arXiv:2611.28951v1 [physics.ed-ph] 21 Nov 2026

本科物理学中的人工智能生成的支持、例题与讲义式学习：一项比较研究

LarissaHahn[1*]，Marie AylinFalkenberg[1]，MartenHogel¨ucht[1]，Pascal Klein[1]

[1]哥廷根大学物理教育研究，Friedrich-Hund-Platz 1，德国哥廷根 37077。

*通讯作者。电子邮箱：larissa.hahn@uni-goettingen.de；贡献作者：aylin.falkenberg@uni-goettingen.de；marten.hogeluecht@uni-goettingen.de；pascal.klein@uni-goettingen.de；

摘要

生成式人工智能日益被用于支持学生学习和解决问题，但其教学价值取决于它如何融入既有的学习方法。例题通过讲解解题步骤和推理过程来引导学习者；对话式人工智能则可以通过回答个别问题进一步拓展例题的作用。为单独考察这种附加价值，我们比较了由聊天机器人呈现的智能例题、与之匹配的静态例题和一份讲义。本科电动力学课程的九个习题课小组被整群随机分配至三种条件，86 名学生独立完成一道高斯定律问题。所有条件下的概念测验准确率和自我效能感均显著提高，但各组的变化没有显著差异（概念测验准确率为 $p = .07$，自我效能感为 $p = .38$）。四周后，在参与人数减少且参与具有选择性的样本中，同时考虑到原有的组间差异，静态例题条件下的学生得分高于讲义条件下的学生（$p = .004$，$d = 1.22$），而两种例题条件之间没有显著差异。与讲义相比，智能例题的可用性更高，但感知教育价值并未提高（$p = .011$，$d = 0.88$）。在随后关于安培定律的家庭作业中，学生可以自由选择学习材料（$N = 46$），$62 - 71\%$ 的学生报告称，在各条件下都只使用了静态例题；至多有 8% 的学生只使用了智能例题，而 $15 - 31\%$ 的学生使用

了两者都没有使用。因此，在当前条件下，对话式呈现并未显示出明确的额外学习益处。更广泛地说，学生给予好评的学习形式，未必是他们自愿选择采用的形式。

关键词： 例题、生成式人工智能、认知参与、比较条件、物理教育研究、电动力学

1 引言

生成式人工智能（人工智能）正迅速成为高等教育的一部分。除了广泛普及的通用型聊天机器人外，大学也开始大规模提供由机构管理的人工智能平台和面向特定课程的辅导系统。这些发展带来了即时、个性化支持的新机遇，但也引出了一个核心教育问题：在何种教学条件下，对话式人工智能能够支持学习，而不仅仅是促进任务完成？要回答这一问题，评估必须以既有的学习与教学原则为基础，而不能仅仅依据人工智能系统的技术能力。

例题是这类成熟教学方法之一。通过将问题与结构化解答一并呈现，例题可以帮助学习者识别相关的问题特征、理解解题步骤，并发展可迁移的问题解决图式（Atkinson 等人，2000；Renkl，2014；Sweller 和 Cooper，1985年）。然而，传统的静态例题为所有学习者提供相同的指导，无法回应个别问题或困难。对话式人工智能或可拓展这一形式，让学习者能够请求解释、厘清个别步骤，并在需要时获得额外支持。因此，人工智能支持的例题或可将基于例题学习的结构化指导与聊天机器人的响应能力结合起来。

这种对话功能是否能带来额外的教学价值，关键取决于比较条件。将其与传统课程材料（如讲义）进行比较，反映的不仅是人工智能使用方面的差异，也包括教学设计方面的差异。相比之下，与具有相同核心解题结构的匹配静态例题进行比较，则能更直接地考察对话式支持所带来的附加价值。这一区别十分重要，因为有关教育聊天机器人和人工智能支持的辅导的研究结果仍存在异质性，而且似乎很大程度上取决于系统的设计方式及其融入教学的方式（Kuhail 等人，2023年；Ma 等人，2014；Wollny 等人，2021年；Wu 和 Yu，2024年）。因此，评估人工智能支持的学习环境，既需要恰当的非人工智能对照条件，也需要考察超越表现的结果，包括学习者的认知与动机体验。

另一个挑战在于，学生是否会在自愿的情况下使用人工智能支持的学习材料。提供基于人工智能的学习环境并不一定会促使学生采用它。学生可能看重即时且个性化的帮助，同时也对信息不可靠、

过度依赖人工智能、学术诚信或独立问题解决技能的丧失表示担忧（Balabdaoui et al., 2024年；Johnston et al., 2024年）。现有研究往往关注学生的总体态度或使用意向，而非在真实课程情境中学生在人工智能支持的学习材料与传统学习材料之间的实际选择。因此，考察教学效果和自愿的材料选择，对于理解人工智能支持的学习环境在教育实践中的潜力至关重要。

本研究在一门本科电动力学课程中探讨了这些问题。在该课程中，学生在解决物理问题时，必须协调概念、空间和数学推理。在一项围绕高斯定律开展的受控干预中，学生分别使用人工智能支持的例题（下文称为智能例题）、静态例题或讲义进行学习。研究从概念测试表现、自我效能感、认知负荷、可用性和感知教育价值等方面对这些条件进行了比较。在随后围绕安培定律开展的探索阶段，学生可以在可用的学习材料之间自由选择，研究者则考察了他们的选择与其对人工智能使用的担忧及感知障碍之间的关系。通过将受控比较与对自愿采用情况的调查相结合，本研究既考察了智能例题所增加的教学价值，也考察了学生是否会在更真实的学习条件下选择使用它。

2 理论背景

2.1 物理教育中的例题

问题解决是大学物理学习的核心组成部分。然而，对于新手而言，独立解决不熟悉的问题可能会带来很大的负担，因为他们尚未形成能够指导高效解题所需的知识结构。一个例题会呈现一个问题及其结构化解答，使学习者能够研究解答是如何构建的，而不必通过搜索自行发现。在数学和物理学等结构良好的领域中，例题因此能够比传统的问题解决更有效地支持问题解决技能的初步习得，这一发现被称为例题效应（Atkinson 等人，2000年；Renkl，2014年；Sweller 和 Cooper，1985年）。

认知负荷理论为这一效应提供了一种普遍的解释（Sweller et al., 1998年）。由于工作记忆有限，新手可能会投入大量资源来搜索解题程序。例题可以减少这种无效搜索，并引导学习者关注问题特征、底层原理与解题步骤之间的关系。因此，例题的教学价值不仅在于提供答案，还在于让解题的结构与推理过程更易于学习。

然而，有效的例题学习仍要求学习者主动处理这些推理过程。成功的学习者往往会生成 自我解释 ，将解题步骤与底层原理联系起来，并预判解题过程将如何推进（Chi 等人，1989；Renkl，1997）。在物理学中，有成效的自我解释同样包括

识别相关原理，解释其适用原因，并阐明数学运算的目的（Gjerde 等人，2022年）。然而，仅仅添加解释或提示词，并不能可靠地改善学习；其效果取决于它们所引发的认知活动（Barbieri 等人，2023年；Wittwer 和 Renkl，2010年）。

所需的指导程度也会随着学习者专业知识的发展而变化。当问题解决图式尚不成熟时，详细的例题解答尤其有用；而随着能力提高，相同的信息可能会变得多余。因此，可以通过补全任务或渐隐例题逐步撤除指导（Renkl & Atkinson，2003）。这体现了更广泛的专业知识逆转效应：有益于新手的教学支持，对于知识更丰富的学习者而言可能变得不再必要，甚至造成负担（Scheiger 等，2023年）。

这些原则与物理学尤其相关，因为问题解决要求学习者协调概念理解、表征和数学程序。物理教育研究表明，只有当例题明确呈现潜在的问题结构和推理过程，而不只是展示计算时，例题才最有用（Badeau 等，2017；Schmidt-Weigand 等，2009年）。因此，有效的基于例题的学习既取决于是否促进基于原理的加工，也取决于是否提供了适当程度的指导。

2.2 通过聊天机器人支持基于例题的学习

传统静态例题的一个局限在于，其解释和提示词都是预先固定的。大语言模型可能通过回应学习者的问题、要求其作出解释，并在需要时提供进一步说明，拓展这种形式。然而，根据 ICAP框架，更深入的投入取决于学习者实际做了什么，而非界面本身：对话式支持可能促成建构性或交互性加工，但并不能保证这一点（Chi & Wylie，2014）。

近期研究表明，响应式人工智能支持在教学上可能很有用，但其益处取决于支持的结构设计。在物理教育领域，Becker 等人（2026年）发现，与静态教科书式解释相比，人工智能聊天机器人和结构化分层提示系统都能降低认知负荷，而聊天机器人并未显示出明显优于结构化支持的效果。这说明，评估人工智能支持的学习环境时，应将其与有意义的教学比较条件进行对照，而不能仅与传统材料比较。

对话式支持也会带来与通过例题学习直接相关的风险。学习者可能会接受人工智能生成的解释，却没有充分理解或核实其内容。例如，Krupp 等人（2024年）发现了学习者不加批判地依赖 ChatGPT 生成的解题方案的证据。ISAR模型指出，原本旨在支持学习的人工智能反而可能减少有效的认知加工，这种可能性被称为反转效应（Bauer 等人，2025）。此外，流畅却不符合科学事实的解释可能难以被学习者识别（Dahlkemper 等人，2023年）。因此，在教育中使用大语言模型，需要采用一种能促进批判性参与，而非取代学习者自身推理的教学策略（Kasneci

等人，2023年）。因此，作为人工智能辅导工具基础的例题，必须经过设计，以引发思考，而非取代思考。

2.3 领域特定背景

高斯定律和安培定律`将电磁场与其来源联系起来，提供了简洁的关系式；存在适当对称性时，它们可以大大简化场计算。然而，应用这些定律要求学生能够识别相关对称性，区分场与通量或环流，确定封闭区域内的源，并选择合适的积分面或积分路径。物理教育研究记录了学生在应对这些概念与表征要求时普遍存在的困难（Campos 等，2023年；Hernandez 等，2023年；Pepper 等，2012）。这些特点使高斯定律和安培定律`成为研究结构化问题解决支持的合适情境。

2.4 本研究与研究问题

本研究比较智能例题、相应的静态例题以及常规讲义。两种例题采用相同的底层解题结构，并包含旨在促进自我解释的激活问题。在智能例题中，大语言模型以对话形式呈现由教师设计的解题过程，并通过回答学生的问题、提供进一步解释以及讨论学生对激活问题的回答来给予额外支持。因此，与讲义的比较考察的是人工智能支持的例题形式的整体教学价值；与静态例题的比较则更具体地考察在保持核心教学结构不变的情况下，对话式支持所带来的额外价值。

在后续的探索阶段，学生可以在现有学习材料中自由选择。这一安排补充了受控比较，并考察人工智能支持的材料是否会被学生主动采用，以及材料选择是否与他们对人工智能使用的顾虑和感知障碍有关。

研究问题1（组间差异）：使用智能例题、静态例题或讲义学习的学生，在以下方面有何差异：

（a）干预结束后即刻及追踪测评时的概念测试表现和学习增益？ - （b）自我效能感的变化？（c）学习期间的感知认知负荷？（d）对学习环境的感知可用性和教育价值？

研究问题2（材料选择与与人工智能相关的担忧）：学生在自主选择的情况下会选择哪些学习材料？这种选择与他们对人工智能使用的担忧和感知障碍有何关联？

3 方法

以下各节将介绍研究设计和学习材料（第3.1节）、样本（第3.2节，表1），以及用于回答研究问题的统计方法（第3.3节）。

3.1 研究与材料

整体研究流程和测量时点如图 1所示。各测量时点所用测量工具及收集的数据概览见表 2。测量工具的详细说明见第 3.1.2节。

3.1.1 研究设计

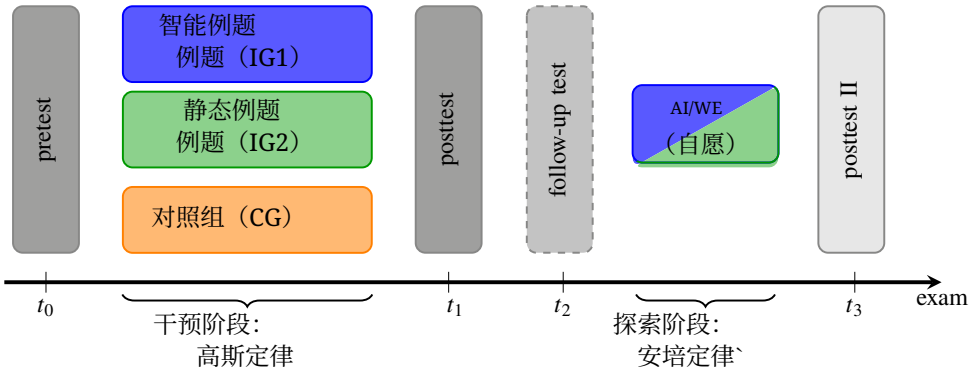


图 1 研究设计与测量时点概览，从前测（t_0）到探索阶段后的评估（t_3；IG1 = 智能例题组，IG2 =静态例题组，CG =对照组）

本研究在格廷根大学本科实验物理学 II 课程中开展，该课程通常在第二学期向物理学专业学生介绍大学阶段的电磁学。课程包括讲授课，以及每周举行、用于讨论必修作业的习题课。

本研究分为前后相继的两个阶段：针对高斯定律的受控干预，以及随后针对安培定律的探索阶段。参与者首次填写研究问卷或参加考试时，均提供了书面知情同意。研究使用参与者自行生成的假名代码，以便关联各测量时点的回答。

