

# 同伴互评在大学英语口语教学中的作用

史天化

(福建工程学院 人文学院, 福建 福州 350118)

**摘要:** 采用定量分析与定性分析相结合的方法,探讨了同伴互评在普通高校英语专业口语教学中的可行性和有效性。结果表明同伴评分虽然也参照评分量表,但更多地依据评分员的主观感受进行评分。学生与教师的分数分布存在显著差异,教师评分的离散程度更大,而学生分数的同质程度更高。同伴评分反馈在一定程度上提升了学生对评分标准的理解,提高了与教师评分的拟合度。

**关键词:** 大学英语; 口语教学; 同伴互评; 同伴反馈

**中图分类号:** G642.475

**文献标志码:** A

**文章编号:** 1672-4348(2016)02-0193-06

## A study of the role of peer assessment employed in college oral English teaching and learning

Shi Tianhua

(School of Humanities, Fujian University of Technology, Fuzhou 350118, China)

**Abstract:** The feasibility and effectiveness of peer-assessment in oral English class in college English teaching is investigated via quantitative and qualitative methods. The results show that peer-assessment is more based on the rater's (individual's) experience rather than the rating scale. The assessing is significantly different between the teachers and the students. The assessing of the teachers is more discreet while that of the students is more homogeneous. Compared to the ranges of marks given by teachers, peer-assessing tends to overrate low performance and underrate high performance, thus displaying the higher central tendency. The feedback promotes to a certain extent the students' understanding of rating criteria and increases the fitting between peer- and teacher-assessment.

**Keywords:** college English; oral English teaching; peer-assessment; peer-feedback

近年来有关同伴评估的实证研究数量处于不断上升趋势。同伴评估将学生看作是有感情、有思想、独立的个体,通过分析、监控和评价同伴的学习过程与效果,充分调动学生的自主性与积极性,并最终对学习产生正面、持久性的影响。同伴评估被认为是非常有效的评估方式,不仅能帮助教师减轻工作负担,而且有助于培养学生的探索和创新精神,促进同伴间的合作学习和交流。因此,如何在普通高校英语口语教学中培养学生的

合作意识与创新能力,使学生意识到“同伴评估”的力量成为一个亟须解决的研究课题。

### 一、文献回顾

国外学者从各个角度阐释了同伴评估对课堂教学以及自主学习产生的影响。同伴互评能够有效促进学习,且教师评分与同伴互评具有统计意义的高度相关。<sup>[1]</sup>同伴互评反馈意见相比教师反馈更易被学生采纳;学生评语多为非命令式的,更

容易转化为持续、复杂的和微观层面上意义的修改,更便于操作。<sup>[2]</sup> Matsuno, S. 借助 Many-Facel Rasch Model 模型研究学生互评和教师评分之间的宽严度,发现学生互评时更为宽容;学生评分比教师评分的内部一致性高且偏差更小。<sup>[3]</sup> 国内自 20 世纪 80 年代开始也有一些学者对形成性评估在外语教学中的作用进行了一系列探索。实证研究自评与互评对非英语专业大学生写作自主能力的影响,并尝试性地提出了英语写作教学研究模型<sup>[4]</sup>,建立了一套形成性评估和总结性评估相结合的口语能力评估体系。<sup>[5]</sup> 目前,同伴评估的研究现状与其在教学中的重要性还不对等。先前研究较少涉及到学生的评价态度、自尊心和自信心等心理因素问题。同伴评价需要在友好和谐的氛围中进行,学生互评中如何相互信任,公正地对己和对人还需要更多的探索。其次,虽然同伴评估是形成性评估中的常用方法之一,但是教学环境以及实验设计的差异导致先前有关同伴互评的可行性和有效性研究呈现不同的结果。Oldfield 和 Macalpine 研究结果显示同伴评估与教师评价相关系数  $r$  值仅为 0.30,相关性甚微,可忽略不计。<sup>[6]</sup>

