

doi:10.13582/j.cnki.1672-7835.2019.04.023

# 大数据时代在线同伴互评提升策略的 审视与展望<sup>①</sup>

赵呈领<sup>1</sup>,李红霞<sup>1</sup>,蒋志辉<sup>2</sup>

(1.华中师范大学教育信息技术学院,湖北武汉430079;2.长沙师范学院电子与信息工程系,湖南长沙410100)

**摘要:**大数据时代在线同伴互评既能增强学习体验,提升在线学习效果,又能创新评价方式,保证在线教育的质量。现有提高同伴互评信效度的策略主要有评阅人的匹配及推荐、提升评阅人的能力和评价分数的校准三种。大数据时代同伴互评有效性的提升应重点探寻如何提高学习者的评价素养、透视同伴互评设计的基本原理及科学利用过程性数据来提高同伴互评的信效度。

**关键词:**大数据;在线学习;同伴互评;匹配

**中图分类号:**G423 **文献标志码:**A **文章编号:**1672-7835(2019)04-0170-08

国务院2017年印发的《国家教育事业发展规划“十三五”规划》明确提出要“积极发挥大数据在数字学习空间应用方面的重要作用,探索未来教育教学新模式,为推动个性化学习和针对性教学提供支持”。以慕课为代表的在线学习依托互联网平台,可将学习者的学习行为数据进行完整系统的跟踪与记录。因此,在线学习平台的广泛应用,教育数据的急剧增加,意味着教育大数据时代已经来临,大数据与在线学习的深度融合应用已成必然趋势。同伴互评作为在线学习中重要的学习活动受到实践者和研究者的日益关注。其信效度的提升策略更是关注的重中之重。学习者在参与同伴互评的过程中,不仅产生了大量的定量评分及定性的文本数据,也产生了大量的行为数据。大数据延伸了人们的洞察力和过程优化能力,利用大数据来优化同伴互评的过程从而改进和提升信效度也是其发展趋势。本文对现有的同伴互评

信效度提升策略进行审视,进而展望大数据时代在线同伴互评后续研究及实践的方向。

## 一 大数据时代在线同伴互评的价值

### (一) 增强学习体验,提升在线学习效果

学生互评,也称同伴互评,其核心是组织学习者对能力相当的其他学习者的学习作品或表现进行水平、价值或质量的考量和判定<sup>①</sup>。同伴互评活动影响评价者与被评价者双方的成果质量。大数据时代,同伴互评对增强学习体验,提升在线学习效果仍具有重要意义。Papathoma的调查表明学习者认可慕课学习中同伴互评的方式<sup>②</sup>。赵宏对国内典型慕课平台进行调查,其中156门课程中同伴互评都作为重要的评价方式<sup>③</sup>。相较于机器评分而言,同伴互评能使参与者在互评过程中提高学习动机、增强社会存在感、促进高阶思维和

① 收稿日期:2019-04-21

基金项目:教育部-中国移动科研基金项目(MCM20170520);湖南省教育教学改革项目(湘教通[2018]436号925);普通高校教学改革研究项目(湘教通[2016]400号413)

作者简介:赵呈领(1956-),男,湖北老河口人,教授,博士生导师,主要从事教育技术学、教育信息处理、教育信息资源设计与开发研究。

① Topping, K.J. "Peer assessment between students in colleges and universities", *Review of Educational Research*, 1998(3):249-276.

② Papathoma, T., Blake, C., Clow, D., & Scanlon, E. "Investigating Learners' Views of Assessment Types in Massive Open Online Courses (MOOCs)", *Design for Teaching and Learning in a Networked World*, 2015.

③ 赵宏,张亨国,郑勤华,等:《中国MOOCs学习评价调查研究》,《中国电化教育》2017年第9期。

元认知能力的发展<sup>①②</sup>、增强学习体验、支持学习者的个性化需求<sup>③</sup>。对于评价者而言,同伴互评涉及相当高阶的认知技能,如比较、对比和沟通,这有助于巩固和加深评价者对该主题的理解。对于被评价者,讨论或协商标准相关的互评反馈可以帮助其来提高作品质量,鼓励自我反思或更准确的自我评估和(如果很快给出反馈)避免累积误差。

