

翻转课堂下不同平均绩点学生的自主学习路径分析：K-means 聚类与过程挖掘

Self-regulated Learning path analysis of students with different Grade Point Average in Flipped classroom: K-means clustering and Process Mining

王婷¹, 龙陶陶^{2*}, 朱晓萌³, 伍冬⁴

¹² 第华中师范大学人工智能教育学部

³⁴ 武汉理工大学材料科学与工程学院

* taotaolong@ccnu.edu.cn

【摘要】 自我调节学习 (SRL) 是学生的一项重要能力, 特别是在像翻转课堂这样的混合学习环境中。本研究采用基于问题的学习 (PBL) 方法, 探讨了翻转课堂中工科学生的自我调节学习行为。本研究以工科学生自主学习行为为视角, 采用 K 均值聚类分析基于学生的平均成绩绩点进行分组, 将学生分为五类。然后, 利用过程挖掘技术对各类学生的学习行为特征进行可视化分析, 旨在为探索和优化工科翻转课堂教学提供基于客观数据的理论支持和实践参考。研究发现, 不同平均成绩绩点段学生的学习行为路径存在较大差异。基于此, 本研究提出了一些可能提高工科学生在翻转课堂中自我调节学习效果的建议。

【关键词】 自我调节学习 (SRL); 工程翻转课堂; 过程挖掘; K 均值聚类分析

Abstract: Self-regulated learning (SRL) is a critical competency for students, particularly in blended learning environments like flipped classrooms. This study explores the SRL behaviors of engineering students in a flipped classroom employing a problem-based learning (PBL) approach. This study takes engineering students' self-regulated learning behavior as a perspective, uses K-means cluster analysis to group students' Grade Point Average, and divides students into five categories. Then, the process mining technology is used to analyze the learning behavior characteristics of all kinds of students, aiming to provide theoretical support and practical reference based on objective data for exploring and optimizing flipped classroom teaching in engineering. It is found that there are great differences in the learning behavior paths of students with different GPA segments. Based on this, this study puts forward some suggestions to improve the self-regulated learning effect of engineering students in flipped classroom.

Keywords: Self-regulated learning (SRL), flipped engineering classroom, process mining, K-Means clustering

1. 引言

自我调节学习 (SRL) 被认为是 21 世纪成功的基石技能, 它使学习者能够掌握自己的认知、动机和行为过程, 以实现教育目标 (Manukaram, K., & Abdullah, M. N. L. Y., 2021; Du, J., Huang, W., & Hew, K. F., 2021)。虽然 SRL 已被证明可以显著提高学生的参与度和学习成绩, 特别是在在线学习环境中 (杨燕清、吴明、林艳芹和洪霞, 2022; 李文昊、任晓瞳、朱希雅和陈冬敏, 2022), 但研究仍在为理解其微妙的行为途径而努力。翻转课堂是一种广泛使用的

混合学习策略，它要求学生参与课前材料，如视频和测验（王翠如和胡永斌，2018），从而要求积极的计划和自我调节。尽管有这种潜力，但之前关于翻转课堂中 SRL 的研究往往依赖于主观的自我报告，限制了研究结果的可靠性和深度。

聚类分析是一种重要的数据挖掘方法，其核心是通过分析数据内部的相似性来发现“自然分组”（Blashfield, R. K., & Aldenderfer, M. S., 1978），即聚类。这种无监督学习方法可以帮助研究人员深入挖掘数据的内在规律，特别是在需要处理大量复杂多维数据的场景中，表现出较强的特征分类能力。因此，K-Means 聚类分析在教育领域经常被用来对等级进行分类（Wu, J., & Wu, J., 2012）。

过程挖掘是一种数据驱动的分析事件日志的方法，它已经成为一种强大的工具，可以客观地探索在线环境中学生的学习行为和模式（Van Der Aalst, W., 2011）。虽然 SRL 的应用已经扩展到分析 mooc 的参与度（邓国民、徐新斐和朱永海, 2021），但在翻转课堂环境下的探索较为有限，特别是在工程学科中，这种方法也许能产生关键的见解。

