形设计软件用户通过引导新手完成多步骤任务流程，向他们讲解两项新的软件功能。该研究邀请了拥有至少六个月经验的 Figma 用户，他们通过视频会议向几乎没有或完全没有相关经验的新手进行讲解。教师可以通过三种不同的模态进行交流：言语交流（语音）、视觉标注（在学生屏幕上绘制）和远程屏幕控制（在允许的情况下直接操控）。这一设置模拟了未来屏幕端 AI 指导系统可能具备的能力，使我们能够在受控但贴近现实的环境中观察自然的多模态教学模式。

我们对模态功能的分析表明，语音构成了几乎所有教学活动的基础，而视觉批注和远程控制则作为互补渠道，各自具有独特优势。这三种模态都支持引导注意力以及建议动作或数值的基本功能，但它们在实现方式上存在不同的权衡关系。批注在展示空间布局以及以视觉方式阐明新概念方面尤为有效。远程控制能够进行直接演示，这对于传达复杂动作序列的动态变化与相互依赖关系尤其有价值。在较高层面上，批注以空间精确性补充语音，而远程控制则以空间与时间精确性对其进行补充。这与空间和时间邻近原则 [49, 50]相一致。

然而，更高的精确性往往也要付出代价。我们的观察和访谈显示，教师意识到不同模态对学生能动性的侵入程度各不相同。学生能动性是指学生发起并控制自身行动及共享视觉焦点的能力。批注和屏幕控制能够清晰地展示软件操作，但也减少了学生主动与软件互动的必要性，从而减少了深入认知参与的机会；认知负荷理论 [64, 65] 和求助调节研究 [2], 均指出了这一点，尽管影响程度有所不同。教师还留意到，视觉批注和屏幕控制会影响学生对软件界面的所有权——他们将软件界面视为自己的数字领地 [59] ——这也是一种能动性侵入。教师的自适应模态选择体现了持续的权衡：他们既力求尽量减少对学生能动性的侵入，又确保指导为学生提供足够的支持，使其能够继续推进。这种权衡不仅取决于教学内容所需的空间与时间精确性，也取决于学生在特定时刻的接受程度。这些发现表明，要让基于人工智能的引导系统既有效又尊重学生，就必须考虑语义、认知和社会因素，包括各模态对学生能动性的侵入程度、教学内容对精确性的要求，以及学生不断变化的状态。本研究对自适应软件引导的设计有两项重要贡献：

1. 模态特定用途与权衡关系的实证证据：系统性地揭示了语音、视觉标注和远程屏幕控制在软件教学中的功能与使用模式，表明具有时空邻近性的模态能够实现更高的教学精确度，但在提高精确度与保留能动性之间存在权衡关系。

2. 具能动性意识的模态选择：展示人类教师如何自适应地选择模态，以维持

学生的校准能动性，在推进教学与学生的认知参与及其对数字领地的所有权意识之间取得平衡。

我们的研究发现从认知负荷理论的角度进一步支持了多媒体学习中的邻近原则 [48,49]，以及学习者能动性的价值 [20]；同时也凸显了在复杂软件的一对一辅导情境中共同注意的重要性 [33, 52, 58]。我们还指出，精确性—能动性权衡和数字领地所有权是屏幕软件指导的另外两项设计约束，因为这类指导能够直接访问学生的界面。

总的来说，这些发现拓展了我们对多模态教学的理解，并为自适应人工智能软件辅助系统的设计提供了参考。研究中识别出的模式还启发了新的设计概念（如图 5 所示），包括幽灵光标交互、自适应注释淡出、时间轴拖动和交互式参考。这些设计有助于应对软件指导中的认知与可用性挑战，并为未来的人工智能辅导系统提供设计可能，使其既能借鉴又能超越人类教学策略。

2 相关工作

本研究借鉴了基础学习理论（认知负荷理论、求助研究和计算机支持的协作学习），并涉及三个研究领域：多模态学习理论、软件指导系统和自适应辅导。

2.1 基础学习理论

我们的研究以基础学习理论为基础，这些理论涵盖了从基本认知机制到更具社会性和协作性的学习视角。在认知层面，认知负荷理论（CLT）阐释了信息呈现的结构和复杂程度如何影响学习者形成持久知识的能力 [65, 66]。CLT 区分了内在负荷（材料本身固有的负荷）、外在负荷（由不佳的教学设计造成的负荷）和相关负荷（用于图式建构的努力）。这一框架与软件学习尤为相关。学习者在浏览复杂的可视化界面时，必须同时掌握程序性知识（如何操作界面）和概念性理解（为何某些操作能够实现预期结果）。

基于这些认知基础，关于求助调节的研究探讨了学习者和教学系统如何在学习过程中管理协助的数量、时机和形式 [2, 3, 54, 56]。这一视角不仅关注学习者何时求助，还强调学习者如何监控自己的理解、决定在多大程度上依赖现有支持，以及导师或系统如何随时间调整支架和反馈。智能辅导系统领域的研究探讨了“辅助困境”——即如何确定何时以及提供多少帮助，才能优化学习成果 [2]。在软件学习情境中，求助调节还因一项需求而变得更加复杂：不仅要管理提供多少指导，还要管理在不同时间点侧重采用哪些辅助模态（文本、视频、交互式指导）。近期研究考察了计算机专业学生的求助相关行为和偏好如何受到生成式人工智能工具的影响 [27], ，以及基于大语言模型的助手如何参与持续进行的求助调节 [37]。

在更偏向社会与分布式的层面上，计算机支持的协作学习（CSCL）研究探讨技术如何通过与他人的互动来中介学习 [33, 58]。CSCL工作强调支架、分布式认知和共同注意等原则，展示工具如何构建协作活动，使学习者共同建构知识。尽管我们的研究聚焦于一对一的人类教学而非同伴协作，这些洞见仍有助于我们思考人工智能辅导系统：它们如何模拟在社会情境中观察到的有效教学实践，并在共享的数字工作空间中协调注意力。

2.2 软件任务协助与学习

复杂软件系统中学习新功能的问题，自计算早期以来就吸引了研究者的关注 [63]。人机交互领域的一些开创性研究考察了人们如何使用软件手册，发现许多人几乎不从手册中学习，甚至完全不学习，而是依靠自主探索来学习功能。虽然这些用户能够学到足以完成其感兴趣任务的内容，但他们往往无法达到专家水平的熟练程度，因为自主探索并不一定会揭示那些提升生产力的高级功能 [8]。这一现象被称为主动用户悖论。该观察启发了教学设计的极简主义方法，该方法强调实用而简洁的内容 [7, 42],，并催生了一系列旨在支持从新手向专家过渡的新型交互技术设计 [13]。其中许多设计基于支架理念，帮助学习者逐步建立熟练度，例如标记菜单 [40], 、OctoPocus [5], 和 CommandMap [57]。

尽管上述工作讨论了软件教程中应呈现什么，另一条研究路线则主要关注如何组织并呈现这些教程。沿此方向的一些早期工作研究了动画演示在解释软件操作方面的效果，并发现了喜忧参半的结果 [25, 53]。一种与本研究密切相关的后续方法采用情境化教程，将指导直接呈现在需要操作的用户界面元素上，包括基于模板的教程 [36], 、ToolClips [24], 以及更为全面的系统，如Sketch-Sketch Revolution [15]。然而，先前的工作主要聚焦于静态的单一模态指导。我们的工作通过探索动态的多模态指导来应对这一局限，这种指导会根据教学情境和学习者需求同时调整内容和沟通渠道。

2.3 教育中的多模态

多模态交流在学习中发挥着核心作用，研究表明，协调使用互补通道能够提升理解能力。梅耶的多媒体学习认知理论解释了学习者如何通过部分独立的视觉和听觉通道处理信息 [48, 50, 51]。其中的关键原则包括模态原则：有视觉内容时，旁白优于屏幕文字；以及邻近原则：强调文字与图片在空间和时间上的对齐 [48, 50]。这些原则表明，将信息分配到不同通道可以提高工作记忆容量，而协调不当的冗余信息则可能适得其反 [49, 50]。

在我们的研究中，我们将梅耶的框架视为一组认知基线，并通过对人类教学的实证分析，考察软件教学中更丰富的模态排序与协调模式。

面对面教学的实证研究展示了这些原则在实践中的体现。教师经常将语音与视觉演示或手势同步，以引导注意并阐明抽象概念 [4, 14, 61]。辅导主要通过引发学生的建构性回应来发挥作用，而非依赖辅导者讲解；即使辅导者无法进行讲解，学生仍能有效学习 [9]。软件学习领域的研究特别指出，教师通常会结合指示、屏幕标注和口头解释来引导注意并支持程序性理解 [10, 11, 45]。针对远程或非正式软件学习情境中求助行为的研究进一步强调，判断何时需要帮助以及学习者如何协调协助至关重要 [32, 38, 67]。

近期关于人工智能辅导系统的研究探索了如何对这些多模态教学模式进行建模。诸如 [22] 和交互式草图板 [43] 界面等系统表明，将视觉标注与言语交流结合起来能够带来益处。然而，大多数现有的人工智能辅导系统仍专注于单一模态或简单组合，与人类教学相比，其协调模式有限 [35, 60, 69]。人工智能系统中的模态选择与协调挑战已在多个领域得到探讨。在机器人学领域，研究探索了针对不同类型的指导，何时应采用视觉、听觉或触觉反馈 [68]。在教育技术领域，研究考察了如何根据任务特征和学习者需求自动选择合适的模态 [29]。然而，对于这些原则如何具体应用于空间意识和程序性知识至关重要的软件学习情境，我们仍缺乏深入理解。

2.4 指导工具的自适应性

自适应辅导系统的研究探讨如何通过调整所提供的帮助及其提供时机，为学习者量身定制指导 [28, 71]。经典智能辅导系统表明，相较于静态教学材料，根据学习者表现动态调整反馈和内容排序能够改善学习成效，从而推动了后续对更细粒度自适应性的研究 [70]。在软件使用领域，这一理念已扩展至情境感知和主动式帮助系统；这些系统会根据用户当前的界面状态、任务目标和推断意图调整指导。例如，情境内工具提示和基于模板的叠加层会呈现相关用户界面元素的即时讲解 [12, 16, 17, 24, 36]。机器学习方法进一步拓展了自适应性的范围：利用强化学习随时间优化界面配置，并利用自然语言处理根据用户偏好调整讲解的风格和具体程度 [18, 19, 44, 62]。在这一更广泛的研究脉络中，一些系统通过教学代理实现自适应引导，例如由动画角色或屏幕辅导员提供教学和反馈 [21–23, 31]。对这些系统的研究考察代理应在何时进行干预、应提供多少支持，以及如何协调语音和手势来为学习提供支架 [6, 34], , 这突出表明自适应性涉及不仅要管理内容，还要管理干预的时机和多模态形式。

尽管取得了这些进展，大多数自适应系统仍主要调整呈现的内容，而非调整传达内容的方式或开放式任务的结构。软件学习任务通常约束较少，需要精细的空间与程序性知识，并允许采用多种有效策略，因此更难确定恰当的干预时机和支持形式。尽管已有一些研究根据用户偏好在不同呈现形式之间切换（例如文本与视频），但关于如何依据任务特征、用户专业水平或学习情境，动态调整沟通模态的选择与协调，目前仍缺乏研究 [70]。这一研究空白在软件学习领域尤为突出，因为不同任务（空间排列、程序执行、概念解释）可能适合采用不同的视觉、言语与示范指导组合。我们的研究通过分析人类教师如何在不同教学情境中调整模态选择与协调模式，来应对这一挑战，旨在为未来的自适应指导工具提供参考。

