用于在提问情境下评估学习助理学科教学知识的开放式作答题的开发

Beth Thacker[1], Stephanie Hart[2], Kyle Wipfli[1], Jianlan Wang[3]德克萨斯理工大学物理与天文学系，*MS 41051*，拉伯克，得克萨斯州[2]西德克萨斯农工大学特里·*B*·罗杰斯教育与社会科学学院，卡尼翁，得克萨斯州[3]德克萨斯理工大学教育学院，拉伯克，得克萨斯州

摘要： 我们设计并开发了一组开放式作答题，用于评估学习助理在提问情境中的教学内容知识（PCK-Q）。我们的目标是开发一组书面开放式作答题，以代表课堂中观察到的学习助理的 PCK-Q。本文讨论了设计、开发和验证过程。这些过程包括课堂互动的视频录制与分析、编码评分标准的开发，以及书面开放式作答题的开发、实施和验证。

I. 引言

随着大学物理入门课程日益重视学生参与，越来越多地依赖学习助理（LA）来降低学生与授课教师的比例，并提高大班课程中与学生的主动互动。这一趋势同时出现在传统课程形式和循证课程形式中，例如 SCALE-UP（面向大规模本科课程的以学生为中心的活动） [1]，ISLE（探究式科学学习环境） [2]，开源教程，例如《大学物理入门教程》 [3],INQ（《基于探究的实验物理学》） [4] 以及建模[5]。

本文聚焦于物理课堂中的循证教学形式。这些形式都以学生为中心，采用小组学习环境，并且通常基于实验。作为一种教学策略，授课教师会运用苏格拉底式提问 [6] 和引导式探究 [7] 提出问题，引导学生经历比传统讲授更有效的学习过程 [8, 9]。

大量研究表明，授课教师需要运用教学内容知识（PCK） [10, 11] ，才能有效实施这些策略。研究人员考察了PCK的若干要素，包括学生理解 [12]， 、数学推理方面的PCK要素 [13]， 以及教学内容知识的影响 [14-17]。

在循证教学形式中，授课者不只有教师，还常包括博士后、研究生；随着课程规模扩大，也会包括本科生学习助理。因此，学习助理和其他授课者一样，需要具备足够的教学内容知识，才能有效引导学生学习。一个亟待回答的重要问题是：“我们有哪些证据表明学习助理具备教学内容知识（PCK）？”

为了回答这一问题，需要研究学习助教的学科教学知识。一种方法是在课堂上对学习助教进行视频录制。负责该课程的授课教师可以与学习助教一起查看录像，讨论并评估他们的教学实践。然而，视频录制以及查看录像是一个耗时且费力的过程。因此，除了通过课堂访问或视频录制进行直接观察之外，还需要一种高效的学科教学知识评估流程。

由于基于证据的策略通常采用苏格拉底式提问作为教学法，我们决定专注于开发一种工具，用于在提问式教学法（PCK-Q）的情境下评估学习助教的学科教学知识。为实现这一目标，我们：1）录制学习助教在课堂上与学生互动的视频；2）根据视频开发用于评估学习助教PCK-Q的编码评分量规；3）开发并验证一组可作为学习助教PCK-Q指标的问题。下文将介绍开发过程、我们的编码评分量规、示例问题以及验证过程。我们还将讨论如何利用这组问题评估学习助教的PCK-Q，并帮助引导学习助教开展更有效的教学。

II. 理论框架

我们采用了 Magnusson 关于科学教学学科教学知识（PCK）的理论框架。Magnusson 将 PCK 描述为“对多种教学知识（包括学科知识）进行转化的结果，因此它代表了教师知识的一个独特领域” [10]。在其研究中，PCK 包括以下五个组成部分：（a）科学教学取向；（b）科学课程知识；（c）学生对科学的理解知识；（d）科学教学策略知识；以及（e）科学素养评估知识。每个组成部分都描述了科学授课教师 PCK 所需的不同知识体系。这些组成部分基于 Grossman [18] 和 Tamir [19] 的研究，对他们最初提出的 PCK 定义进行了拓展。

Magnusson 所提出的“科学教学取向”（取向）这一概念，指的是授课教师对于特定阶段教学目的和目标的知识。取向指塑造课程中不同类型相互作用和评估的总体教学实践。取向的示例包括让学生通过引导式探究进行学习，让学生阅读不同论文以发现某一概念的相关思想，或采用讲授方式。取向是 PCK 其他组成部分的基础，并影响不同教师知识的形成方式。