针对上述问题，文章采用过程挖掘技术，对翻转课堂环境下不同平均成绩绩点的工科学生的自主学习路径进行了深入探究。最后根据研究结果，有针对性地提出了一些改进教学的建议，以期工程教育提供有价值的见解。

2. 文献综述

2.1. 自我调节学习 (SRL)

SRL 起源于 Bandura 提出的社会认知理论，随后 Zimmerman 等人在此理论的基础上对 SRL 进行了具体的解释。Zimmerman (1989) 认为 SRL 是指学习者能够从元认知、动机和行为等方面积极参与学习过程，主动获取知识和技能，并强调自我调节策略、自我效能感和学业目标是 SRL 的三个要素（Zimmerman, B. J., 1989）。SRL 经过几十年的发展，其模型也逐渐丰富起来。Zimmerman (2009) 认为 SRL 本应分为三个过程，即计划阶段、执行阶段、自我反思阶段。Pintrich (2004) 将 SRL 分为四个阶段：预见、计划和激活；监测；控制；反应和反思（Pintrich, P. R. 2004）。

研究强调了 SRL 在翻转课堂的课前活动中的重要性，如视频讲座和测验，这些活动对课堂学习效果有显著影响。例如，最近的研究强调设计混合和翻转学习策略，以提高工程本科生的参与度和技能习得（Obada, D. O., Bako, R. B., Ahmed, A. S., & Obada, I. B., 2023）。然而，在深入探索翻转课堂环境下的 SRL 行为方面，特别是在工程教育中，仍然存在显著的差距。虽然像 mooc 这样的在线教育环境已经得到了广泛的研究，但针对工程学生在翻转环境中所面临的具体 SRL 模式和挑战的研究仍然有限。SRL 在翻转课堂中的应用见证了后疫情时代的创新（刘云凤, 阎小青 & 乔增杰, 2021；Alsuraihi, A. A., 2022），融入了互动工具和丰富的学习活动，以提高参与度和成果。然而，这些研究主要使用学生自述的数据来调查 SRL 行为，很难做到客观。

2.2. 工程翻转课堂

作为近年来的热门议题之一，翻转课堂是一种学生在课前独自学习视频、文本等资源，课堂中再与教师、同学互动交流的教学模式（Bergmann, J., & Sams, A., 2012; Kim, M. K., Kim, S. M., Khera, O., & Getman, J., 2014）。研究表明，翻转课堂能够促进学生学业进步以及创新思维等高阶技能发展（王翠如和胡永斌, 2018; Rodríguez, G., Díez, J., Pérez, N., Baños, J. E., & Carrió, M., 2019）。将翻转课堂融入高等工程教育中已被证明是一种有效促进学生学习的方法（孟军、刘冰璇、翟洪江和兰洪, 2018），然而，Nielsen, K. L. (2023) 指出在使用翻转课堂的过程中，学生可能会感到挫败与沮丧。

为提高工程翻转课堂的实施效果, 研究者们尝试将项目式学习 (Chua, K. J., & Islam, M. R., 2021)、基于问题的学习 (Gómez-Coma, L., Díaz-Sainz, G., Fallanza, M., Ortiz, A., & Ortiz, I., 2023) 等策略融入到教学设计中, 且相关研究结果表明这些策略对于促进学生学习是有效的。与其他学习策略相比, 基于问题的学习导向的工程翻转课堂强调构建工程实践中的真实问题 (Chen, X., Long, T., Cheng, L., Gan, X., & Zhu, X., 2023), 在引导学生解决工程实践问题的过程中, 提高对专业知识的理解并促进工程能力的发展。

2.3. K 均值聚类

K 均值聚类是一种基本的无监督学习算法 (Blashfield, R. K., & Aldenderfer, M. S., 1978), 广泛用于根据相似度对数据点进行分组。该算法由 MacQueen 于 1967 年首次提出, 通过最小化每个数据点与其最近的聚类质心之间的欧几里德距离来迭代数据点到聚类的分配。麦昆不仅详细解释了算法的过程, 还从数学的角度论证了其原理 (Morissette, L., & Chartier, S., 2013)。K 均值聚类的本质是在给定的数据集 $\{x_1, x_2, \dots, x_n\}$ 中识别 K 个聚类质心 $\{a_1, a_2, \dots, a_K\}$, 使得每个数据点与其最近的质心之间的总平方距离最小。自引入以来, K 均值聚类以其简单性、计算效率和处理大型数据集的鲁棒性获得了认可。它基于固有模式有效地将数据划分为不同的集群的能力已经在自然语言处理、土壤科学和考古学等不同领域得到广泛采用, 这些优点使得该方法在分类和分组研究中得到了广泛的应用。