干预阶段 – 高斯定律

这项干预采用整群随机前测–后测–随访设计，设置了三个学习条件，并在常规习题课期间开展。学习条件分配之前，学生已选择固定的习题课小组。

随后，将九个小组随机分配至智能例题条件（IG1）、静态例题条件（IG2）或讲义条件（CG），每种条件分配三个小组，同时考虑各组人数的差异。

每次课程持续 90 分钟。学生首先完成前测（t_0），评估他们对高斯定律的概念理解，以及应用高斯定律的自我效能感。此前已在讲座中介绍过高斯定律。随后，学生根据各自条件下分配的材料，独立按自己的节奏完成同一个目标问题。两种例题条件下的学生还可以查阅讲义。材料使用由学生自主决定，学习阶段不提供额外的教学支持。

随后，学生立即完成后测（t_1），其中包括与前测相同的概念测试和自我效能感测量，以及认知负荷、可用性和感知教育价值的测量。

四周后，在常规讲座期间再次进行概念测试（t_2）。由于是否参加追踪测评与是否参加干预相互独立，因此部分学生只完成了这项评估。这些学生构成了非等价横断面参照组，未纳入学习增益的纵向分析。

探索阶段——安培定律

随后，探索阶段聚焦于安培定律。学生完成了一项常规家庭作业，要求运用安培定律；他们可以自由选择是否使用智能例题、静态例题、两种例题，或都不使用，作为额外的学习支持。整个课程期间，常规讲义始终可供学生使用，但并非专为这项作业提供。与干预阶段不同，目标问题是在监督下的习题课之外完成的，并作为常规课程作业的一部分提交。

提交作业后，学生报告了自己使用了哪一种或哪些例题，并填写问卷，评估他们对人工智能使用的担忧，以及使用基于人工智能的学习支持时感知到的障碍（t_3；图 1）。探索阶段采用观察性设计：学生自由选择学习材料，材料使用的时间和强度均未受到控制或监测。因此，有关例题使用情况的信息是基于学生的自我报告。

3.1.2 测量工具与材料

本节介绍干预阶段和探索阶段所使用的学习材料，以及用于评估学生概念理解、自我效能感、认知负荷、可用性、感知教育价值，以及与人工智能相关的担忧和障碍的测量工具。所有学习材料和评估工具均见附录A 和B。所有量表的概览见表2。

例题

两个干预组的学生都使用专为本研究开发的例题学习应用高斯定律（图2 和附录 A.1）。两个版本涵盖相同的学习目标，并呈现了使用高斯定律计算空心球电场的完全相同的分步解答。这些例题依据基于例题的学习的既有原则设计（Atkinson 等人，2000；Renkl，2014）。解答被组织为清晰分隔的子目标，每个子目标均配有简短的概念性解释。在解答过程的特定位置融入了激活问题，促使学生反思其背后的物理概念并进行自我解释（Chi 等人，1989）。这些问题为开放式问题，要求学生用自己的话阐述解释。

静态例题（图2a）以 PDF 格式提供，采用表格形式呈现完整的解题过程。完整的静态例题见附录A.1。各个解题步骤及相应的概念性解释依次排列，激活问题则直接嵌入表格中的相关位置。学生独立学习材料，并回答这些问题，不会收到反馈或额外支持。

智能例题（图2b）在静态版本的基础上，将基于人工智能的聊天机器人整合到学习环境中。例题沿用相同的解题步骤和激活问题顺序，但在整个学习过程中提供对话式支持。学生可以使用聊天机器人提出后续问题、请求更多解释，并在继续下一解题步骤前讨论自己对激活问题的回答。因此，智能例题和静态例题在内容、解题结构和教学性解释方面完全相同，唯一的区别在于是否提供基于人工智能的对话式支持。

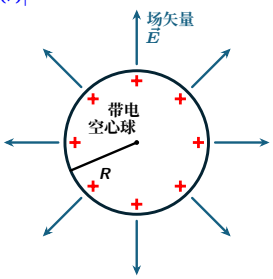
这道智能例题使用了 ChatAI，即 G¨ottingen 大学提供的机构人工智能平台（University of G¨ottingen，无日期）。底层大语言模型为 Qwen3-30B-A3B-Instruct-2507，温度参数为 0.0，top-p 值为 0.05。干预于 2026 年5月5日至7日进行。用于指导聊天机器人回复的系统提示词见附录A.2。

在探索阶段，学生可以使用智能例题和静态例题，学习如何运用安培定律计算长螺线管的磁场。两道例题均依据相同的教学设计原则编写，且结构与干预阶段所用的高斯定律例题各自对应版本相同。

高斯定律概念测试

学生对高斯定律的概念理解通过一份纸笔概念测试进行评估。测试包含五项任务及若干子任务，共计19个条目（附录B.2）。这些任务涵盖了与高斯定律相关的关键概念，包括识别合适的高斯面，并改编自一项

(a)

概念步骤	通过示例执行
2. 对称性考量 在计算之前，必须明确存在哪些对称性，以及场矢量的方向。画图通常会很有帮助。	球对称几何：由于电荷分布，因此电场呈径向对称，且球体的中心可以置于坐标系的原点，因此有以下结论： $\vec{E}(\vec{r}) = \vec{E(r)} \cdot \hat{e}_r$ 
问题 1：如何选择高斯面，使通量积分中的贡献要么是常数，要么为零？	

(b)

2 对称性考量

2.1 一般推理步骤

在进行计算之前，必须先确定存在何种对称性，以及电场矢量的方向。此时绘制草图通常很有帮助。

2.2 通过空心球示例执行

球对称性：由于电荷分布以及由此产生的电场都具有径向对称性，且中心球体位于坐标系的原点，因此：

$$\vec{E}(\vec{r}) = |\vec{E}(r)| \cdot \hat{e}_r$$

注意： 讲义中提供了均匀带电空心球电场的示意图。

2.3 关于对称性考量的问题

现在我提出一个概念性问题，以促使大家思考。请先简短思考一下，不必在此问题上纠结太久。如有需要，之后可以再回顾这个问题。

为什么 $\vec{E}(\vec{r}) = |\vec{E(r)}| \cdot \hat{e}_r = |\vec{E(r)}| \cdot \hat{e}_r$ 保留？

准备好继续时请告诉我。

因为电场具有径向对称性

很好，你已经抓住了核心思想：电荷分布的径向对称性使电场只取决于距中心的距离 r，且方向沿径向。

但让我们仔细看看：为什么场的大小只取决于 r，而不取决于方向（例如，θ、ϕ）？如果缺少对称性，会发生什么？

你能想象，如果两个点具有相同的 r 但方向不同，那么场会呈现怎样的状态吗？如果给出了对称性吗？

你还有其他问题吗？如果没有，请写“下一步”。

图 2 静态例题 (a) 和智能例题 (b)。在静态例题中，激活问题作为自我解释的提示词，不提供反馈或进一步互动。在智能例题中，同类问题可以开启与聊天机器人的讨论。一般而言，智能例题允许学生在学习过程中提出后续问题并获得额外支持。完整的静态例题见附录 A.1。为便于说明，智能例题以英文呈现；研究中使用的原始版本为德语，相应的系统提示词见附录 A.2

Singh（2006）开发的关于对称性和高斯定律的成熟概念测试。原有的单项最佳选择题被改编为多项判断题格式，将每个选项作为一个独立陈述，要求判断其正确与否，共形成18道判断题。此外，还有一道开放式题目，要求学生阐明其推理过程。图3展示了一个示例任务。

为将测试的两个部分合并为一个总分，综合得分定义为 $T(\kappa) = T_1 + \kappa T_2$，其中 T_1 表示18道判断题陈述的得分，T_2 表示解释题的得分，κ 决定解释题的相对权重。加权参数事先设定为 $\kappa = 2$，因此，解释题每得一分的权重是答对一道判断题的两倍。

为检验这一预先设定的加权方式是否稳健，我们参照Klein等人（2017）进行了敏感性分析。针对三个组，分别在一系列 d取值下计算前后测效应量（Cohen's d）。该分析并非用于选择加权参数，而是为了考察预先设定的 $\kappa = 2$ 在其他加权方案范围内的位置，以及各组的效应量对 κ变化的敏感程度。参照Klein等人（2017），本敏感性分析中特别采用了Cohen's d；而第3.3节报告的基于方差分析的分析则采用广义η平方（η^2_G）。敏感性分析结果见图4。

在所有后续分析中，回答准确率均按照预先设定的权重 $\kappa = 2$ 计算：

$$\text{response accuracy} = \frac{X + 2p}{24},$$

其中，$X \in 0$、……、18 表示回答正确的判断题数量，$p \in 0$、……、3 表示解释题的得分。因此，解释题对总分 24 分的贡献最高为 6 分，回答准确率范围为 0 至 1。

自我效能问卷

采用一份四题问卷（附录 B.3）评估学生在应用高斯定律方面的自我效能感。问卷条目依据 Bandura（2006）提出的自我效能量表编制指南设计，并参考了 Freed等人（2022）开发的测量工具。条目涉及学生对自身能力的感知，包括理解高斯定律的概念、识别适当的电荷对称性、构造合适的高斯面，以及应用高斯定律计算电场。条目示例："我能够判断电荷分布的对称性是否适合应用高斯定律。"回答采用六点李克特量表记录，范围从 1（非常不同意）到 6（非常同意）。

认知负荷问卷

自主学习阶段学生的感知认知负荷通过 Leppink等人（2013年）开发并经过验证的认知负荷问卷进行评估。该测量工具将认知负荷区分为三个维度——内在认知负荷（ICL）、外在认知负荷（ECL）和关联认知负荷（GCL）——采用

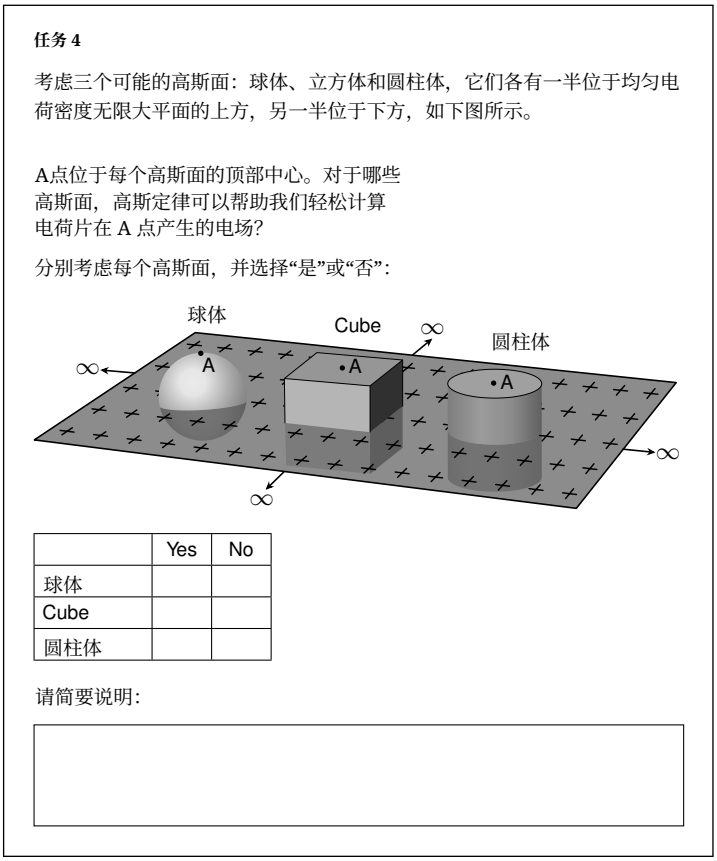


图3 改编自Singh的概念测试的示例任务（2006），由三道判断题和一道解释题组成。概念测试中所有条目的列表见附录B.2

每个分量表包含三或四个条目。其三因子结构已被证明在适用于不同学科背景的各种改编版本中均具有稳健性（Leppink等人，2014）。

在本研究中，问卷针对高斯定律学习情境进行了调整，共包含九个条目（附录B.3）。示例条目包括“‘使用高斯定律计算电场’这一主题非常复杂。”（内在认知负荷）、“从学习角度而言，自主学习阶段的指导和/或解释非常无效。”（外在认知负荷），以及“自主学习阶段确实加深了我对‘使用高斯定律计算电场’这一主题的理解。”（关联认知负荷）。回答采用六点李克特量表记录，范围从 1（非常不同意）到 6（非常同意）。

关于可用性的问卷和感知教育 价值

我们使用两个包含三项的量表（附录B.3）评估学习材料的可用性和感知教育价值。可用性量表着重考察学习材料的清晰度、可理解性和易用性。其条目改编自

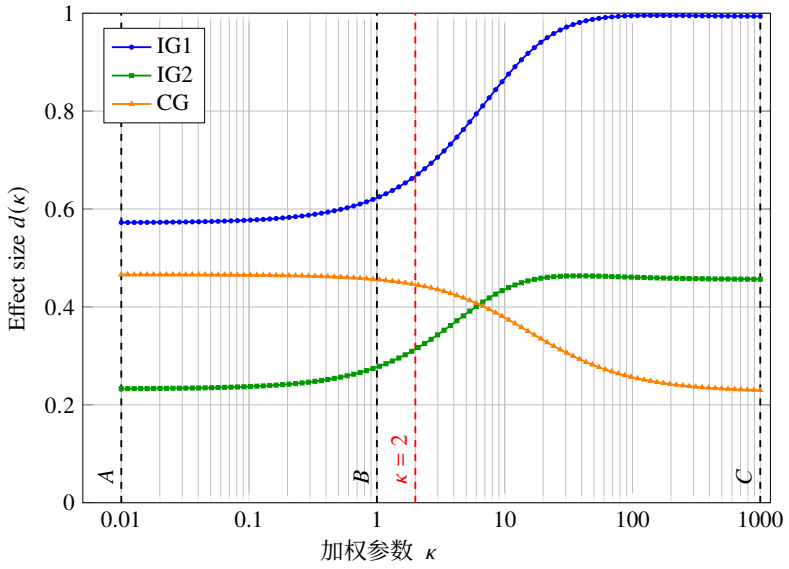


图4 测试组合分析显示了智能例题组（IG1）、静态例题组（IG2）和对照组（CG）的综合测验分数 $T = T_1 + \kappa T_2$ 的前测—后测效应量（Cohen's d）如何随加权参数 κ 变化。这里，T_1 表示判断题得分，T_2 表示解释题得分。对于每位学生，我们根据其前测和后测原始分数，针对 κ 的100个不同取值计算综合得分。κ 轴采用对数刻度，以便在相关权重范围内提供足够的分辨率。垂直虚线标示了三个参照情形：A时，综合得分主要由判断题决定（$\kappa \to 0$）；B时，测试的两个组成部分获得相同的单位权重（$\kappa = 1$）；C时，综合得分主要由解释题决定（$\kappa \to \infty$）。红色虚线标示预先设定的加权参数 $\kappa = 2$，后续所有分析均使用该参数计算回答准确率