“科学课程体系知识”（课程知识）主要包括两个部分：（a）目标与目的；以及（b）课程中使用的特定课程项目与材料。课程的目标与目的不仅包括预期学生应学习的学科知识，还包括从以往学习经历中获得的知识

以及学习学科知识不同方面背后的推理。对特定课程项目与材料的知识，重点在于了解不同的课程设计及相关材料，例如由华盛顿大学开发的 Physics Tutorials[3] 和 Modeling Instruction [5]。科学课程体系知识的重要性有助于区分同时擅长教学和学科内容的授课教师与仅精通学科内容的专家。

“学生对科学理解的知识”（学生知识）指授课教师对学生学科知识的了解，以及如何提升学生理解的认识。学生过去的学习经历会影响其当前知识。了解学生的知识和先入观念会影响授课教师选择与学生互动的方式。了解学生常见的困难，有助于教师为学生提供学习支持。学科教学知识的这一部分关注的是对学生观念的了解，而非课程知识本身。

“教学策略知识”（教学）指了解用于科学教学的学科特定策略和主题特定策略。学科特定策略是指特定实验、对话或任务的组织方式，以帮助学生获取知识。主题特定策略是指对内容或原理的表征方式，以帮助学生获取知识。

科学中的“评估知识”（评估）包括两个相互关联的部分：用于评估学生的方法，以及评估课程不同维度的相对重要性。评估方法既包括对既有评估的了解和实施，例如 BEMA [20] 和 FCI [21], ，也包括课程内部使用的评估。实施是指评估的形式或时间安排。了解不同维度的重要性，是指知道哪些课程目标值得评估。例如，在面向 STEM专业学生的基于微积分的物理课程中，掌握使用不同数学方法求解动能可能更为重要，而掌握如何使用能量守恒解决不同的简单情境，对于面向非STEM专业学生的基于代数的物理课程可能更为重要。

在本研究项目中，我们对 Magnusson 关于科学教学的教学内容知识（学科教学知识）及其相关组成部分的定义进行了阐释，以适应我们的研究情境。鉴于循证课程形式在物理学中的普及，我们选择将研究重点放在采用引导式探究而非讲授式教学的课程上。本研究具体考察提问情境中的学科教学知识，因此将 Magnusson 的取向（O）概念缩小为考察学习助教在与学生讨论时采用引导式探究还是直接教学。我们将 Magnusson 的科学课程知识（C）解释为学习助教展示如何实施符合课堂材料目的和设计的材料。为了有效使用循证课程，授课教师需要了解课程的设计方式及其预期用途。我们将学生知识（S）定义为学习助教在参与讨论时识别学生优势和困难的能力。我们将 Magnusson 的教学（I）概念缩小为学习助教在与学生互动时运用经证明有效的引导技巧。由于学习助教通常不编制或评分评估，我们未将 Magnusson 的评估（A）组成部分纳入本项目。我们将对 Magnusson 研究的改编称为 OCSI（取向／课程／学生／教学）框架。

III. 设置与参与者

我们在两所大学收集了数据：德克萨斯理工大学（TTU）的代数基础课程，以及佛罗里达国际大学（FIU）的微积分基础课程。TTU的数据收集时间为2019年秋季和2020年春季。由于新冠疫情以及线下课程取消，TTU的数据收集于2020年3月停止。2020年秋季和2021年春季，我们在FIU收集了数据。这些课程通过Zoom在线授课。

组成部分	Magnusson 的定义	应用于我们语境中的 Magnusson 定义
取向	对总体教学 实践的认识，这些实践塑造了不同类型的 课程中的相互作用与评估。	学习助教在与学生讨论时采用引导式探究还是直接 教学。
课程	对目标与具体目标的了解，以及 所使用的特定课程项目与材料 在课程中。	学习助教展示了如何实施 与目的相符的课堂材料 以及这些材料的设计。
学生	了解学生的学科知识以及 如何提高学生的理解力。	学习助教识别学生优势和 参与讨论时所遇困难的能力。
教学	知识 of 特定学科的 and 用于教授科学的主题特定策略。	学习助教所使用的、已被证明在与 学生互动时有效的引导技巧 。
评估	了解用于评估的方法 学生以及……的相对重要性 评估……不同维度 课程	不适用

