

如何科学地构造教育评价方案中的权集合

林丰勋

权集合是教育评价方案中所有可以相加的权重系数集。它反映的是人们对教育活动中各个教育因素与达到教育目标程度之间的关系、重要性程度以及相对价值的认识。权集合构造的是否合理,直接影响着教育评价的质量,进而影响着教育的质量。纵观我国目前的基础教育评价,特别是各学校自己内部搞得各种评价,之所以整体水平不高,原因当然很多,但其中一个主要原因就是在制定评价方案时,不能科学地构造权集合。因此,本文拟对如何科学地构造教育评价方案中的权集合作一探讨,以期对广大教育评价工作者有所裨益。

1 权集合的主要构造方法

权集合的构造过程,是人们对各项指标间联系形式的认识过程,也是人们对各项指标的重要性和相对价值的认识过程,更是统一人们对这些问题的认识过程。那么,采用什么样的方法才能使人们对于这些问题的认识趋于统一呢?根据已有研究,主要可采用以下几种方法:

1.1 特尔斐法

特尔斐法是美国兰德公司赫尔默于本世纪50年代创造并首先在现代未来研究中被广泛地用作直观预测,之后又引入教育评价中用来设计指标体系和权集合构造的一种方法。其基本做法是:将初拟的评价指标系统或拟提请咨询的问题,以分发问题的形式,征求、汇集并统计所请咨询专家们的意见,以使专家对各指标价值的判断达成一致。具体步骤为:

第一步,选择熟悉被评教育客体并有代表性的专家至少数十人作为咨询人。选出的专家最好是与被评客体没有直接利害关系。

第二步,咨询工作组织将指标系统的权重咨询表用函调的方式与专家联系。咨询表中问题的表述要明确清晰,以避免专家作出模棱两可的理解。在整个咨询过程中,专家彼此互不见面,完全以“背靠背”的形式进行,以尽可能减少权威、资历、口才及人数优势等无关因素的负面影响。

第三步,第一轮咨询表回收后,做以下两项统计处理:

一是求出第*i*项指标在第一轮咨询中专家所给权重的平均数 $\bar{X}_{wi}$,

$$\bar{X}_{wi} = \frac{1}{n} \sum_{j=1}^n w_{ij} \quad (1)$$

式中 w_{ij} 为第*j*项专家对第*i*项指标权重系数的给出值。

二是求出每一位专家的给出值与 $\bar{X}_{wi}$ 的偏差 Δ_{ij} ,

$$\Delta_{ij} = w_{ij} - \bar{X}_{wi} \quad (2)$$

再将 Δ_{ij} 反馈给各位专家,请他们对照 Δ_{ij} 作出新的判断。经过几轮反复咨询后,一般说来 Δ_{ij} 会逐渐趋于零,即大家的意见趋于一致。这时,评价方案设计人可将专家咨询的最后意见进行归一化处理,以得到各项指标的权重系数。

由于特尔斐法是一个根据大多数咨询人的认识去统一所有参加咨询人员价值认识的过程,该过程的收敛性保证了参加咨询人员的价值认识能够趋于一致且不受各种干扰,因此,它是目前较受大家欢迎

的一种方法。

1.2 层次分析法

层次分析法是采用多目标、多准则的两两比较的方法对一个指标系统列出其各项指标的优先顺序进而确定权重系数的一种方法。它是由美国学者萨蒂(T. L. Saaty)首先引入教育评价领域中的。层次分析法的具体操作步骤为:

第一步,列出待求权重系数的子指标系统。

第二步,对该系统中各项指标按相对重要等级(见表1)进行两两比较。

第三步,将两两比较的结果写成 n 阶倒数矩阵(见图1)。该矩阵中 x_{ij} 表示第 i 项指标对第 j 项指标的相对重要性等级, $x_{ij} = \frac{1}{x_{ji}}$;矩阵的对角线元素均为1,即第 i 项指标自身相比同等重要; n 阶表示指标有 n 项。

表1 相对重要性等级表

相对重要程度	定 义	说 明
1	同等重要	两者对目标的贡献相同
3	略为重要	根据经验,一个比另一个评价稍有利
5	重要	根据经验,一个比另一个评价更为有利
7	非常重要	一个比另一个评价更为有利,且其优势已在实践中证实
9	绝对重要	重要的程度为最高
2,4,6,8	两个相邻程度的中间值 需要折衷时采用	