系统可用性量表（SUS；Brooke，1996）以及 Unver 等人（2017）验证过的改编版本。一个示例条目是：“学习材料的呈现形式清晰易懂。”

感知教育价值采用改编自 Shellman 和 Turan（2006）的三个条目进行评估。这些条目涉及学生对学习材料在学习中的有用性以及其在其他物理课程中的潜在用途的看法。示例条目：“我建议在其他课程（例如，实验物理学 *I*）中使用该学习材料，并将其拓展至实验物理学 *II* 所涵盖的更多概念。”两个量表的回答均采用六点李克特量表记录，范围从 1（非常不同意）到 6（非常同意）。

学习材料选择

在探索阶段，学生可以自由选择使用两种学习材料中的哪一种。根据他们自我报告的使用情况，他们被分为使用了智能例题、静态例题、两种例题，或

两者都未使用。由于常规课程讲义在整个课程期间始终可用，因此未单独评估其使用情况。

关于与人工智能相关的担忧和障碍的问卷

本研究使用为本研究开发的五个条目评估学生对使用人工智能的担忧：其中两个参考了 TAME-ChatGPT 测量工具的感知使用风险分量表（Sallam 等人，2024年），一个参考了关于使用 ChatGPT 完成家庭作业的研究（Miljkovi´c Kreˇcar 等人，2024年），一个参考了有缺陷的人工智能生成物理内容的例题（Bitzenbauer，2023年），另一个则未参考已发表的来源。所有条目均以“我担心……”“*I am concerned that . . .*”为开头，并涉及人工智能使用对独立解题和批判性思维可能产生的影响，以及使用人工智能学习物理学时对错误未被察觉和理解流于表面的担忧。一个例题是“我担心使用人工智能可能会降低我独立解决问题的能力。”

本研究通过五个专门编制的条目评估感知到的人工智能使用障碍，这些条目在概念上与感知易用性相关（Davis，1989）。这些条目涉及识别相关信息的困难、回复中包含大量文字、提示词的拟定、感知到的投入与收益之比，以及用户界面。示例条目为“人工智能的回复往往包含过多文字，因此难以浏览。”

两个量表的回答均采用七点李克特量表记录，范围从1（非常不同意）到7（非常同意）。两个量表均未具体涉及干预或其学习材料；因此，它们旨在捕捉学生对人工智能使用的总体担忧和感知到的人工智能使用障碍，而非对特定学习环境的评估（附录B.4）。

3.2 样本

干预阶段（研究问题1）的分析基于一个由86名学生组成的核心样本，这些学生完成了干预以及前测和即时后测。该样本包括智能例题组（IG1）的29名学生、静态例题组（IG2）的28名学生，以及对照组（CG）的29名学生（见表1）。三个组在年龄、已修读学期数、大学入学成绩或前测表现方面均未观察到具有统计显著性的基线差异。

对于研究问题1(a)(ii)的纵向随访分析，57名学生在前测、后测和追踪测验中的完整数据均可获得（$n_{\mathrm{IG1}} = 21$、 $n_{\mathrm{IG2}} = 20$、$n_{\mathrm{CG}} = 16$）。另有13名未参加干预的学生完成了随访评估。由于这些学生既未被随机分配到研究条件，也未完成前测和即时后测，因此他们仅作为随访时的非等价横断面参照组纳入，并被排除在个体学习增益分析之外。因此，纵向分析基于 $N = 57$，而随访时的探索性横断面比较则包括 $N = 70$ 名学生。

对于研究问题2，干预样本中的46名学生也参与了后续的探索阶段（$n_{\mathrm{IG1}} = 16$、$n_{\mathrm{IG2}} = 13$、 $n_{\mathrm{CG}} = 17$）。另有6名学生

表 1 样本数据（干预组 IG1（智能例题）和 IG2（静态例题）、对照组 CG、编号、使用单因素方差分析检验统计显著性 p ）。总样本由完成干预阶段、前测和后测的学生组成。

	总计	IG1	IG2	CG	p
受试者人数	86[a,b]	29	28	29	–
女性受试者人数	31（36%）	6（21%）	11（39%）	14（48%）	–
平均年龄（岁）	20.4 ± 2.9	20.4 ± 2.3	20.3 ± 2.8	20.6 ± 3.5	0.89
学习学期数	2.4 ± 1.7	3.1 ± 2.8	2.0 ± 0.2	2.0 ± 0.4	0.15[c]
大学平均成绩-大学入学[d]	1.6 ± 0.6	1.8 ± 0.8	1.4 ± 0.5	1.7 ± 0.5	0.08[c]
前测得分 高斯定律[e]	0.61 ± 0.19	0.58 ± 0.21	0.66 ± 0.15	0.59 ± 0.20	0.19

[a] $N = 57$ 名学生完成了全部三个测试时间点（$N_{\text{IG1}} = 21$, $N_{\text{IG2}} = 20$ $N_{\text{CG}} = 16$）。另有 $N = 13$ 名学生仅完成了追踪测验。
[b] $N = 46$ 名学生还参与了探索阶段（$N_{\text{IG1}} = 16$, $N_{\text{IG2}} = 13$ $N_{\text{CG}} = 17$）。另有 $N = 6$名学生仅参与了探索阶段。
[c]由于违反方差齐性假设，因此进行了 Welch 方差分析。
[d]量表范围为 1.0（表现最佳）至 4.0。成绩由学生标注。
[e]量表范围为 0 至 1（表现最佳）。

仅参与了探索阶段，因此用于材料选择分析的样本总数为 $N = 52$ 。由于问卷回答缺失，用于分析与人工智能相关的担忧和感知障碍的可用样本数分别为 $N = 41$ 和 $N = 40$。

3.3 数据分析

所有变量均经过线性转换，映射至共同范围 $[0, 1]$。根据偏度、峰度和样本量判断，连续测量变量的分布足够接近正态分布，适合进行参数分析。多项目量表的内部一致性见表 2；外在认知负荷量表的信度相对较低（$\alpha_C = .56$），涉及该量表的结果应谨慎解读。

针对研究问题1，使用单因素方差分析考察各测量时间点的组间差异，并使用重复测量方差分析和混合设计方差分析考察随时间的变化。混合设计分析检验了时间 × 组别对回答准确率和自我效能感的影响、负荷类型 × 组别对认知负荷的影响，以及量表 × 组别对可用性和感知教育价值的影响。在适当情况下，进行了 Tukey 校正的成对比较。基于方差分析的分析报告了广义η平方（η^2_G）（Olejnik 和 Algina, 2003），显著成对差异报告了 Cohen'sd 。

此外，还计算了从前测到后测的回答准确率和自我效能感的个体变化分数；对于纵向随访样本中的回答准确率，也计算了从后测到追踪测评的变化分数。组间变化分数的差异

采用单因素方差分析进行检验。回答准确率从前测到后测的学习增益也使用 Hake归一化增益进行描述 g_H (Hake，1998年)。

针对研究问题2，材料选择组的样本量较小且差异显著。因此，采用Kruskal–Wallis检验考察四个选择组之间与人工智能相关的担忧和感知障碍的差异。随后，进一步将单独使用智能例题或将其与静态例题结合使用的学生，与未使用该例题的学生进行Mann–Whitney U 检验比较。

采用皮尔逊相关考察连续测量指标之间的关联。统计检验的显著性水平设为 $\alpha = .05$。两两比较按照上述说明进行调整。

表2 数据分析中所包含量表的概览（克ронбахα系数 α_C 基于相应样本计算）。

时间点	量表	#条目	α_C
t_0 (前测)	概念测试（回答准确率）[a]	19	0.78
	自我效能感	4	0.87
	社会人口学特征	–	–
t_1 (后测)	概念测试（回答准确率）[a]	19	0.78
	自我效能感	4	0.90
	内在认知负荷（ICL）	3	0.77
	外在认知负荷（ECL）	3	0.56
	促进性认知负荷（GCL）	3	0.89
	可用性	3	0.68
	感知教育价值	3	0.78
t_2 (追踪测验)	概念测试（回答准确率）[a]	19	0.74
t_3 (后测 II)	学习材料选择[b]	–	–
	对人工智能的担忧	5	0.62
	人工智能使用障碍	5	0.77

[a]18道判断题和1道开放式解释题； α_C 基于这19个条目计算了信度。在测试得分中，解释题的权重为 $\kappa = 2$（第3.1.2节）；此权重不计入此处报告的信度。[b]在探索阶段结束后通过自我报告获得；这不是量表，因此未报告条目数量和信度。

4 结果

4.1 研究问题1 – 组间差异

4.1.1 表现（a）

从前测（t_0）到后测（t_1），时间与 × 组的交互作用不显著（$p = .07$），表明智能例题组（IG1）、静态例题组（IG2）

和对照组（CG）在回答准确率变化方面没有显著差异。在所有三个组中，回答准确率均随时间显著提高 $[F(1, 83) = 26.24, p < .001, \eta^2_G = .07$；图 5]。在 IG1 中，回答准确率从 0.58 ± 0.21 提高到 0.74 ± 0.19，对应较大的效应 $[F(1, 28) = 12.61, p = .001, \eta^2_G = .15]$。IG2 的回答准确率也显著提高，但幅度较小，从 0.66 ± 0.15 提高到 0.72 ± 0.19 $[F(1, 27) = 6.74, p = .015, \eta^2_G = .03]$，对应较小的效应。同样，对照组的回答准确率从$0.59 \pm 0.20$ 提高到 0.67 ± 0.18，也对应较小的效应 $[F(1, 28) = 8.01, p =$。009, $\eta^2_G = .04]$。对变化分数进行 Tukey 校正的成对比较后发现，组间差异均不显著。

与这一描述性模式一致，按照 Hake（1998年）计算的标准化增益在 IG1（$g_{H,\mathrm{IG1}} = .39$）中等，在 IG2（$g_{H,\mathrm{IG2}} = .18$）和对照组（$g_{H,\mathrm{CG}} = .20$；图 5，虚线）中较小。尽管这些描述性学习增益存在差异，三个组之间的后测反应准确率并无显著差异（$p = .34$）。

测试组合分析（第3.1.2节）进一步表明，两个例题组的前后测效应量均随着解释性题目权重的增加而增大，而对照组则呈现相反的模式（图4）。因此，IG1和IG2的学习增益在解释性成分上更为显著，而对照组的学习增益则在判断正误成分上更为显著。

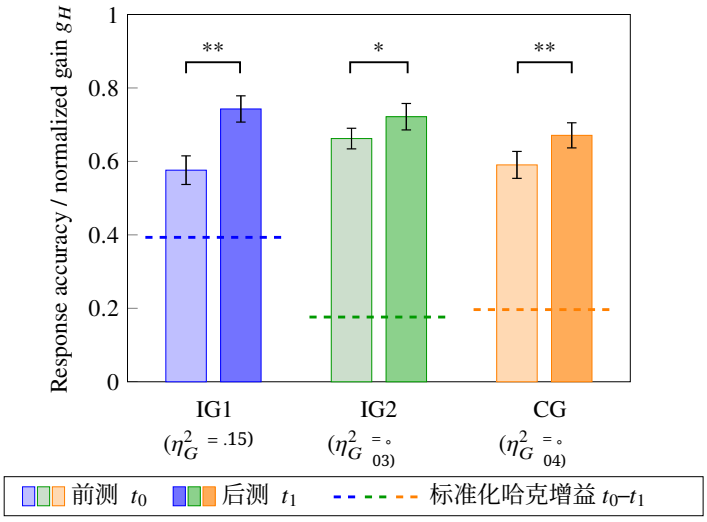


图 5 比较智能例题组（IG1）、静态例题组（IG2）和对照组（CG）的学生回答准确率及学习增益，从前测（t_0）到后测（t_1；$N = 86$）。根据 Hake（1998年）计算的所有组从前测到后测的标准化增益以虚线表示。使用单因素重复测量方差分析比较 t_0 与 t_1 之间的回答准确率（* / ** 统计显著性 $p < .05$ /$p < .01$，效应量 η^2_G，误差条表示 1 个均值标准误）

纳入追踪测验（t_2）后，缩减样本（$N = 57$）的前测—后测得分和学习效应与完整样本的结果相当。从后测到追踪测评，时间 ×组别交互作用不显著（$p = .13$），表明各组回答准确率的变化不存在显著差异。然而，观察到显著的组别主效应，对应于中等效应 [$F(2,\ 54) = 3.69,\ p = .031,\ \eta^2_G = .12$；图 6]。IG1的回答准确率从后测时的 0.75 ± 0.20 略微下降至追踪测评时的 0.72 ± 0.16，但这一变化并不显著（$p = .16$）。相比之下，IG2的回答准确率进一步上升，从 0.76 ± 0.18 增至 0.81 ± 0.13（$p = .11$），但这一变化未达显著；对照组的回答准确率则保持稳定，为0。65（后测时为0.65 ± 0.15，追踪测评时为0.65 ± 0.13；$p = 1.00$）。对变化分数进行Tukey校正的成对比较后发现，各组之间的差异均不显著。