那么,同伴互评时对不同的评分标准的态度是否一致、同伴互评与教师评分的相关性及具体差异在哪里,本文将依据实证研究方法获得的数据,对以上问题进行探究。

## 二、研究设计

### (一) 研究对象

本研究从福建工程学院英语专业中随机选取两个自然班,进行 5 周的对比实验。授课时间为每周 2 学时,共计 17 周,选前 5 周数据进行分析。为了降低不同英语水平的干扰作用,用 2013 ~ 2014 学年第一学期期末成绩进行对比。从表 1 可以看出,两个班无显著性差异,符合实验要求。

### (二) 实验程序

教师把事先设计好的同伴互评表和反馈表发给每个人。学生根据指定话题进行口头演讲,每人三分钟。学生对同伴在小组讨论以及口头演讲的表现作出评价。侧重评估中的准备、组织、设计以及构思等具体过程。参照 Burnett 和 Cavaye<sup>[7]</sup>,制定包含两部分、九个细项的评分量表。教师 and 所有学生根据评估表对学生口语分项评分,并填

写反馈表。之后,分组讨论评分和反馈情况,并布置接下四周的口头演讲话题。控制组按完全相同的程序和方法进行,但是没有最后的小组讨论和反馈,口头演讲结束后直接布置新话题。

表 1 2013 ~ 2014 学年英语专业两个班第一学期期末成绩  
Tab.1 The means and variance of two classes majored in English at the first term during 2013 and 2014

班级	人数/人	均分/分	方差
控制组	23	76.640 0	6.755 18
实验组	25	75.760 0	6.989 75

## 三、结果与讨论

搜集五周所得的数据,研究者更感兴趣的是学生在评分过程中评分标准使用的具体情况,以了解评分过程中各个不同维度倾向以及容易产生的晕轮效应(评分者评分中对某些特征或维度的过度使用,不能很好地区分各个维度)。学生在互评中逐渐形成了一套包含固定的维度和尺度的评分图式,不同的学生在头脑中形成的评分图式具有鲜明的个体差异性,下面详细讨论同伴评分过程中评分标准使用的异同以及与教师评分的差异特征。

### (一) 教师与学生评分比较

在形成性评估过程中,学生的互评与教师的评价是紧密联系和动态影响的。课堂评估的任务由教师根据课程内容以及学生的学习现状设计。同伴提供促进学习的反馈,激发彼此作为相互的教学资源和评价成功的标准。学生互评也需要教师提供榜样,展示优秀的评分案例。因此,首先比较教师评分与同伴互评的一致性。

对比教师与学生在六个评分细项(其余三个涉及小组贡献的只用于学生评分)的使用情况,分析教师评分与同伴互评的差异。零假设为  $H_0: \mu_1 = \mu_2$ ,即控制组和实验组的教师评分与学生互评分数平均差异为零。表 2 中控制组在内容表达、准备和内容以及语言水平(口语)的  $T$  值均大于 2.70(显著性水平为 0.01 时的临界值为 2.704,  $P < 0.01$ ),因此拒绝无差别假设,也就是说,控制组中教师评分与同伴互评在上述三种情况下都存在显著差异。而实验组除了口语表达情况的  $T$  值为 4.59( $P < 0.01$ ),其余  $T$  值均小于 2.70,说明实验组的教师评分与同伴互评基本不

存在显著差异。两组唯一差别在于同伴评分后是否进行小组讨论与反馈,因此,小组讨论与评分反馈在评分过程中起作用,对同伴互评中的评分行为有显著影响。通过“反馈-理解-调整-反馈”

表 2 配对 T 检验  
Tab.2 Paired T-tests

		T(CG)	P(CG)	T(EG)	P(EG)
讨论	准备和过程	-1.52	NS	-1.21	NS
	内容表达	2.85	<0.01	1.37	NS
	语言水平	-1.78	NS	2.61	NS
口语	准备和内容	3.12	<0.01	1.87	NS
	表达情况	2.66	NS	4.59	<0.01
	语言水平	2.89	<0.01	2.19	NS

