

无领导小组讨论， 如何检验评委可信度

■ 郭朝晖

无领导小组讨论是目前应用较为广泛的人才测评技术。然而，在无领导小组讨论中评委可信度该如何检验，目前方法研究仍显匮乏。通常采用的 Kendall 和谐系数可以用来检验全体评委评价结果的一致性，但却无法反映单个评委的可信度。文章运用 Spearman 相关系数弥补以上不足，对无领导小组讨论过程中的评委可信度进行实证研究，得出结论与建议。

一、评委可信度检验的方法

评委可信度，是指多个评委给一组被评价者评分，所评分数之间的一致性程度。评委可信度主要用于揭示测评结果中个体主观误差的大小，评委可信度越高，说明测评结果就越可靠。对评委可信度进行检验，有利于规范评委的评价行为，改进人才测评工作，提高人才选拔质量。

目前，对评委可信度进行检验，主要运用统计学中的相关分析方法。如果无领导小组讨论安排 2 名评委，可以采用 Pearson 积差相关系数；如果是多个评委，则采用 Kendall 和谐系数来估计。Kendall 和谐系数最早由英国统计学家肯德尔提出，又称 W 系数，主要用来衡量多个评委的评价意见是否一致，它的使用条件是：(1) 多个评委同时评价多个对象；(2) 评价结果以等级记录（或以分数记录，转换成等级）。Kendall 和谐系数的数学表达式为：

$$W = \frac{\sum_{i=1}^n R_i^2 - \frac{(\sum_{i=1}^n R_i)^2}{n}}{\frac{1}{12} K^2 (N^3 - N)} \quad (\text{公式 1})$$

其中， R_i 为每一个被评价者的 K 个

等级之和； n 表示被评价者的人数， K 表示评委人数。 W 的取值范围为 $[0, 1]$ ，计算所得的 W 越大，全体评委的评价结果较一致，所得评价结果的精确度也越高；反之，就说明评委之间的分歧较大，所得评价结果的准确度也越低。

但是，Kendall 和谐系数只能反映全体评委评价结果的一致性程度，不能反映单个评委的评价结果的可靠性。而事实上，单个评委评价结果的可靠性往往会影响到全体评委评价结果的可靠性，所以还需要运用 Spearman 相关系数对单个评委评价结果的可信度进行分析。Spearman 相关系数在使用时，对数据条件的要求没有 Pearson 积差相关系数严格，不论评委和被评价者的总体分布形态、样本容量的大小如何，只要无领导小组讨论中评委对被评价者评分结果表示成等级（或者等级排序）的形式，都可以使用。其数学表达式为：

$$r = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n^3 - n} \quad (\text{公式 2})$$

其中， $D_i = x_i - y_i$ ， $x_i (i=1, 2, \dots, n)$ 表示单个评委对每个被评价者的打分结果的顺序值， $y_i (i=1, 2, \dots, n)$ 表示全体评委对每个被评价者的打分结果的顺序值， n 表示被评价者（评分结果）的数量。Spearman 相关系数 r 的取值范围为 $[-1, 1]$ ，其值越小，说明单个评委的可信度也就越差。如果 $r < 0$ 表示单个评委的评价结果与全体评委评价结果负相关， $r > 0$ 表示单个评委的评价结果与全体评委评价结

果正相关。一般情况下， $r > 0.7$ 表示单个评委评价结果与全体评委评价结果的相关性好，即单个评委与大多数评委评价的结果较一致。

二、实例分析

由于无领导小组讨论具有情景模拟性强、人际互动性好、能够贴近工作实际等特点，因此在各类组织的人才选拔中应用越来越广。最近，武汉某大型国有企业面向社会公开招聘人力资源总监 1 名，经过前期的履历分析、笔试和面试，从 32 位符合任职资格条件的应聘者中筛选出 7 位（分别用 A、B、C、…、F、G 表示）进入无领导小组讨论环节，由无领导小组讨论决定最终拟录用人选。为慎重起见，公司在无领导小组讨论之前进行了周密的准备，制订了详细的无领导小组讨论测评方案，设计了评分标准，并对无领导小组讨论的程序进行了统一规定。最后由公司高层管理人员和外聘管理专家 9 人组成测评委员会，进行现场打分。评分结果如表 1 所示。