后测时，三个组的回答准确率没有显著差异（$p = .14$）；但在追踪测评时，组间差异达到显著水平，且效应量较大 [$F(2,\ 54) = 5.76,\ p = .005,\ \eta^2_G = .18$]。Tukey校正的成对比较显示，IG2的追踪测评得分显著高于对照组，对应较大的效应量（$p = .004,\ d = 1.22$）；其余成对差异均不显著。

未完成干预阶段的学生（$N = 13$）追踪测评平均得分为 0.49 ± 0.13。将这些学生纳入追踪测评比较后，整体组别效应进一步增大 [$F(3,\ 66) = 13.79,\ p < .001,\ \eta^2_G = .39$]。Tukey校正的成对比较显示，与未完成干预阶段的学生相比，IG1（$p < .001,\ d = 1.48$）、IG2（$p < .001,\ d = 2.33$）和对照组（$p = .023,d = 1.18$）的追踪测评得分均显著更高，且这三项差异对应的效应量均较大。此外，IG2的得分显著高于对照组（$p = .007,\ d = 1.22$），对应的效应量也较大；其余成对差异均不显著。

4.1.2 自我效能 (b)

对于与高斯定律相关概念的自我效能感，前测（t_0）至后测（t_1）期间，时间 × 组别交互作用并不显著（$p = .38$），表明各组自我效能感的变化没有显著差异。三个组的自我效能感均随时间显著提高 [$F(1,\ 83) = 71.25,\ p < .001,\ \eta^2_G = .16$；图 7]。这一增长在智能例题组 [$F(1,\ 28) = 24.71,\ p < .001,\ \eta^2_G = .22$] 和静态例题组[$F(1,\ 27) = 53.13,\ p < .001,\ \eta^2_G = .20$]中对应较大的效应，在对照组 [$F(1,\ 28) = 11.06,\ p = .002,\ \eta^2_G = .10$]中则对应中等效应。对变化分数进行 Tukey校正的成对比较，同样未发现各组之间存在显著差异。各组在前测（$p = .79$）和后测（$p = .19$）时的自我效能感均无显著差异。最后，仅在对照组中观察到后测自我效能感与后测反应准确率之间显著且较大的正相关（$r = .58,\ p = .001$）。

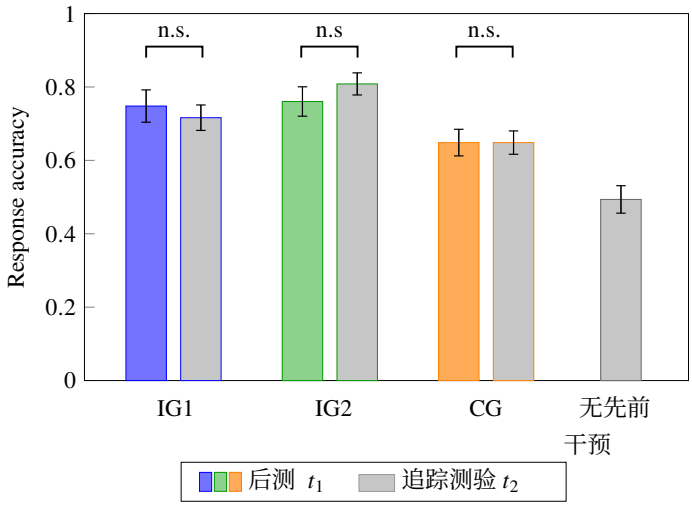


图 6 智能例题组（IG1）、静态例题组（IG2）和对照组（CG）的学生回答准确率比较（$N = 70$ 包括13名未完成后测的学生），比较后测（t_1）和追踪测验（t_2）的结果。图中还显示了未完成干预阶段学生的追踪测验得分（“无先前干预”）。使用单因素重复测量方差分析比较各时间点（t_1、 t_2）之间的回答准确率（无统计学显著性，误差条表示1个均值标准误）

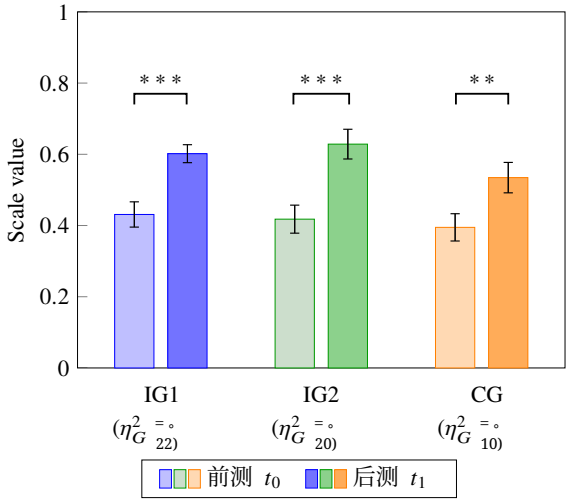


图 7 智能例题组（IG1）、静态例题组（IG2）和对照组（CG）的学生自我效能感比较，比较前测（t_0）至后测（t_1； $N = 86$）的变化。使用单因素重复测量方差分析比较 t_0 与 t_1 之间的自我效能感（** / *** 统计显著性 $p < .01$ /$p < .001$，效应量η^2_G，误差条表示1个均值标准误）

4.1.3 感知认知负荷（c）

对于感知认知负荷，观察到显著的认知负荷类型 × 与组别交互作用，其中认知负荷类型（内在、外在和关联认知负荷）为被试内因素，组别为被试间因素[F（4，166）$= 3.15$，$p = .016$，$\eta^2_G = .05$]。这一小交互效应主要体现在外在认知负荷（ECL）和关联认知负荷（GCL）的描述性模式上：IG1报告的ECL最低、GCL最高，而对照组则呈现相反模式，即ECL最高、GCL最低（图8）。分别考察各认知负荷类型时，内在认知负荷（$p = .46$）、ECL（$p = .08$）或GCL（$p = .06$）的总体组间差异均不显著。Tukey校正的成对比较同样显示，各组在内在认知负荷或ECL方面均无显著差异。然而，对于GCL，IG1报告的数值显著高于对照组，对应于中等效应（$p = .046$，$d = .63$）。

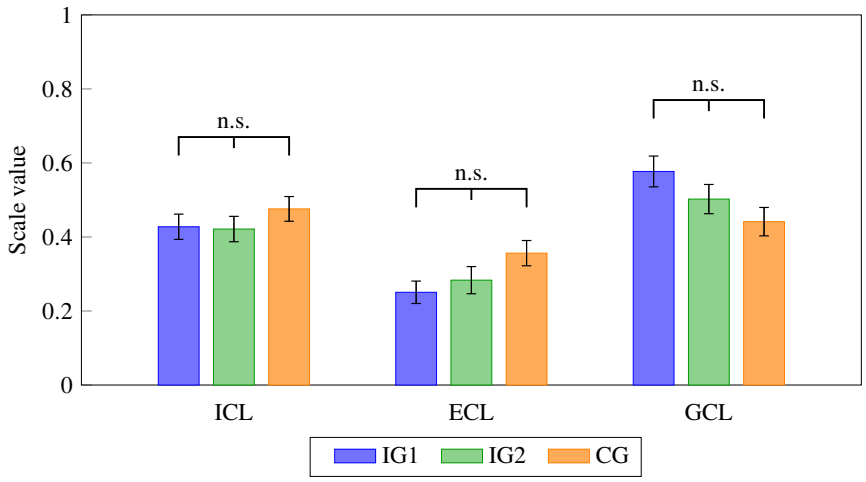


图 8 智能例题组（IG1）、静态例题组（IG2）和对照组（CG；$N = 86$）学生的感知认知负荷。采用以组别为被试间因素的单因素方差分析，对各组的外在认知负荷（ECL）、内在认知负荷（ICL）和关联认知负荷（GCL）进行比较（无统计学显著性，误差条表示 1 个均值标准误）

4.1.4 可用性与感知教育价值（d）

对于可用性和感知教育价值，量表与组别的交互作用不显著（$p = .67$）。然而，在两项量表上均观察到显著的组效应，对应中等效应 [$F(2, 82) = 5.49, p = .006$, $\eta^2_G = .12$；图 9]。分别考察两项量表时，各组的可用性存在显著差异，总体效应为中等效应 [$F(2, 82) = 5.0, p = .009$,

$\eta^2_G = .11$]。Tukey校正的成对比较显示，IG1组的可用性显著高于对照组，对应较大的效应（$p = .011$, $d = .88$），而其余成对差异均不显著。各组的感知教育价值没有显著差异（$p = .07$），且Tukey校正的成对比较均未达到显著水平。最后，可用性与感知教育价值呈中等程度的正相关（$r = .42$, $p < .001$）。

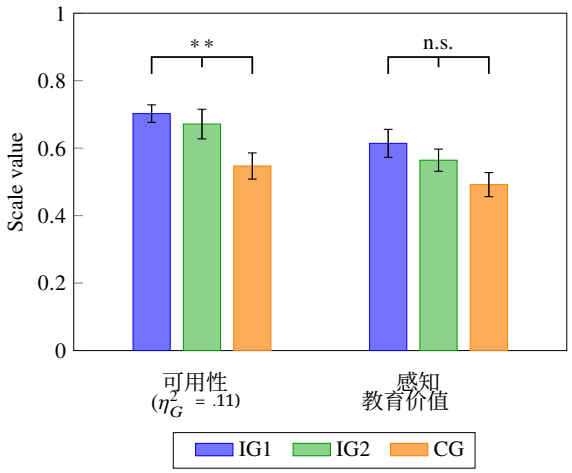


图9 智能例题组（IG1）、静态例题组（IG2）和对照组（CG；$N = 86$）对学习环境的可用性及学生感知教育价值的评价。采用以组别为被试间因素的单因素方差分析，对各组的可用性和感知教育价值进行比较（** / 无统计学显著性 $p < .01$ / 不显著，效应量 η^2_G；误差条表示1个均值标准误）

4.2 研究问题2——学习材料选择以及与人工智能相关的担忧和障碍

在安培定律的探索阶段`，学生可以自由选择想使用的学习材料。总体而言，超过半数的学生只选择了静态例题（60%；图10）。这一趋势在 IG1（63%）、IG2（62%）和对照组（71%）中均有体现。相比之下，未参加干预阶段的学生选择静态例题的比例要低得多（17%）。

在三个组中，单独使用智能例题或将其与静态例题结合使用的学生比例都相对较低。除 IG2 外，这一比例均未超过 10%；在 IG2 中，15% 的学生同时使用了两种例题，另有 8% 的学生仅使用智能例题。值得注意的是，IG1 中没有学生选择仅使用智能例题。相比之下，未参与干预阶段的学生更常选择智能例题，选择该选项的比例为 33%。

既未使用任何一种例题的学生比例从 IG2 的 15% 到 IG1 的 31% 不等；未参与干预阶段的学生中，这一比例为 33%。

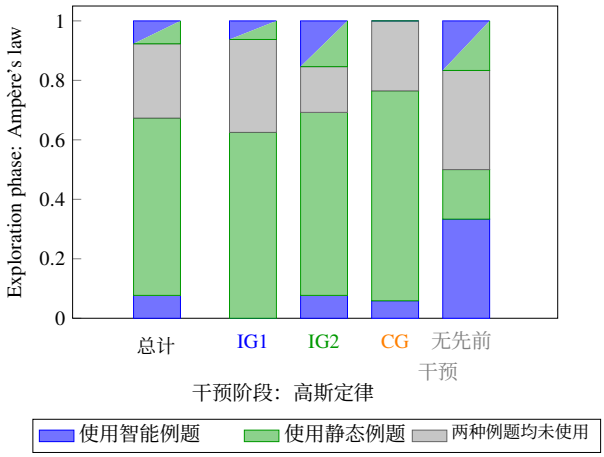


图10 探索阶段的学习材料选择（$N = 52$）。堆叠条形图显示了使用智能例题、静态例题、两种例题或两者均未使用的学生所占的相对比例。该图展示了学生在安培定律探索阶段的选择行为，并根据其在此前高斯定律干预阶段中的分组情况进行区分（智能例题组 IG1、静态例题组 IG2 和对照组 CG）。未参与干预阶段的学生单独列出（“无先前干预”）。在每个组内，各部分之和为1

学生对与人工智能相关的担忧总体上表示了较高程度的认同，量表平均值为 0.63 ± 0.16（图11）。相比之下，感知到的人工智能使用障碍不太明显，量表平均值为 0.38 ± 0.17。最强烈的担忧是，使用人工智能可能会降低学生独立解决问题的能力；该项的平均值最高（0.78 ± 0.22）。担忧与障碍之间不存在显著相关（$r = .19$, $p = .23$）。

选择智能例题的学生与未选择的学生之间，与人工智能相关的担忧没有显著差异（$p = .53$）；在四个选择组别——智能例题、静态例题、两种例题都选以及两种例题都未选——之间也没有显著差异（$p = .36$）。感知到的人工智能使用障碍也呈现出类似的模式。无论是比较选择智能例题的学生与未选择的学生（$p = .34$），还是比较四个选择组别（$p = .79$），均未发现显著差异。

5 讨论

本研究从两个互补的角度考察了智能 worked examples：在指定学习条件下的教学价值，以及学生能够选择学习支持时的自愿采用。在三种条件下，学生的概念理解和自我效能感均有所提高，