注:CG: 控制组; EG: 实验组。

的持续、循环过程,同伴反馈在一定程度上提升了学生对评分标准的理解,提高了学生与教师评分的拟合度以及评分的信度和效度。配对 T 检验的结果在讨论和口语两个层面没有表现出明显的差异。

(二)教师与同伴评分值域对比

为了解教师与同伴评分的异同,比较了教师与学生评分的最大值和最小值。表 3 中只有实验组讨论层面的“准备和过程”及“内容表达”的最小值小于教师评分的最小值,其余所有学生评分的最小值与最大值都位于教师的分数区间范围以内,也就是说学生评分最小值都大于教师最小值,而学生最大值都小于教师最大值。即教师评分较学生评分有更大的离散程度。

表 3 评分范围对比  
Tab.3 Maximum and minimum marks awarded for each of the assessment criteria

		教师评分(CG)		学生互评(CG)		教师评分(EG)		学生互评(EG)	
		最小值	最大值	最小值	最大值	最小值	最大值	最小值	最大值
讨论	准备和过程	4.00	9.00	5.00	9.00	5.00	9.00	4.00	8.00
	内容表达	4.00	9.00	4.00	10.00	4.00	8.00	3.00	7.00
	语言水平	3.00	8.00	4.00	8.00	2.00	9.00	4.00	8.00
口语	准备和内容	4.00	10.00	4.00	8.00	4.00	9.00	4.00	10.00
	表达情况	4.00	8.00	3.00	7.00	4.00	8.00	4.00	6.00
	语言水平	3.00	9.00	5.00	8.00	3.00	9.00	4.00	9.00

而后,对比教师和学生使用整体评分法的评分结果差异。教师和学生根据对一个口语片段整体特征的印象对其做出判断,而不是按先前的六个细项分别评分。其评分结果如下:

控制组:学生(60~86),教师(55~88)

实验组:学生(60~85),教师(45~90)

学生分数无论控制组还是实验组均位于教师评分区间范围之内,且学生最低分均为 60 分及合格标准线。也就是说学生互评的分数相比教师评分具有更高的同质程度。之后通过访谈了解到学生在同伴评分时,通常考虑“面子”首先预设了这一及格线,然后凭感觉很快地评出分数。

以上分项和整体评分法的结果都表明同伴互评在极值和总分两方面均位于教师分数区间范围内。由此,得出以下结论:学生与教师评分存在显著差异,教师评分的离散程度更大,而同伴互评的同质程度更高。

为了解同伴互评和教师评分的变量关系以及同伴评分中的评分倾向,进一步引入线性回归并建立回归模型。

表 4 给出的标准系数  $R$  为多元相关系数,是教师评分和同伴互评的皮尔逊相关系数的绝对值( $R$  取值范围为 0~1)。  $T-P$  相关系数分别为 0.773(控制组)和 0.884(实验组),表明学生在评价小组成员时与教师的评分高度相关(系数在 0.70~0.90 为高相关)。两者唯一差别在于是否进行小组讨论与反馈,因此笔者认为小组讨论和反馈提高了同伴互评与教师评分的相关性。 $R$  Square 是  $R$  的平方,当乘以 100% 时,理解为因变量的总变异性被自变量解释的百分比。其中实验组对应值为 0.781,表示在实验组中  $T-P$  能够 78.1% 地预测学生评分的总变异,即同伴评分是可靠的,与教师评分高度相关。实验组(78.1%)远高于控制组(59.7%),进一步证明学生对同伴

口语表达的评判能力通过小组讨论和评分反馈得到明显提升,且评分一致性提高了 18.4 个百分点。实验组与教师评分的相关系数明显高于控制组,并且只有实验组具备了较高的校标关联度(EG 的  $R^2$  大于 0.7)。此结果不同于 2002 年 Mrudula 的实验结论——即学生评分之后的讨论和评分反馈对评分结果没有显著影响,教师与学生评分的相关系数( $R$  分别为 0.50 和 0.49)均较低。<sup>[9]</sup>