## (二) 创新评价方式,保证在线教育质量

要建设高质量的在线课程,其评价机制是根本保障,没有规范合理的评价机制,就无从实现真正意义上的有效在线学习。在理论层面,无论是教师还是研究者都认识到在线教学像传统教学一样需要利用高信效度的学习评价促进学与教的改进与优化。从某种意义上而言,行之有效的评价机制是在线学习的关键与核心所在。在实践层面,近年来各大在线学习平台的参与人数呈爆发式增长,由此带来了数以万计的待评阅作业,对传统课程中教师批改评价作业的模式带来了巨大冲击。随着大数据和人工智能的发展,机器评价被引入。对于客观题而言,机器评分无疑是效率最高、占用资源最少的评分方式。但课程学习中还存在相当一部分的主观题,主要考察学习者的分析能力、表达能力、批判思维、创新思维等高阶思维能力。同伴互评模式的创新性引入,为主观题的评价提供了新思路。同时,大规模学习者间的互评,是协同学习与集体智慧生成的重要途径。有效的同伴互评不仅是完善在线课程“教-学-评”体系的重要举措,还可切实提升在线教育质量,更可全面提升在线教育的社会认可度,架构在线教育与传统教育学分互认的桥梁。

## 二 大数据时代在线同伴互评提升策略审视

同伴互评作为形成性评价在学习者进行在线学习后的期末成绩中占有一定比例。现有在线学习平台提供的互评形式,其基本程序如图 1 所示。

在线平台公布课程的评分标准,学习者依照

要求参与学习,提交课程作业,开展作业互评,并将评价结果反馈给相应学习者。与此同时,同一门课程根据课程需要,可能会展开多轮互评活动,学习者的最终成绩也是多轮互评成绩的加权得分。因此,为了确保同伴互评的信效度,研究者尝试运用结合多种评价策略,并贯穿于同伴互评的全过程。本研究试图从具体在线课程中的各个阶段与环节,针对现有的策略与模式进行全面梳理、归纳、总结。

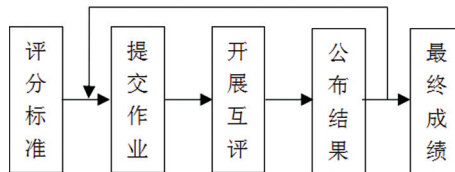


图 1 同伴互评的一般过程

### (一) 评阅前:评阅人的匹配及推荐

评阅人的匹配是对评价结果准确性具有决定性影响的一个环节,同伴互评受学习者自身特征的影响。先前研究表明学习者的知识背景、互评经验、作业的相似度、评价意愿、学习风格等影响学习者评价的信效度。因此,有研究者通过学习分析技术构建学习者模型,将待评作业与学习者特征匹配从而精准推送,以提高同伴互评信效度。例如,基于对待评作业与学习者的若干特征,建立了基于互评效果的推荐模型<sup>④</sup>。也有研究者基于学习者评阅能力的模型进行待评作业与评价者的匹配。随着研究不断深入,各类算法也被应用于解决此匹配问题<sup>⑤</sup>。具体分为基于学习者分组模型、学习者特征、作业相似度等三种匹配策略。

#### 1. 基于学习者分组的匹配

在修订性作业中,参与同伴互评人员的“学术多样化”具有优势,能力较强的学习者通过向较弱的学习者解释概念加强理解,较弱的学习者也可获到更多个人关注。Gehring 基于学习者的分组模型提出了八种匹配策略<sup>⑥</sup>,如表 1 所示。

①Lu, J& Law, N. "Online peer assessment: Effects of cognitive and affective feedback", *Instructional Science*, 2012(2): 257- 275.

②Srijbos, J.W., &Sluijsmans, D. "Unravelling peer assessment: Methodological, functional, and conceptual developments", *Learning and Instruction*, 2010 (4): 265- 269.

③Vickerman, P. "Student perspectives on formative peer assessment: An attempt to deepen learning?" *Assessment & Evaluation in Higher Education*, 2009 (2): 221-243.

④许云红,王如:《MOOC 背景下基于推荐机制提高同伴互评研究效果的研究》,《现代远程教育》2014 年第 5 期。

⑤Crespo, Raquel, Pardo Abelardo, Kloos Carlos Delgado. "An adaptive strategy for peer review", *Frontiers in Education*, 2004.

⑥Gehring Edward F. "Assignment and quality control of peer reviewers", *American Society for Engineering Education*, 2001.

表1 同伴互评匹配策略

策略
策略 R: 随机; 每次运行算法时, 都会获得不同的映射(评价者-被评价者配对集)。
策略 P: 伪随机; 为每份作业随机分配评价者, 但是每份作业每次都会分配相同的评价者。
策略 G: 学习者分组; 组内随机分配评价者。
策略 G+: 从学习者所在小组分配部分评价者, 从其他小组分配部分评价者。
策略 S: 自我选择; 学习者个性化选择评阅作业。
策略 S+: 学习者个性化选择部分评阅的作业, 系统随机分配部分作业。
策略 V: 学习者评阅所在小组的所有成员的作业(每个成员做一个不同的项目)。
策略 T: 学生评阅所在小组的所有成员(合作同一个项目)。