2.4. 过程挖掘

流程挖掘, 也被称为工作流挖掘, 是一种从系统内部事件日志中提取信息的方法, 用于发现、监控和改进实际操作过程 (徐晓青、赵蔚、姜强、刘红霞和乔丽方, 2022)。近年来, 流程挖掘被视为探索学生自我调节学习行为的新技术。例如, 乔丽方、赵蔚、段红和徐晓青 (2022) 使用过程挖掘技术探索不同类型的学习者的自我调节学习过程差异。过程挖掘技术能够直接分析学生真实的在线的学习日志, 从而将学生的自我调节学习较为全面的可视化的进行呈现。然而, 在现有的研究里, 研究者们更常运用滞后序列分析、分层聚类分析等方式分析学生的学习行为 (Chen, W., & Cheng, H. N., 2020; Han, F., & Ellis, R. A., 2023), 而不是应用过程挖掘技术发现学生的学习行为。

2.5. 本研究

尽管一些研究已经使用了过程挖掘技术探索在线学习情境下学生的自我调节学习行为, 但进一步使用 K 均值聚类和过程挖掘技术探索不同类型课堂里学生的自我调节学习行为仍有价值。本研究旨在深入探究翻转课堂环境下不同平均成绩绩点的工科学生的自主学习路径, 以期为解决当前工程翻转课堂中学生自我调节过程中的问题提供一些参考。

3. 方法

3.1. 背景

本研究在华中地区的某工程高校进行。课题组选择工程课程《计算机在材料工程中的应用》进行翻转课堂教学实践, 并根据基于问题的学习策略设计课程活动, 包括在线学习和课堂学习。所有学生必须登录在线学习平台——智能学习管理系统“小雅”。平台可进行教与学, 如发布讨论、测试、问卷、作业、课堂习题等课程任务, 智能完成学习分析。学生需要在课前完成相关的学习任务, 然后在课堂上与老师和同学进行交流和互动。

3.2. 参与者

本研究共选取 194 名材料工程专业的大三学生, 其中高分子专业 70 名, 金属材料专业 124 名。这些学生自入学以来, 没有参与过工程翻转课堂和系统的基于问题的学习活动, 也从未接受过任何与自我调节学习相关的训练。

3.3. 工具

本研究采用 Pintrich 提出的四阶段自我调节学习模型对学生的学习行为进行分类。该模型包括以下阶段：(a)预见、计划和激活，(b)监测，(c)控制，(d)反应和反思。此外，在 SRL 理论的框架下，学生的在线学习活动进一步被划分为 10 种不同的事件类型，详见表 1。

表 1 SRL 阶段与学习行为事件对应表

SRL 行为阶段	学习行为事件	编码	学习行为的具体解释
预见、计划和激活	总览大纲	V-S	浏览课程大纲或课程章节
	生成计划	G-P	编辑节点、更新资源生成学习计划
	分析任务	A-T	打开任务导航栏或任务导图
监测	观看资源	V-R	查看具体学习资源，如：视频、文档等
	查看任务	V-T	查看任务说明或任务提交界面
	移动资源	M-R	整理与移动小组资源
控制	下载资源	D-R	下载任务数据
	提交作业	S-W	提交4个pbl学习任务、小测和思考题
反应和反思	提交反思	S-R	提交个人反思报告
	查看论坛	V-F	查看论坛或在论坛发表观点