表 4 线性回归模型  
Tab.4 Linear regression model

回归模型	非标准化系数		标准系数	$R$	$R$ Square	sig
	$B$	标准 误差	$R$			
$T-P$ (CG) 常量	1.888	0.328				
	0.448	0.087	0.773	0.597	0.000	
$T-P$ (EG) 常量	1.176	0.331				
	0.735	0.091	0.884	0.781	0.000	

注:  $T-P$ :教师-同伴; CG:控制组, EG:实验组。

回归系数表给出了建立回归方程需要的值。非标准化系数包括回归方程中的  $Y$  轴截距和斜率,以及标准误差(自变量不能预测因变量值的程度)。将表 4 数据代入方程  $Y = bX + a$ ,得到如下两个回归方程:

回归方程 1: $Y(SCG) = 0.448 X + 1.888$

回归方程 2: $Y(SEG) = 0.735 X + 1.176$

注:SCG: 学生控制组; SEG: 学生实验组;  $X$ : 教师评分。

最理想的系数为 1,截距为 0,说明该方程对样本数据点的拟合度完全一致。以上两组中的第二个方程,实验组  $Y$  轴截距为 1.176,斜率为 0.735,更趋于完美线性正相关( $a = 0, b = 1$ ),即 SEG 与教师评分的拟合度更高。上述方程清晰地展示了教师与学生的具体评分差异倾向:对英语口语能力弱的同伴评分时,学生分数总体高于教师分数;而针对英语口语水平较高的同伴时,学生分数则低于教师分数。以方程 2 为例, $Y = 1.176 + 0.735 X$ ,当  $X$  为 4.438 时,两者评分完全相同,该回归线与理想回归线相交;当  $X$  小于 4.438 时,同伴评分高于教师评分;当  $X$  大于 4.438 时,学生评分开始低于教师评分。

(三) 同伴互评时对不同的评分标准的态度

表 5 和表 6 用五级量表(1 = 非常同意,5 = 不

同意)反映了受试对问卷中关于态度问题的回答情况。总体而言,两表中的均值都比较集中,变化幅度较小,全距  $R$  分别为 0.60(2.49 ~ 3.09)和 0.45(2.32 ~ 2.77)。数据显示同伴评估时对评分标准持有不同态度。研究者关注的是九项评分细则,尤其是“语言水平”的排序情况。表中  $F$  值为平均数差别的显著性检验,因此  $F$  值越大, $P$  值则越小,结果就越有统计意义差异。表 5 中,语言水平维度分别排第 7 位(小组讨论)和第 9 位(口头演讲),也就是说这 2 个维度得分让学生心里纠结,感觉压力大。尤其是口头演讲的语言水平

表 5 评分态度之“轻松愉快”分析表  
Tab.5 Analysis of students' responses to the question:  
“I felt comfortable in assessing the performance of my peers on each of the assessment criteria.”  
(1 = strongly agree, 5 = strongly disagree)

评分标准	$M$	SD	$F$	Sig.
1. 组员贡献:想法和建议	2.49	0.81	3.53	0.05 *
2. 组员贡献:讨论过程	2.52	0.75	0.57	0.63
3. 口头演讲:准备和内容	2.54	0.77	0.62	0.58
4. 口头演讲:表达情况	2.54	0.72	0.54	0.65
5. 小组讨论:准备和过程	2.57	0.80	3.81	0.04 *
6. 小组讨论:内容表达	2.59	0.84	1.36	0.29
7. 小组讨论:语言水平	2.68	0.86	0.31	0.87
8. 组员贡献:口头演讲	2.71	0.77	1.08	0.33
9. 口头演讲:语言水平	3.09	0.61	2.34	0.12

表 6 评分态度之“公正负责”分析表  
Tab.6 Analysis of students' responses to the question:  
“I think I assessed my peers fairly and responsibly on each of the assessment criteria.” (1 = strongly agree, 5 = strongly disagree)