通常情况下,策略 R 优先于策略 P,这是因为每份作业固定评价者容易产生评分偏见。但策略 P 在一个作业建立在前一个作业的基础上具有优势。策略 G,学习者随机评价同一项目中其他学习者的作业,由于学习者之间的背景知识相同,从而能够得到更专业的评价。但是如果评价目的是让学习者从其他项目中汲取知识,策略 G+既能保障学习者作业得到专业评价,也能从其他项目中获得知识共享。在线环境下,学生有不同的背景和经验,学习者开展不同的项目学习,有些学习者可能特别适合评价某些项目,在这种情况下,学习者(策略 S)的个性化选择是合理的。为了保证每份作业都有一定数量的评价者,确保评价内容的可读性,同时非专家型的学习者评价意见对被评价者也有一定的帮助,这种情况下,策略 S+更为适合。策略 V 中,学生评价他们小组的所有其他成员(每个成员完成不同的项目)。策略 T 是最常用的同伴互评模式之一。项目由教师评分,但团队成员评估彼此所做的贡献,采用这种同伴互评策略可以有效避免教师对学生贡献率的忽视,给出不公平分数。

通过上述分析,不难发现,以上匹配策略仍存在共性问题:第一,仅考虑学习者是否参与同一任务但未考虑学习者个人特征。第二,大部分策略是基于传统课堂教学实践总结归纳,面向在线课程复杂的动态学习环境的适切性仍值得考量。第

三,匹配方式多为人工操作,在线课程中“大规模”的学习群体意味着由少数教师执行的传统对等匹配方法并不完全适用。与此相反,这些环境需要引入基于大数据的自动且有效地将评价者与被评价者匹配的“智能机制”,以便为学习者提供更高质量的评估,从而提高互评的信效度。

## 2. 基于学习者特征的匹配

Crespo 等在智能导师系统的应用背景下提出通过构建学习者模型为每份作业精准推荐评阅人<sup>①</sup>。研究中,Crespo 获取学习者的学习分数、学习风格等特征构建学习者模型,采用基于模糊聚类 and 遗传算法对学习者的进行匹配。但是,自适应匹配仅为一对一的,未能实现一对多或者多对多的评审机制。Giannoukos 等提出一种融合前馈神经网络的自适应算法进行评阅人的推荐。该算法主要是基于学习者对该领域的熟练程度、平均严格度、平均有用率、互评意愿等相关信息进行动态推荐。夏雪提出一种基于学习者显性特征的作业互评模式,通过在线学习平台获取参与互评学生的先前成绩、所在地域、学习环境等,将背景大致相同的学生分成一组,然后在组内进行随机互评<sup>②</sup>。Raquel M 等根据学习者模型和基于模糊分类技术来匹配学习者,遗传算法用于优化学习者和评论者的分配<sup>③</sup>。自适应推荐评阅者的同伴互评具体过程如图 2 所示。首先,学习者或小组处理给定任务并提交作业,然后根据先前存储的学习者文档,将每份作业分配给一组合适的评阅者。每个评阅者单独或成组地产生一组评论发给作者。其次,学习者评估收到的评论的有用性和一致性程度。最后,更新学习者文档,从而实现学习者模型的动态更新,匹配机制的不断优化。

在研究者构建的以上模型中,实证研究显示,互评效果均有一定程度提升。但是在线课程情境下应用仍存在一定局限性。由于慕课的“开放性”,意味着学习者可以灵活地参与和退出课程,学习者数量也是动态变化的。因此,造成第一次参与课程的学习者其用户文档中特征数据不足或者其评论的有效性数据为零,导致推荐算法的不适用。另一方面,需要学习者对评阅者评论的有

<sup>①</sup>Crespo, R. M., Pardo, A., Juan Pedro Somolinos Pérez, & Kloos, C. D. "An algorithm for peer review matching using student profiles based on fuzzy classification and genetic algorithms", *18th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, 2005.

<sup>②</sup>夏雪:《基于蚁群聚类算法的 MOOC 作业互评系统的分组研究》,成都理工大学硕士论文,2017 年。

<sup>③</sup>Crespo, R. M., Pardo, A., & Kloos, C. D. "An adaptive strategy for peer review. *Frontiers in Education*", IEEE, 2004.