图1. Magnusson 的学科教学知识组成部分与我们在分析学习助教时对学科教学知识的使用方式之比较。

A. 代数基础课程

TTU的代数基础物理学入门课程序列包含两个学期，第一学期主要讲授力学，第二学期主要讲授电学与磁学（E&M）。学生主要是健康科学专业学生，包括医学预科、牙医学预科、物理治疗预科等。第一学期通常约有400名学生注册，第二学期约有250名学生注册。每学期设有两个INQ班级，每班60名学生。其余学生则分布在讲授式班级中，每班60至200人，具体人数取决于授课教师的安排和课堂资源情况。本研究仅关注INQ班级。

INQ课程单元于20年前在美国国家科学基金会（NSF）的资助下开发，此后每学期一直作为代数基础课程的一个或多个单元开设。该课程专门为健康科学专业学生开发，充分考虑了他们的需求、学习风格、背景和学习动机。课程在工作室式教学环境中采用苏格拉底式提问教学法授课，不要求使用教科书。学生以小组形式学习各单元，通过观察和推理学习构建定量模型和定性模型，并运用这些模型进行预测和解决问题。

课程的目标是像科学家一样从实验室开始学习，并通过引导式实验掌握课程内容。材料包括实验单元、前测、阅读材料和练习。此外，还有作业集、考试和测验。该课程涵盖的内容与班级其他单元大致相同，但更加注重根据实验构建模型，以及培养观察、分析和思考能力。学生可以选择参加INQ单元或传统单元，但通常取决于课程安排是否适合他们，以及根据毕业要求他们何时需要修读这门课程。学习助教由之前以INQ形式修读过该课程的学生中选拔。每门课程配备2至3名学习助教、1至2名研究生（助教）和一名授课教师。

B. 微积分基础课程

与 TTU 类似，FIU 的两门物理课程序列包括第一门力学课程和第二门电磁学课程。FIU 开设的课程序列是基于微积分的建模[5] 课程，学生分组合作构建物理学知识。在我们收集数据的两个学期中，

授课教师和学习助教通过 Zoom 在线与学生会面，学生在分组讨论室中开展学习。动手物理实验被虚拟实验室所取代。学习助教在不同的分组讨论室之间巡回，以促进学生进行探索。实验接近结束时，所有学生都回到 Zoom 的主房间，反思他们的探索并总结结果。

C. 学习助教

i. TTU

在 TTU，参与者是基础物理课程序列中的四名学习助教和一名研究生助教（TA）。LA-1 是一名西班牙裔本科男生，修读过 INQ 课程序列，且此前没有教学经验。LA-2 是一名白人本科女生，也修读过 INQ 课程序列，此前没有教学经验。LA-3 是一名西班牙裔本科女生，主修物理学，修读的是传统的基于讲授的课程序列，而非 INQ 课程序列。LA-4 是一名白人本科男生，修读过 INQ 课程序列，此前没有教学经验。TA 是一名中东裔研究生女生，有 1 年教授 INQ 课程的经验。INQ 课程的教学团队每周都会开展准备课程，内容涵盖学科内容和教学法。

ii. FIU

在 FIU，参与者是基础物理课程序列中的三名学习助教。LA-1 是一名西班牙裔本科女生，此前没有担任学习助教的经验。LA-2 是一名黑人本科女生，有 1.5 年在大学物理入门课程中担任学习助教的经验。LA-3 是一名西班牙裔本科男生，有 1.5 年在大学物理入门课程中担任学习助教的经验。在本项目开始前，三名学习助教都修读过一门教学方法课程，学习探究式教学的教学法，例如支架式教学和提问。在这门课程期间，学习助教每周都会在线与授课教师参加准备课程，预习下周要讲授的物理学科内容知识。

IV. 方法

本节将介绍数据收集和数据分析的多个阶段。为此，我们采用了以下框架的修改版本：Adams & Weiman [22]。首先，我们记录了入门物理课堂中的学生与学习助理的互动。其次，我们进行了视频分析，并制定了用于描述学习助教与学生之间互动的评分量规。第三，我们验证了评定者间信度。第四，我们根据视频中的真实互动设计了开放式作答题。最后，我们使用相同的评分量规对开放式作答题和视频进行编码，从而验证了开放式作答题。