按照平均分排名，应聘者 B 最终被录取。然而仔细分析数据，可以发现各位评委给应聘者的打分排序不尽相同。评委 2 给出的最高分是应聘者 C，评委 5 和评委 8 给出的最高分是应聘者 A，评委 6 给出的最高分是应聘者 D。也就是说，九位评委在谁是最佳人选的看法上存在分歧，其中五位认为是 B，两位认为是 A，一位认为是 C，一位认为是 D。那么，这种评分内部所存在的不一致性，是否超出了统计

表 1 无领导小组讨论评分结果

被评价者 评价者	应聘者 A	应聘者 B	应聘者 C	应聘者 D	应聘者 E	应聘者 F	应聘者 G
评委 1	91	97	95	86	89	81	84
评委 2	90	92	96	95	88	85	83
评委 3	92	95	89	90	87	82	85
评委 4	90	96	88	93	85	83	81
评委 5	94	92	89	90	84	86	80
评委 6	90	86	88	92	80	82	84
评委 7	88	95	90	84	81	78	75
评委 8	94	88	89	82	90	84	81
评委 9	90	96	88	83	85	80	76
平均分	91	93	90	88	85	82	81

表 2 Spearman 相关系数计算结果

被评价者 评价者	应聘者 A	应聘者 B	应聘者 C	应聘者 D	应聘者 E	应聘者 F	应聘者 G	相关系数 r
评委 1	3	1	2	5	4	7	6	0.89
评委 2	4	3	1	2	5	6	7	0.71
评委 3	2	1	4	3	5	7	6	0.93
评委 4	3	1	4	2	5	6	7	0.89
评委 5	1	2	4	3	6	5	7	0.89
评委 6	2	4	3	1	7	6	5	0.54
评委 7	3	1	2	4	5	6	7	0.96
评委 8	1	4	3	6	2	5	7	0.57
评委 9	2	1	3	5	4	6	7	0.96
总评排序	2	1	3	4	5	6	7	

原则可以接受的限度范围？各位评委的评分结果是否可靠？从表 1 的数据我们无法进行直观的判断，必须通过计算和检验评委的可信度才能确定。

将各位应聘者的得分按照高低顺序进行排名，计算出排名的顺序差 D_i 后，已知应聘者 $n=7$ ，根据公式 2，很容易计算出九位评委的 Spearman 相关系数 r ，见表 2。

无领导小组讨论的评分主要依靠评委的主观判断，不同的评委其观察角度、注意力和对被测评者的偏见等都会使评价的结果不一致，评委之间的差异也就成了误

差的来源之一。通常情况下，大多数评委评价所得的结果往往接近真实结果。按照这一原则，从表 2 可知，评委 6 和评委 8 的 Spearman 相关系数分别只有 0.54 和 0.57，说明这两位评委的评价结果与大多数评委评价结果的相关性较低，可信度较差。

此时，根据公式 1 计算出全体评委的 Kendall 和谐系数 $W_1=0.55$ 。如果舍弃评委 6 和评委 8 的评价结果，重新计算的 Kendall 和谐系数 $W_2=0.82$ ，结果明显增大。这说明，舍弃与大多数评委评价结果相关性较差的极少数单个评委的评价结果，可

以大大提高总评结果的可靠性。这种检测方法简便易行，可操作性强，能够对每一个评委以及全体评委的可信度进行评估，还可以避免因少数评委评价结果不一致对总评结果可信度所产生的不良影响。

三、结论与建议

通过结合实例的分析，我们得出四点结论：一是采用无领导小组讨论选拔人才，进行评委可信度检验是有必要的；二是 Kendall 和谐系数与 Spearman 相关系数相结合，可以弥补单个评委信度无法检验的不足；三是评价结果存在一定分歧是正常的，关键是分歧必须控制在统计原则可以接受的限度之内；四是尽管评委可信度检验揭示的是相关关系，而不是因果关系，但对于评估人才选拔是否可靠仍然具有十分重要的参考价值。

为了进一步提高无领导小组讨论中评委的可信度，笔者建议：

第一，严格遴选评委。目前，我国许多企业运用无领导小组讨论进行人才选拔时，尚未建立起一整套关于评委遴选原则、程序、条件及方法的管理制度，对评委的选择过于随意，要么唯职务、要么唯职称，使得运用无领导小组讨论方法“选对人”的概率，远远低于预期。由于无领导小组讨论的操作程序较为复杂，很大程度上依赖于评委对候选人的分析与判断，最好由经过专业训练、有测评经验的人士担任评委。评委个人需要具备良好的职业道德和强烈的责任心，为人处世公道正派，具有较强的观察能力、具有较强的分析概括能力、判断能力与言语表达能力，文化程度较高。尤其值得注意的是，遴选评委时，相对于评委的职务、学历、年龄等出身背景而言，评委参加各种人才测评的实际经验显得更为重要。

