

面对面小组同伴互评中插电与不插电评分方式如何影响评分结果

How Plugged-in and Unplugged Scoring Methods Affect Scoring Results in Face-to-face

Group Peer-to-Peer Assessment

刘欣悦, 娄舒婷, 黄洛颖*

河北大学教育学院

* huangly@hbu.edu.cn

【摘要】 本研究探讨基于电子评分表的插电方式与基于纸质评分表的不插电方式在面对面小组同伴互评中的效果差异。选取小组作业汇报与互评为情境, 以 12 个小组为样本, 基于 16 位评分员在两种评分方式下的评分结果进行多维分析。结果表明, 两种评分方式在总体及各维度评分、改进建议字数和评分时长上均无显著差异; 两种评分方式下评分员均更易于在客观结果评分上达成共识, 而非过程和情感维度; 插电方式在改进建议的字数与内容质量、与专家评分一致性等方面表现更佳。未来将优化实验设计, 以进一步验证和深化上述结论。

【关键词】 同伴互评; 插电评价; 不插电评价; 面对面; 小组任务

Abstract: This study investigates the differences in the effectiveness of the plug-in approach and the unplugged approach in face-to-face group peer-to-peer assessment. Group work debriefing and peer assessment were the context, and a sample of 12 groups was analyzed based on the scoring results of 16 raters under the two scoring modalities. The results showed that there were no significant differences between the two grading methods in scores, word count of suggestions, and grading duration; graders were more likely to agree on objective outcome scores than process and affective dimensions; and the plug-in method performed better in word count and content quality of suggestions, and consistency with expert. The experimental design will be optimized to validate and deepen the above findings.

Keywords: peer-to-peer evaluation, plugged-in evaluation, unplugged evaluation, face-to-face, group tasks

1. 引言

同伴互评能够增强学生群体协同互动(马宁等, 2022)、激发动机和自我效能感、促进思维发展(张涛等, 2022)以及提高评价效益等。尽管已探索不同互评方式的效果差异, 但现有研究多关注在线环境, 鲜少探讨面对面互评及其不同方式的效果, 缺乏实证证据和规律探究。本研究聚焦于面对面同伴互评场景, 以小组作业汇报与互评为情境, 比较电子评分表(插电)与纸质评分表(不插电)在面对面同伴互评中的效果差异, 为传统课堂同伴互评提供实证指导。

2. 文献综述

2.1. 同伴互评有效性研究

研究者从定量评分和定性评论建议两方面探究了同伴互评的有效性。前者聚焦于同伴互评的可靠性和准确性, 采用相关性检验、内部一致性检验以及配对样本 t 检验等方法, 将教师评分作为参考(张义兵等, 2018), 对比其与同伴互评结果的一致性(Vista et al., 2015); 后者聚焦于评语内容质量(张涛等, 2022)并对其进行质性等级划分, 等级越高表明同伴互评质量越好。

2.2. 小组协作任务评价

研究者主要从任务完成质量、小组协作过程以及反思与收获三个维度评价小组协作任务。在任务完成质量维度, 主要从成果内容的准确性、深度与完成度, 以及产出成果的过程和步

骤是否科学两方面评价。在小组协作过程维度上,任务分工、沟通交流以及主动参与等是普遍关注的评价指标。在反思与收获维度,主要关注了小组成员自我反思及收获的深度两方面。

2.3. 研究问题

鉴于现有研究鲜有涉及面对面同伴互评中不同互评方式的效果差异,本研究旨在比较电子评分表(插电)与纸质评分表(不插电)在面对面互评中的效果,包括以下研究问题:1)两类评分方式的结果是否存在差异?三个维度是否存在差异?2)两类评分方式的改进建议的字数和质量是否存在差异?3)两类评分方式的评分时长是否存在差异?4)两类评分方式的评分结果的内部一致性是否存在差异?5)哪类评分方式的评分结果与专家评分结果更接近?