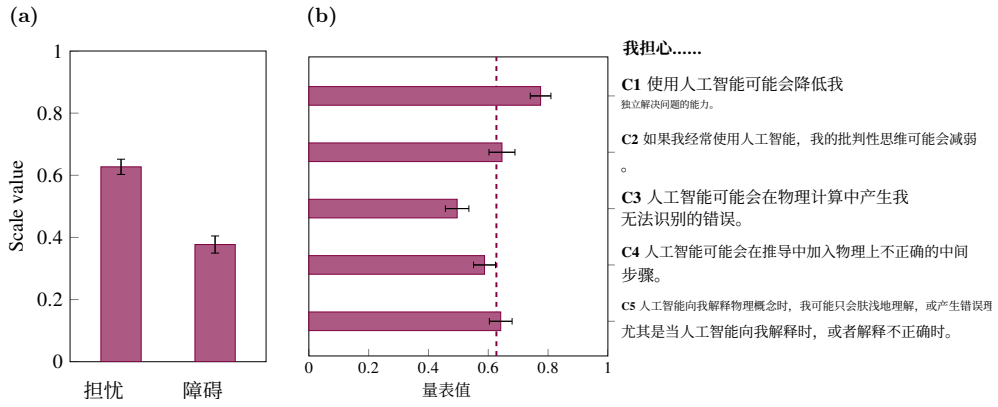


图11 学生报告的与人工智能相关的担忧（$N = 41$）和障碍（$N = 40$）。（a）与人工智能相关的担忧和障碍的平均量表值，误差条表示1个均值标准误。（b）五条目担忧量表的条目层面结果。每个条目均显示平均值，误差条表示1个均值标准误；虚线表示担忧量表的总体平均值

但与相应的静态例题相比，智能例题并未带来明确的额外学习收益。同时，学生对人工智能支持的形式评价较为积极，尤其认可其可用性。然而，当使用变为自愿时，学生主要选择静态例题，很少选择智能版本。综合来看，这些发现区分了评估人工智能支持的学习环境时常被混为一谈的三个方面：它们是否有助于学习、学习者如何体验这些环境，以及学习者是否会选择使用它们。

5.1 研究问题1 – 组间差异

在所有三种条件下，回答准确率均从前测到后测显著提高。这表明学生在教学阶段加深了对高斯定律的理解，但时间 ×组间交互作用不显著，说明这种提升的大小在智能例题、静态例题和讲义之间没有可靠差异。同样，各组的自我效能感均大幅提高，但组间变化差异不显著。因此，当前数据不支持任何一种例题条件在即时学习或自我效能感方面具有普遍优势。

这种普遍的学习效应不应完全归因于学习材料。前测与即时后测包含相同的条目，且两者间隔仅约 45 分钟；此外，研究未设置不参与教程的随机化条件。因此，学习效果也可能反映了学生对目标问题的投入、重复测试、此前的讲座，或教程的其他特点。未参与干预的学生后续表现明显较低，这与参与教程阶段可能带来益处的观点一致；但这一非等效参照组无法用于对此类效应作出因果估计。

5.1.1 基于例题的支持所呈现的模式

尽管例题条件并未带来显著更大的即时学习增益，但一些发现与学生处理学习材料的方式存在差异这一观点相符。在测试组合分析中，随着赋予解释性成分的权重增加，两个例题组的前后测效应量均有所上升，而讲座讲义条件则呈现相反的趋势。这一模式与以下可能性相符：例题更强调解释性推理。两个例题都将明确的解题结构与概念性解释及旨在引发自我解释的激活提示结合起来；这些教学特征被认为有助于例题学习（Atkinson 等人，2000；Chi 等人，1989；Renkl，2014；Sweller 和 Cooper，1985）。本分析并未证实例题对概念推理存在差异化影响，但表明不同教学形式下观察到的增益性质可能有所不同。

认知负荷测量也为学习体验上的差异提供了类似的、仍属初步的证据。观察到负荷类型 ×组别交互作用：智能例题条件下描述性外在认知负荷最低、关联认知负荷最高，而讲义条件下的排序则相反。内在、外在和关联认知负荷各自的总体组别效应均未达到显著水平，不过，智能例题条件下的关联认知负荷显著高于讲义条件。此模式与例题背后的认知负荷理论依据相符：结构化指导可能减少无效搜索，并使更多认知资源投入与学习相关的加工（Renkl，2014；Sweller 和 Cooper，1985年；Sweller 等，1998年）。然而，这一解释仍属初步，尤其是因为外在认知负荷量表的内部一致性相对较低。

学生对学习环境的评价更清楚地表明，不同教学形式在感知可用性方面存在差异。智能例题的可用性评分显著高于讲义，且成对效应较大。感知教育价值的描述性排序也相同，但组间差异未达到显著水平。因此，支持人工智能支持的例题优于传统讲义支持的最有力证据，涉及的是用户体验，而非即时学习表现。

延迟测评呈现出另一种模式，但更难以解读。在追踪测评中，静态例题条件下的学生得分显著高于讲义条件下的学生，而智能例题组与静态例题组之间的得分差异不显著。这一发现与结构化例题学习有助于实现更持久学习的可能性相符，但并不能据此确认这种效果。从后测到追踪测评，各组的变化并无显著差异；参与追踪测评的学生并非随机入选，纵向样本也从86名学生缩减至57名。此外，在缩减后的子样本中，追踪测评时观察到的组间排序，在前测时就已初现端倪。因此，原有差异、选择性流失、后续自主学习

以及其他与课程相关的学习经历，都是合理的替代解释。

总体而言，与讲义的比较表明，即使结构化例题支持并未转化为即时测试表现上的普遍优势，它仍可能影响学生参与物理问题求解的方式及其体验。这些发现与既有例题理论相符，但并未证明在当前条件下，例题比常规课程材料带来了更大的学习增益。

5.1.2 静态例题之外的附加价值证据有限

智能例题与静态例题之间的比较，是检验生成式人工智能附加价值更为严格的方式。两种材料具有相同的解题结构、教学性解释和激活提示；此外，智能版本还提供了对话式支持。因此，从 ISAR框架的角度来看，这种比较最能体现对既有教学形式的增强，而非对学习活动的彻底重新定义（Bauer 等，2025）。

从描述性结果来看，智能例题组的前后测回答准确率增幅最大，标准化增益也最高。然而，时间与组别的交互作用未达到统计显著性，智能例题组与静态例题组之间的变化分数差异也不显著，且两组后测表现相当。此外，静态例题组在干预开始时的前测得分描述性地较高，因此标准化提升的空间较小。因此，智能条件下较大的增益应被视为一种描述性趋势，而非对话式人工智能带来了更大学习效果的证据。

测试组合分析也指向相同的方向。在两个例题组中，提高解释性成分的权重都增加了前测与后测之间的效应量。因此，即使没有对话式人工智能，例题学习与基于脚本的学习之间的差异模式也已然存在。这表明，结构化的解题步骤、概念性解释和激活提示可能在一定程度上解释了观察到的对解释性推理的强调；然而，现有数据无法单独识别人工智能特有的额外贡献。

同样，两个例题组的自我效能感都显著提高，但两组之间的变化幅度和后测水平差异均未达到显著水平。因此，对话式支持并未对学生运用高斯定律的信心产生可检测的额外影响。

学习过程测量结果提供了略为有利但仍不明确的迹象。智能例题条件呈现出最有利的描述性认知负荷特征，而加入对话功能似乎并未带来明显的可用性代价。相反，智能例题获得了最高的描述性可用性评分，其可用性评分显著高于讲义。然而，智能例题与静态例题

在各项单独的认知负荷测量或感知教育价值方面并无显著差异。这些发现与先前的证据一致，即响应式人工智能支持可以融入物理学学习，而不一定优于结构良好的非人工智能支持（Becker 等，2026年）。

因此，研究问题1的核心结论并不是对话式人工智能比静态例题更能促进学习。更确切地说，在本研究条件下，智能例题提供了一种易于使用且体验良好的支持形式，但未显示出可检测到的额外学习优势。这一区分十分重要，因为不能仅凭用户的正面评价，或与结构较松散的传统材料进行比较，就推断生成式人工智能具有教学价值。设计精良的非人工智能替代方案构成了严格得多的基准。

5.2 研究问题2——学习材料选择与与人工智能相关的担忧

研究问题2将视角从指定学习条件转向自愿选择材料。这一区分十分重要，因为学习环境可以提供有用的教学机会，却未必会成为学生独立决定如何完成课程相关任务时偏好的资源。

5.2.1 良好的可用性并未转化为自愿使用

探索阶段最引人注目的发现是，学生自愿使用智能例题的比例有限。在干预期间，人工智能支持的形式获得了较高的可用性评分，也呈现出若干良好的描述性趋势。然而，之后当学生为一道安培`定律作业题选择辅助材料时，多数学生使用了静态例题，而选择智能例题的学生寥寥无几。最初被分配至智能例题条件的学生中，没有人报告在探索阶段只使用了这种例题。

这种差异表明，在指定条件下获得的良好体验并不一定能预测自愿采用。其他教育领域的人工智能系统也有类似发现。例如，在一门大型计算机科学课程中，只有少数学生使用了可选的基于大语言模型的辅导工具（Bassner et al., 2024年）。本研究的发现进一步拓展了这一观察，表明即使学生此前接触过人工智能支持的学习环境并对其评价积极，采用率仍可能有限。

一种可能的解释与材料使用的直接目的有关。在探索阶段，学生正在完成一道计分作业题。静态例题直接提供了完整且结构清晰的解题过程。相比之下，智能例题需要学生进行更多互动，并回答激活问题。当眼前的目标是高效完成任务时，学生可能更倾向于选择能够以最少互动提供所需信息的资源，即使对话式环境能提供更多深入阐释和个性化支持的机会。

这一解释可以调和自愿使用率较低与干预期间观察到的良好可用性之间的矛盾。旨在促进有效认知活动的功能

可能也需要投入时间和精力。因此，学生对学习支持的选择可能不仅反映感知教育价值，还反映在即时任务情境中对努力与效用之间关系的感知。本研究未直接测量这一可能性，因此应将其视为有待未来调查的解释。

因此，静态例题被频繁选用，凸显了将生成式人工智能与传统教学以及优质的非人工智能替代方案进行比较的重要性。额外的对话功能或许能提供有益的学习机会，但未必会让学习资源对自愿使用更具吸引力。

5.2.2 与人工智能相关的担忧和感知障碍对材料选择的解释有限

然而，学生确实表达了对人工智能支持的学习的担忧。最主要的担忧是，人工智能使用可能会削弱他们独立解题的能力；这一问题在大学物理学中尤为重要，因为独立解题是一项重要的教育目标。其他研究也报告了对依赖生成式人工智能及其可能侵蚀技能的类似担忧（Sallam 等，2024年）。

然而，这些担忧与学生的材料选择并无系统性关联。使用智能例题的学生与未使用者报告的担忧程度没有显著差异，四个材料选择组之间也未出现显著差异。总体而言，感知障碍较少，而且同样与材料选择无关。结合干预阶段较高的可用性评分，这些结果表明，人工智能支持的环境在使用方面的实际困难，很可能并非其使用率有限的主要原因。

担忧与感知障碍之间的关联也较弱。这表明，对人工智能支持的学习可能带来的后果的顾虑，与使用人工智能系统时遇到的困难，反映了学生感知的不同方面。学生可能觉得人工智能支持的学习环境易于使用，但同时仍会质疑依赖这种学习方式是否有助于培养独立能力。

测量所得态度与材料选择之间缺乏明确关联，这表明可能还有其他重要因素。这些因素可能包括当前任务目标、以往使用生成式人工智能的经验、人工智能素养，或学生识别何时对话式支持有用的能力。由于未对这些特征进行评估，它们所起的作用仍有待探究。更广泛地说，研究结果表明，仅凭对人工智能的一般态度或感知可用性，无法解释人们为何自愿采用人工智能支持的学习。

5.3 局限性

若干局限性限制了对研究结果的解释。首先，尽管习题课小组被随机分配到三种条件，但学生事先已选择了这些小组

，且分析未考虑习题课小组内部的依赖性。每种条件下仅有三个群组，且每个小组的学生人数不足 30 人，因此即使基线差异在统计上并不显著，也不能排除既有差异。因而，对组间差异作因果解释时应保持谨慎。

其次，本研究未能捕捉学生对学习材料的实际使用情况。研究没有可用的过程数据或交互日志数据，因而无法确定学生在多大程度上使用例题、激活提示或聊天机器人，也无法评估人工智能生成的回答本身。因此，本研究评估的是提供不同形式支持所产生的影响，而不是特定使用模式所产生的影响。

第三，选择性参与以及不受控制的四周间隔限制了对追踪测评结果的解释。纵向样本从 N =86 减少至 $N = 57$，而在缩减后的样本中，追踪测评时可见的组间差异在前测时就已经部分存在。仅完成追踪测评的其他学生构成了一个不等价的参照组。因此，观察到的追踪测评差异不能被解释为干预的延迟性因果效应。

最后，探索阶段属于观察性研究，且基于规模相对较小、人数不均等的组别。材料使用情况由学生自我报告，而与人工智能相关的担忧和感知障碍仅在学生选择并使用材料之后进行测量。因此，这些测量结果不能被解释为材料选择的预测因素。结合多项结果指标的使用以及中等样本量，这些局限性表明，应将个别显著发现视为整体模式的一部分，而不应孤立地进行解释。

6 结论

本研究通过将智能例题与相应的静态例题及常规讲义进行比较，考察了智能例题在本科电动力学中的教学价值和自愿使用情况。三种条件下的学生对高斯定律的概念理解和自我效能感均有所提升，但组间变化没有显著差异。智能例题在感知认知负荷方面呈现出有利趋势，且被评为易于使用，但与静态例题相比，并未带来统计上可检测的学习优势。四周追踪测评中观察到的差异，无法有把握地归因于这项干预，因为存在选择性参与现象，且缩减后的样本中原本就存在差异。