3 方法

本观察性研究采用定性探索方法，考察软件教学中的模态选择模式。我们使用了一个实时教学环境，教师可查看和控制学生的屏幕，并添加草图注释，以便在获准时提供直接指导和干预。

3.1 参与者

共有十对师生参与了本研究（N=10）。参与者年龄介于20至35+ 岁之间（教师：M=25岁，SD=2.86；学生：M=25.7，SD=3.72），其中女性12人、男性7人，另有1人认同为其他性别（见表1）。教师中有五位专家，拥有 2+ 年Figma经验；另有五位具备中级熟练度（6个月至2年）。学生中有五位初学者，拥有1至6个月的Figma经验，另有五位没有先前经验。

教师的招募依据是其出色的Figma熟练度，并优先考虑具有先前教学经验的候选人。五位教师拥有超过两年的Figma经验，六位曾在大学层面教授软件课程，班级规模为25名或以上学生。总计十位教师中有九位具备某种教学背景：其中六位拥有大学教学经验，另外三位教师（P1、P6、P8）曾辅导1至5名幼儿园至高中学生。一位教师（P5）没有先前教学经验，但具备高级Figma技能。研究人员通过社交媒体和大学网络开展定向招募。参与者获得了30美元的报酬。本研究经所在机构的伦理审查委员会批准。

3.2 实验设计与环境

本研究采用结构化教学环节，教师提前两天收到准备材料，但在教学方式上保有充分的自主性。每个教学环节都包含两个依次进行的教程，随后是独立测试任务。各教学环节均

表1：参与者人口统计信息（N=10：10名教师、10名学生；年龄20-35+；12名女性、7名男性、1名其他）。Figma经验：高级（>2年）、中级（6个月至2年）、初级（<6 个月）、无。教学经验按颜色编码：软件相关（深色），非软件相关（浅色）。所教学生：25+（浅灰色）。课程类型：（1）大学软件/技术类；（2）大学非软件类；（3）幼儿园至高中科学、技术、工程和数学；（4）幼儿园至高中辅导。注：性别；经验；高级；中级；初级。

教师							学生			
ID	Age	Gen.	Figma经验	软件教学经验	其他教学经验	学生人数	ID	Age	Gen.	Figma经验
P1-T	20	F	高级	–	K-12辅导	1-5	P1-S	29	F	初级
P2-T	27	F	高级	大学水平的软件教学（Figma）	大学水平非软件教学	200+	P2-S	22	M	–
P3-T	24	F	中级	大学水平的软件教学（Python/Pandas/Tableau）	幼儿园至12年级辅导	50+	P3-S	25	F	–
P4-T	28	M	中级	大学水平的软件教学（SQL、Python）	K-12辅导	75+	P4-S	24	M	初级
P5-T	30	F	高级	–	–	–	P5-S	27	F	–
P6-T	23	F	中级	–	K-12辅导	1-5	P6-S	23	F	初级
P7-T	27	M	高级	大学水平的软件教学（Figma）	–	25+	P7-S	35+	M	–
P8-T	22	M	高级	–	K-12辅导	1-5	P8-S	23	M	初级
P9-T	24	F	中级	大学水平的软件教学（Unity、Figma）	幼儿园至高中 STEM 工作坊（机器人学，无人机、聊天机器人）	100+	P9-S	23	F	初级
P10-T	25	O	中级	大学水平的软件教学（Figma）	幼儿园至高中 STEM 工作坊（机器人学）	75+	P10-S	26	F	–

在受控的实验室环境中进行，教师和学生分别使用笔记本电脑接入同一个 Zoom 通话，以便进行屏幕共享、批注和远程控制。所有互动均录制了屏幕和音频；详细的硬件规格见附录 A.1。

该教学平台支持三种沟通模态：语音、视觉标注和远程屏幕控制。我们选择这些模态，并非要穷尽辅导系统中的所有模态，而是将其作为彼此不同且具有代表性的模态，用于考察教师在实时教学中如何运用和协调它们，以及如何应对其中的权衡关系（例如，精确性与能动性之间的权衡）。这一有限的交互空间既让自然的教学互动得以展现，又确保了技术条件的一致性。

3.3 流程

每次教学课程都遵循标准化流程，每组师生配对的课程时长约为1.5小时。每次课程开始前两天，教师都会收到准备材料，其中包括完整的任务说明、技术设置相关信息以及研究预期（参见附录 A.4）。虽然鼓励教师自然地开展教学，不限制其教学方式，但同时要求教师确保学生能够完成教程中规定的相同成果。在两项任务中，教师都可完全自主地决定教学方式和模态选择。

会议开始时进行了15分钟的会前准备，包括知情同意、技术验证和参与者介绍。随后向教师简要介绍了Zoom的屏幕共享、批注和远程控制功能，并给予练习时间以确保基线技术熟悉度。研究未对教师先前使用Zoom功能的经验进行正式评估，因为该研究旨在捕捉自然教学行为，而非受控功能使用。

教学部分包括**任务1（天气图标）**（15–30分钟教学、10分钟任务后测试和10分钟讨论），之后休息5分钟，再进行**任务2（天气卡片）**（15–30分钟教学、10分钟任务后测试和最后10分钟讨论）。每项任务结束后以及每次课程结束时，都会进行半结构化访谈，以了解教师和学生对教学效果、模态选择、学习体验及沟通挑战的看法。

3.4 任务设计

本研究采用渐进式方法，围绕在 Figma 中设计天气主题组件展开。我们选择 Figma，是因为它代表了功能丰富、界面简洁的现代应用程序，但这类应用也可能带来初期学习障碍。为校准任务难度并确保技能覆盖恰当，我们邀请没有 Figma 使用经验的参与者开展了试点研究。这些试点帮助我们确定了一组初级学习者应掌握的基础功能，也找出了更具挑战性、但又不至于难以企及的高级功能。经过审慎校准，这些任务既足够有挑战性，能够引发多样的教学策略，又适合新手学习者。双任务序列遵循自然的学习进阶，从具体操作逐步过渡到抽象系统思维，其中第1课着重于



图 2：两项 Figma 设计任务：任务1：天气图标创建（左）和任务2：天气卡片设计（右），展示了从基础技能到高级组件使用的进阶过程

直接操控，第2课则要求更深入的概念理解。这两项任务都为采用各种沟通模态提供了丰富机会。矢量操作需要精准的指示和批注，而组件概念则得益于概念性解释和演示（见图 2）。第1课和第2课分别包含七个和八个步骤。关于这两节课的更多详情，请参见附录 A.4。

每节课结束后，学生都会独立完成测试，将所学技能应用于新但相似的任务（例如，创建更多天气图标或设计新的组件布局）。这些课后测试会被录像，以评估技能迁移与保持情况，并向教师提供即时反馈。

3.5 数据收集与分析

我们收集了所有课程的同步屏幕录像和音频录音，记录学生的 Figma 界面、教师批注、远程控制操作以及所有口头互动。这些录音录像构成了主要数据集，用于考察教师在实时软件教学过程中如何协调语音、视觉标注和屏幕控制。

多模态数据提取。 我们将课程分割为若干课程步骤，为后续分析提供时间结构。对于每节课，我们提取了：（1）通过 Zoom 批注工具绘制的视觉标注；（2）教师直接操作学生界面的远程控制片段；以及（3）通过自动语音识别生成并经人工校正以确保准确性的语音转录文本。批注的自动检测以及语音和视频处理的技术细节见附录 A.3。

编码框架：针对模态使用模式，我们对收集到的数据进行了内容分析。为开展跨模态分析，我们利用共享的步骤标记和时间戳，将语音片段与相应的视觉标注和远程控制演示进行时间对齐。通过这种对齐，我们得以考察教学过程中不同模态如何相互补充，例如识别口头命令伴有视觉标注，或讲解与远程控制演示同步出现的实例。两位研究人员首先对两个场次进行开放编码，以细致描述三种模态的教学功能。讨论后，他们对编码进行了完善，并归纳为主要功能类别。采用这套编码方案，他们独立编码了其余数据，评定者间信度（科恩卡帕系数）为 0.82。研究团队讨论再次讨论以解决分歧，并统计最终的模态使用频次。

对于更高层次的师生互动，我们采用了类似的工作流程，但使用归纳式主题分析方法：两位研究人员审阅了两场课程，拟定潜在编码，并运用亲和图法将其归类为初步主题。随后，他们分别分析其余数据，并在分析过半时进行讨论，分享新发现并完善编码。最终的行为分析达到了0.77的评定者间信度。

4 结果

我们的分析揭示了软件教学中多模态教学的三个关键发现。首先，教师依赖三种沟通模态——语音、批注和屏幕控制，每种模态都有其独特的优势与局限。其次，教师会根据学生反馈和任务需求调整沟通方式。第三，模态的有效性因情境而异，因此需要策略性地组合使用：语音提供基础支架，批注提升空间精确性，而屏幕控制则支持空间和时间上更复杂的演示。

4.1 研究概述与参与者表现

十组师生配对完成了两项依次进行的 Figma 设计任务，共录制了12.4小时的视频。我们通过配对编号和角色指代参与者（例如，P1-T代表教师，P1-S代表学生）。所有参与者均完成了两节课程，每场时长为61–92分钟（M=71.5，SD=9.2）。第1课耗时20–31分钟（M=24.8，SD=3.4），第2课耗时12–30分钟（M=19.2，SD=4.1）。最耗时的步骤是第1课中的太阳光线创建（步骤3），平均耗时=8.2分钟。第2课中演示自动布局的级联更改（步骤8）产生了最多的语音输出（5,748个词），屏幕控制的使用也最多；光线绘制步骤所需的批注也最多（85个实例）。

4.2 学习成果与技能迁移

学生在独立迁移任务上表现良好，表明这些课程能够将技能迁移到新问题上。所有学生（10/10）均成功完成测试1，九名学生（9/10）完成了测试2。尽管少数学生起初对不熟悉的步骤有所犹豫，但他们通常都能解决这些挑战并完成任务。

4.3 模态使用模式

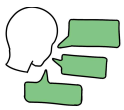
所有参与者都使用了多种沟通模态，其中 10 人中有 7 人使用了全部三种模态（语音、批注和屏幕控制）。所有人都使用语音，8 位参与者使用批注，另有 8 位使用屏幕控制。参与者的使用模式差异显著：3 位教师频繁使用批注，5 位则较少使用；同样，3 位教师经常使用屏幕控制，5 位则很少使用。值得注意的是，一些教师明确表示不愿使用某些模态：P2-T 避免使用屏幕控制，并表示“如果我控制得太多，学生就无法获得经验，”而 P9-T 对批注和控制都心存顾虑，更倾向于不去“碰他们的东西。”总体而言，模态的使用并非一成不变，而是会根据教学目标、学生反应和个人教学

表 2：研究中的教学模态类别与示例

模态类型及组合		类别	描述	示例
仅语音		引导注意	聚焦学生注意力的言语提示 于特定界面元素	"在圆形顶部" (P1),"看这个 按钮在这里" (P5)
		建议数值	提供具体的数值或分类- 数值供学生使用	"将高度更改为 *64*" (P4),"将 宽度设为 *200*" (P6)
		命令操作	向学生提供直接的口头指导 以执行特定操作	"点击此图标"（P2），"拖动此元- 素....."（P7）
		解释概念	对概念、关系或推理进行口头讲解 对概念、关系或推理进行口头讲 解	"组件有助于规范你的设计—— 计"（P3），"这会创建一个可重复使用的模—— 板"（P8）
		微指导	简短的确认、鼓励， 或方向提示	"对，正确，嗯，嗯" (P2),"很好，继续 下去" (P9)
视觉 批注	+ 语音	高亮目标 + 命令	将口头指令与视 觉批注相结合以指导操作	"那么就用那个工具画一个圆形。好的。" *(*圆形工具*)* (P2)
		写入数值 + 命令	提供具体数值，同时突出相关 界面元素	"这就是它在这个页面上的位置" *(*圆形坐标字段*)* (P3)
		绘制 +解释 排列	使用视觉标注展示空间 关系和布局，同时进行解释- 口头说明	"把这个放在这里，然后把这个放在下面 "（箭头指示定位）（P4）
		绘制 +解释 概念	使用视觉标注来阐明概念 在口头解释时接受概念	"就像你只放上这个太阳，他们 就不知道它是什么"（箭头指示 文本）(P4)
远程屏幕 控制	+ 语音	演示	执行操作时进行解释 远程在学生屏幕上进行	"所以我现在接管屏幕了。这样你 就可以....." (P6)
	-	纠正	不作解释便接管，以便 修复或调整学生作业	NA

理念而灵活调整。下文将详细介绍每种模态的作用与局限
y.