3. 研究设计

3.1. 样本选择与分组

本研究的样本为某地方综合性大学教育相关专业的硕士一年级学生,共75人。基于学生的个人意愿和自由随机分组的方式,分为12个小组(5-6人/组),有8人未加入任何小组。评分员共18位,含每组选出的一员(共12位)、2位专家及4位未入组志愿者。

3.2. 研究过程

本研究以一门研究生课中的小组文献综述汇报与同伴互评环节为研究情境。研究过程包括:1)讲解互评流程;2)抽签确定顺序,评分员随机选择评分方式并分组交替使用两种评分表;3)分发纸质评分表和二维码;4)小组汇报,评分员评分;5)汇报结束后回收评分表。

3.3. 数据收集工具

基于已有文献自主开发了小组协作文献综述汇报评分表,分为纸质评分表和电子评分表两种形式,涵盖综述过程与结果(5题)、小组协作(3题)和反思与收获(2题)三个维度,另设置了一道“请提出完善建议”的开放性试题。此外,还采集了评分员的基础信息。

3.4. 数据分析方法

采用了多种分析方法,包括:描述性统计用于揭示互评分数、修改建议和评分时间的特征和分布;推断性统计用于分析两种评价方式在总评分、各维度评分、改进建议字数和评分时长等方面的差异;内容分析用于评估建议质量,两位编码员一致性较高($\kappa=0.772, p<0.01$);组内相关系数用于分别分析不同评价方式评分结果内及其分别与专家评分结果间的一致性。

4. 研究结果

4.1. 两种评分方式在总体及各维度评分上的比较

相较于插电方式,评分员在不插电方式时的评分相对更高且更加集中(见表1)。

表1 插电/不插电评分方式在总体评分上的比较

评价方式	N	最小值	最大值	均值	标准差
插电	96	2.8	4.9	4.05	0.460
不插电	96	3.2	5	4.09	0.406

分析三个维度的评分,结果如表2所示。两种评分方式在三个维度上的关系与总体评分趋势相一致。此外,两种评分方式在综述过程与结果维度评分的差距较小。

表2 插电/不插电评分方式在各维度评分上的比较

评价方式	综述过程与结果		小组协作		反思与收获	
	均	标准	均	标准	均	标准
	值	差	值	差	值	差
插电	4.06	0.439	4.08	0.648	3.96	0.689
不插电	4.09	0.422	4.16	0.507	4.01	0.603

独立样本t检验显示,两种评分方式在总评分和三个维度评分上均无显著差异($p>0.05$)。

4.2. 两种评分方式在改进建议上的比较

相较不插电方式,评分员在插电时的字数更多且更离散(见表3),但两者间无显著差异。

表3 插电/不插电评价方式在建议字数上的描述性统计

评价方式	N	最小值	最大值	均值	标准差
插电	96	1	87	20.78	17.04
不插电	96	0	63	16.87	14.56

其次,在建议的质量上,采用插电方式时给出的建议质量相对更高,如表4所示。

表4 插电/不插电评价方式在评价建议质量上的频次和百分比数据

等级	插电		不插电	
	频次	百分比(%)	频次	百分比(%)
优	3	3.1	1	1
良	12	12.5	11	11.5
中	61	63.5	59	61.5
差	20	20.8	25	26