A. 课堂视频录制

在 TTU，我们对 INQ课程序列进行了两个学期的课堂视频录制。我们使用视频采集系统（Swivl） [23] 在课堂期间录制学生与学习助理的互动。学习助教和授课教师都佩戴了麦克风，Swivl 设备会转动跟随学习助教在教室内移动，并录制学生与学习助理的互动视频。

在 FIU，我们通过 Zoom 对建模课程序列进行了视频录制。学习助教录制了各自分组讨论室的课程，并将录制内容提交给我们。

通过录制视频，我们能够分析正常课堂期间产生的各种类型的相互作用，并根据学习助教提出的问题类型，对学习助教的不同能力进行分类。

B. 视频分析与评分量规开发

我们每周分析每位学习助教的多段相互作用视频。我们逐句检查视频记录，并识别学习助教与学生互动时所作陈述的不同类型，将其归入不同类别。我们将此称为微观编码系统。在将微观编码系统应用于多段视频记录时，我们发现了学习助教陈述中的模式，并将其归入与学科教学知识各组成部分相对应的不同层级。我们将此称为宏观编码系统。

i. 微观编码系统

我们观察到，学习助教通常采用辅导型互动或提问型互动。辅导型互动包括学习助教向学生讲解概念，或通过提供信息，逐步引导学生完成问题解决。提问型互动则主要通过提问而非陈述，引导学生正确理解概念或解决问题。我们识别出

三种提问型互动，并将其标记为引导性问题、探究性问题和核查性问题。

引导性问题是通过帮助学生在评估实验性证据、解决问题或形成正确的概念理解时进行推理并得出结论，从而引导学生思考的问题。我们尤其关注一系列引导性问题，因为它通常展示了PCK-Q的全部四个要素：学科知识、课程知识、学生理解知识，以及应对学生困难的教学策略知识。

探究性问题是用于收集有关学生的理解的信息的问题。

我们观察到，相互作用通常始于探究性问题。例如："你为什么从那个高度把它放下？"或"你能告诉我为什么把磁铁放在这里吗？"例如，这些问题用于进一步收集有关实验装置或可能存在的学生困难的信息。

核查性问题（也称为温度检查）是关于实验装置或学生理解的简单问题，通常只需回答"是"或"否"。示例包括"你是否将传感器归零了？"或"你能理解这个吗？"

微观编码方案如图2所示。

类别	**Code**	**描述**	**示例**
引导性问题	gq	学习助教提出问题，引导学生思考，帮助他们在评估实验性证据的同时，推理得出结论评估实验性证据，解决问题或得出正确的概念理解。	如果你用恒定的力，比如说1单位的橡皮筋拉力，会发生什么？如果你在无摩擦轨道上用一单位的橡皮筋拉力拉小车，会发生什么？你能让相同的情况发生吗？使用蹦极绳产生的运动？如何产生？如何比较这两者？
探究性问题	pq	学习助教提出问题，以收集有关学生的理解的信息。	1) 你为什么从那个高度把它放下？ 2) 你能告诉我你为什么把磁铁放在这里吗？
检查问题	cq	学习助教提出一般性问题以收集信息（温度检查）。	1) 你将传感器归零了吗？ 2) 你能理解这是什么意思吗？
辅导/讲授	le	学习助教直接向以下对象传授信息学生。	引力不会相同，因为质量不同。

图2. 微观编码方案，列出了每种编码的描述和示例。上表包含编码类别、编码标签、编码本身的描述，以及每个类别中的一个示例陈述或问题。

ii. 宏观编码

在最初的（微观）编码系统中，我们对学习助教提出的单个问题/陈述进行编码。随着时间推移，我们观察到学习助教经常会使用一系列通过提问和/或陈述，引导学生正确理解概念或帮助他们解决问题。我们将"情境片段"一词定义为一段时间，即从