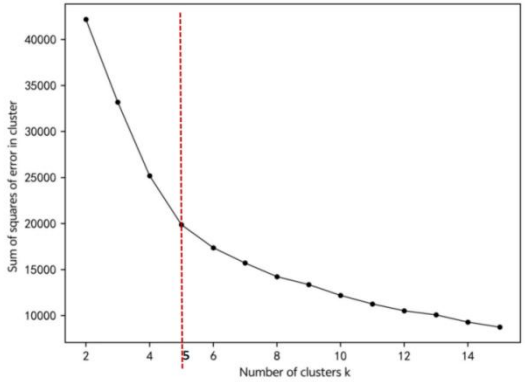
3.4. 数据收集与分析

参与者完成了一个学期的工程类翻转课堂学习。在这种学习模式下，参与者需要在课前通过“小雅”智能学习平台完成所需的知识。研究团队收集了参与者在学习平台上产生的所有事件的日志数据，共计 875,100 个事件。每个事件日志都包括参与者与学习平台的交互，以及其他一些必要的信息，如学生 ID、所采取的操作和时间戳。如表 2 所示，我们展示了某一时刻学生行为事件日志的数据抽取。

表 2 事件日志的示例

Student ID	SRL phase	Detailed activity
123	Monitoring	View Resources
123	Monitoring	View Resources
123	Monitoring	View Resources
123	Monitoring	View Resources
123	Control	Download Resources

在分析数据之前，对数据进行预处理，排除了一些与研究无关的行为事件数据，如学生登录和注销，得到有效事件 797995 个。技术人员使用 pycharm 平台（由 JetBrains 为 Python 语言开发的集成开发环境（IDE））来标记数据。该研究还收集了所有学生的学习成绩数据。采用 k-means 聚类分析，根据平均绩点对学生进行分类，根据生成的聚类肘图将学生分为 5 个类别（图 1）。在本研究中，根据 GPA 将这 5 个类别命名为 1~5。最后，利用 ProM 6.12 对五类学生的背景行为数据进行过程挖掘，并将学习过程可视化。本分析主要使用 iDHM 工具中的依赖图流模型，该模型基于 ProM 软件中的启发式挖掘算法。



4.结果

4.1. 五类学生在pbl 导向的工程翻转课堂中的自我调节学习行为

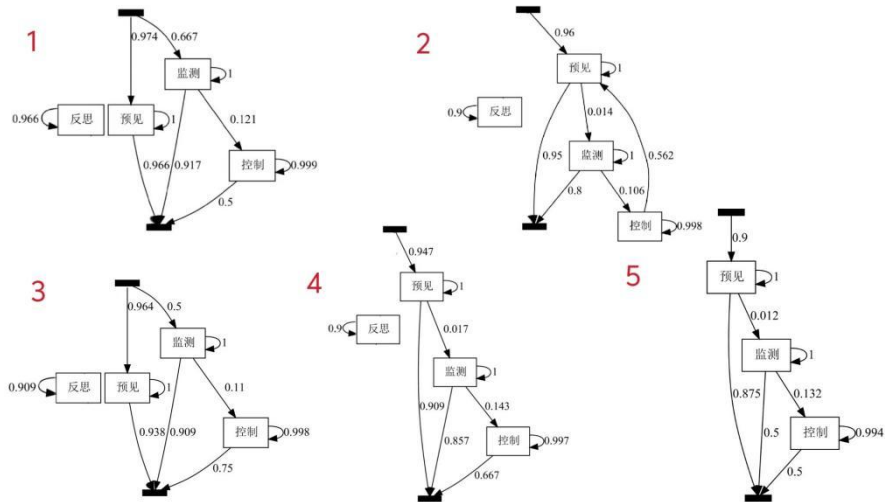


图2 五类学生在 SRL 阶段Dependency Graph 过程模型

如图 2 所示，Dependency Graph 过程模型被用于挖掘学生的自我调节学习行为过程，其结果显示了四个 SRL 阶段之间的依赖关系。图 2 中的矩形表示行为事件，其中，两个黑色矩形分别代表开始与结束。矩形之间的有向箭头表示事件之间发生的顺序关系，箭头上的值表示事件之间的依赖强度。五类学生在根据 GPA 聚类后的自我调节学习路径如图 2 所示，从 1 类到 5 类成绩依次递减。可见，前 4 类学生均存在反应和反思行为(1：依赖度=0.966；2：依赖度=0.900；3：依赖度=0.909；4：依赖度=0.900)，而第 5 类，即 GPA（平均成绩绩点）最低的学生，则没有这种行为。第 2 类学生在完成“控制行为”（提交作品、下载资源等）后并没有选择直接完成任务，而是再次进行“预习、规划、激活”（重新打开知识点学习）（控制(C)→预习、规划、激活（F-P-A），依赖度=0.562）。