用性进行评价,增加了额外工作,同时评价的有用性也受到学习者个人特征影响,从而使得整体推荐机制准确性降低。

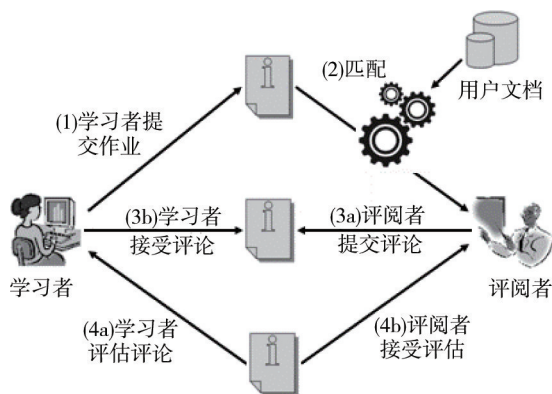


图2 同伴互评过程

### 3. 基于作业相似度的匹配

多位研究者提出通过计算学习者之间作业的相似度来进行评阅人的匹配。如王如提出从评阅人的评阅意愿、评阅能力和作业相关度三维度构建学习者特征集和作业特征集,并基于以上三维度目标最大化约束条件下建立推荐模型<sup>①</sup>。学习者特征集的构建从学习者直接表露的信息、学习者与学习系统的交互活动信息和从学习者的行为活动中推断出的信息。学习者的作业相似度通过TF-IDF算法进行关键词计算和余弦相似度语义分析得到,作业相似度相近的学习者对彼此的作业领域具有更高的熟悉度,若将他们进行同伴互评匹配有助于减少评价偏差。

但笔者认为,通过关键词计算和语义分析得到的作业相似度其意义区别于知识领域匹配。将作业相似度高的学习者间进行同伴互评匹配不仅有可能因为主观上的认同感造成评分偏高,也不利于学习者的观点交互和互评过程中的再学习,

还有可能不利于培养学习者对于同伴互评的兴趣和意愿提升。

## (二) 评阅中:提升评阅人的能力

### 1. 模拟训练提高评价技能

同伴互评的成功在很大程度上取决于学生是否具有批判性评估技能并对同伴的工作提供有效判断<sup>②</sup>。学生应清楚地了解评价标准以及如何评估同伴的表现以及他们自己的表现<sup>③</sup>。但是,学生可能有困难理解评价标准,或者他们可能会以不同于教师的方式解释标题中使用的术语<sup>④</sup>。或者学生可能缺乏批判性评估技能<sup>⑤</sup>。因此,在同行评估之前,培养学生的评估技能引起了人们的注意。目前培训形式包含:(1)提供同伴互评的评审标准和参与互评的意义价值<sup>⑥</sup>。如在参与互评前,在线学习平台在公告中说明学习者参与同伴互评的必要性,提供课程团队研发的评价量规供学习者参考,从而使学习者了解评价的具体要求;(2)范例式学习,即提供有专家打分的范例供学习者学习<sup>⑦⑧</sup>。(3)试评练习,用评价样例供学习者进行练习,直到学习者的评分符合系统要求才允许参加同伴互评。

为了探究哪种训练方式学习者的评价技能提升更快,李菲茗等比较了“练习+反馈”(学习者依据课程团队提供的评分规则对作品进行试评活动,并在完成后收到专家的评分反馈,可以对照发现自己的评分不足,进行改进)和范例式学习(学习者直接学习专家打分的相同作品,不参加试评活动)这两种培训效果的差异<sup>⑨</sup>。结果表明,“练习+反馈”比范例学习评分误差较低,准确性较高。虽然范例式学习操作简单易学,但是通过学习者实际操作练习并获得反馈,学习者能够进行更为深刻的训练,从而评分准确性更高。Moccozet等提出在同伴互评中增加一个关键环节

①王如:《基于信息整合方法的网上同伴互评评阅人推荐系统分析与设计》,昆明理工大学硕士论文,2016年。

②Li, L., X. Liu, and A. L. Steckelberg. "Assessor or Assessee: How Student Learning Improves by Giving and Receiving Peer Feedback", *British Journal of Educational Technology*, 2010 (3): 523-536.

③Liu, X., & Li, L. "Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment", *Assessment & Evaluation in Higher Education*, 2013(3): 275-292.

④Orsmond, P., Merry, S., & Reiling, K. "The importance of marking criteria in the use of peer assessment", *Assessment & Evaluation in Higher Education*, 1996(3): 239-250.

⑤Svinicki, M. D. "Encouraging your students to give feedback", *New Directions for Teaching & Learning*, 2001(87): 17-24.

⑥白丽茹:《基础英语写作同伴互评反馈模式的可行性及有效性检验》,《解放军外国语学院学报》2013年第1期。

⑦马志强,王雪娇,龙琴琴:《基于同侪互评的在线学习评价研究综述》,《远程教育杂志》2014年第4期。

⑧Liu, X., & Li, L. "Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment", *Assessment & Evaluation in Higher Education*, 2013(3): 275-292.