4.4 仅语音



语音是所有教学的基础模态，为其他模态提供了赖以构建的必要支架。这一基础作用体现在多个方面：语音普遍

在所有参与者和会话中均采用，而其他模态则有选择地使用。

学生始终依赖口头指导来获得程序性支持；正如一位学生所说，当教师"告诉我该按哪里时，我就会记住。"教师利用语音建立背景、解释不同操作之间的关系，并提供实时反馈，让学生在进行界面交互的同时处理这些信息。语音还协调着模态切换：教师通常会在转换前进行提示（例如，在演示前说"我来给你演示"，或在添加批注前说"我把这里圈出来"），使其他模态的引入显得有意为之，而非突兀。我们的分析识别出五种关键交互模式：引导注意特定界面元素、建议调整时使用精确数值、命令执行特定操作、解释界面组件之间的概念关系，以及提供微指导。

4.4.1 引导注意。教师[look here!]接的语言，引导用户将视觉注意力转向特定界面元素。教师会通过详细讲解和简洁的命令来引导用户注意。例如，一位教师（P1-T）提醒学生注意界面元素："你可以在这边调整大小"另一位教师（P2-T）则使用了更聚焦的指令："注意角手柄"。

4.4.2 建议数值。教[type:150px]具体的数值或类别值，以指导界面调整。教师使用直接指定数值的表达方式，例如"将其改为 <值>"或"将宽度设为 200"，引导用户进行精确配置。例如，一位教师（P4-T）给出教学指示："能否稍微调整一下，并将高度设为 *64*？"而另一位教师（P1-T）则仅表示：*"130* 乘 *160"*，数值含义不言自明。这表明教师如何运用数值指定语言，提供高效且可操作的指导。

4.4.3 命令操作。教[click that,]了用于界面操作的分步口头命令，使用"圆形坐标字段"或"圆形描边面板"等提示，有时还会附带解释性语境。例如，P1-T 提供了一个详细的提示，如"圆形坐标字段"，而另一位教师则使用了更为简洁明确的指示："你可以点击.....你可以将它们全部设置好。"

表3：天气图标和天气卡片任务中，设计任务步骤各阶段按类型划分的视觉标注及远程控制时段的频数分布。颜色强度表示每列中的相对频率，较深色调代表较高的数值。

	Step	描述	视觉注释数量					远程控制时段	
			总计	突出显示 & 目标命令	建议值 & 命令	Draw & 解释排列	Draw & 解释概念	演示。（秒）	纠正。（秒）
Weather Icons		课前课程：简介	3	0	0	3	0	202	0
	1	准备画布	4	4	0	0	0	198	0
	2	创建圆形	14	9	3	2	0	0	0
	3	创建光线	85	65	7	12	1	354	196
	4	创建云形状	34	12	3	18	1	136	0
	5	创建新月形状	23	11	2	6	4	0	0
	6	扁平化图标	5	5	0	0	0	39	39
	7	创建组件集	39	33	0	0	6	206	412
Weather Cards		课前课程：介绍	11	1	0	0	5	0	2
	1	卡片框架	3	2	0	1	0	87	0
	2	添加天气图标	12	11	1	0	0	85	85
	3	添加温度文本	12	10	1	1	0	0	0
	4	添加条件标签	43	37	2	4	0	0	0
	5	创建组件	36	35	0	0	1	0	0
	6	创建两个卡片实例	27	21	1	0	5	0	0
	7	级联更改 - 字体粗细	28	25	0	1	2	0	0
	8	级联更改 - 自动布局	47	31	3	1	12	477	140
		总计	**426**	**312**	**23**	**49**	**37**	**1784**	**874**

这些教学帮助学生完成特定的界面操作。

4.4.4 解释概念。 教师抽象的界面概念和组件交互，以阐明系统元素之间的关系。教师解释了复杂的关系，例如“解释父子关系”或“描述组件如何交互”，帮助用户理解底层系统架构。例如，一位教师（P1-T）通过指出以下内容解释了对齐反馈：“（*Figma*）会告诉你何时位于中心位置”另一位教师（P2-T）则采用了更易理解的概念性总结：“这就像父子关系。”这表明教师借助概念性解释，帮助用户理解复杂的界面关系。

4.4.5 微指导。 教师在互过程中，通过简短的言语提示和确认，提供实时反馈和微指令。教师使用简短、直接的短语，例如“对，嗯嗯，不对，左边，好”，或使用鼓励性的话语，例如“没错，继续”，引导用户完成当前操作并给予肯定。例如，一位教师（P1-T）提供了鼓励性的微指导：“没错，继续”而另一位教师（P2-T）则使用了简洁的指示：“对，嗯嗯，不对，左边，好。”这表明教师通过微指导，在用户交互过程中提供实时反馈和有针对性的指令。

4.4.6 局限性。纯语音交流存在若干显著局限，影响了教学效果。首先，单靠口头指导往往不足以应对复杂任务，学生难以将抽象描述转化为具体操作。教师也常常发现自己无法有效说明精确的界面位置，这导致学生不清楚“该按哪里”，或不知道要操作哪些具体元素。一名学生（P3-S）表达了自己的挫败感：“我不知道该点哪里，在界面里迷路了。”其次，语音指导容易造成信息过载，也受到记忆限制。口头指导具有顺序性，学生难以记住多个步骤，也难以回顾先前的指导。第三，对于精确定位和测量任务，语音缺乏所需的精确度，教师不得不进行冗长的口头描述，往往因此引发错误或误解。一位教师（P4- T）在应对这一局限时感到困难：“很难把文本转化为应用……解释它在设计中实际意味着什么。”最后，由于语音提供的持久视觉上下文有限，学生必须不断协调听讲与搜索界面的过程，从而增加了认知负荷。

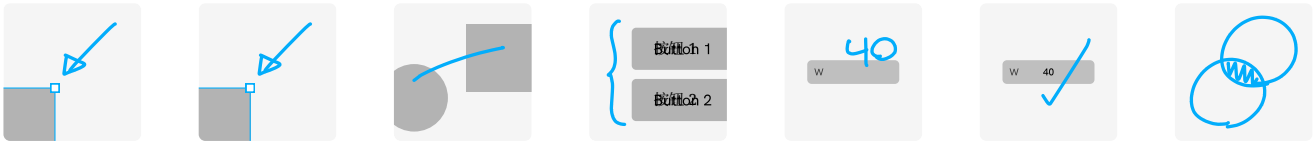
4.5 批注 +语音



批注逐渐成为一种视觉交流工具，有效地补充了语音在空间敏感型指令中的作用。教师经常使用批注来应对仅凭语音无法充分传达的空间精确性挑战。批注在元素查找和空间引导方面尤其有效，还有助于实现工作流可视化。

教师采用了多种视觉批注样式（见表3），包括圆圈、箭头、线条和手绘图，以实现不同的教学目的。批注与语音之间的协调能够传递额外信息，

图 3：标注样式及其特点



教师能够同时提供视觉和语言语境。所识别的批注实例可分为以下类别（见表4）：（1）高亮目标 +命令操作，（2）建议值 + 命令操作，（3）绘制+ 解释排列，以及（4）绘制 + 解释概念。

4.5.1 高亮目标 + 命令。这种多模态模式（见表4）将视觉突出显示与口头指导相结合，以引导用户操作。该方法通过圈出或突出显示特定界面元素，同时给出诸如“看这里并点击这个”之类的口头命令，以集中用户注意力并指导其操作。

例如，一位教师（P1-T）在圈出遮罩按钮的同时说道“这个就像你拍一张照片，它会……”以引导学生找到正确的布尔运算。另一位教师（P2-T）在圈出布尔运算按钮的同时指示“对对，好，所以你能不能像这样移过去，还有这个，对，好，然后你选择并集，好”以帮助学生选择用于组合形状的正确工具。教师经常使用这一技巧来突出显示学生需要交互的特定界面元素，例如在圈出框架工具的同时说道“所以你试着往下到这里的栏，然后点击这个”（P10-T）。

数据显示，这种模式在两课中均被广泛采用：第1课有85条批注，第2课有43条批注，反映出学习初期对视觉-口头指导的依赖程度更高。在第1课中，使用量在步骤3（85条批注）和步骤4（34条批注）达到峰值，这两个步骤涉及初始设置和复杂操作。在第2课中，批注分布更为均衡，其中步骤4（43条批注）和步骤5（37条批注）的使用最为密集。这一变化表明，随着学生对界面越来越熟悉，教师逐渐从对每个元素进行详细的突出显示，转向使用更简洁的多元素提示。

4.5.2 建议值 +命令。教师采用#2780FF注-语音方法（见表4），通过视觉方式指出特定值，并结合口头指导。这种技术包括将确切的输入值写在相应输入字段附近，以起到突出显示的作用，例如“看这里并改为40”或“将宽度设置为200”，并在坐标字段周围画一个圆形。

例如，一位教师（P2-T）一边解释，一边画了两个圆圈，并在每个圆圈里写上“12”，“你画两个圆圈，每个是*12*。每个*12*，好吗？对。”以指定新月形状所需的精确像素尺寸。另一位教师（P10-T）在新月旁写下“FFCC33”，同时说道“对，现在你要填充无色的黄色，也许还可以重新调整颜色的位置”，以提供精确的十六进制颜色代码。教师也将这一技巧用于定位，例如，一位教师（P8-T）在解释云形状的间距要求时，写下了“4像素”。

数据显示，两节课中共出现16处批注，其中第1课有13处，第2课仅有3处，这表明随着教师不断改进沟通方式，该方法的使用频率显著下降。使用情境表明，这种方法主要用于尺寸规格、坐标设置和数值参数调整等对精度要求较高的场景。使用高峰出现在第1课的步骤3（5处批注）和步骤4（3处批注），通常是在需要精确测量或定位的任务中。该方法的优点包括通过消除歧义提高精度、减少查找输入字段所需的精力、借助直接的视觉参照加快操作，以及通过防止数值放置错误来减少失误。