4.2. 自我调节学习视域下，五类学生在pbl 导向的工程翻转课堂中的具体学习过程

如图3所示, Dependency Graph 过程模型被用于分析 SRL 视域下五类学生在 PBL 导向的工程翻转课堂中的学习行为过程。可以看出, 第1类学生在浏览学习资源 (V-R, 依赖度=1.000) 中表现出较高的重复性和碎片性。此外, 在任务完成过程中, 1类学生在分析任务之后更频繁地返回到查看资源的行为 (A-T→V-R, 依赖度=0.119)。第2类学生则更频繁地进行移动资源操作 (M-R, 依赖度=0.955), 这表明该类学生在小组讨论中更活跃。同时可以看到该类学生从查看资源到查看任务的路径较少 (V-R→V-T, 依赖度=0.006), 更多学生选择在下载资源后直接开始完成作业任务 (V-R→D-R, 依赖度=0.111), 这说明这类学生更加注重任务本身的完成。3类学生在“查看资源”行为后倾向于转向“查看论坛” (V-R→V-F, 依赖度=0.400), 表现出对他人意见的习惯依赖。并且他们分析任务的行为是孤立进行的 (A-T, 依赖度=0.997), 这表明该类学生可能没有很高的协作水平。查看教学大纲后, 第4类学生直接进入移动资源部分 (V-S→M-R, 依赖度=0.500)。很明显, 他们的主要目标是完成老师布置的任务。第5类学生一般在查看教学大纲后选择直接提交作品 (V-S→S-W, 依赖度=0.500), 说明他们的操作大多是以完成任务为导向的, 并且该类学生自我调节学习路径中的反思行为较少, 他们生成计划的行为也与后续的学习任务相隔离 (G-P, 依赖度=0.917)。