第二，控制评委的自由裁量权。在无领导小组讨论中，由于评委打分不可能做到完全客观化，其主观性的存在使有些评

把握奖励性绩效工资激励倾斜力度

■ 罗双平

如何把握激励倾斜力度,是事业单位绩效工资实施中一个难点。

事业单位绩效工资包括两个部分:一部分是基础性绩效工资,按月发放;一部分是奖励性绩效工资,各单位自主发放。奖励性绩效工资激励重点是向关键岗位倾斜,向业务骨干倾斜,向做出突出业绩的人员倾斜。这就需要在设计奖励性绩效工资方案时,要明确重点激励的岗位,重点激励的人员,重点激励的项目,体现多劳多得,优绩优酬,体现岗位价值,不让干活的人吃亏,不让有贡献的人吃亏。但重点激励的突出,分配差距拉大,如果把握不好,可能会引起部分职工心理不平衡。要把握好激励倾斜力度这个难题,需要对“关键岗位”、“业务骨干”和“突出业绩”

做出科学的分析和界定。

一、如何理解向关键岗位激励倾斜

什么是关键岗位,有的人理解为关键岗位就是一线上的部门,一线上的岗位。向关键岗位倾斜,就是向一线部门倾斜,向一线上的岗位倾斜。这种理解不全面。按照企业的界定,通常一线是指企业的生产制造部门、服务部门、经营部门、技术开发部门;二线是指技术支持部门,如工厂的机器修理维护部门,它们为生产经营提供技术服务,创造生产条件;三线是指后勤保障部门、综合管理部门,如生产管理科、组织人事科、行政后勤等管理部门。按照事业单位的界定,通常分为一线和二线。一线是指一个单位的业务部门,二

线是指一个单位的管理部门(职能部门)。从企业和事业单位的界定可以看出,将一个组织划分为一线、二线或三线,是根据工作职能划分的,并不是根据工作重要程度划分的。不能简单地,一线重要,二线不重要,也不能认为二线重要,一线不重要。因为,任何一个单位的工作,都是一个整体,缺少哪一部分,工作都会出问题。那么,如何理解和把握向关键岗位倾斜呢?笔者认为有以下几个方面需要引起思考。

一是对关键岗位实施激励倾斜时,要注意一线、二线各部门的激励平衡,不能因为部门职能的不同,或由于岗位所在部门的区别,出现大幅度的待遇差别。这不利于职工的团结,不利于工作积极性的调

委利用自己的权力打出“人情分”、“关系分”,扰乱了人才测评的秩序,使得人才选拔的整体质量受到影响,而且也易引起一些无谓的纷争和不正之风,损害组织形象、败坏组织声誉。因此,在不打击评委会工作积极性的前提下,要对评委会的裁量权进行必要的控制,主要做法是:1.评分与统分相分离,评委会负责严格按照评价指标与标准评分,另外安排工作人员总分,评委会评定结果不得无故调整和改动;2.及时对各位评委打分结果进行统计分析,对于偏差较大的评委,要进一步分析原因,涉及徇私舞弊的要坚决予以弃用;3.坚持多位评委独立打分、集体评价的测评方式,

防止因评委数量过少而导致人事决策上的独断专行。

第三,评委用来打分的评价指标数量不宜过多。评价指标是衡量人才素质高低的标尺,也是各位评委用来打分的依据。评价指标是否合理关系到整个测评的质量,指标的数量太少,对人才素质的反映往往就不够全面,数量太多,又会增加测评过程中评委的认知负担。现有研究已经发现,评委用来打分的指标数量会影响到评委评分结果的可靠性。当使用3个或6个评分指标时,各位评委打分的内部一致性都很高,而评分指标是9个的时候,评委的评分内部一致性就明显地下降。所以,

使用无领导小组讨论这种人才测评技术进行人才选拔时,建议评分指标以6个左右为宜。

第四,进行必要的数据处理。对评分结果进行必要的数据处理有助于提高无领导小组讨论结果的可信度。最为简单有效的做法就是去掉两极端分数求平均。这种方法在体育比赛中得到了广泛使用,但在无领导小组讨论以及人才选拔评分中使用较少。采用这种数据处理的方式,能够最大限度上避免评委由于各种主观因素偏离客观事实甚至背弃客观事实的行为发生。^[5]

(作者单位:武汉科技大学管理学院)