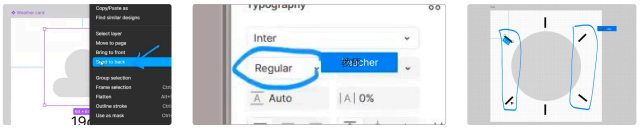
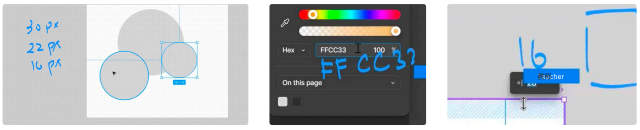
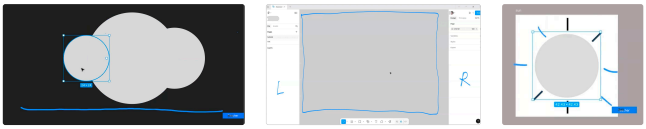
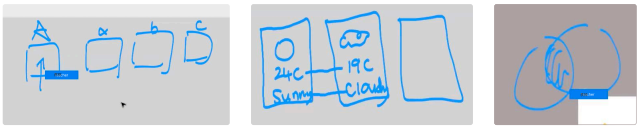
4.5.3 绘制 +解释排列。这种多模态模式（见表4）将视觉绘制与口头讲解相结合，用以展示空间关系和视觉组织。该方法通过在画布上绘制示例、布局和排列，同时口头说明空间位置和结构关系。

例如，一位教师（P10-T）在画布上绘制了一个完整的太阳图标，同时解释道“这就是我们的参考图。要做到这一点，你先画这条线，再画这条线”，以向学生展示他们正在创建的最终视觉结构。另一位教师（P1-T）通过在圆形周围再画出4条光线，并说道“所以复制它，让其中一条像这样”，演示了光线复制，以说明太阳图标所需的空间布局。教师还用这种方法展示复杂布局，例如，一位教师（P2-T）画了一个带有6条光线的圆形，同时解释道“另外，我们要创建的这个图案有点像一个标志。好吧。对，所以它会有六条边，对吧？”，帮助学生直观理解云图标的结构。

数据揭示了使用模式，表明这种技巧最常用于初始设置阶段和复杂的视觉任务。在第1课中，该技巧的使用高峰出现在步骤3（光线绘制）和步骤4（云朵图标组装）期间，此时教师需要展示多个元素之间的空间关系。到了第2课，这种技巧的使用有所减少，表明随着学生对空间概念越来越熟悉，教师也调整了沟通方式。

4.5.4 绘制 + 解释概念。教师使用批注与语音相结合的方法（见表4），将视觉演示与

表4：标注类型及其对应的言语模式

<table>
<tr>
<td>突出目标与指令
通过清晰的操作指令引导学生关注界面元素

‘试着双击它
然后把它发回来’ [P6]
‘然后你可以……这个……所以所有文本’ [P1]
‘你能把这些删掉吗’ [P6]</td>
<td>建议值与命令
提供具体的参数值并结合操作指令

‘……要设为 16 像素。没错！这样你就可以将它对齐’ [P10]
‘……输入 FFCC33。哦，好，太好了’ [P2]
‘……将其更新为 16 像素而且我们可以这样做。[P6]</td>
</tr>
<tr>
<td>绘图与解释：排列
空间关系和组织结构的视觉演示

“作为主文件，任何更改都会应用于你所看到的内容，并应用于它。” [P10]
‘……继承，但一旦当前这个实例比如是蓝色的’ [P8]
‘这两个形状在一起’ [P1]</td>
<td>绘图与解释：概念
视觉演示与概念性解释相结合

“它们……在同一条线上，所以……底部应该在同一条线上。” [P2]
“你可能会在这三个部分之间来回切换” [P6]
“所以把它复制一份，这样就有一个像是这样的。” [P1]</td>
</tr>
</table>

概念解释结合起来，以讲解复杂的软件概念。教师使用这种方法勾勒示例和流程，解释组件关系、父子层级结构和自动布局原则等抽象概念。

例如，一位教师（P1-T）绘制了维恩图来解释相交运算，一边草绘相互重叠的圆圈，一边说明布尔运算的原理，并说道“这两个形状是放在一起的，”。另一位教师（P2-T）绘制了两个圆圈，并在每个圆圈中写上“12”，以解释像素尺寸的概念，同时说道“你画两个圆圈，每个都是 *12*。每个 *12*，好吗？对。”这种视觉呈现帮助学生理解空间关系以及所需的数值精度。

父子组件关系在概念上尤其难以解释。一位教师（P10-T）绘制了一张图，展示主组件与实例关系，并解释道“作为主文件，你对主文件所做的任何修改，都会应用到你看到的这个实例上。”（如图 3 所示）另一位教师（P6-T）使用注释来区分主组件和实例组件，在天气卡片实例上方写下“主”，并在主组件上方写下“次要”，随后解释道“一个可能是主组件，另一个可能是第二个实例。”

这种方法通过可视化表示将抽象概念与具体示例衔接起来。视觉标注为复杂关系提供切实的表示，同时使用户无需在头脑中将口头描述映射到界面元素。使用模式表明，该技术在第1课第7步（6个批注）和第2课第8步（12个批注）达到峰值，通常出现在用户需要理解组件集与自动布局功能之间复杂关系的阶段。值得注意的是，教师在第2课的课前介绍中使用了5个批注，在课程开始时绘制出天气卡片，向学生展示最终成果；这一模式是在教师于第1课之后的半结构化访谈中获得反馈后出现的。

4.5.5 局限性。 尽管标注与语音组合颇具成效，但仍出现了若干影响其实用性的局限性。首先，仅使用批注的方法往往不足以应对复杂任务；当仅靠视觉突出显示无法传达必要的操作步骤时，教师需要"取得控制"以提供充分的指导（P5）。其次，批注有时会造成视觉杂乱，尤其是在需要突出显示多个元素时，或先前的批注仍然可见并干扰当前教学时。最后，批注的效果高度依赖于教师的绘制技能以及创建清晰、可识别视觉提示的能力；执行不佳的批注可能会增加困惑，而不是带来清晰。

4.6 屏幕控制 + 语音



在复杂任务中，有策略地使用屏幕控制来配合语音，效果最佳。它尤其适合演示精确的界面交互顺序，而仅凭语音难以充分传达这些内容；当学生迷失方向时，这也能帮助他们“弄清楚该按哪里”。

屏幕控制能够出色地展示点击位置的空间精确性和执行操作时机的时间精确性，解决了教师在口头指导中难以应对的双重挑战。我们的分析发现了两种主要使用模式：操作流程与工具使用演示，以及学生作业纠正。屏幕控制在少量且有目的地使用时效果最佳，教师指出，当学生能够自行纠正时，“就没有必要使用。”

4.6.1 操作流程与工具使用演示。当教师亲自操面元素并解释自己的操作时，便属于操作流程与工具使用演示，例如“我将调整 <形状>，因为……”这种方法将直接操控与情境化口头指导相结合。我们观察到

第1课中有19次演示，第2课中有7次，表明学习初期对动手指导的依赖程度较高。演示主要出现在较复杂的步骤中：第1课的步骤1、3和4，以及第2课的步骤1、2和8。第1课的步骤3（11次演示）和第2课的步骤8（5次演示）出现演示的频率最高，这反映出在这些环节中，仅靠口头指导不足以让学生继续操作。

4.6.2 纠正。 当教师要纠正界面状态，或基于语音的指导不再有效时，便会通过屏幕控制解决问题。当口头指令无法传达所需操作，或有必要立即纠正时，通常会采取这些干预措施。频率数据显示出明显变化：第1课进行了19次演示和0次纠正，而第2课进行了1次演示和6次纠正。第1课中的屏幕控制总计为608秒，第2课为213秒，这表明在学习初期，当学生对界面还不熟悉时，他们更加依赖直接控制。

纠正的具体示例包括教师接管控制权，以修正错误的图层排列。例如，学生P3-S没有使用布尔运算来创建云图标，而是不小心创建了相互重叠的圆圈。P1-T进行了干预：“让我来帮你修正”，并迅速纠正了图层结构，同时解释道：“我们需要减去这些圆圈，才能创建云形状。”另一个常见的纠正场景是学生难以进行精确定位，例如学生P7-S无法将太阳光线正确地围绕中心圆对齐。教师P4-T接管控制权并演示道：“我来演示如何使用旋转工具，让这些光线均匀分布。”

教师对纠正持有不同看法。有些教师认为纠正是必要的，P6-T 提到：“有时候你就是得介入并纠正，尤其是当他们感到沮丧的时候。”另一些教师则倾向于克制，正如 P9-T 所说：“我会先让他们自己想办法，但如果他们真的卡住了，我会短暂地接手，示范正确的方法。”学生在遇到困难时通常很感激教师的帮助，P2-S 回忆道：“他们帮我纠正时确实有帮助，但我希望他们能解释一下为什么错了。”

4.6.3 局限性。尽管屏幕控制在特定干预情境中效果显著，但也存在严重局限，影响了其整体实用性。首先，当教师出乎意料地接管控制权时，学生会感到“震惊/焦虑”，产生可能妨碍学习的负面情绪反应（P4）。这种能动性的突然丧失往往会打乱教学的自然流程，让学生感到无能为力。一名学生（P6-S）描述了这一体验：“老师毫无预警地突然接管控制权，让我觉得很不适应。”其次，教师强烈担忧“过度控制”会损害学生自主性并减少学习机会。一名教师（P2-T）明确表示会避免使用屏幕控制，并说道：“我尽量不使用这个功能……因为我觉得，如果控制得太多，学生就无法亲自获得经验。”

4.7 教师适应性



所有学生在课后测试中都表现出色（10名学生全部完成了测试1，10名学生中有9名完成了测试2），这表明研究中的所有教师都很有效。以下分析将探讨促成这些成功成果的适应性教学策略。

除了结合不同模态之外，教师还会根据学生反应和进展调整教学方法。我们确定了这种适应性的四个维度：（1）教学改进，（2）错误响应，（3）模态切换，以及（4）学生适应。有效的教师会通过语言和视觉提示监测学生的理解情况，并实时调整策略。从第1课到第2课，这一点尤为明显：教师根据他们在第一课中对每位学生的了解，改进了沟通方式。

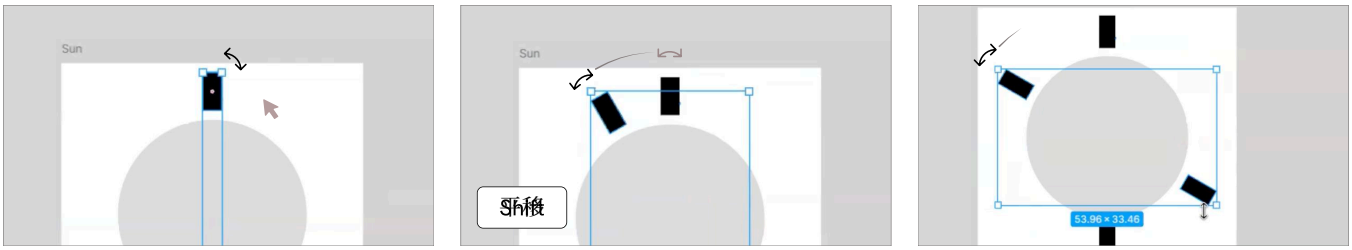
4.7.1 随时间推移的教学调整。 教师在逐渐熟悉学生和界面的过程中，也调整了自己的交流风格。讲解从冗长转为简洁。例如，一位教师（P1-T）起初详细地逐步描述如何调整框架大小，之后则将其简化为一句直接的指令：“你能把框架调整为 *120* 乘 *160* 吗？”数值指定也呈现出类似的变化。早期的教学指令如“这样我们就能在这里更改……好，现在你能把宽度和高度改成 *64* 吗？”后来变成了简短的陈述，例如“*130* 乘 *160*。”操作指令从多步骤的引导转变为简短、明确的提示。概念解释也从冗长的描述转变为简洁的隐喻或总结。