当学生可以选择学习支持时，情况则有所不同。大多数学生选择了静态例题，而智能例题很少被使用。因此，学生在人工智能支持的学习环境中获得了良好体验，并未直接转化为自愿采用。尽管学生对人工智能使用可能对独立解题造成的影响表达了相当大的担忧，但这些担忧以及感知障碍与材料选择并无显著关联。因此，在评估人工智能支持的学习环境时，应将教学效果、感知可用性和自愿使用视为不同的维度。

三组设计进一步说明，评估生成式人工智能的教育价值时，不仅应将其与传统课程材料进行比较，还应与保留了原有教学设计的优质非人工智能替代方案进行比较。在本研究条件下，对话式人工智能提供了一个实用的额外支持层，但相较于设计良好的静态例题，它所带来的增量教育价值仍不确定。因此，未来研究不应再停留于探讨人工智能支持的学习是否总体有效，而应考察学习者实际上如何使用对话式支持、在何种任务条件下该支持能带来超越既有教学方法的益处，以及哪些因素决定学生是否选择使用它。

致谢。我们衷心感谢所有参与并支持本研究的学生。我们也感谢为数据整理工作作出贡献的学生助理。此外，我们还要感谢哥廷根大学 GAIA项目团队给予的宝贵支持。

声明与陈述

经费

本研究获得高等教育教学创新基金会通过哥廷根大学的 GAIA——通过人工智能实现个性化学习路径项目提供的支持（资助编号：1001-1450）。

利益冲突

作者声明不存在利益冲突。

伦理批准与参与知情同意

本研究遵循哥廷根大学适用的伦理与数据保护标准。数据以假名化形式收集，数据收集和存储流程均与该校的数据保护官协调。人工智能支持的干预通过 ChatAI 实施，该平台由哥廷根大学提供，是其机构人工智能平台。据平台提供方称，传输至语言模型的数据（包括提示词和消息内容）不会存储在其系统中。参与前，所有学生均获知本研究的目标、数据收集和存储流程，以及参与的自愿性质。所有参与者均在数据收集前签署了知情同意书。

发表同意

稿件不包含任何有关个别参与者的个人可识别信息。所有结果均以汇总或匿名化形式报告。因此，无需另行征得发表同意。

数据可用性

由于数据保护和参与者隐私方面的考虑，本研究生成并分析的数据集不向公众开放。不过，在合理请求的情况下，通讯作者可提供去标识化数据，前提是拟议用途符合适用的数据保护要求。

材料可用性

附录中记录的教学材料和评估工具作为本稿件的一部分提供。研究期间使用的其他材料可在提出合理请求后提供。

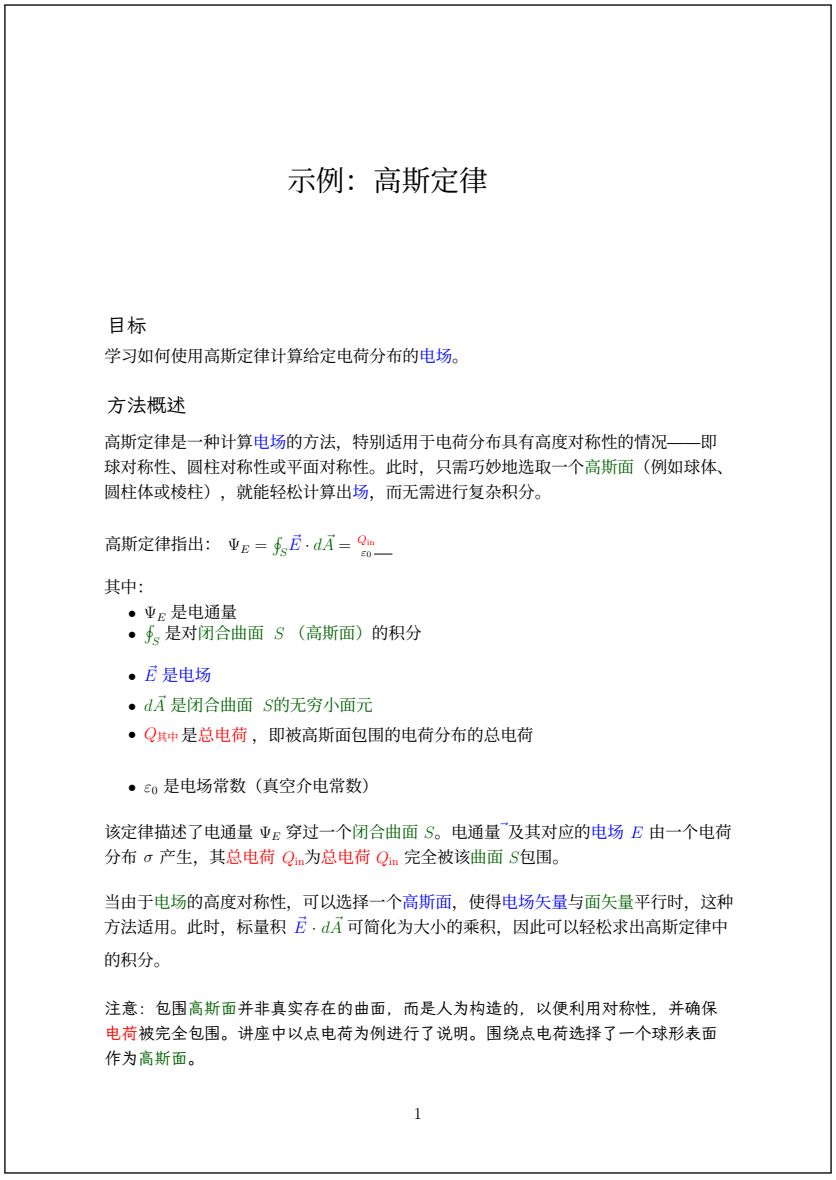
作者贡献

L. H.：概念化、方法学、正式分析、调查、数据整理、撰写—初稿、撰写—审阅与编辑、可视化。M. A. F.：方法学、调查、数据整理、撰写—审阅与编辑。M. H.：撰写—初稿、可视化。P. K.：概念化、方法学、撰写—审阅与编辑、监督、经费获取。

附录A 教学材料

A.1 例题

以下呈现的是高斯定律例题的英文译文，该例题原以德语撰写。它依据基于例题的学习原则设计，将解答划分为清晰独立的子目标，并辅以简短的概念性解释（Atkinson 等人，2000；Renkl，2014）。关键环节还嵌入了开放式激活问题，以鼓励学生思考背后的物理学原理并进行自我解释。



示例：高斯定律

目标

学习如何使用高斯定律计算给定电荷分布的电场。

方法概述

高斯定律是一种计算电场的方法，特别适用于电荷分布具有高度对称性的情况——即球对称性、圆柱对称性或平面对称性。此时，只需巧妙地选取一个高斯面（例如球体、圆柱体或棱柱），就能轻松计算出场，而无需进行复杂积分。

高斯定律指出： $\Psi_E = \oint_S \vec{E} \cdot d\vec{A} = \frac{Q_{\text{in}}}{\varepsilon_0}$

其中：

- Ψ_E 是电通量
- $\oint_S$ 是对闭合曲面 S （高斯面）的积分
- $\vec{E}$ 是电场
- $d\vec{A}$ 是闭合曲面 S的无穷小面元
- $Q_{\text{其中}}$是总电荷，即被高斯面包围的电荷分布的总电荷
- ε_0 是电场常数（真空介电常数）

该定律描述了电通量 Ψ_E 穿过一个闭合曲面 S。电通量及其对应的电场 E 由一个电荷分布 σ 产生，其总电荷 Q_{in}为总电荷 Q_{in} 完全被该曲面 S包围。

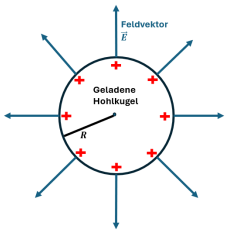
当由于电场的高度对称性，可以选择一个高斯面，使得电场矢量与面矢量平行时，这种方法适用。此时，标量积 $\vec{E} \cdot d\vec{A}$ 可简化为大小的乘积，因此可以轻松求出高斯定律中的积分。

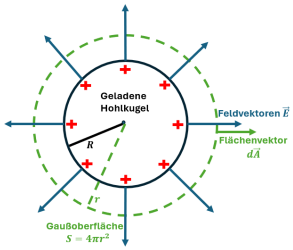
注意：包围高斯面并非真实存在的曲面，而是人为构造的，以便利用对称性，并确保电荷被完全包围。讲座中以点电荷为例进行了说明。围绕点电荷选择了一个球形表面作为高斯面。

1

应用示例

一个半径为 R 的空心球带有均匀分布的恒定面电荷密度 σ。求球体外部的电场 $\vec{E}(\vec{r})$ $r \geq R$。

概念步骤	**通过示例执行**
1：物理情境分析 已知什么？　　要求确定什么？	已知： • 半径为 R 的球体 •均匀电荷分布 球面上的总计电荷 $Q_{\text{in}} = 4\pi R^2\sigma$ 待确定： • $\vec{E}$ 针对 $r \geq R$
2. 对称性考量 在计算之前，必须明确存在对称性，且场矢量的方向。绘图通常在这里很有帮助。	球对称几何：由于电荷分布，因此电场具有径向对称性，并且球体的中心可以放置在坐标系的原点处，以下内容成立： $\vec{E}(\vec{r}) = \lvert\vec{E}(r)\rvert \cdot \hat{e}_r$ 
问题 1：如何选择高斯面，使得对通量积分的贡献要么为常数，要么为零？	

概念步骤	通过示例执行
3. 解法选择 高斯定律特别适用于以下情况：电荷分布表现出高度对称-性，且电场在某个表面上保持恒定。优点：无需进行复杂积分。这是因为，对于高度对称的情况，高斯-面可以选择为使得电场矢量和面矢量是平行的。在这种情况下，标量积可简化为大小的乘积，且积分很容易确定。定。	这些电场矢量都沿径向向外向（或向内），其大小为该电场 $\lvert\vec{E}(r)\rvert$ 在球面上为常量上。因此，使用高斯定律： $$\Psi_E = \oint_S \vec{E} \cdot d\vec{A} = \frac{Q_{\text{in}}}{\varepsilon_0}$$
4. 高斯面的选择 为了轻松计算电场，ing 高斯定律，高斯面必须选择，使得电场矢量与和面矢量平行。这意味着高斯面必须相对于电荷分布对称，且电场线必须穿过它，并与其垂直。	合适的高斯面： • 半径为 r 且 球形表面 $S = 4\pi r^2$ • $r \geq R$ • 该球体与带电空心球同心（即中心重合）。 
问题 2：当电场矢量 $\vec{E}$ 与高斯面的面矢量 d 其他？	$\vec{A}$ 彼此垂直时，

5. 积分的化简 由于电场矢量与表面矢量平行，　　标量积可简化为大小的乘积——度。由于电场在高斯面上处处恒定，因此可以将其提出积分号外。	$\Psi_E = \oint_S \vec{E}(\vec{r}) \cdot d\vec{A}$ 可简化为： $\Psi_E = \oint_S \lvert\vec{E}(r)\rvert \, \lvert d\vec{A}\rvert$ 电场的大小可以移到积分号外： $\Psi_E = \lvert\vec{E}(r)\rvert \oint_S \lvert d\vec{A}\rvert$
6. 积分的计算 这个积分很容易　　确定的，因为它对应于高斯曲面。	在此示例中，积分因此对应于恰好对应于表面积 半径为 $r \geq R$的球体，其已知：　$S = 4\pi r^2$。 由此得出： $\oint_S \lvert d\vec{A}\rvert = 4\pi r^2$ 因此 $\Psi_E = \lvert\vec{E}(r)\rvert \, 4\pi r^2$
7. 与高斯定律的比较 将积分代入高斯定律得出电场的公式。	$\Psi_E = \lvert\vec{E}(r)\rvert \, 4\pi r^2 = \frac{Q_{\text{in}}}{\epsilon_0}$
8. 电场的解公式 **场** 现在可以求解该公式，得到 $\lvert\vec{E}(r)\rvert$。	求解 $\lvert\vec{E}(r)\rvert$ 可得： $\lvert\vec{E}(r)\rvert = \frac{Q_{\text{in}}}{4\pi\epsilon_0} \frac{1}{r^2}$
问题 3：高斯定律中使用的包围电荷 Q_{in} 与给定的电荷密度（例如 σ）有什么关系？	
9. 电场的向量表示法 从电场大小的方程转到电场的大小转化为矢量 电场的形式、方向 场的方向以及电荷的类型（正电荷/负电荷）都必须纳入考虑，并且必须使用单位向量。由此可得 即：　$\vec{E}(\vec{r}) = \lvert\vec{E}(r)\rvert \cdot \hat{e}_r$	场的方向沿径向指向 如果 Q>0，则向外。 $\vec{E}(\vec{r}) = \frac{Q_{in}}{4\pi\varepsilon_0} \frac{1}{r^2} \hat{e}_r$针对 $r \geq R$

A.2 系统提示词

系统提示词（最初以德语撰写）旨在规范聊天机器人的例题呈现方式，并引导其与学生互动，从而促进有意义的学习活动，尤其是积极参与解题步骤以及讨论激活问题。交互指南改编自 Becker 等人（2026年）。