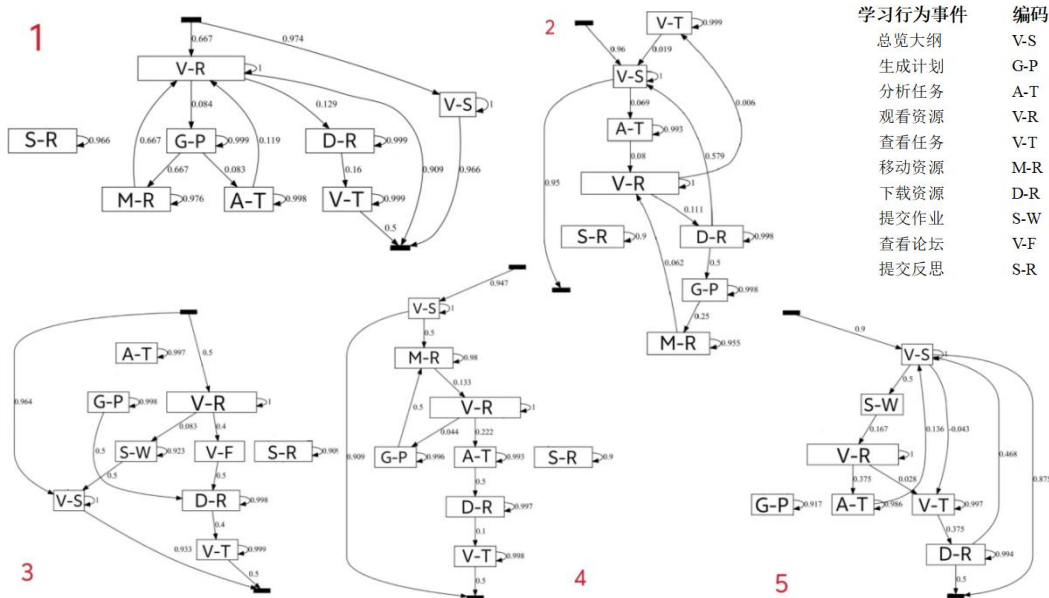


图3 自我调节视域下五类学生工程翻转课堂学习行为Dependency Graph 过程模型

5.讨论和启示

本研究采用聚类分析和过程挖掘的方法, 对工程类本科生在翻转课程中的自我调节学习行为模式进行了研究。结果表明, 根据学习成绩聚类的不同学生群体在自我调节过程中采用的学习行为模式存在较大差异。这表明不同学习成就水平的学生在翻转课堂学习中采用了多种不同的 SRL (自我调节) 策略。

本研究的一个重要发现是高考成绩群体学生的学习表现通常具有较高的自我调节能力和学习策略的灵活性。研究发现, 最高成绩组学生在“浏览资源”方面表现出较高的重复和碎片化学习行为。他们多次复习学习材料以检查遗漏和填补空白, 这种策略表明他们具有高度的自主性和完成任务的内在动机。值得注意的是, 最高成绩组的学生在完成作业时, 往往会采取深入分析任务并回溯到考核标准的策略, 以确保任务的准确性和高质量完成。这种高效的学习策略不仅可以帮助学生在学习过程中优化自己的学习路径, 还可以提高完成任务的质量。

本研究的另一个重要发现是,低成绩群体学生在学习过程中往往面临着更困难的挑战,特别是在处理复杂的工程问题和开放式任务时。研究发现,低成绩群体的学生不仅缺乏反思行为,而且在学习过程中往往表现出被动的任务完成和低质量的参与。在粗略地看了一下课程资源后,他们往往选择直接提交作业,对任务要求和学习计划的实施情况缺乏深入的了解。此外,在面对开放性问题时,低成绩群体学生往往缺乏深入的思考,给出的答案过于肤浅,无法通过应用其他学科的知识来有效解决问题。这种现象反映出了学习困难的学生在解决复杂问题时的认知局限性和学习策略的不足。

基于研究发现,文章提出了一些建议,以期进一步提升翻转课堂学生的学习效率,实现深度学习。首先,鼓励其他学生采用高成绩群体的学习策略将有助于提高他们的学习效果。教师可以通过案例研究或学习实例来鼓励其他学生遵循高成绩群体的学习方式,例如深入分析任务要求或反复学习资源。其次,面对学习有困难的学生,教师可以通过逐步引导和提供更多的支持来帮助学生克服困难。例如,可以设计更简单的学习任务或小步骤指南来帮助低成绩群体的学生逐步培养解决问题的技能。最后,在自我调节学习中,高成绩群体拥有更多的策略,与低成绩群体的学习者存在较大的差距。因此,有必要加强对自主学习策略的训练。

6.结论和局限性

本研究采用过程挖掘和K均值聚类技术探索了工科学生在翻转课堂环境下的自主学习行为路径。研究结果显示,按平均成绩分类的五类学生在学习策略和行为上存在显著差异。高成绩群体的学生展示出重复性和资源密集型的策略,有效地利用了自我反思和团队合作。中等成绩群体的学生表现出较多对外部意见的依赖,合作能力相对较弱,而低成绩群体的学生表现为任务导向式的学习习惯,且小组合作也展现出缺乏反思行为的肤浅参与。

这项研究存在几个局限性。首先,研究仅选择了一门课程的数据进行收集,其代表性可能不足。此外,过程挖掘算法较多,研究只选择了一种算法,可能不太全面。未来的研究还可以选用更多的工程课程来设计翻转课堂,从而拓展数据收集的来源。除此,还可以运用更多的算法或者过程发现模型来挖掘学生的学习行为。