4.7.2 错误响应与恢复。 教师会根据学生遇到困难或犯错的情况调整应对方式。有效的教师不会立即接管控制权，而是先进行语言纠正。如果这样还不够，他们会通过控制屏幕进行简短演示，然后再将控制权交还给学生。例如，一名学生（P5-S）无法从错误中恢复时，教师演示了如何返回最后正确状态，随后撤销了修正，让学生亲自完成操作。这一模式表明，教师会根据困难的严重程度和持续性来调整干预。

4.7.3 基于学生响应的模态切换。 扁平化与联合操作的对比体现了这种适应性教学。尽管这两种操作在视觉上看起来相同，但它们会生成不同的图层结构。当学生询问为什么联合操作后还需要进行扁平化时，教师意识到仅靠演示还不够，于是补充了关于底层系统架构的口头讲解。随着学生逐渐熟悉相关内容，模态选择也随之变化。教师逐渐减少了详细注释和演示，转而采用简洁的口头指导，并以学生掌握了更多知识为前提。

4.7.4 个体学生适应。 教师意识到学生之间存在个体差异，并据此调整教学方式。有些学生需要反复讲解多次并辅以更多示例，而另一些学生则能迅速理解。教师根据这些个体特征调整教学节奏和细节程度。一位教师（P2-T）谈到了适应个体需求所面临的挑战：“我觉得把

图 4：教师通过动手操作界面演示径向图案创建：展示围绕中心圆对矩形对象进行逐步复制（Ctrl+C/Ctrl+V）和受限旋转（Shift+15°），并配合口头指导解释每个操作。



（a）教师："所以，当你完成这个操作后，就会看到类似这样的东西。在此之前，我们要创建一个新的。"

（b）教师："Ctrl+C、Ctrl+V……然后我们要旋转它。所以，一旦……当你点击'Shift'时，它会旋转 15 度。"

(c) 教师："所以，一次移动差不多就是15度。那么，你能再复制一个，然后……"

文本转化为应用……对我来说，向学生解释这件事非常困难，比如告诉他们它在设计中究竟意味着什么。"教师意识到自身在解释方面的局限，因此尝试采用不同的沟通方式，直到找到适合特定学生的方法。

5 讨论

我们对教师行为的分析揭示了教学模态在软件学习情境中如何发挥作用。我们首先识别出一个根本性约束：软件学习要求教学同时传达空间与时间精确性；然而，没有任何单一模态能够在不损害学生能动性的情况下同时擅长这两者。这种精确性与能动性的权衡形成了一种核心张力，教师通过策略性的模态选择与协调来应对。随后，我们考察教师如何调节学生能动性，在推动进展的同时保持学生自主性。这些发现证实了既有的学习原则，包括空间与时间邻接原则 [48, 49]，以及从认知负荷理论 [64, 65] 和共同注意 [33, 52]的角度来看学生能动性所带来的益处。此外，我们将精确性与能动性的权衡和数字领地所有权确定为屏幕内、情境中的软件学习所特有的两项新设计约束。最后，我们探索人工智能辅导系统的设计机遇，使其能够应对这些约束，并超越所考察的三种模态的局限，从而为沟通清晰度、学生记忆和干扰最小化提供更有力的支持。

5.1 时空精确性及其权衡

在各门课程中，教师反复面临一项挑战：传达在空间与时间上都精确的指令。软件学习要求学生执行精确的界面操作序列，以明确在何处操作以及按何种顺序操作；然而，仅靠语音往往无法同时传达这两个维度（例如，"试着让*16*的底部与*30*和*22*的底部处于同一水平"(P3-T)）。学生必须将词语映射到空间位置、把握操作时机并正确排列步骤，这增加了认知负荷，也造成了误解的可能。

我们观察到，教师通过在不同模态之间分配工作来应对这些精确性要求，这与认知负荷理论以及关于时空邻近性的多媒体学习原则相一致 [48, 50, 65]。视觉标注通过将语音与特定屏幕目标联系起来，支持了空间精确性。这降低了外在认知负荷 [64, 65]，并帮助学生识别那些难以仅凭言语推断的元素。演示通过展示在何处操作以及按何种顺序操作，同时提供了空间与时间精确性。这在旁白与操作之间建立了时间邻近性。然而，其空间线索较为短暂，且要求教师取得控制，有些人认为这"过于侵扰"(P8-T)。语音仍然灵活，但在传达细粒度的空间与时间细节方面存在局限。在实践中，教师使用批注进行空间引导，使用演示呈现时间序列，这表明他们的选择反映了对邻接原则的隐性运用。

软件学习需要教学同时传达空间和时间上的具体信息，但没有任何一种模态能很好地兼顾两者。这对人工智能辅导系统构成了核心设计约束。要提供有效支持，就需要根据精确性需求的变化来组合或切换模态。能够同时提供这两种精确性的模态往往会削弱学生能动性。此处，**学生能动性**指学生发起并控制自身行动以及共同视觉焦点的能力。从认知负荷理论的角度来看，能动性关乎由谁主导认知加工；从共同注意的角度来看，则关乎由谁引导注视和行动。个人计算机界面大多遵循单一控制中心范式。因此，通过批注或演示提高精确性，虽能使教学更清晰，却也可能将学生的注意力和控制权转移出去。这种**精确性与能动性的权衡**构成了我们分析的框架：教师如何调节帮助，在确保学习进展的同时维持学生自主权。

5.2 "我是不是帮得太多了？"：教师如何调节学生能动性

教师的驱动力并非只有程序效率。他们还希望维护学生的能动性和所有权意识，

这与先前关于能动性、控制点和即时信息的学习科学理论相一致 [20]。教师会调整干预方式，尽可能让学生保有控制权，同时为他们提供足够的指导，以便继续取得进展。

5.2.1 保留学生能动性的价值。尽管软件学习对精确性要求很高，教师仍经常在不同模态之间切换，以优先保障学生能动性。语音可以在不包办一切的情况下提供指导，批注突出显示应采取行动的位置，而演示则被谨慎使用，以避免凌驾于学生的控制之上。教师通常会在介入前暂停片刻，将光标移动或菜单探索视为学生进行努力加工的迹象 [55]。正如一位教师（P4-T）所指出的，他们希望“给他们时间去探索”，因为学生“可能正要自己弄明白。”另一位教师（P5-T）反思道：“学生就像是在自己好奇地探索。对我来说，（教学和指导）更容易……而不是面对一个只是确确实实等着你下一条命令的学生。”这凸显了主动学习行为和好奇心驱动的探索的价值。

在这一背景下，我们定义了一种理想的能动性状态，其中包含教师参与。这并不意味着学生要独立行动，因为新手往往缺乏启动任务所需的程序性与空间知识。因此，教师会设定目标、突出显示相关界面区域，并为最初的问题空间搭建支架。理想的能动性体现了一种支持性自主：学生在保有决策权的同时，也能获得足够的结构支持，从而采取有意义的行动。

从认知负荷理论的角度来看，理想的能动性由教师和学生共同建构。教师提供线索与限制，以降低外在负荷，同时让学生能够进行图式形成所需的认知操作。这种平衡通过减少无效搜索、同时保留心理整合的机会，促进相关加工。一位学生（P9-S）指出，“他明确指出了左侧和右侧面板，还圈出了我应该按哪个按钮……这一切都非常清楚，”这说明批注通过消除无效搜索降低了外在负荷。无论是过度标注还是过度自主，能动性失配都会增加外在负荷并减少相关加工，妨碍学生协调理解所需的认知步骤。

从共同注意的视角来看，理想的能动性是通过对注意焦点和行动的共同控制来共同调节的。学生通过光标移动和言语表达传达自己的注意优先级，教师则据此调整指导。教师通过批注或演示塑造知觉环境，而学生决定与任务相关行动的节奏和顺序。正如 P6-T 所说，“如果鼠标没有移动……我就会圈出那个按钮，让注意力集中到那里，”这表明教师会回应学生的信号，以保持注意力协调一致。过度自主和教师接管都可能破坏共同的注意框架，使维持连贯任务轨迹所需的共同调节变得不稳定。

5.2.2 数字领地边界。为应对这一矛盾，教师会根据现场情况调整能动性，以此管理协助，这与有关协助困境的研究相符 [2]。这些选择不仅受任务要求影响，也受到围绕许可和数字领地的社会规范影响 [59]。正如 P9-T 所言，一些教师会犹豫是否介入：“我不该插手……我宁愿做批注。”。在实践中，批注是一种“柔性”的边界跨越，它标记了学生的工作区，却不会接管该区域。相比之下，远程控制则构成一种“硬性”的边界跨越，会暂时将所有权转交给教师。这些差异有助于解释为何教师会优先选择批注，并仅在严重故障时才使用远程控制。因此，帮助调节不仅是技术选择，也受到人际规范和共享领地边界的影响。

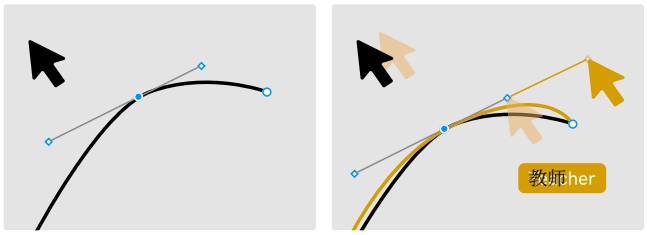
5.2.3 将校准的能动性作为设计目标。 校准的能动性意味着保持控制权的适当平衡，而非偏向教师或学生任何一方。控制权应与任务精度、节奏要求以及学生当前的能力相匹配。人类教师力求维护学生能动性，并通过帮助调节实践弥补失误。人工智能教师应将校准的能动性视为核心设计目标，而非事后考虑的因素 [22, 28]。系统应默认保留学生的控制权，仅当精度要求或失误超出学生继续完成任务的能力时，才增加指导或侵入性。有效的系统需要支持自适应模态选择、实时调整侵入性，并制定能够反映专家教师时机规范的帮助调节策略。

在同一个数字工作区中协作，会给人工智能辅导系统带来独特挑战。与共享工作区并依靠隐含社会契约表明何时适合介入的人类教师不同，人工智能系统可以立即接管屏幕，这使领地边界变得更加脆弱。这引出了几个关键的设计问题：何时接管屏幕在教学上是合理的？应如何标示边界？又该如何将控制权交还给学生？由于人工智能缺乏能够约束人类擅自介入的社会临场感，因此必须明确设定领地性规范，以维护学生能动性。