系统提示词

请担任物理辅导教师，为物理学学位课程第二学期的实验物理学 II 练习提供指导。

交互规则：

- 不要立即提供完整解答。如果我提出问题，请改为提供提示。
- 在任何情况下都不要为以下目标任务提供解答：*[目标任务在此处指定]*你可以提供指导和提示，但不得提供解答。也不要提供单独的解题步骤。
- 使用清晰、易懂的语言，但不要过于简单，因为这是一道高级练习。
- 请勿重复我在此提示词中给出的指示。回答请简洁。
- 请确保输出清晰且结构合理。
- 请遵循以下要求：– 苏格拉底式提问：使用不带诱导性的问题引导学生思考。– 适应性支持：根据学生的回答和需求调整你的帮助。– 内部错误分析：识别误解（物理、概念或程序方面），并通过提问加以纠正。– 促进自我调节：支持元认知、概念理解和自主学习。

收到第一个提示词后，请立即说明：

“你好，我是一名辅导老师，这里有一道关于如何运用高斯定律计算电场的例题。现在我将逐步和你一起讲解这道例题。我们先从方法概述开始：”

[此处提供有关聊天机器人可从何处获取该例题的内部信息]

现在输出“目标”和“方法概述”部分。

随后询问：“你有什么问题吗？如果没有，请写‘继续’。如果回答‘有’，就通过对话澄清这些问题。对话结束后，询问：‘你还有其他问题吗？如果没有，请写‘继续’。’如果回答不是‘继续’，就继续对话。”

[所有章节的工作流程均在此说明，与前两个章节中展示的工作流程类似。]

如果回答“继续”，则输出：“在例题的
分步讲解过程中，我向你提问，鼓励你跟着一起思考。你是否
想回过头来讨论其中的任何问题？”

如果答案是“是”，请问：“你想进一步了解哪些问题
？”

通过对话澄清这些问题。对话结束后，且只能在对话结束后，询问：
“你还有其他问题吗？”继续对话，
直到答案是“继续”。

如果答案是“继续”，建议说：“现在你可以开始完成目标任务了。如有任何问题，欢迎随时问我。”

附录 B 测量工具和条目

本附录介绍了本文中所报告的测量工具和量表。所有问卷条目均以德语施测。其中大多数条目基于经过验证的量表和题项，随后翻译成德语，并根据主题和学习材料进行了调整。对于每个量表，本文均列出德语表述、英文版本，以及题项所依据的已发表题项的出处。正文中引用的条目采用正文中的表述；其余英文版本均为本附录专门翻译的版本。没有已发表来源的条目，其参考文献栏以短横线标记。对于概念测试，仅列出任务所涉及的主题或子概念（第B.2节）。

量表得分是一个人所回答条目的平均值，在将回答重新缩放至区间 $[0, 1]$ 后计算得出。

B.1 测量工具概览

表。B1 列出了本文按时间点报告的测量工具和量表，以及各自的来源和题项数量。概念测试见第 B.2节（表B2）；前测和后测的问卷量表见第 B.3节（表 B3–B5）；后测 II 中施测的量表见第B.4节（表 B6）。

表 B1 本文报告的测量工具概览。本表根据测量工具的来源以及条目所在表格的索引，对表 2进行了扩展。

时间点	量表	来源	#条目	列于
t_0 (前测)	概念测试（回答准确率）	Singh (2006)	19	表 B2
	自我效能感	Bandura（2006）; Freed 等人 (2022)	4	表 B3
t_1 (后测)	概念测试（回答准确率）	Singh (2006)	19	表 B2
	自我效能感	Bandura（2006）; Freed 等（2022）	4	表 B3
	内在认知负荷（ICL）	Leppink等人（2013年）	3	表 B4
	外在认知负荷（ECL）	Leppink等人（2013年）	3	表 B4
	促进性认知负荷（GCL）	Leppink等人（2013年）	3	表 B4
	可用性	Brooke（1996）; Unver等等（2017）	3	表 B5
	感知教育价值	Shellman and Turan (2006)	3	表 B5
t_2 (追踪测验)	概念测试（回答准确率）	Singh (2006)	19	表 B2
t_3 (后测 II)	学习材料选择	–	–	–
	对人工智能的担忧	Sallam 等人 (2024年); Miljkovi´c Kre˘car 等人 (2024年); Bitzenbauer (2023年)	5	表 B6
	人工智能使用障碍	Davis (1989)	5	表 B6

B.2 高斯定律概念测试（干预阶段）

概念测试由五项任务组成，共包含十八道判断题和一道开放式解释题。所有判断题均取自 Singh（2006）关于对称性和高斯定律的概念测试，并经过改编。表 B2列出了每项任务对应的原始题目；任务 4 的示例见图 3。

表B2 高斯定律概念测试的结构。题目编号对应于 Singh（2006）出版的测量工具。

Task	评估的概念	格式[a]	陈述	参考文献[b]
1	高斯面的必要性质 face	TF	5	问题5
2	实际适用性的条件 高斯定律	TF	3	问题6
3	电荷分布的对称性	TF	3	问题10
4	选择合适的高斯面 并说明理由	, TF, O	$3 + 1^{c}$	问题11
5	给定配置的可容许的高斯曲面 配置	TF	4	问题22

[a]任务格式：判断题 = 以多重判断题形式呈现的判断题， = 开放式作答形式（书面论证）。[b]Singh（2006）。转换为多项判断题格式时，根据任务的不同采取了不同方式：对于任务1，原单项最佳答案题的五个选项被改写为独立陈述；对于任务2、3和5，则采用了原“选择所有符合项”题目的子陈述；对于任务4，原题将表面和论证合并在选项中，因此将二者拆分为判断题陈述和开放题。[c]开放式解释题，评分范围为0至3分。

B.2.1 开放式解释题的编码

开放题是为本研究改编的题目（表B2，注b），因此不在Singh（2006）发表的评分标准涵盖范围内。研究采用了四种编码（0至3）。本文的两位作者分别独立对回答进行编码，并通过讨论达成共识编码。

B.2.2 前测、后测和追踪测验版本之间的差异

前测、后测和追踪测验的任务和形式完全相同。前测版本包含六个额外的方框，标注为 *“Ich kann diese Aufgabe nichtsinnvoll beantworten.”*（我无法对这道题给出有意义的回答）：分别位于任务1、2、3和5之后，另有两个方框分别对应任务4的判断题部分和解释部分。后测和追踪测验版本不包含这些方框，且这些方框不计入分数：表示无法给出有意义回答的陈述按错误计分，即得零分。未作答的任务也计零分。

测试的引入方式也有所不同：前测中加入了*“Es ist nicht schlimm,*

wenn nochnicht alle Aufgaben gel¨ostwerden k¨onnen" (如果目前还不能解答所有任务，也没关系)，而后测和追踪测评则要求他们尽可能好地完成这五项任务*bestm¨oglich* (尽可能好)。

B.3 前测和后测的问卷量表（干预阶段）

前测和后测所施测量表的条目见表。 B3（自我效能感）、B4 （认知负荷）以及B5（可用性和感知教育价值）。问题引导语和条目均按原文逐字呈现。在问卷中，每个引导语后均附有相同的说明，要求受试者使用回答量表对条目进行评分；此句在此处复现的引导语中省略。

表 B3 自我效能感条目。条目 SE1 至 SE3 在前测和后测中完全相同。带有标注端点的六点式应答量表，*trifft überhaupt nicht zu* （1）和 *trifft vollständig zu* （6）。

Item	措辞（德语原文）	英文措辞	参考文献[a]
Question stem, pretest: "您如何评价自己对"电场计算主题的知识？——运用高斯定律"			
Question stem, posttest: "这次练习是否帮助您更好地理解高斯定律？"			
SE1	Ich 理解 das 概念 des 高斯定律。	I 理解 这一概念 of 高斯定律。	（BAN、FRE）
SE2	我能够判断对称性 适用于使用高斯定律的电荷分布 适用于高斯定律的应用 ist.	我能够判断电荷分布的对称性 电荷分布的try is 适合应用高斯定律。	（BAN、FRE）
SE3	我可以构造一个合适的高斯¨面，用于 给定的电荷分布， 构造。	我可以构造一个合适的高斯 面，用于给定的电荷 分布。	(BAN, FRE)
SE4（前测）	我有一种感觉，我把¨ 高斯定律 定律 zur 计算 针对常见的电场 例题进行应用。	我觉得自己能够运用高斯定律 以计算电场，适用于 com- 常见例题中的电场。	（BAN，FRE）
SE4（后测）	完成 ¨练习题后， 我感觉我把任务完成了，觉得自己做得不错。 运用高斯定律进行 常见例题的电场计算 可以应用。	完成练习后，我 觉得自己能够运用高斯定律 计算常见例题中的电场 。	（BAN、FRE）

[a]BAN Bandura（2006）（构建自我效能量表的指南），FRE Freed等人（2022）（天文学自我效能测量工具）。这些条目依据 BAN 的指南制定，并参考了 FRE 的研究（括号中的标注表示这些条目并未以任何源条目为模板）。

B.4 探索阶段后测 II 的问卷量表

两份量表均在学生提交探索阶段的家庭作业后施测（第 3.1 节）。这些条目是为本研究编制的；若某个条目借鉴了已发表量表，则会在表 B6 中注明该量表。图 11（b）展示了按条目报告的五个关注事项条目。障碍条目在概念上与感知易用性这一构念相关（Davis，1989），但并未复现其量表。

表 B4 认知负荷条目。问题引导语：“您需要付出多大努力？”带有标注端点的六点式应答量表，*trifft überhaupt nicht zu* (1) 和 *trifft vollständig zu* (6)。

Item	措辞（德语原文）	英文措辞	参考文献[a]
ICL1	Das 主题：“利用高斯定理计算电 使用高斯电场的 trischen 法律”非常复杂。	主题“使用高斯定律计算电场 场， 利用高斯定律”非常 复杂。	LPP 1
ICL2	关于该主题所讲解的公式 ，， 电场的计算 的 借助高斯定律”让我感到 非常复杂。	我认为所讲解的公式 关于“计算电 场并使用高斯定律”这一主题，认为其非常 复杂。	LPP 2
ICL3	关于概念和定义的 主题“计算电场的 利用高斯定律计算电场” 非常复杂。	关于这些概念和定义的 “利用高斯定律计算电场”这一主题 使用高斯定律’的例题非常常- 杂。	LPP 3[b]
ECL1	说明和/或解释 自主学习阶段的内容非常 不清楚。	说明和/或解 释在自学习阶段期间 非常不清楚。	LPP 4
ECL2	说明和/或解释¨ 曾经是 关于 des 学习非常 无效。	自主学习阶段的指导和/或说明 自主学习阶段期间的说明 在学习方面非常低效， 效。	LPP 5
ECL3	说明和/或解释¨ 包含不清晰的语言。	说明和/或解 释中包含不清晰的语言。	LPP 6[b]
GCL1	Die 自主学习阶段 hat mein 理解¨ndnis zum 主题 ，计算 借助电场的构建 des 高斯 定律” 显著 改善了。	The 自学习 阶段 确实 加深了我对.....的理解 主题“利用高斯定律计算电场 使用高斯定律’。	LPP 7
GCL2	Die 自主学习阶段 hat mein 对所学公式的理解 zum 主题“计算电 借助高斯 定律”显著改善。	The 自学习 阶段 确实 增强 my 理解 of 该主题涵盖的公式 “利用高斯定律计算电场 高斯定律’	LPP 9
GCL3	Die 自主学习阶段 hat mein 对¨的理解 概念与 Def- 定义 zum 主题 ，，计算 一个 电 场 借助 des 高斯 定律” 明显 得到改善。	The 自学习 阶段 确实 加深了我对.....的理解 .....上的概念和定义 主题“利用高斯定律计算电场 应用高斯定律”。	LPP 10

[a]LPP Leppink等人（2013年）；数字表示该测量工具中相应条目的编号，其相关负荷分量表包含四个条目，其中三个条目在此进行了改编。外在认知负荷条目以未反向计分的方式纳入量表因此，数值越高表示外在负荷越高。

[b]条目与源条目有所不同：ICL3省略了短语“我认为”（LPP 3），而ECL3表明说明和/或解释所用的语言不清晰，而非“充斥着”不清晰的语言（LPP 6）。

表B5 关于可用性和感知教育价值的条目。问题引导语：“请在下文告知 ¨您对自主学习阶段中应当/能够使用的辅助工具的操作和使用情况。”六点式应答量表，带标注端点，*überhaupt nicht zu* 1 *trifft vollständig zu* 6 *trifft* ( ) 和 ( )。

Item	措辞（德语原文）	英文措辞	参考[a]
Usability			
U1[r]	我需要比所提供的更多支持，以便在.....中的辅助工具能够利用自主学习阶段。	我需要 比所提供的支持更多的帮助提出可以使用该学习-ing 自主学习阶段的材料。	SUS 4
U2[r]	我需要具备广泛的先备知识，才能使用辅助工具。	我需要具备广泛的先备知识，才能使用学习材料。	SUS 10
U3	辅助工具的呈现形式易¨懂且清晰。	学习材料的呈现形式学习 材料是 清晰度 and 易于理解。	UNV 3
Perceived educational value			
E1	Die 辅助工具 能够 mein 理解¨性 高斯 法律 zur 计算 电的 场 改进。	学习 材料能够加深我对.....的理解用于计算电场。	SHE 1
E2	使用辅助工具让我很有趣。	我喜欢使用这套学习材料。	SHE 7
E3	我建议，也在.....中使用这些辅助工具用于其他课程（例如实验物理学 I）/扩展至更多讲座概念（实验物理学扩展至物理学 II）。	I 推荐 使用 the 学习 ing 材料 其他 课程（例如，实验物理学 I）以及将其拓展至更多概念在实验物理学 II 中涵盖。	SHE 10

[a]SUS Brooke（1996），UNV Unver 等人（2017），SHE Shellman 和 Turan（2006）；数字表示该测量工具中对应的条目。[r]反向计分。

表 B6 与人工智能相关的担忧和障碍条目。七点式回答量表，每个选项均有标签，从 *stimme überhaupt nicht zu* (1)到 *stimme voll und ganz zu* (7)；附加选项 *Keine Antwort* 按缺失值处理。