⑨李菲茗,李晓菲,黄亚平,等:《训练对同伴互评评分准确性的影响——以“三维动画设计与建模”课程为例》,《中国远程教育》2018年第5期。

为理解和应用评价量规的培训活动<sup>①</sup>。但是他们不建议通过直接提供评价量规来进行培训。在线同伴互评的研究表明,更好的同伴互评培训应该更具互动性,学生有机会阐明和协商他们对评价量规的理解并参与话语过程<sup>②</sup>。众所周知,它可以促进更深入的学习,例如解释、争论和提问。因此在参与在线同伴互评之前,向学习者提供关于评分标准的小组同步讨论。

然而,通过培训来提高学习者评阅能力的相关研究仍处于初级阶段。现有研究大多是在有培训(某一种培训方式)和无培训之间进行对照比较<sup>③</sup>,或者是培训前后比较,无对照组。我们应该将研究重点关注于培训的实质,即在什么样的情境下,采用什么样的培训方式,提供什么培训内容更有助于培训效果的提升及学习者评估技能的提升。

## 2. 评价量规提供支架式辅助

评价量规是在同伴互评中,对于作业的评价指标及对应权重的标准性说明,以及作业最终分数的计算。对于同伴互评而言,评价量规的制定和实施是降低评价主观随意性,确保互评结果客观、统一、公平的重要工具<sup>④</sup>。量规在学生自评与互评中也有显著成效。Panadero 等人发现与专家评价相比,无量规的同伴互评中容易给予同伴作品过高的评价,而基于量规的实验组具有更高的信效度<sup>⑤</sup>。设计良好的评价量规在一定程度上提高了同伴互评的有效性(即作业互评的平均分数与教师批改该作业所给出的分数更为接近)和准确性(学生们给出的分数之间也更为一致)<sup>⑥</sup>。基于量规的在线同伴互评中学生认知投入度显著高于无量规的在线同伴互评,但是情感投入度显著低于无量规的在线同伴互评。量规应该在什么情境下使用?在线课程中如何设计量规?有研究者进行了更为细致的探究。

在量规使用情境方面,张屹的研究表明在学

习活动前给予量规对元认知水平的提升具有促进作用<sup>⑦</sup>。在一定程度上可以降低学生认知负荷。学习活动中使用量规对学生学习活动的完成、评价和完善具有积极作用。学习活动后使用量规,可以帮助学生反思自身问题和不足之处,进行自我评价和反思。但在学习活动前给予量规,在一定范围内可能限制学生的思路,应根据不同的学习活动类型和学习目标确定量规内容,以及给予的时间和数量。

在评价量规的设计方面,Moskal BM 指出评价量规可分为整体量规和分项量规<sup>⑧</sup>。整体量规提供对作品或过程的整体而全面的评价。而分项量规则包括有若干评价指标,需要逐一进行评价,最终计算总分。同伴互评中的评价量规应该如何设计,已有研究还没确定的结论。有研究发现,整体量规指导下的打分更接近教师给出的分数<sup>⑨</sup>。有趣的是也有研究发现,让学生分别从不同评价指标打分,计算的总分能较好地吻合教师评分<sup>⑩</sup>。但是,达成共识的是:整体量规能够提高评价的效率,因此,教师评分较多使用整体量规。分项量规的多项评价指标能够更为具体地呈现需要学生掌握和应用的知识点或技能点,提供较为清楚的评分结构,便于学生阅读和理解,更客观地评价他人作品或作业。但是需要注意评价指标的数量不能太多,已有研究发现,使用较多评价维度的量规会增加评价难度进而影响评分有效性。评价量规的等级设计需要尽量避免重合性,也需要有较为明确的辨识度,采用是非选择型的表达方式,减少阅读难度。

综合先前研究,更多关注在于验证同伴互评中使用量规的有效性及其对学习者的影响。但是较少研究说明如何使用量规或者怎么设计量规,且实践较少。在线学习中,学习者的背景复杂,什

①Moccozet, L., Tardy, C. "An assessment for learning framework with peer assessment of group works", *International Conference on Information Technology Based Higher Education & Training*, 2015.

②Kollar, I., Fischer, F. "Peer assessment as collaborative learning: a cognitive perspective", *Learning & Instruction*, 2010(4).

③Scott Ashton, Randall S. Davies. "Using scaffolded rubrics to improve peer assessment in a MOOC writing course", *Distance Education*, 2015.

④范逸洲,冯菲,刘玉:《评价量规设计对慕课同伴互评有效性的影响研究》,《中国远程教育》2018年第11期。

⑤Panadero, E., Romero, M., & Stribos, J. W. "The impact of a rubric and friendship on peer assessment: effects on construct validity, performance, and perceptions of fairness and comfort", *Studies in Educational Evaluation*, 2013(4):195-203.

⑥范逸洲,冯菲,刘玉:《评价量规设计对慕课同伴互评有效性的影响研究》,《中国远程教育》2018年第11期。

⑦张屹,陈蓓蕾,陈珍,等:《智慧教室中的教学对大学生研究能力和元认知的影响——基于量规的干预》,《电化教育研究》2017年第7期。

⑧Moskal, B. M. "Scoring rubrics: what, when and how?" *Practical Assessment Research & Evaluation*, 2000(3):1-5.