相互作用开始，到一名学习助理结束与该组学生的相互作用为止。

在每个情境案例中，我们通过微观编码识别出四种主要模式，并可将其组织成不同层级。我们观察到，有些情境案例中的学习助教会反复使用一系列探究性问题和引导性问题。学习助教通过一连串引导性问题，引导学生经历某一思维过程或完成一项实验，其中穿插着探究性问题和核查性问题。我们还观察到，有些情境案例中的学习助教主要使用探究性问题和核查性问题，其间穿插着讲授环节。看起来学习助教的意图是通过提问引导学生，但在某些时刻，学习助教转而进行辅导或讲授。通常只有一两个引导性问题，但同时也包含相当多的讲授或辅导。由于这两组都采用了提问方式，我们将其归类为提问取向。第一组展现出最高水平的提问技能，因此我们将该模式标记为 Qa。第二组较少使用提问，因此我们将该模式标记为 Qb。

层级	描述	编码模式
Qa	学习助教通过思维过程或 实验引导学生，并使用一系列引导性问题 穿插探究性问题和核查性问题。这是 PCK-Q 最直接的证据。	pq-gq-(pq/tr/cq)-gq-(pq/tr/cq) 很少 le
Qb	学习助理似乎意在引导学生 通过提问进行引导，但在某些时候，学习助理会转而采用 辅导或讲授的方式。通常会提出一两个 引导性问题，但同时也会进行相当多的 讲授/辅导。这为以下内容提供了一些证据： PCK-Q，但不如 Qa 多。	(pq/cq)-gq-le- 是一般形式。这 可能会重复。
Qe	Qa 或 Qb，其中存在可识别的内容错误。	er
Da	辅导——带有一定探究和检查的讲授 问题。这些学习助教可能具备学科教学知识，但不具备 PCK-Q。	一系列 pq-le-cq
Db	基本上是在讲授。学习助理没有考虑学生的 理解情况。几乎没有证据表明其具备学科教学知识 或 PCK-Q，C（学科知识）除外。	cq-le
De	Da 或 Db，并且存在可识别的内容错误。	er

图 3. 带有描述和编码模式的宏观编码方案。上表列出了每个层级的标签、该层级本身的描述，以及一个表明应按该层级对相互作用进行编码的模式示例。

与提问组不同，我们还观察到一些情境片段，其中学习助教主要通过讲授或辅导来指导学生，同时提出一些探究性问题和核查性问题。我们将这种模式标记为 Da。有时，我们观察到一些情境片段，其中学习助教只通过讲授或辅导，并提出一两个核查性问题。学习助理似乎意在传授信息，而没有考虑学生是否理解。我们将这种模式标记为 Db。

如果某个情境片段原本会被标记为 Qa 或 Qb，但其中包含可识别的物理学错误，则将其编码为 Qe。同样，如果某个情境片段原本会被

标记为 Da 或 Db，但其中包含可识别的物理学错误，则将其编码为 De。

每段录像都包含多个情境片段。每个情境片段都被编码为 Qa、Qb、Da、Db、Qe 或 De。在同一段录像中，学习助理展现了一个或多个层级。宏观编码方案如图 3 所示。

iii. 重新审视 ***OCSI***（取向/课程/学生/教学）框架

我们认识到，Qa、Qb、Da 和 Db 这些层级与 Magnussen 的学科教学知识组成部分在本研究情境中的应用相关。为阐明这种关系，我们还使用 OCSI 框架对录像进行了编码。我们将 OCSI 框架与宏观编码方案的各个层级相对应，如图 4 所示。

标记为 Qa 的层级代表这样一种情形：学习助理通过一系列引导性问题，并穿插探究性问题和核查性问题，引导学生进行思维过程或实验。这是四个学科教学知识组成部分——取向、课程知识、学生知识和教学策略知识（OCSI）——最直接的证据。它至少包含一个引导性问题，但通常是一系列引导性问题。

标记为 Qb 的层级代表这样一种情况：学习助理似乎有意通过提问引导学生，但在某些时刻转而进行辅导或讲授。通常会出现一两个引导性问题，但同时也包含大量的讲授或辅导。这表明具备 PCK 的三个组成部分：取向、内容知识和学生知识（OCS）。Qb 体现了提问取向，但没有显示出有效教学策略的证据，例如一系列引导性问题（I）。

标记为 Da 的层级代表这样一种情况：学习助理主要进行辅导或讲授。学习助理可能偶尔使用探究性问题和/或核查性问题，但谈话主要由陈述构成。Da 所体现的 PCK 组成部分是内容知识和学生知识（CS）。与 Qa、Qb 和 Qe 层级不同，Da、Db 和 De 层级没有显示出提问取向的证据（O 和 I）。