4.3. 两种评分方式在评分时长上的比较

描述性统计分析结果如表5所示,插电评分方式用时更短,但是离散程度相对更大。

表5 插电/不插电评分方式在评价时长上的描述性统计

评价方式	个案数	平均值	标准差
插电	96	605	215.778
不插电	96	614.53	148.208

进一步进行差异性检验,结果显示两种评价方式在评价时间上不存在显著差异($p>0.05$)。

4.4. 两种评分方式评分结果内部一致性分析及比较

组内相关系数分析结果如表6所示。两种评分方式各自的内部一致性均较差($p>0.05$)。

由于项间平均协方差为负,不插电的同类相关性系数为负,后续将进一步探究其原因。

表6 插电/不插电评分方式的评分结果内部一致性分析及比较

数据	插电	不插电
同类相关性	0.211	-0.726
显著性	0.277	0.771

4.5. 两种评分方式评分结果与专家评分结果的比较

两名专家和两种评分方式的分数的分析结果如表7所示。插电方式评分结果和专家评分结果的一致性较强,在一定程度上说明评分员采用插电评分方式给出的评分结果相对更可靠。

表7 插电/不插电评分方式的评分结果与专家评分结果的一致性分析

数据	专家之间	插电方式和专家	不插电方式和专家
同类相关性	0.667	0.681	0.023
显著性	0.041*	0.035*	0.485

进一步对插电方式的三个维度的评分和专家评分进行一致性分析,结果如表8所示,其一致性水平由高到低依次为综述过程与结果、反思与收获和小组协作,且在综述过程与结果维度的一致性达到显著水平。

表8 插电评分方式的三个维度评分结果与专家评分结果的一致性分析

数据	综述过程与结果	小组协作	反思与收获
同类相关性	0.735	0.472	0.614
显著性	0.019*	0.152	0.065

5. 结论与讨论

5.1. 采用两种评分方式的评分结果、改进建议字数和评分时长均不存在显著性差异

这一发现对评分系统的设计与改进具有重要意义,即在考虑评分方式时,可以更多地关注评分员的偏好和操作便利性。然而,鉴于样本量有限、实验周期较短,研究结果有局限性。

5.2. 采用不插电方式时评分员普遍更倾向打高分,提出的改进建议字数相对更少、质量更低

评分员同伴在旁、评分表公开、评分过程透明的多重原因可能致使打分更高;同时可能由于评分时间紧迫、书写记录方式不便等,导致了改进建议的详尽程度和质量下降。后续将在更加独立和安全的评分环境(Yu & Lee, 2016)、更为充裕的评分时段开展研究,探究以上原因。

5.3. 采用插电方式的评分结果、建议字数及评分时长均更离散,但效率更高、结果更可靠

尽管插电评分方式在评分结果上表现出一定的离散性,但其在效率和可靠性方面的优势使其成为一种有效的评分工具。后续将通过引入更先进的分析工具和方法,并扩充访谈、观察等质性数据,来更准确地分析评分结果,从而提高研究结论的信度和效度。

5.4. 采用两种方式的评分员都易于在客观结果评分上达成共识,而在过程和情感方面更离散

这一结果反映了评分员对客观结果的评估标准更为明确和统一,而在评估主观和定性的过程和情感因素时,可能更加依赖个人经验和价值判断。这启发我们在设计评分表时需对主观性较强的评价维度提供更明确的评分指导和标准,以降低评分的主观性。

参考文献

- 马宁,路瑶,郭佳惠,& 刘春平.(2022). 评价支架对教师在线同伴互评质量的影响研究. *电化教育研究*, 43(2), 34–41. <https://doi.org/10.13811/j.cnki.eer.2022.02.005>
- 张义兵,孙俊梅,& 木塔里甫.(2018). 基于知识建构的同伴互评教学实践研究. *电化教育研究*, 39(7), 108–113. <https://doi.org/10.13811/j.cnki.eer.2018.07.016>
- 张涛,张思,高倩倩,& 王佳慧.(2022). 促进在线同伴互评中学习者批判性思维发展研究. *电化教育研究*, 43(6), 53–60. <https://doi.org/10.13811/j.cnki.eer.2022.06.008>
- Vista, A., Care, E., & Griffin, P. (2015). A new approach towards marking large-scale complex assessments: Developing a distributed marking system that uses an automatically scaffolding and rubric-targeted interface for guided peer-review. *Assessing Writing*, 24(4), 1–15. <https://doi.org/10.1016/j.asw.2014.11.001>
- Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, 49(4), 461–493. <https://doi.org/10.1017/S0261444816000161>