⑨Falchikov, N., & Goldfinch, J. "Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks", *Review of Educational Research*, 2000(3):287-322.

⑩Bouzidi, L., & Jaillet, A. "Can online peer assessment be trusted?" *Journal of Educational Technology & Society*, 2009(4):257-268.

么样的量规使用既能发挥学习者的多元评价,又能让学习者的分数更为合理准确?也是后续研究应该关注的重点。

### 3. 匿名评价避免社交效应

同伴互评中的三大主体为评价者、被评价者和教师。三者间人际变量的相互作用会影响评估结果,人们常常认为通过匿名可以减少负面社会影响<sup>①②</sup>。通过匿名学生对同伴的知识和关系的偏见较小<sup>③</sup>,可能提供更客观的反馈。当评估员的匿名性得到保证时,学生对同伴评估有更积极的态度。Yu and Sung 指出,匿名能会为学生提供更多的心理安全,但同时也可能导致不良行为<sup>④</sup>。匿名环境下,学习者感知评价者的专业知识与自己越相关,其更愿意改进作业。因此,同伴互评的匿名性似乎影响了反馈(评语或分数)的接收和处理方式。Panadero 和 Brown 最近的一项调查研究表明西班牙教师大部分认同匿名同伴互评<sup>⑤</sup>。

以上研究证明了同伴互评中匿名形式可以提高同伴互评的信效度,但是多数研究集中在匿名与否状态下同伴互评信效度的差异。匿名具有不同的形式。评价者和被评价者都有署名、匿名两种状态。同时,评价者与被评价者也可选择是否对教师匿名。现在研究者开始关注不同的匿名形式是否会对互评的信效度产生影响。例如,张倩发现单向匿名(评价者匿名)与署名(评价者署名)相比,单向匿名的批判性强,但是关注点涣散,评论质量低;完全匿名(评价者与被评价者双向匿名)与单向匿名相比,互评的信效度高于单向匿名<sup>⑥</sup>。韩庆年从研究中表明女生倾向于匿名评价,男生更倾向于实名评价<sup>⑦</sup>。Miyazoe 等的研

究也说明不同匿名形式下不同性别的学习者参与同伴互评的质量有差异,匿名效果下,女生的表现更佳<sup>⑧</sup>。学习者参与匿名评价时,焦虑感降低,更倾向于表达真实意见。署名状态下,评价者更为认真负责,谨慎,但是可能存在虚假评论<sup>⑨</sup>。因此可以看出,不同匿名方式下,学习者的评分准确性会受到影响。

但是现有研究对教师匿名与否会影响学习者评分准确性的研究较少。同时,究竟什么情境下使用哪种匿名方式,暂无研究者进行系统的研究,因此实践者需要根据自己的情境自行选择匿名方式。

### (三) 评阅后:评价分数的校准

#### 1. 基于同伴评分的信效度的校准

同伴互评通常用作学习者作业的总结性评估,鉴于同伴互评在这些情况下的重要性,学习者和教师都表示担心部分学生可能没有足够的动力或知识足以提供与同龄人和教师一致的评分<sup>⑩</sup>。一些在线同行评估系统通过使用评阅者权重来有效避免这些问题。例如,Aropä、SWoRD 系统等通过算法来计算多个评阅者对同一作业的评分一致性。评阅者所打分的权重基于其评分的一致性程度。Song 等证明,现有算法产生的加权评分比同伴互评原始评分的平均值更为准确<sup>⑪</sup>。因此,使用评阅者权重来计算成绩可能会降低动机不足或有困惑的学生的影响。一些研究者建议利用评阅者评分的准确性来进行奖励。一个评阅者其评分越准确,可给予额外的奖励分数,评分准确性越低,反而会从其原始作业成绩中扣分,这种互评准

①Howard, C. D., Barrett, A. F., & Frick, T. W. "Anonymity to promote peer feedback: Pre-service teachers' comments in asynchronous computer mediated communication", *Journal of Educational Computing Research*, 2010(1): 89-112.

②Dijks, M., Brummer, L., & Kostons, D. "The anonymous reviewer: the relationship between perceived expertise and the perceptions of peer feedback in higher education", *Assessment & Evaluation in Higher Education*, 2018(3): 79-99.

③Schellens, T. "Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education", *Studies in Higher Education*, 2015(1): 178-193.

④Yu, F. Y., & Sung, S. "A mixed methods approach to the assessor's targeting behavior during online peer assessment: Effects of anonymity and underlying reasons", *Interactive Learning Environments*, 2015(1): 1-18.

⑤Panadero, E., & Brown, G. T. L. "Teachers' reasons for using peer assessment: Positive experience predicts use", *European Journal of Psychology of Education*, 2017(1): 133-156.