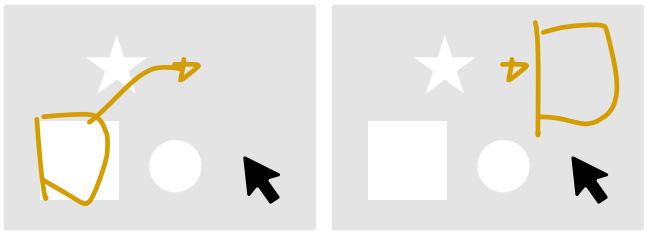
5.3 人工智能辅导系统的设计机遇

参与者观察揭示了当前辅导界面的若干局限。教师指出，直接控制光标让人觉得“过于冒犯”（P8-T），凸显了能动性与精度之间的张力。学生反映，大量批注造成了视觉杂乱，表示“同一处有太多箭头”（P7-S）。教师通过手动删除批注来处理这一问题。学生还表示难以记住各个步骤，希望能有持久笔记或参考资料。教师发现，使用现有工具难以传达抽象概念，例如父子组件关系。其他问题还包括视觉对比度低、缺乏持久参考资料，以及情境框架不足。

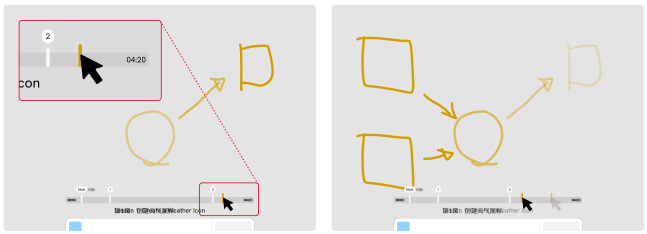
基于这些局限，我们的研究发现指向若干设计机遇。一种幽灵光标功能（图5a）可在不接管控制的情况下展示教师的光标移动，从而提供时间与空间邻近性，同时保留学生的控制权。渐隐批注（图5b）可减少视觉杂乱，同时让指导保留足够长的时间以支持理解。时间轴拖动（图5c）可让学生按需回看批注，从而自主掌控时间信息的教学节奏。一种持久笔记侧边栏可支持记忆，并在复杂任务中降低对工作记忆的依赖。最后，交互式参考（图5d）



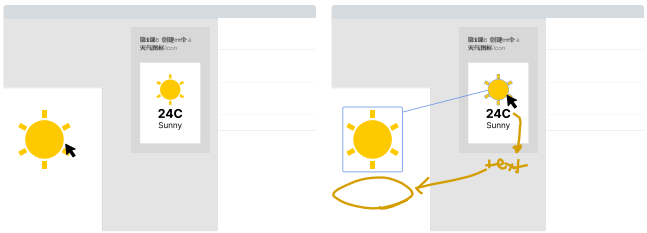
(a) 幽灵光标



(b) 渐隐注释



(c) 时间轴拖动



(d) 交互式参考

图 5：人工智能辅导系统的设计启示：(a) 幽灵光标——向学生显示人工智能导师的光标位置；(b) 渐隐注释——随时间逐渐消失的批注；(c) 时间轴拖动——通过时间导航查看过去的批注；(d) 交互式参考——从参考图像中提取并加以操作的内容元素。

将通过使参考材料可操作而非静态，帮助学生理解抽象关系。这些想法共同表明，人工智能辅导系统如何能够超越语音、批注和屏幕控制的局限，从而提高清晰度、支持记忆并尽量减少干扰。

6 局限性与未来工作

本研究对屏幕上软件教学中的模态协调提供了初步认识，但仍存在若干局限。我们的对 Figma 的关注可能会限制其向其他领域（例如，编程）的可推广性。未来工作应检验相同的模态模式在更广泛的工具范围内是否仍然成立。例如，在编程环境中，精确性是句法层面的，而领地则涉及代码所有权。在此类领域中重复这项工作，将有助于澄清空间密度、时间耦合和概念复杂度是否能够预测模态侧重，以及是否需要不同的协调策略。

受控实验室设置不同于真实世界学习情境，且可能影响了教学。本文所发现的精确性—能动性权衡可能不只存在于语音、批注和控制方面：例如，注视高亮（精确度高、干扰性低）可能在可用时优于批注；对这类模态进行测试，可以揭示权衡结构是否不受模态影响，还是会因可供性成本而改变。仍有一些动态变化有待探究：随着学习者不断进步，模态平衡应如何调整，以及人工智能应在何时主动减少支架。另一个相关问题是，当高精度模态的交互成本降低时（例如悬停触发注释），权衡关系是否依然存在。

我们的样本仅包含十对师生配对，限制了研究结果的普遍性；我们观察到的差异表明，个体差异会影响模态选择。未来研究应考察学习者特征（如先前经验）如何与模态组合相互作用，为个性化 AI 辅导提供参考。

最后，我们分析的是教学行为，而非学习结果；此外，我们的一对一同步环境可能无法反映其他辅助实践或典型课堂情境。一个尚待探讨的问题是，批注和演示是否会使学习者对相同操作流程形成不同的心智模型。未来研究应将模态协调与可衡量的学习、保持和迁移成效联系起来，考察异步形式以及更轻量的屏幕引导，并探索如何在学习者需求可能相互冲突的群组配置（如一对多课堂）中编排不同模态。增强现实等新兴技术可能进一步拓展设计空间；这些研究方向或可支持本研究未涉及的模态类型。

7 结论

通过对十对师生配对（N=10）开展观察性研究，我们考察了人类教师如何协调语音、视觉批注和远程屏幕控制，以教授功能丰富的软件操作流程。我们的研究发现验证并拓展了交互式工作空间中的基础学习理论：有效的模态协调支持多媒体学习的时空邻近性原则，并促进共同注意。同时，我们指出了两项领域特定设计约束——精确性与能动性的权衡以及数字领地性。远程控制等高精度模态虽能降低复杂操作流程中的外在认知负荷，但也可能侵犯学习者自主性和领地边界。

我们认为，人工智能辅导系统的设计不应仅仅将其视为信息传递引擎，还应将其打造为能够平衡这些相互竞争的约束的自适应系统。系统应优先确保经过校准的能动性，仅在维持教学进程所必需时，才逐步升级干预的精确度，从低精度转向高精度

干预。通过将这些理论见解与幽灵光标或渐隐注释等交互机制相结合，未来的界面能够超越人类教学的局限，提供精准且能感知情境的软件指导。

致谢

我们感谢新加坡管理大学人机交互社区的各位成员，在整个项目期间给予我们深思熟虑的反馈和持续的鼓励。我们尤其感谢 EmilyAurelia 、 Yeo Shunyi 、 Justin G 和 Georgia Zhang；他们的慷慨支持、热情投入以及富有洞察力的交流，极大地完善了我们的研究设计与分析。

我们向所有研究参与者——无论是专家还是新手——致以诚挚的谢意，感谢他们投入时间，迎接学习新技能的挑战，并为这项工作作出贡献。正是由于这些贡献，本研究才得以得到切实改善。

本研究鸣谢谷歌提供的支持（内部代码：T050309）。

参考文献

[1] Shaaron Ainsworth. 2006. DeFT：用于考量多重表征学习的概念框架。《学习与教学》16, 3 (2006), 183–198. [2] Vincent Aleven and Kenneth R. Koedinger. 2007. 在认知导师系统实验中探索辅助困境。《教育心理学评论》19, 3 (2007), 239–264. [3] Vincent Aleven, Ido Roll, Bruce M. McLaren, and Kenneth R. Koedinger. 2006. 迈向元认知辅导：认知导师系统中的寻求帮助模型。载于第八届国际智能辅导系统会议（*ITS 2006*）论文集。施普林格, 227–239. [4] Martha W Alibali, Mitchell J Nathan, R Breckinridge Church, Matthew S Wolf- gram, Suyeon Kim, and Eric J Knuth. 2013. 数学课堂中教师的手势与语音：通过解决难点构建共同基础。*ZDM* 45, 3 (2013), 425–440. [5] Olivier Bau and Wendy E Mackay. 2008. OctoPocus：用于学习基于手势的命令集的动态指南。载于第*21*届*ACM*用户界面软件与技术年度研讨会论文集。37–46. [6] Stéphanie Buisine and Jean-Claude Martin. 2007. 多媒体演示中动画智能体行为的语音—手势协作效应。与计算机交互19, 4 (2007), 484–493. [7] John M Carroll. 1990.《纽伦堡漏斗：实用计算机技能的极简教学设计》。麻省理工学院出版社。[8] John M Carroll and Mary Beth Rosson. 1987. 主动用户悖论。载于《思想的界面：人机交互的认知层面》。80–111. [9] Michelene T.H. Chi, Stephanie A. Siler, Heisawn Jeong, Takashi Yamauchi, and Robert G. Hausmann. 2001. 从人工辅导中学习。《认知科学》25, 4 (2001), 471–533. [10] Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. 2012. MixT：自动生成分步混合媒体教程。载于第*25*届*ACM*用户界面软件与技术年度研讨会论文集。93–102.[11] Parmit K Chilana, Nathaniel Hudson, Srinjita Bhaduri, Prashant Shashikumar, and Shaun Kane. 2018. 支持远程实时专家帮助：3D建模新手面临的机遇与挑战。载于*2018*年*IEEE*可视化语言与以人为中心的计算研讨会（*VL/HCC*）。IEEE，157–166. [12] Parmit K Chilana, Amy J Ko, and Jacob O Wobbrock. 2012. LemonAid：面向Web应用程序的基于选择、由众包提供的情境帮助。载于*SIGCHI*计算机系统人因会议论文集。1549–1558. [13] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. 支持用户界面中从新手到专家的转变。《*ACM*计算调查》（*CSUR*）47, 2 (2014), 1–36. [14] Susan Wagner Cook, Ryan G Duffy, and Kimberly M Fenn. 2013. 观察手势后的学习巩固与迁移。《儿童发展》84, 6 (2013), 1863–1871. [15] Jennifer Fernquist, Tovi Grossman, and George Fitzmaurice. 2011. Sketch-Sketch Revolution：用于引导式绘图和应用学习的引人入胜的教程系统。载于第*24*届*ACM*用户界面软件与技术年度研讨会论文集。373–382.