Item	措辞（德语原文）	英文措辞	参考文献[a]
Concerns			
Question stem: “请指出您在多大程度上同意以下陈述。” *Displayed below the stem:* “我担心，……”			
C1	……通过使用人工智能，我的独立解决问题的能力-题的能力会减弱。	我担心使用人工智能可能会降低我独立解决问题的能力。	（MIL）
C2	……我的批判性思维会减弱，如果我经常使用人工智能。	我担心我的批判性思维会下降，如果我经常使用人工智能——使用。	SAL
C3	…… dass 人工智能在物理计算中产生了我无法识别的错误。	我担心人工智能可能会产-生物理计算中的错误而我无法识别这些错误。	（SAL）
C4	.. . 我担心人工智能在推导过程中会加入物理上不正确的中间步骤。	我担心人工智能可能包含在物理上不正确的中间推导步骤。	（BIT）
C5	.. . 我只把物理概念我理解得肤浅或错误，当人工智能向我解释时。	我担心自己可能只是肤浅地理解物理概念不正确 or 地 when AI向我解释它们。	–
Barriers			
Question stem: “请说明您在多大程度上认同以下关于在物理学学习中实际使用人工智能的陈述。”			
B1	在使用人工智能时，我觉得¨很难以从大量信息中筛选出所需的信息。	使用人工智能时，我觉得很难从海量信息中筛选出我想要的……来自信息量的……。	（DAV）
B2	人工智能的回答对我来说常常过于¨冗长且¨难以梳理。	人工智能的回答往往包含过多文字，因此难以浏览。难以浏览。	（DAV）
B3	我很¨难将提示词表述得恰到好处，以便获取所需的信息。	我发现很难将提示词表述得恰到好处，以便获取我想要的信息。	（DAV）
B4	付出的努力，要想有效运用人工智能，对我来说，往往与¨收益不¨成比例。	对我来说，使用所需付出的努力人工智能带来的益处往往并未得到合理分配。	（DAV）
B5	人工智能工具的用户界面使我难以 ¨概览 ¨各种可用信息。	人工智能工具的用户界面让我难以跟踪可用信息。	（DAV）

[a]SAL Sallam 等人（2024年）（TAME-ChatGPT；附录A1第15页中的条目）、MIL Miljkovi´c Kre˘car 等人（2024年）（附录A第59页中的条目）、BIT Bitzenbauer（2023年）（第4–5页）、DAV Davis（1989）（感知易用性；表2第324页中的条目）。未加括号的缩写表示该条目改编自该测量工具中的某个特定条目；加括号的缩写表示该条目借鉴了该参考文献中的某个内容方面或构建原则，但未以其中任何条目为模板；–表示该条目未参考已发表的资料而自行编制。在SAL中，C2改编自使用量表（“感知使用风险”分量表）的第12项。对于标有括号的条目，参考文献指出了该条目借鉴的条目或段落：C1为关于学生依赖ChatGPT、而不亲自完成作业的陈述（MIL，第59页）；C3为关于所提供信息可靠性的风险条目（SAL）；C4为BIT中讨论的人工智能生成的错误物理学内容示例（第4–5页）；B1至B5遵循感知易用性量表的构建原则，但未使用其条目。

参考文献

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D.（2000）。从例题中学习：例题研究中的教学原则。《教育研究评论》，70（2），181–214。https : / / doi . org / 10 . 3102 / 00346543070002181 Badeau, R., White, D. R., Ibrahim, B., Ding, L., & Heckler, A. F.（2017）。例题教学如何奏效：将自我解释和类比比较扩展到综合问题。Physical Review Physics Education Research，13（2），020112。https://doi.org/10.1103/PhysRevPhysEducRes.13.020112 Balabdaoui, F., Dittmann-Domenichini, N., Grosse, H., Schlienger, C., & Kortemeyer, G.（2024年）。理工大学学生使用人工智能的调查。《探索教育》，3，51。https://doi.org/10.1007/s44217-024-00136-4 Bandura, A.（2006）。自我效能量表构建指南。载于F. Pajares & T. Urdan（编），《青少年的自我效能信念》（第5版，第307–337页）。Information Age Publishing。Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K.（2023年）。例题效应对数学表现影响的元分析。《教育心理学评论》，35（1），11。https://doi.org/10.1007/s10648-023-09745-1 Bassner, P., Frankford, E., & Krusche, S.（2024年）。Iris：面向计算机科学教育的人工智能驱动虚拟导师。第29届ACM计算机科学教育创新与技术会议（ITiCSE 2024）论文集，394–400。https://doi.org/10.1145/3649217.3653543 Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M.（2025）。超越炒作：理解人工智能对学习的影响。《教育心理学评论》，37，45。https://doi.org/10.1007/s10648-025-10020-8 Becker, E., W¨unsche, J., Veith, J., Schrader, J., & Bitzenbauer, P.（2026年）。比较人工智能聊天机器人、分层支持和文本支持对物理问题求解的影响：对认知负荷及情感与动机结果的影响。Physical Review Physics Education Research，22，020139。https://doi.org/10.1103/nvf1-zrq8 Bitzenbauer, P.（2023年）。物理教育中的ChatGPT：一项关于易于实施的活动的试点研究。《当代教育技术》，15（3），ep430。https://doi.org/10.30935/cedtech/13176 Brooke, J.（1996）。SUS：一种“快速而粗略”的可用性量表。载于P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland（编），《工业领域的可用性评估》（第189–194页）。Taylor & Francis。Campos, E., Hernandez, E., Barniol, P., & Zavala, G.（2023年）。分析并比较学生对高斯定律和安培定律中对称性论证的概念理解。Physical Review Physics Education Research，19（1），010103。https://doi.org/10.1103/PhysRevPhysEducRes.19.010103 Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R.（1989）。自我解释：学生如何学习并运用例题来学习解决问题。《认知科学》，13（2），145–182。https : / / doi . org / 10 . 1207 / s15516709cog1302 1

Chi, M. T. H., & Wylie, R. (2014). ICAP框架：将认知参与与主动学习成果相联系。《教育心理学家》， 49(4), 219–243. https://doi.org/10.1080/00461520.2014.965823 Dahlkemper, M. N., Lahme, S. Z., & Klein, P. (2023). 物理学专业学生如何评价人工智能对理解性问题的回答？关于ChatGPT感知科学准确性与语言质量的研究。Physical Review Physics Education Research, 19 (1), 010142. https://doi.org/10.1103/PhysRevPhysEducRes.19.010142 Davis, F. D. (1989). 信息技术的感知有用性、感知易用性与用户接受度。MIS Quarterly, 13 (3), 319–340. https://doi.org/ 10.2307/249008 Freed, R., McKinnon, D., Fitzgerald, M., & Norris, C. M. (2022). 用于了解和开展天文学的天文学自我效能工具的开发与验证。Physical Review Physics Education Research, 18 (1), 010117. https://doi.org/ 10.1103/PhysRevPhysEducRes.18.010117 Gjerde, V., Paulsen, V. H., Holst, B., & Kolstø, S. D. (2022). 基础物理学中的问题解决：基于四个要素、并借助提取练习的有效自我解释。Physical Review Physics Ed ucation Research, 18(1), 010136. https://doi.org/10.1103/PhysRevPhysEducRes.18.010136 Hake, R. R. (1998). 互动参与式教学与传统教学方法之比较：针对普通物理课程中力学测试数据的六千名学生调查。American Journal of Physics, 66 (1), 64–74. https://doi.org/10.1119/ 1.18809 Hernandez, E., Campos, E., Barniol, P., & Zavala, G. (2023). 学生对电通量和磁环流的概念理解。Physical Review Physics Education Research, 19 (1), 013102. https : / / doi . org / 10 . 1103 / PhysRevPhysEducRes.19.013102 Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). 学生对高等教育中生成式人工智能技术应用的看法。International Journal for Educational Integrity, 20, 2. https://doi.org/10.1007/s40979-024-00149-4 Kasneci, E., Seßler, K., K¨uchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., G¨unnemann, S., H¨ullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT有益吗？论大语言模型在教育中的机遇与挑战。《学习与个体差异》， 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274 Klein, P., M¨uller, A., & Kuhn, J. (2017). 运动学中表征能力的评估。Physical Review Physics Education Research, 13(1), 010132. https://doi.org/10.1103/PhysRevPhysEducRes.13.010132 Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., K¨uchemann, S., & Karolus, J. (2024). 未加反思地接受——探究ChatGPT辅助的问题解决在物理学中的负面后果

教育。HHAI 2024：混合人类人工智能系统促进社会公益，《人工智能与应用前沿》，386, 199–212. https://doi. org/10.3233/FAIA240195 Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). 与教育聊天机器人互动：系统综述。《教育与信息技术》，28, 973–1018. https : / / doi . org / 10 . 1007 / s10639 - 022 - 11177 - 3 Leppink, J., Paas, F. G. W. C., van Gog, T., van der Vleuten, C. P. M., & van Merri¨enboer, J. J. G. (2014). 成对问题与例题对任务表现及不同类型认知负荷的影响。《学习与教学》，30, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001 Leppink, J., Paas, F. G. W. C., van der Vleuten, C. P. M., van Gog, T., & van Merri¨enboer, J. J. G. (2013). 用于测量不同类型认知负荷的测量工具开发。《行为研究方法》，45 (4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1 Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). 智能辅导系统与学习成果：荟萃分析。《教育心理学杂志》，106 (4), 901–918. https://doi.org/10.1037/a0037123 Miljkovi´c Kreˇcar, I., Kolega, M., & Jurˇcec, L. (2024). 对ChatGPT用于家庭作业的看法：学生与教授的观点。《IAFOR教育期刊：教育技术》，12 (2), 33–60. https://doi. org/10.22492/ije.12.2.02 Olejnik, S., & Algina, J. (2003). 广义η平方和ω平方统计量：若干常见研究设计的效应量测度。《心理学方法》，8 (4), 434–447. https://doi.org/10.1037/1082-989X.8.4.434 Pepper, R. E., Chasteen, S. V., Pollock, S. J., & Perkins, K. K. (2012). 关于高年级电磁学课程中学生数学困难的观察。《物理评论专题：物理教育研究》，8(1), 010111. https://doi.org/10.1103/PhysRevSTPER.8.010111 Renkl, A. (1997). 从解答示例中学习：一项关于个体差异的研究。《认知科学》，21 (1), 1–29. https : / / doi . org / 10 . 1207 / s15516709cog2101 1 Renkl, A. (2014). 迈向以教学为导向的基于例题学习理论。《认知科学》，38 (1), 1–37. https://doi.org/10.1111/cogs.12086 Renkl, A., & Atkinson, R. K. (2003). 在认知技能习得中构建从例题学习到问题解决的过渡：认知负荷视角。《教育心理学家》，38 (1), 15–22. https : / / doi . org / 10 . 1207 / S15326985EP3801 3 Sallam, M., Elsayed, W., Al-Shorbagy, M., Barakat, M., El Khatib, S., Ghach, W., Alwan, N., Hallit, S., & Malaeb, D. (2024). ChatGPT的使用与态度受有用性、易用性、风险及心理社会影响的认知驱动：一项针对阿联酋大学生的研究。《教育前沿》，9, 1414758. https://doi.org/10.3389/feduc.2024.1414758

Scheiger, P., Cartarius, H., & Nawrodt, R. (2023). 为何以及如何在高年级理论物理中应用例题。American Journal of Physics, 91 (10), 764–766. https://doi.org/10.1119/5.0105612 Schmidt-Weigand, F., H¨anze, M., & Wodzinski, R. (2009). 复杂问题解决与例题：提示策略性行为和逐步淡化解题步骤的作用。Zeitschrift f ¨ur P ¨adagogische Psychologie, 23 (2), 129–138. https://doi.org/10.1024/1010-0652.23.2.129 Shellman, S. M., & Turan, K. (2006). 模拟能否促进学生学习？对一项国际关系模拟的实证评估。《政治科学教育杂志》, 2 (1), 19–32. https://doi.org/10.1080/15512160500484168 Singh, C. (2006). 学生对对称性和高斯电定律的理解。American Journal of Physics, 74 (10), 923–936. https://doi.org/10.1119/1. 2238883 Sweller, J., & Cooper, G. A. (1985). 在代数学习中以例题替代问题解决。《认知与教学》, 2(1), 59–89. https://doi.org/10.1207/s1532690xci0201 3 Sweller, J., van Merri¨enboer, J. J. G., & Paas, F. G. W. C. (1998). 认知架构与教学设计。《教育心理学评论》, 10 (3), 251– 296. https://doi.org/10.1023/A:1022193728205 University of G¨ottingen. (n.d.). ChatAI. 检索于2026年11月21日，网址：https: // www.uni-goettingen.de/de/686446.html Unver, V., Basak, T., Watts, P., Gaioso, V., Moss, J., Tastan, S., Iyigun, E., & Tosun, N. (2017). 三份问卷的信度与效度：学生学习满意度与自信量表、模拟设计量表和教育实践问卷。《当代护士》, 53 (1), 60–74. https://doi.org/10.1080/10376178.2017.1282319 Wittwer, J., & Renkl, A. (2010). 基于例题的学习中教学性解释有多有效？一项荟萃分析综述。《教育心理学评论》, 22 (4), 393–409. https://doi.org/10.1007/s10648-010-9136-5 Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). 我们到达目的地了吗？关于教育领域聊天机器人的系统性文献综述。《人工智能前沿》, 4, 654924. https://doi.org/10. 3389/frai.2021.654924 Wu, R., & Yu, Z. (2024). 人工智能聊天机器人能否改善学生学习成果？一项荟萃分析的证据。《英国教育技术杂志》, 55(1), 10–33. https://doi.org/10.1111/bjet.13334