评分标准	$M$	SD	$F$	Sig.
1. 组员贡献:想法和建议	2.32	0.67	0.71	0.65
2. 口头演讲:准备和内容	2.35	0.58	1.23	0.31
3. 组员贡献:讨论过程	2.36	0.62	0.88	0.48
4. 小组讨论:准备和过程	2.42	0.62	3.51	0.05 *
5. 小组讨论:内容表达	2.50	0.73	2.78	0.19
6. 口头演讲:表达情况	2.51	0.71	3.82	0.04 *
7. 组员贡献:口头演讲	2.60	0.73	1.18	0.37
8. 小组讨论:语言水平	2.64	0.75	1.23	0.31
9. 口头演讲:语言水平	2.77	0.79	1.46	0.29

最为困难。表6中,同样两个语言水平标准分别排第8和第9,说明这两个标准评分时最不负责任、最不公正的。使用方差分析(ANOVA)了解以上九个变量中哪些具有显著性效应,结果只有四个评分标准是具有统计意义差别。其中三项都是表达情况,一个为想法和建议,没有一个涉及语言指标。也就是说,同伴评分过程中虽然也参照评分量表,但更多地是依据自己的主观感受和经验进行评分。

为了进一步获取学生评分过程中的大脑思维情况,了解差异产生的具体原因,研究者对评分过程的心理活动进行分析。评分反馈表的数据采用话语分析这种传统的定性方法,并使用对文本内容编码并通过归类的分析框架。当问及为什么会在评分过程中感到困难和压力时,24%的受试认为评分是有难度的。其中,提及最多为“避免伤害彼此感情”“不想让同伴尴尬”等涉及面子、情感的表达。此外,有近一半的受试认为他们不能准确地按照评分标准进行评分是因为自身语言水平不足以胜任评分工作。该结果与香港理工大学调查一年级学生对同伴评估的态度相似,学生对同伴评估持消极态度,不能对同伴语言水平进行准确评价。<sup>[9]</sup>反馈表中给出的典型理由如下:

No. 13:“我自身口语表达很差,他说得比我流利多了,不知道怎么评价同伴”。

No. 23:“如果对方口语比自己好,很难评分”。

No. 4:“如果能匿名评分会比较好,不要考虑对方“面子”,更容易按照评分标准给分,评分更准确”。

No. 1:“不太理解评分标准,有时会忽略它,按照整体印象反而更简单”。

No. 20:“评分中存在很多偶然因素,比如先前分数对后续评分的影响,不太可能完全按评分标准”。

总之,同伴评分时学生评分员内部一致性较差,倾向于依据自己的主观感受和经验,使用自己习惯的评分图式,比如,流利性和整体印象,而不是严格按照评分量表中的具体标准进行评分。结合访谈归纳同伴评分较小离散程度的原因如下:(1)同伴互评时对评分活动缺乏重视,评分不够严谨,有时不能严格遵照评分要求进行;(2)学生对评分标准的理解存在偏差,几乎所有的学生在

访谈中都提到会根据自身的主观感受和经验加入量表中没有的非语言特征进行评分,导致评分一致性较差;(3)小组内没有采取匿名评分,相互间的亲密程度及“面子问题”使评分存在难以预测的偏差;(4)学生对分数过于敏感和重视,更多地把评分理解为“对学习的评估(assessment of learning)”,而不是真正意义上的“为学习的评估(assessment for learning)”。

## 四、启示

基于本次实验设计,后续形成性口语课堂评估研究可以结合以下三点进行:首先,鉴于同伴互评中“轻松愉悦”和“公正负责”这两个维度在评分过程中最难实现,后续研究中有必要探究愉快、互助、和谐的课堂氛围是否有助于构建学生积极的评分态度,并导致更为客观准确的评分。其次,评分标准向来都是权威人士,如专家、学者或教师制定。作为评估的主体,应该让学生参与评估标准的制定和完善,不仅有利于激发学生的参与度和学习热情,而且有助于准确地理解和把握评分标准,提高同伴评估的效度和信度。最后,受试随机选自福建工程学院英语专业,48个受试中只有6个男生,女生占总人数的91.2%,性别严重失衡。未来研究中有必要充分考虑性别因素,分析男女生在同伴互评中是否有统计意义上的不同。