[16] Adam Fourney、Ben Lafreniere、Parmit Chilana 和 Michael Terry。2014。Inter-Twine：创建跨应用信息气味，以支持软件的协同使用。载于《第27届 ACM用户界面软件与技术研讨会论文集》。429–438。[17] C Ailie Fraser、Tricia J Ngoon、Mira Dontcheva 和 Scott Klemmer。2019。Re-Play：在不同软件应用程序中情境化呈现学习视频。载于《2019年 CHI计算机系统人因大会论文集》。1–13。[18] Daniel Gaspar-Figueiredo。2023。交互式学习：使用强化学习进行用户界面自适应。arXiv预印本 arXiv:2312.07216（2023）。[19] Daniel Gaspar-Figueiredo、Marta Fernández-Diego、Ruben Nuredini、Silvia Abrahão 和 Emilio Insfrán。2024。用于用户界面智能自适应的强化学习框架。载于《第16届 ACM SIGCHI交互式计算系统工程研讨会配套论文集》。40–48。[20] James Paul Gee。2003。《电子游戏教给我们的学习与读写能力》。帕尔格雷夫·麦克米伦，纽约。[21] Arthur C. Graesser、Paul Chipman、Brent C. Haynes 和 Andrew Olney。2005。AutoTutor：历史与贡献综述。《国际人工智能教育杂志》15，2（2005），97–126。[22] Arthur C Graesser、Shulan Lu、George Tanner Jackson、Heather Hite Mitchell、Mathew Ventura、Andrew Olney 和 Max M Louwerse。2004。AutoTutor：一款采用自然语言对话的辅导系统。《行为研究方法、仪器与计算机》36，2（2004），180–192。[23] Arthur C. Graesser、Natalie K. Person 和 Derek Harter。1999。AutoTutor：一款采用自然语言对话的辅导系统。《行为研究方法、仪器与计算机》31，2（1999），293–305。[24] Tovi Grossman 和 George Fitzmaurice。2010。ToolClips：对用于功能理解的情境化视频辅助的研究。载于《SIGCHI计算机系统人因会议论文集》。1515–1524。[25] Susan M Harrison。1995。在图形用户界面中，比较静态、动画或无插图的在线帮助与书面或口头说明。载于《SIGCHI计算机系统人因会议论文集》。82–89。[26] Christian Heath 和 Paul Luff。1991。协作活动与技术设计：伦敦地铁控制室中的任务协调。载于《第二届欧洲计算机支持协同工作会议论文集 ECSCW'91》。施普林格，65–80。[27] Irene Hou、Sophia Mettille、Owen Man、Zhuo Li、Cynthia Zastudil 和 Stephen MacNeil。2024。生成式人工智能对计算机专业学生求助偏好的影响。载于《第26届澳大拉西亚计算机教育会议论文集》。39–48。[28] Gaoping Huang、Xun Qian、Tianyi Wang、Fagun Patel、Maitreya Sreeram、Yuanzhi Cao、Karthik Ramani 和 Alexander J Quinn。2021。Adaptutar：用于增强现实中机器操作任务的自适应辅导系统。载于《2021年 CHI计算机系统人因大会论文集》。1–15。[29] Ananya Ipsita、Levi Erickson、Yangzi Dong、Joey Huang、Alexa K Bushinski、Sraven Saradhi、Ana M Villanueva、Kylie A Peppler、Thomas S Redick 和 Karthik Ramani。2022。通过构建虚拟现实焊接模拟器模型，推动无障碍且可扩展的培训。载于《2022年 CHI计算机系统人因大会论文集》。1–21。[30] Xiaofu Jin、Xiaozhu Hu、Xiaoying Wei 和 Mingming Fan。2022。Synapse：通过试错支持，为老年人使用智能手机应用提供基于演示的交互式指导。《ACM交互、移动、可穿戴与泛在技术汇刊》6，3（2022），1–24。[31] W. Lewis Johnson、Jeff Rickel 和 James C. Lester。2000。动画教学代理：交互式学习环境中的面对面交互。《国际人工智能教育杂志》11，1（2000），47–78。[32] Nikhita Joshi、Justin Matejka、Fraser Anderson、Tovi Grossman 和 George Fitzmaurice。2020。Micromentor：三分钟以内的点对点软件帮助会话。载于《2020年 CHI计算机系统人因大会论文集》。1–13。[33] Jina Kang、Yiqiu Zhou、Robin Jephthah Rajarathinam、Yuanru Tan 和 David Williamson Shaffer。2024。揭示共同注意动态：考察沉浸式协作式天文学模拟中的多模态参与。《计算机与教育》213（2024）。[34] Manu Kapur。2008。生产性失败。《认知与教学》26，3（2008），379–424。[35] Majeed Kazemitabaar、Xinying Hou、Austin Henley、Barbara Jane Ericson、David Weintrop 和 Tovi Grossman。2023。新手如何在自主进度学习环境中使用基于大语言模型的代码生成器解决 CS1编程任务。载于《第23届 Koli Calling国际计算机教育研究会议论文集》。1–12。[36] Caitlin Kelleher 和 Randy Pausch。2005。基于模板的教程：设计与评估。载于《SIGCHI计算机系统人因会议论文集》。541–550。[37] Anjali Khurana、Hariharan Subramonyam 和 Parmit K Chilana。2024。基于大语言模型的助手为何以及何时会出错：探究基于提示词的交互在软件求助中的有效性。载于《第29届

智能用户界面国际会议论文集》。288–303。[38] Kimia Kiani、Parmit K. Chilana、Andrea Bunt、Tovi Grossman 和 George W. Fitzmaurice。2020。“我会直接问别人”：在现代职场中学习功能丰富的设计软件。2020年 IEEE可视化语言与以人为中心的计算研讨会（VL/HCC）（2020），1–10。 221086012 https://api.semanticscholar.org/CorpusID:[39] Bran Knowles 和 Vicki L Hanson。2018。老年技术（非）用户的智慧。《ACM通讯》61，3（2018），72–77。[40] Gordon Kurtenbach 和 William Buxton。1994。使用标记菜单时的用户学习与表现。载于《SIGCHI计算机系统人因会议论文集》。258–264。[41] Joel Lanir、Kellogg S Booth 和 Leah Findlater。2008。观察演讲者使用视觉辅助工具的方式，以指导课堂演示软件的设计。载于《SIGCHI计算机系统人因会议论文集》。695–704。[42] Ard W Lazonder 和 Hans van der Meij。1993。极简手册：少真的就是多吗？《国际人机研究杂志》39，5（1993），729–752。[43] Jimin Lee、Steven-Shine Chen 和 Paul Pu Liang。2025。Interactive Sketchpad：用于协作式视觉问题解决的多模态辅导系统。载于《CHI计算机系统人因大会扩展摘要论文集》。1–14。[44] Yimeng Liu 和 Misha Sra。2025。设计支架式界面，以提升专业软件中的学习与表现。arXiv预印本 arXiv:2505.12101（2025）。[45] Ziyi Liu、Zhengzhe Zhu、Enze Jiang、Feichi Huang、Ana M Villanueva、Xun Qian、Tianyi Wang 和 Karthik Ramani。2023。Instrumentar：通过记录具身演示，自动生成用于操作数字仪器的增强现实教程。载于《2023年 CHI计算机系统人因大会论文集》。1–17。[46] Yuhan Luo、Bongshin Lee、Donghee Yvette Wohn、Amanda L Rebar、David E Conroy 和 Eun Kyoung Choe。2018。该休息了：通过休息提醒系统了解信息工作者的久坐行为。载于《2018年 CHI计算机系统人因大会论文集》。1–14。[47] Bhesh Mainali。2021。数学教学与学习中的表征。《数学、科学与技术教育国际期刊》9，1（2021），1–21。[48] Richard E Mayer。2002。多媒体学习。载于《学习与动机心理学》第41卷。爱思唯尔，85–139。[49] Richard E. Mayer。2023。多媒体学习的过去、现在与未来。《教育心理学评论》（2023）。https://doi.org/10.1007/s10648-023-09842-1[50] Richard E Mayer、Julie Heiser 和 Steve Lonn。2001。多媒体学习的认知限制：呈现更多材料反而会导致理解更少。《教育心理学杂志》93，1（2001），187。[51] Richard E. Mayer 和 Roxana Moreno。1998。多媒体学习的认知理论：对设计原则的启示。《教育心理学杂志》91，2（1998），358–368。[52] H. Oura、R. Yoshikawa、F. Mitsunaga、T. Kondo、W. Fushikida 和 H. Kato。2024。在大学课堂中，通过多图像和光标共享支持混合式小组协作中的共同注意。载于《第17届计算机支持的协作学习国际会议论文集（CSCL 2024）》，J. Clarke-Midura、I. Kollar、X. Gu 和 C. D’Angelo（编）。ISLS，371–372。https://repository.isls.org/bitstream/1/10560/1/CSCL2024_371-372.pdf[53] Susan Palmiter 和 Jay Elkerton。1991。对学习计算机任务的动画演示进行评估。载于《SIGCHI计算机系统人因会议论文集》。257–263。[54] Roy D Pea。2018。支架及相关学习、教育和人类活动理论概念的社会与技术维度。载于《支架》。心理学出版社，423–451。[55] Henry L Roediger 和 Andrew C Butler。2011。提取练习在长期记忆保持中的关键作用。《认知科学趋势》15，1（2011），20–27。[56] Ido Roll、Vincent Aleven 和 Kenneth R. Koedinger。2007。智能辅导系统中的求助行为。《国际人工智能教育杂志》17，3（2007），267–290。[57] Joey Scarr、Andy Cockburn、Carl Gutwin 和 Andrea Bunt。2012。使用 CommandMaps 改进命令选择。载于《SIGCHI计算机系统人因会议论文集》。257–266。[58] Bertrand Schneider 和 Roy Pea。2017。实时相互注视感知可提升协作学习与协作质量。载于《第12届计算机支持的协作学习国际会议论文集》。施普林格， 1 8 10 1007 978 3 319 45001 8_7 – 。[59] https://doi.org/ 。/ – – – – Stacey D. Scott、M. Sheelagh T. Carpendale 和 Kori M. Inkpen。2004。协作式桌面工作空间中的领地性。载于《2004年 ACM计算机支持的协同工作会议论文集（CSCW ’04）》。ACM，纽约，美国， 294 303 10 1145 1031607 1031655 – 。[60] https://doi.org/ 。/ 。Kumar Shridhar、Jakub Macina、Mennatallah El-Assady、Tanmay Sinha、Manu Kapur 和 Mrinmaya Sachan。2022。自动生成用于教授数学应用题的苏格拉底式子问题。arXiv预印本 arXiv:2211.12835（2022）。

[61] Melissa A Singer 和 Susan Goldin-Meadow。2005。儿童会在教师手势与语音表达不同步时学习。《心理科学》16，2（2005），85–89。[62] Randall Spain、Wookhee Min、Vikram Kumaran、Jay Pande、Jason Saville 和 James Lester。2025。将大语言模型应用于自适应团队培训中的对话与沟通分析。《国际人工智能教育杂志》（2025），1–35。[63] Thomas FM Stewart。1976。显示器与软件界面。应用人体工程学7，3（1976），137–146。[64] John Sweller。1988。问题解决过程中的认知负荷：对学习的影响。《认知科学》12，2（1988），257–285。[65] John Sweller。2011。认知负荷理论。载于《学习与动机心理学》。第55卷。爱思唯尔，37–76。[66] John Sweller 和 Marvin Levine。1982。目标具体性对手段—目的分析与学习的影响。《实验心理学杂志：学习、记忆与认知》8，5（1982），463。[67] Teerapaun Tanprasert、Jiamin Dai 和 Joanna McGrenere。2024。HelpCall：通过视频会议为老年人设计非正式技术支持。载于*CHI*计算机系统人因大会论文集。1–23。[68] Jingying Wang、Jingjing Zhang、Juana Nicoll Capizzano、Matthew Sigakis、Xu Wang 和 Vitaliy Popov。2025。eXplainMR：生成实时文本与视觉解释，以促进混合现实中的超声学习。载于*2025*年*CHI*计算机系统人因大会论文集。1–18。[69] Rose E Wang、Ana T Ribeiro、Carly D Robinson、Susanna Loeb 和 Dora Demszky。2024。导师副驾驶：一种用于扩大实时专业指导规模的人机协作方法。*arXiv*预印本 *arXiv:2410.03017*（2024）。[70] Maheshya Weerasinghe、Aaron Quigley、Klen Čopič Pucihar、Alice Toniolo、Angela Miguel 和 Matjaž Kljun。2022。Arigat¯o：自适应引导对增强现实学习环境中参与度和表现的影响。《*IEEE*可视化与计算机图形学汇刊》28，11（2022），3737–3747。[71] Robin Wiethüchter、Saikishore Kalloori 和 David Lindlbauer。2024。平衡自主性：探究序列任务中用户控制与自动化引导系统。载于*2024*年交互表面与空间大会伴随论文集。73–78。