参考文献

- 王翠如和胡永斌(2018)。翻转课堂真的能提升学习成绩吗?——基于38项实验和准实验研究的元分析。开放教育研究(04), 72-80。
- 李文昊、任晓瞳、朱希雅和陈冬敏(2022)。认知弹性设计下自我调节学习行为影响学习成绩的机制研究。电化教育研究(02), 64-71。
- 杨燕清、吴明、林艳芹和洪霞(2022)。自我调节学习在高职护理专业学生在线学习准备度与学习投入间的中介作用。中华护理教育(06), 535-539。
- 邓国民、徐新斐和朱永海(2021)。混合学习环境下学习者的在线自我调节学习潜在剖面分析及行为过程挖掘。电化教育研究(01), 80-86。
- 刘云凤,阎小青和乔增杰.(2021).以培养自我调节学习能力为导向的分析化学实验混合教学模式.大学化学,36(01),124-128.
- 徐晓青、赵蔚、刘红霞和姜强(2021)。自我调节学习中学习分析的应用框架和融合路径研究。电化教育研究(06), 96-104。
- 徐晓青、赵蔚、姜强、刘红霞和乔丽方(2022)。过程挖掘赋能教育数据分析:三种挖掘算法的应用探析。远程教育杂志(03), 45-55。

- Blashfield, R. K., & Aldenderfer, M. S. (1978). The literature on cluster analysis. *Multivariate behavioral research*, 13(3), 271-295.
- Wu, J., & Wu, J. (2012). Cluster analysis and K-means clustering: an introduction. *Advances in K-Means clustering: A data mining thinking*, 1-16.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational psychology review*, 16, 385-407.
- Obada, D. O., Bako, R. B., Ahmed, A. S., Anafi, F. O., Eberemu, A. O., Dodoo-Arhin, D., ... & Obada, I. B. (2023). Teaching bioengineering using a blended online teaching and learning strategy: a new pedagogy for adapting classrooms in developing countries. *Education and Information Technologies*, 28(4), 4649-4672.
- Alsuraihi, A. A. (2022). The effect of implementing mind maps for online learning and assessment on students during COVID-19 pandemic: a cross sectional study. *BMC medical education*, 22(1), 169.
- Blashfield, R. K., & Aldenderfer, M. S. (1978). The literature on cluster analysis. *Multivariate behavioral research*, 13(3), 271-295.
- Morissette, L., & Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 9(1), 15-24.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of educational psychology*, 81(3), 329.
- Zimmerman, B. J. (1989). *Models of self-regulated learning and academic achievement*. In Self-regulated learning and academic achievement: Theory, research, and practice (pp. 1-25). New York, NY: Springer New York.
- Zimmerman, B. J., & Moylan, A. R. (2009). *Self-regulation: Where metacognition and motivation intersect*. In Handbook of metacognition in education (pp. 299-315). Routledge.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational psychology review*, 16, 385-407.
- Obada, D. O., Bako, R. B., Ahmed, A. S., Anafi, F. O., Eberemu, A. O., Dodoo-Arhin, D., ... & Obada, I. B. (2023). Teaching bioengineering using a blended online teaching and learning strategy: a new pedagogy for adapting classrooms in developing countries. *Education and Information Technologies*, 28(4), 4649-4672.
- Bergmann, J., & Sams, A. (2012). Before you flip, consider this. *Phi Delta Kappan*, 94(2), 25-25.
- Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *The Internet and Higher Education*, 22, 37-50.
- Rodríguez, G., Díez, J., Pérez, N., Baños, J. E., & Carrió, M. (2019). Flipped classroom: Fostering creative skills in undergraduate students of health sciences. *Thinking Skills and Creativity*, 33, 100575.
- Chua, K. J., & Islam, M. R. (2021). The hybrid Project-Based Learning–Flipped Classroom: A design project module redesigned to foster learning and engagement. *International Journal of Mechanical Engineering Education*, 49(4), 289-315.

- Chen, X., Long, T., Cheng, L., Gan, X., & Zhu, X. (2023). *Using WPBL to Improve Engineering Undergraduates' Computational Thinking Performance in Flipped Classroom*. In International Conference on Blended Learning (pp. 151-162). Cham: Springer Nature Switzerland.
- Chen, W., & Cheng, H. N. (2020). *Analyzing Graduate Students' Behaviors of Self-regulated Learning in a Blended Learning Environment*. In Proceedings of the 2020 8th International Conference on Information and Education Technology (pp. 53-57), New York: Association for Computing Machinery.