⑥张倩:《网络环境下同伴匿名评价实证研究》,《电化教育研究》2007年第7期。

⑦韩庆年,柏宏权,杨晓敏:《移动学习环境下同伴互评对大学生学习动机的影响效应研究》,《中国远程教育》2018年第11期。

⑧Miyazoe T, Anderson T. "Anonymity in Blended Learning: Who Would You like to Be?" *Educational Technology & Society*, 2011(2): 175-187.

⑨马玉慧,赵乐,刘晴:《在线同伴评价的影响因素及其促进策略》,《电化教育研究》2006年第3期。

⑩Kaufman, J. H., & Schunn, C. D. "Students' perceptions about peer assessment for writing: their origin and impact on revision work", *Instructional Science*, 2011(3): 387-406.

⑪Song, Y., Hu, Z., & Gehringer, E. F. "Pluggable reputation systems for peer review: A web-service approach", *IEEE Frontiers in Education Conference*, 2015.



确度等级将迫使学生认真对待评级任务。

## 2. 基于评阅者偏见的校准

影响同伴互评信效度的一个重要因素是评论者的偏见。例如,如果某些学习者宽松而其他学习者严格,则学习者可能会给出有偏见的评估。因此,在计算学习者同伴互评的成绩时应该考虑降低评论者偏见的影响,从而提高同伴互评的信效度。Piech等基于学习者的信息(学习者的评论方式和学习者的学习表现等)来赋予评阅者不同的权重,从而降低评阅人在评阅他人作业时的偏见<sup>①</sup>。孙力构建了三个同伴互评分数校准模型,模型1:仅考虑评分偏差和评分者可靠度;模型2:进一步考虑评分者偏差的时间相关性;模型3:增加了学习者的评分能力(通过其作业计算)。模型3的信效度高于前两个模型。因此,其建议基于模型3考虑影响因素和评分合成原则生成最终得分<sup>②</sup>。

## 三 大数据时代在线同伴互评提升策略展望

本研究探索了大数据时代同伴互评对在线学习的价值,审视了在线同伴互评提升策略。同伴互评机制有效性的提升策略较为全面,但是现阶段有许多可行的理论成果由于平台、对象的限制无法进行普遍性的验证,在线互评过程中产生的大量数据还未能及时用来反馈指导评价行为。基于此,大数据时代在线同伴互评后续研究及实践将侧重三个方面。

### (一) 提高学习者评价素养是提高同伴互评信效度的内核所需

同伴互评的评价结果受学习者自身素质的影响较大,因而同伴互评的结果常常被在线学习者质疑。获得公正、客观的学习结果评判是学习者的基本诉求,如果获得与他真实水平不符的成绩,则会削弱其学习的积极性。提高同伴互评信效度最有效的方法为提高学习者的评价素养。学习者的评价素养包含评价态度、评价知识及评价技能。目前的策略也都以提高学习者的评价知识和技能为核心。除了评价中提升学习者的评价技能外,评价前基于大数据精确推荐评阅人和评价后分数的校准都是基于学习者的评价技能及知识不足来开展实施。策略的本质即为通过其他技术特征来

弥补学习者评价技能不足而造成的评分偏差。但是对于评价意识的相关策略研究较少。在线学习者对同伴互评的目的与功能认识不清,导致评价态度与意识不到位。正确积极的评价态度与意识能够提高学习者的互评参与度,激发学习动机,培养学生的自主学习和反思能力。因此,在线学习应在学习者参与互评前引导学习者了解同伴互评的目的与意义,激励学习者积极参与互评活动,通过训练及量规等提高学习者的评价知识与技能,综合提高学习者的评价素养是提升同伴互评信效度的内核所需。

### (二) 透视同伴互评设计的基本原理是提高同伴互评信效度的本源所在

跟其他任何一种教学情境下的任何类型的学习评价(比如教师评价、学习者自评)一样,在线课程中的同伴互评的设计必须遵循学习评价设计的基本原理,大数据及其技术为认识探索教育规律提供了前所未有的手段,对改变教与学的方式提供了强有力的支持。但是研究者对此探究还远远不够。为了提高同伴互评的信效度,在同伴互评前,众多研究者设计了自动推荐评阅人机制,在完成同伴互评后,提出互评分数的校准。但是却存在诸多问题,许多基本原理却模糊不清,如推荐评阅人和校准分数时,有研究者基于学习者的评价能力,也有研究者基于学习者的学习风格等,除此以外,还有许多因素影响着同伴互评结果的准确性。又如,在线课程中学习者之间的交互行为,学习者在课程论坛内的社交情况,学习者对于评分量规的认可程度等等。相比之下,这些因素对在线课程中同伴互评信效度的影响主要是通过影响学习者在课程学习以及评价中的心理、行为等感性认知达到的,这些因素具有极强的耦合性和复杂性。因此,究竟什么因素影响在线学习者互评的信效度?同伴互评中学习者模型应该如何构建?仅考虑单个因素来推荐评阅人及校准分数是否存在一定偏见?部分研究者综合考虑多种因素,但是各个因素之间的权重如何计算仍未给出科学合理的答案。学习者之间的匹配策略仍比较模糊。现有研究中基本赞成“高-高、高-低”,但是这种匹配策略仍存在争议,其虽然在评价效度上有所提高,但是“同伴互评不仅仅是为评价而评价”,这在一定程度上限制了低能力