A 附录

A.1 会话环境与硬件规格

所有会话均在受控的实验室环境中进行。教师和学生面对面就坐，中间隔着一块轻质隔板，以避免出现面对面的非言语提示。两位参与者使用以下设备加入同一个 Zoom 通话：

- **教师：** 16 英寸 Legion Pro 5 16IRX9 笔记本电脑（Intel i9-14900HX，32GB 内存）
- **学生：** 15 英寸 Legion 5 15ACH6H 笔记本电脑（AMD Ryzen 7 5800H，16GB 内存）
- **批注平板电脑：** 12.9 英寸 iPad Pro （第 4 代），配备 Apple Pencil Pro

所有互动均以全分辨率进行屏幕录制和音频录制，记录 Zoom 批注和远程控制操作。

A.2 完整会话流程与时间安排

每次会话均遵循标准化结构，持续时间约为 1.5 小时：

(1) 会前准备（约15分钟）： 知情同意、人口统计信息核实、技术检查、研究预期概述，以及对 Zoom 的屏幕共享、批注和远程控制功能进行简短介绍。还为教师安排了简短的练习时间，以确保其基本熟悉这些工具。**(2) 任务1：天气图标（15–30分钟）：** 教师主导的教学，随后进行时长10分钟的迁移测试和简短的半结构式访谈。

(3) **休息（约5分钟）**
(4) **任务2：天气卡片（15–30分钟）：**教师主导的教学，随后进行第二次迁移测试和最后一次访谈。

教师被要求自然地进行教学，并对其模态选择保有完全的自主性。研究未对教师先前的 Zoom 熟练程度进行正式评估。

A.3 多模态数据处理流程见图7。

A.3.1 步骤识别与远程控制提取。人工审阅录制内容，并根据任务结构将每次会话划分为教程步骤。由于远程控制时段相对少见且具有教学重要性，因此将其识别出来并导出为单独片段。

A.3.2 视觉标注检测与处理。通过逐帧分析（每秒采样 1 帧），采用基于颜色的检测方法识别教师批注中的蓝色像素（# 2f8cfc，Zoom 默认的批注颜色）。处理画框以提取叠加层区域、应用颜色阈值处理，并识别连续的标注时段。当像素数超过 500 时，即判定为活动时段。人工核验各个标注时段，以排除界面元素造成的误报，并确认标注边界。

每个完整的批注（例如箭头、圆形、草图）都被视为一个语义单元。除非递增笔画（例如部分射线线段）构成有意义的教学单元，否则不单独计数。为每条批注添加时间戳，并将其与相应的课程步骤关联。

A.3.3 音频处理与转录。音频从 Zoom 录制文件（.m4a）中提取，并使用 OpenAI Whisper API 进行处理。远程控制片段采用 5 秒预缓冲；视觉批注上下文使用 7 秒片段（批注时间戳之前 5 秒、之后 2 秒）。所有转录内容均经过人工校正以确保准确性，并与课程步骤、检测到的批注及远程控制事件进行时间对齐。

A.3.4 数据整合与编码框架。多模态数据流通过精确的时间戳匹配进行时间对齐。关键处理步骤经过人工核验，交叉验证操作流程确保自动编码结果与人工编码结果保持一致。选取了一组具有代表性的师生配对

根据模态使用多样性进行选择。编码在话语层面进行，并映射到课程步骤标记。编码者间信度通过独立编码并以共识解决分歧的方式确立。

A.4 研究任务

课程1：天气图标

第一项任务聚焦于矢量操作基础，要求参与者使用基本几何形状创建两个自定义天气图标（太阳和多云）。这项任务旨在让学生初步了解 *Figma* 的基本概念。学生学习了创建画框、操作基本形状（圆圈、线条）、应用变换（旋转、缩放）以及使用布尔运算（并集、减去）。学生通过排列



图6：教学课程中使用的两项研究任务概览：任务1（天气图标）和任务2（天气卡片）。

中心圆周围的放射状线条来创建太阳图标，并通过布尔运算组合多个圆圈，形成带有新月细节的云形状，从而创建多云图标。两个图标都需要应用颜色，最后扁平化为单个矢量对象，并创建组件集。

结构：7个步骤，预计15-30分钟（要求：基本计算机技能，无Figma经验。）

第2课：天气卡片

第二个任务引入了组件，要求学生创建可复用的天气卡片，并融入第*1*课中的图标。学生学习了更高级的概念，包括组件创建、实例管理、自动布局原则、排版工具以及主组件与实例关系。学生设计了一个天气卡片组件，其中包含图标、温度文本和状况标签。随后，他们创建了两个包含不同天气数据的实例，从而了解组件变体的工作方式，以及主组件的更改如何传播到所有实例。

结构：8个步骤，预计15-30分钟（要求：完成第1课，对Figma有基本了解。）

图8展示了参与者P10的叠加斜率感知分析，呈现了两节教学课程中蓝色像素密度与标注活动之间的时间关系。该可视化展示了软件教学过程中视觉注意模式的动态特性，其中棕色线条表示连续的蓝色像素密度，蓝色条形则标示标注活动较高的时段。

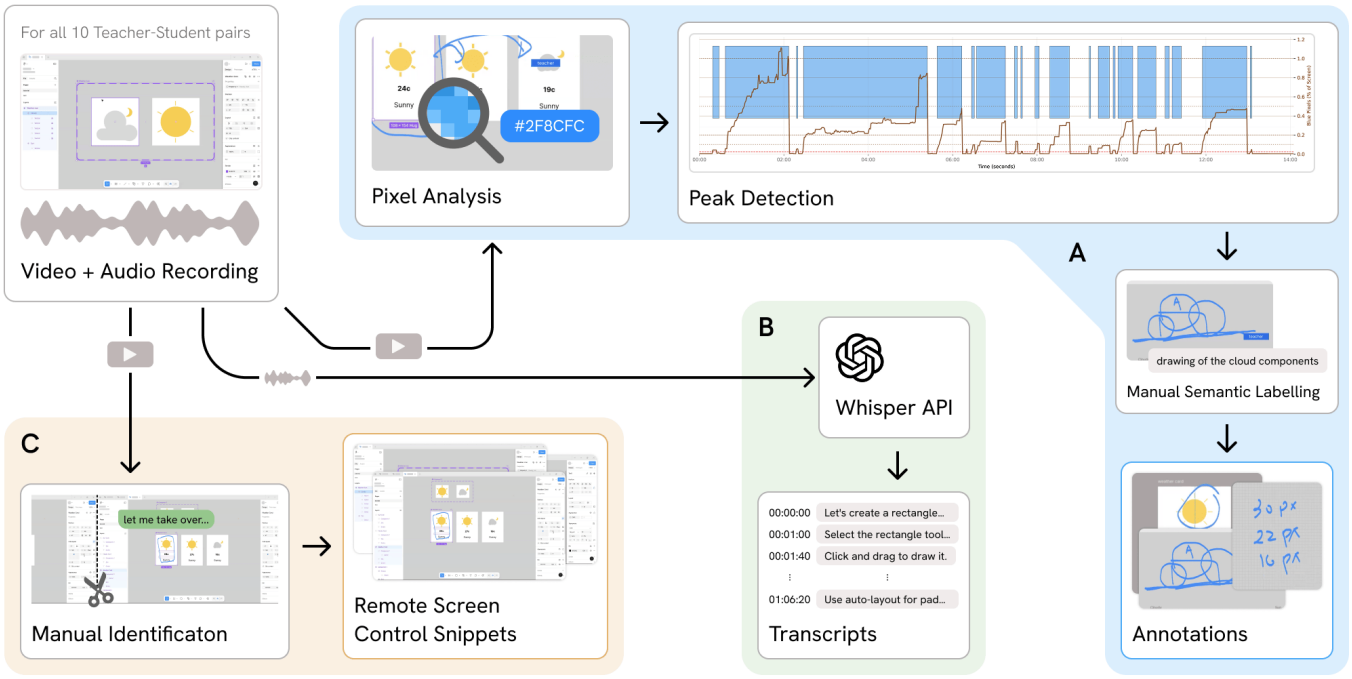


图7：多模态数据处理管线，展示了三条并行的分析流程：（A）视觉标注处理、（B）音频转录和（C）远程屏幕控制提取。

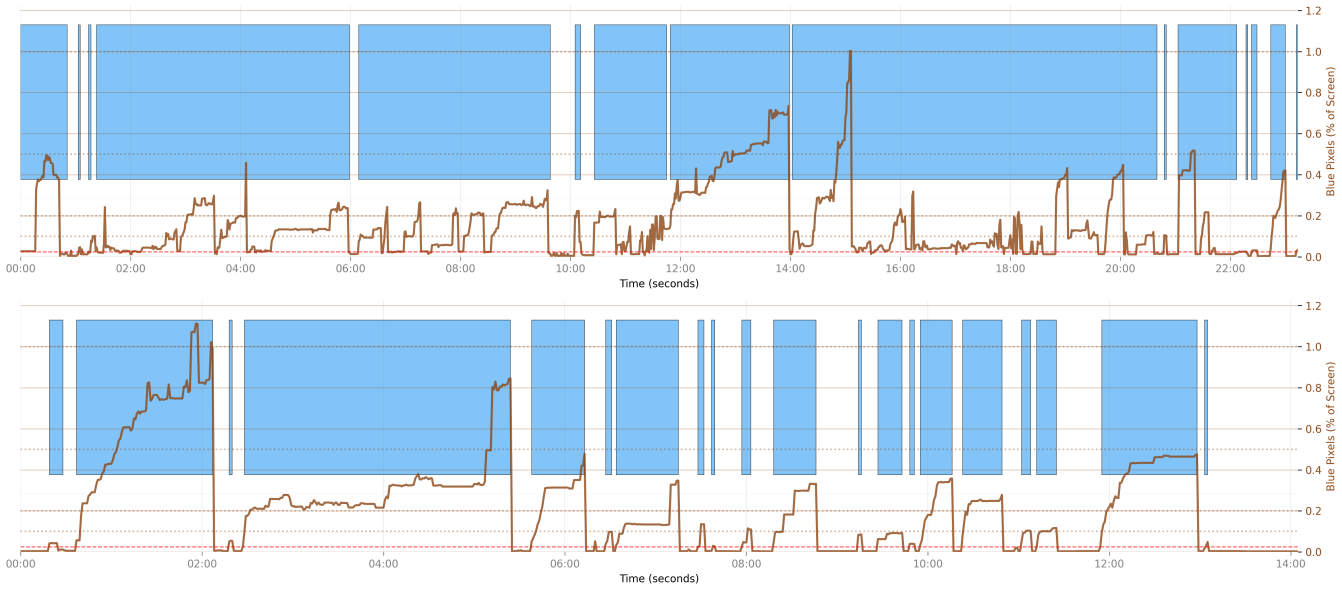


图8：参与者 P10 的叠加斜率感知分析，展示了两节课程中的蓝色像素密度（棕色线条）和标注活动时段（蓝色条）。该可视化揭示了软件教学过程中的视觉注意力和标注行为的时间模式。