标记为Db层级的情况表示，LA在辅导或讲授时没有考虑学生的理解。唯一 Db 所体现的 PCK 组成部分是内容知识（C）。

	O	C	S	I	
Qa	X	X	X	X	x = 更有可能相关
Qb	X	X	X	-	- = 不太可能相关
Da	-	X	X	-	
Db	-	X	-	-	

图 4. 宏观编码层级与 OCSI 的组成部分（取向/内容/学生/教学）之间的相关性。

C. 视频分析的评定者间信度

我们从TTU 2020年春季一周的录音中选取数据，对Qa、Qb、Da和Db层级的编码进行了评分者间信度检验。为确定评分者间信度，三名评定者使用宏观编码系统，分别对两名学习助教的情境案例进行了编码。任意两名评定者之间的未加权科恩卡帕系数始终高于0.75，加权科恩卡帕系数始终高于0.86。这些数值表明一致性达到较高水平至近乎完美的一致性。

D. 自由回答问题开发

这些问题基于课堂互动开发。每个问题题干都包含背景情境、对学生与学习助理互动的描述和/或对话，以及供答题者作答的开放式问题。为撰写题干，我们从视频记录或课堂观察中选取能够体现学生常见困难的学生与学习助理互动。我们在问题题干中描述这些互动，并提供足够的细节来说明学生的困难，有时还会使用直接引语。这些问题旨在反映课堂中真实的学生互动，因此包含学生作业、学生对话、错误和错误观念。图5和图6展示了问题示例。其他样题见附录A。

我们采用迭代过程来完善开放式问题，使学习助教的回答与其在课堂中的视频互动

保持一致。起初，学习助教的回答过于模糊，未能反映视频中观察到的学科教学知识水平。因此，我们通过明确询问学科教学知识的具体组成部分来完善开放式问题，从而更好地 eliciting 视频中所展示的学科教学知识组成部分。

最终结果是一个两部分的问题提示，如图4和图5中的示例所示。第（a）部分要求学习助教根据问题题干中的情境推断学生的优势和困难。该部分专门用于测量学生知识，即OCSI中的S部分。第（b）部分要求受访者描述他们将如何回应学生以及原因。该部分评估PCK-Q的其余组成部分：取向（O）、内容知识（C）和教学策略知识（I）。

背景：学生已经学习了牛顿定律。现在，他们正在学习碰撞，并尝试推导一个将冲量与动量变化联系起来的方程。学生在无摩擦轨道上放置两辆小车，并使它们相互碰撞。每次实验中，两辆小车的质量都不同。他们在小车上安装了探针传感器，以测量碰撞时作用在每辆小车上的力。

学生的实验结果如下表所示。

小车 1 的质量	小车 2 的质量	作用在小车 1 上的力	作用在小车 2 上的力
1.00 千克	1.00 千克	7.41 牛顿	7.41 牛顿
1.25 千克	1.00 千克	11.38N	12.38 N
1.50 kg	1.00 kg	15.10 N	16.10 N

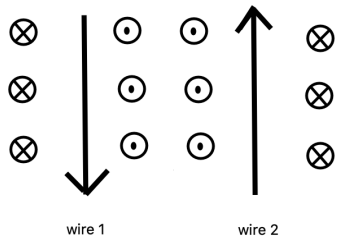
学生得出结论：小车的质量越大，它在相互作用中施加的力就越大。

a. 根据所提供的信息，你能对学生的学科知识得出什么结论？学生有哪些优势和困难？b. 你会如何回应学生？请使用你会说的话的直接引语。你这样回应的理由是什么？

图 5. 经典力学问题示例。

内容 **上下文：学生正在学习磁学。他们正在学习磁场和力** .