<sup>①</sup>Piech Chris, Huang Jonathan, Chen Zhenghao, Do Chuong, Ng Andrew, Koller Daphne. "Tuned models of peer assessment in MOOCs", *Educational Data Mining (EDM 2013)*, 2013.

<sup>②</sup>孙力, 钟斯陶:《MOOC评价系统中同伴互评概率模型研究》,《开放教育研究》2014年第11期。

的学习者向高能力学习者学习的机会。因此如何设计匹配策略在保证信效度的情况下又能提高学习者在同伴互评中的过程性学习,后续研究应给予足够的重视。

在同伴互评中,为了使学习者同伴互评结果更可信,现有研究提出的策略有训练学习者的评价知识与技能、量规的使用避免评价的盲目性、匿名评价减少同伴互评过程中的社会性因素影响。但是现有研究更多集中在以上策略的有效性探究上,后续研究应将研究重点转到在特定情境下实施如上策略的相关探究。如,在线课程分为 cMOOC、xMOOC、SPOC 等,互评作业类型有文档型、作品型等。不同的情境下,采用什么样的培训方式,提供什么培训内容更有助于培训效果的提升及学习者评估技能的提升。量规应该如何设计,在什么时候使用量规更能提高同伴互评的信效度?匿名也分为多种形式(单向匿名、双向匿名),在何种教学情境中使用哪种匿名方式更能提高互评的信效度。科学合理的设计同伴互评需要遵循一定的基本原理,因此,大数据时代,基于大数据技术透视同伴互评设计的基本原理是提高同伴互评信效度的本源所在,也是研究的重中之重。

### (三) 科学分析过程性数据是提高同伴互评信效度的关键保障

大数据能够为教育决策提供支持和服务,基于大数据分析的在线学习过程性评价是发挥大数

据优势、提高在线学习质量的有效途径。基于大数据的同伴互评,过程性与目标性并重,既注重同伴评价的信效度,也注重培养和发展学习者对自我学习评价行为、态度和价值观等的内省,其更能推进对学习者在在线互评动机、过程和效果的三位一体评价。因此,如何加强对在线学习同伴评价数据的深度分析与利用,从而提高同伴互评的信效度研究非常值得教学实践者与研究者的深入关注。在线学习中,同伴互评数据不但包括了在线学习者相互之间的评价分数,也包括了在线平台记录的同伴评价频率、时间、数量等海量数据。如何利用学习分析技术去对同伴互评所产生的大规模隐形数据进行分析从而挖掘出更为便捷的提高同伴互评信效度的策略将是研究的重点。如,同伴互评中,学习者不仅仅给出分数,同时也要对分数给出合理地解释,并提出修改意见等。因此,利用语义分析将学习者的定性与定量评价相结合,对学习者的所给出分数进行纠正从而提高信效度不失为一种策略。研究可以通过隐形数据挖掘出同伴互评有效性模型:在什么时间段、具有什么特征的学习者经历多长时间评阅一份作业其信效度更高。对于系统自动识别出来信效度较低的评分,给予纠正。总之,研究者可以基于大量行为数据以挖掘出在线同伴互评更多更深的价值以及其中蕴含的教育启示与建议。

## Review and Prospect of Online Peer Assessment Strategy in Big Data Era

ZHAO Cheng-ling<sup>1</sup>, LI Hong-xia<sup>1</sup> & JIANG Zhi-hui<sup>2</sup>

(1. School of Educational Information Technology, Central China Normal University, Wuhan 430079, China;

2. Department of Electronic and Information Engineering, Changsha Normal University, Changsha 410100, China)

**Abstract:** Online peer assessment enhances the learning experience and the effectiveness of online learning. Innovative evaluation methods can ensure the quality of online education. The existing strategies for improving the reliability and validity of peer assessment mainly include the matching and recommendation of reviewers, the improvement of reviewers' ability, and the calibration of evaluation scores. Based on the above, in order to further enrich and enhance the effectiveness of peer assessment in the era of big data, follow-up researches will focus on how to improve the evaluation literacy of learners, the basic principles of peer assessment design, and the scientific use of procedural data to improve the reliability and validity of peer assessment.

**Key words:** big data; online learning; peer assessment; adaptation

(责任校对 朱正余)