综上所述,语言水平是评估同伴时最难做到公正的一个维度。同伴评分过程中虽然也参照评分量表,但更多地是依据自己的主观感受进行评分。 $T$ 检验显示同伴评分反馈在评分过程中起作用,对同伴互评中的评分行为有显著影响。线性回归模型中,实验组增加的小组讨论和反馈环节使学生评分准确性提高了18.4个百分点,进一步验证了 $T$ 检验的结果。从极值和整体评分两方面比较教师与学生分数区间范围,发现分数分布均存在显著差异,教师评分的离散程度更大,而学生分数的同质程度更高。

课堂形成性评估的巨大潜在作用吸引了越来越多的学者和教师,但它也面临着巨大的挑战。当前课堂评估如何真正从“对学习的评估”转向“为学习的评估”,如何让学生真正愿意参与课堂评价,并通过参与评估激发学习热情,促进自主学习,提升课堂教学效果,任重而道远。

## 参考文献:

- [1] Pope N K L. The impact of stress in self- and peer assessment[J]. Studies in Higher Education,2005,30(1),51-63
- [2] Cho K, MacArthur C. Student revision with peer and expert reviewing[J]. Learning and Instruction,2010,98(4):891-901.
- [3] Matsuno S. Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms[J]. Language Testing,2009,26(1):75-100.
- [4] 曹荣平,张文霞,周燕. 形成性评估在中国大学非英语专业英语写作中的运用[J]. 外语教学,2004(5):82-87.
- [5] 陈美华,徐小燕. 大学英语口语能力形成性评估实证研究[J]. 东南大学学报,2008(3):119-123.
- [6] Oldfield K A, Macalpine M K. Peer and self-assessment at tertiary level: An experimental report[J]. Assessment & Evaluation in Higher Education,1995,20(1):125-132.
- [7] Burnett W, Cavaye G. Peer assessment by fifth year students of surgery[J]. Assessment in Higher Education,1980(5):73-78
- [8] Mrudula P. The influence of peer feedback on self-and peer-assessment of oral skills[J]. Language Testing,2002,19(2):109-131.
- [9] Winnie C, Martin W. Peer assessment of language proficiency[J]. Language Testing,2005,22(1):93-121.

(责任编辑:王明秀)

(上接第 187 页)

党组织通过微博、微信公众号等新媒体传播理想信念话语,必须将教育内容转化为明易的话语表述、平和的话语态度以及可亲的图文形象,超越学生党员对信息的无意识筛选后进入知觉层面,引起他们关注、了解和选择的兴趣,使学生党员在媒介阅读中得到满足,达到理想信念教育入脑入心的目的。

总之,理想信念从来都不是空洞的存在,必须回应当下的困惑,超越时下的迷茫。学生党员理想信念教育应全面融入当前的社会现实环境之中,在尊重和把握学生党员的群体行为特性的基础上,构建有效教育语境。只有正视理想信念问题,真诚帮助学生党员解决理想信念困惑,才能提升教育话语吸引力,赢得青年一代特别是学生党员们的真正认同。

## 参考文献:

- [1] 习近平. 紧紧围绕坚持和发展中国特色社会主义学习宣传贯彻党的十八大精神[N]. 人民日报,2012-11-19(1).
- [2] 吕绍明. 主体间性视角下的学生党员理想信念教育[J]. 丽水学院学报,2012(4):90-93.
- [3] 辛斌. 批评性语篇分析方法论[J]. 外国语,2002(6):34-36.
- [4] 王荣华. 浅议中国梦背景下的大学生理想信念教育[J]. 新课程,2013(9):180-181.
- [5] 中共中央办公厅. 中共中央办公厅印发《关于培育和践行社会主义核心价值观的意见》[N]. 人民日报,2013-12-24(1).

(责任编辑:许秀清)