学生看到两根电流方向相反的导线，并被要求指出纸面上导线 1 在导线 2 处产生的磁场方向，然后据此确定作用在导线 2 上的磁力方向。他们画出了以下图示：



他们感到困惑，因为他们无法从图中判断导线 1 在导线 2 处的磁场方向，因此也无法确定作用力是向左还是向右。学生遇到了什么困难？学习助理应如何应对？

a. 根据所提供的信息，你能对学生的学科知识得出什么结论？学生有哪些优势和困难？ b. 你会如何回应学生？请用直接引语写出你会说的话。你这样回应的理由是什么？

图 6. 电学与磁学题示例。

E. 自由作答题分析

在 TTU，每学期大约进行 4—5 次开放式作答题测验，每次包含 3—5 道题。在 FIU，每周进行 1—2 道题的测验。开放式作答题的实施时间大致与课堂上讲授相关概念的时间一致。我们依据宏观编码量规中的模式对作答进行编码。随后，为每位学习助教的作答分配一个宏观编码量规水平。

F. 开放式作答题分析的评分者间信度

我们针对开放式作答题的 Qa、Qb、Da 和 Db 层级编码建立了评分者间信度。为此，我们使用了从 TTU 收集的 2020 年春季 数据。为确定评分者间信度，三名评定者使用宏观编码系统，分别独立编码了 15 名学习助教的书面开放式作答答案。随后，我们计算了成对评定者之间的科恩卡帕系数。任意两名评定者之间比较所得的未加权科恩卡帕系数始终高于 0.69，而加权科恩卡帕系数始终高于 0.78。这些数值表明一致性达到较高水平至近乎完美的一致性。

IV. 验证

我们通过比较每位学习助理的视频编码结果与开放式作答编码结果，验证了开放式作答题的效标效度。对于视频，我们每周随机选择一天的情境案例，并记录当天每个学科教学知识水平出现的百分比。然后，我们对所有周次中各水平出现的百分比取平均值。对于所实施的所有开放式作答题，我们确定每个水平出现的百分比。随后，我们比较视频与书面开放式作答结果。TTU 的结果见图7，FIU 的结果见图8和图9。

我们发现，视频和书面开放式作答题的回答总体上较为一致。例如，对于学生 AF，书面回答大多为 Qa 和 Qb，视频回答也大多为 Qa 和 Qb。对于学生 TZ，视频和书面开放式作答题的回答都大多为 Qb 和 Da。2020年秋季，学习助教只完成了 2-3 道书面开放式作答题，因此由于书面开放式作答题的数据有限，趋势并不那么显著。

TTU		Qa	Qb	Da	Db	Qe/De
AF	视频	29%	51%	17%	3%	0%
	书面	50%	25%	0%	0%	25%
TZ	视频	4%	29%	60%	4%	2%
	书面	0%	60%	20%	20%	0%
HB	视频	6%	10%	84%	0%	0%
	书面	0%	20%	40%	40%	0%

图7。TTU学生在多个情境案例中的PCK-Q水平平均出现频率。

FIU 春季		Qa	Qb	Da	Db	Qe/De
VF	视频	41%	3%	50%	3%	3%
	书面	63%	0%	25%	0%	12%
DD	视频	11%	19%	51%	11 %	8%
	书面	0%	0%	73%	0%	27%

图 8. 2021年春季，FIU学生在多个情境案例中PCK-Q水平的平均出现频率。

FIU Fall		Qa	Qb	Da	Db	Qe/De
SL	视频	3%	10%	57%	14%	14%
	书面	33%	67%	0%	0%	0%
DD	视频	0%	11%	78%	11%	0%
	书面	0%	0%	50%	0%	50%
AM	视频	17%	33%	17%	33%	0%
	书面	33%	67%	0%	0%	0%

图 9. 2020年秋季 FIU 学生针对多道开放式作答题的 PCK-Q 水平出现频率百分比。

V. 结论与未来展望

在本项目中，我们创建了一组可用于测量PCK-Q的问题，并通过效标效度验证表明，这些问题能够预测在课堂视频中观察到的学习助教的PCK-Q水平。这满足了对一种高效测量PCK-Q方法的需求。这些问题可用于向授课教师提供有关其学习助教课堂中PCK-Q水平的信息。

在本研究中，由于样本量较小以及所回答问题的数量有限，我们遇到了一些局限性。当前研究受到入门级实验类物理课程中学习助教样本量较小的限制。我们观察到的另一项局限是，对于回答少于三个问题的学习助教而言，我们预测其PCK-Q的能力受到限制。

我们设计的问题可用于评估或学习助理培训。问题可以一次性全部发放，也可以每周发放 1–2 道。授课教师可以用这些问题评估学习助教。或者，授课教师也可以在课堂准备环节中用这些问题培训学习助教。

此外，为了广泛推广，采用多项选择题或李克特量表式问题等形式，设计一组易于评分的问题可能会更有益。这些形式所需的实施时间更少，评分速度更快，也不需要受过培训的编码员。目前，我们正在开发李克特量表式问题。

VI. 致谢

本研究得到美国国家科学基金会本科科学教育改进项目资助（1838339）的支持。

参考文献

1. R. J. Beichner、J. M. Saul、D. S. Abbott、J. J. Morse、D. Deardorff、R. J. Allain、S. W. Bonham、M. H. Dancy 和 J. S. Risley，“面向大规模本科课程的以学生为中心的活动（SCALE-UP）项目”，《基于研究的大学物理教学改革》，1 (1)，2-39，(2007)。2. D. Brookes、E. Etkina 和 G. Planinsic，“实施以认识论真实取向开展以学生为中心的探究式学习”，《物理评论：物理教育研究》16 (2)，020148 (2020)。3. L. C. McDermott 和 P. S. Shaffer，《大学物理入门教程》（培生，波士顿，MA，2013）。4. Beth Thacker、Ann Marie Eligon 和 Abel Diaz，《基于探究的实验物理学》第一卷和第二卷（Stipes, LLC，2003）。5. J. Jackson、L. Dukerich 和 D. Hestenes，“建模教学：一种有效的科学教育模式”，《科学教育者》17(1)，10-17 (2008)。6. A. B. Arons，《大学物理入门教学》（威利，1997）。7. S. Shaffer 和 C. P. Constantinou，“通过探究为教师开展物理和物理科学教学做好准备”，《物理教育》35(6)，411–416 (2000)。8. J. Li 和 C. Singh，“探究并改善大学物理入门学生对电场和叠加原理的理解：以连续电荷分布为例”，《物理评论：物理教育研究》15，010116 (2019)。9. A. Werth、C. G. West 和 H. J. Lewandowski，“远程、大规模、基于课程的本科物理研究体验对学生学习、信心和情感的影响”，《物理评论：物理教育研究》18(1)，(2022)。10. L. Shulman，“那些理解的人：教学中的知识发展”，《教育研究者》15(2)，4-14，(1986)。11. S. Magnusson、J. Krajcik 和 H. Borko，载于《考察教学内容知识》，95-132，（施普林格，多德雷赫特，1999）。12. J. R. Thompson、W. M. Christensen 和 M. C. Wittmann，“在物理学、教学法与教育研究研究生课程中培养未来教师预判学生在物理学习中困难的能力”，Phys. Rev. ST Phys. Educ. Res. 7，010108 (2011)。13. J. H. Van Driel、N. Verloop 和 W. de Vos，“发展科学教师的教学内容知识”，《科学教学研究杂志》35 (6) (1998)。

14. H. C. Hill 和 S. G. Schilling，“剖析教学内容知识：教师关于学生的主题特定知识的概念化与测量”，JRME，**39**(4), 372–400, (2008)。15. E. Etkina，“教学内容知识与高中物理教师的培养”，Phys. Rev. STPhys. Educ. Res. **7**, 020110 (2010)。16. L. Seeley、E. Etkina 和 S. Vokos，“教师学科知识与科学实践知识在支持科学学习共同体中的交织作用”，科学、技术、工程和数学领域的教学内容知识，（Springer Cham，2018）。17. E. Etkina、D. Gitomer、C. Iaconangelo、G. Phelps、L. Seeley 和 S. Vokos，“用于探测教师教学内容知识的评估设计：以高中物理中的能量为例”，《物理评论：物理教育研究》**14**，(2018)。18. P. L. Grossman，《教师的形成：教师知识与教师教育，(教师学院出版社，纽约，1990）。19. P. Tamir，“教师教育中的学科内容及相关教学知识”，Teach Teach Educ，**4** (2), 99-110 (1988)。20. L. Ding、R. Chabay、B. Sherwood 和 R. Beichner，“电学与磁学评估工具的评估：简明电学与磁学评估”，Phys. Rev. ST Phys. Educ. Res. 2, 010105 (2006)。21. I. Halloun 和 D. Hestenes，“大学物理学生的初始知识状态”，Am. J. Phys. **53**, 1043 1985( )。22. W. Adams 和 C. Wieman，“用于测量类专家思维的学习的工具之开发与验证”，Int. J. Sci. Educ. **33**，1289-1312，(2011)。23. Swivl机器人，https://www.swivl.com/robot/，12/16/2022。