$$\begin{matrix}
 & \begin{matrix} a & b & \cdots & i & \cdots & j & \cdots & n \end{matrix} \\
 \begin{matrix} a \\ b \\ \vdots \\ i \\ \vdots \\ j \\ \vdots \\ n \end{matrix} & \begin{pmatrix} 1 & x_{ab} & \cdots & x_{ai} & \cdots & x_{aj} & \cdots & x_{an} \\ x_{ba} & 1 & \cdots & x_{bi} & \cdots & x_{bj} & \cdots & x_{bn} \\ \cdots & \cdots & 1 & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{ia} & x_{ib} & \cdots & 1 & \cdots & x_{ij} & \cdots & x_{in} \\ \cdots & \cdots & \cdots & \cdots & 1 & \cdots & \cdots & \cdots \\ x_{ja} & x_{jb} & \cdots & x_{ji} & \cdots & 1 & \cdots & x_{jn} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & \cdots \\ x_{na} & x_{nb} & \cdots & x_{ni} & \cdots & x_{nj} & \cdots & 1 \end{pmatrix}
 \end{matrix}$$

图1 n 阶倒数矩阵

第四步,对图1所示的 n 阶倒数矩阵进行相容性检验。检验方法是先用公式3求出最大特征值 λ_{max} ,并根据公式4计算出一致指标 $C.I.$,然后再求 $C.I.$ 与同阶矩阵的随机指标 $R.I.$ (见表2)的比值 $C.R.$ 。若 $C.R. \leq 0.10$,表示该 n 阶倒数矩阵作出的权重系数是合理且相容的。若 $C.R. > 0.10$,则说明作出的权重系数不够合理或不相容,应当重新修正。

$$\begin{vmatrix} (1-\lambda) & x_{ab} & \cdots & x_{an} \\ x_{ba} & (1-\lambda) & \cdots & x_{bn} \\ \vdots & & \ddots & \vdots \\ \vdots & & & \vdots \\ x_{na} & \cdots & \cdots & (1-\lambda) \end{vmatrix} = 0 \quad (3)$$

$$C.I. = \frac{\lambda_{max} - n}{n - 1} \quad (4)$$

表2 R.I. 值

n	2	3	4	5	6	7	8
R.I.	0	0.58	0.90	1.12	1.24	1.32	1.41

第五步,按公式5求出各项指标的权重系数 w_i ,

$$w_i = \frac{1}{n} \sum_{i=1}^n (x_{ij} / \sum_{i=1}^n x_{ij}) \quad (5)$$

第六步,按照各级指标权重系数间的关系通过归一化处理求出指标系统中全部指标的权重系数。

层次分析法虽较特尔斐法麻烦一些,但由于其科学性更强一些,因而应用也较为广泛。

1.3 已定指标两两比较法

该方法是通过一定数量(通常20~40人)的专家对已经研究的某一子指标系统中的指标的重要性程度进行两两比较打分,进而确定权重。打分时可采用(0,1)打分法,或(0,1,2,3,4)打分法。若专家认为作比较的两项指标A的重要性大于B,则记A为1,B为0;若作比较的两项指标C、D的重要性分不出高低,则都记为0.5;各指标不作自我比较。最后将每一项指标的所有专家给的得分相加求和,再除以所有专家给出的所有指标的累计总分,即可得到每一项指标的权重系数。如表3所示。

表3 N位专家两两比较汇总表

$P_2 \backslash P_1$	A	B	C	D	行值相加	归一化	W
A		28	22	24	74	74/180	0.41
B	6		25	10	41	41/180	0.23
C	8	7		12	27	27/180	0.15
D	9	15	14		38	38/180	0.21
合计					180	1.00	1.00

采用(0,1,2,3,4)打分法时,0为不重要,1为稍重要,2为较重要,3为显著重要,4为极显著重要;同样,若两项指标的重要性分不出高低时各取0.5。两种打分法相比,(0,1)打分法比较粗糙但简便;(0,1,2,3,4)打分法求得的每位专家所给的权重系数 w 值往往很分散。

1.4 Q分类法

Q分类法是一种简单易行的决定指标优先顺序的方法。其做法是调查人把一套写有指标名称的卡片交给被调查专家,然后让其按照指标的重要性排序。排序的具体步骤为:

第一步,将卡片权重高低分为“高权重”和“低权重”两部分,两部分的卡片数不必相等。

第二步,从两部分中选出“中权重”卡片,形成第三部分。

第三步,从“高权重”卡片中选出“最高权重”卡片,从“低权重”卡片中选出“最低权重”卡片;这样指标就被分成最高权重、高权重、中权重、低权重和最低权重五部分(见图2)。一般说来到此为止就可以了,如有需要仍可继续进行。

第四步,检查分类结果,进行局部调整直至满意。

第五步,根据不同类卡片给出相应指标的权重系数,同类指标的权重系统相同。

Q分类法的缺点是准确性不高,但在指标个数很多时常采用之。

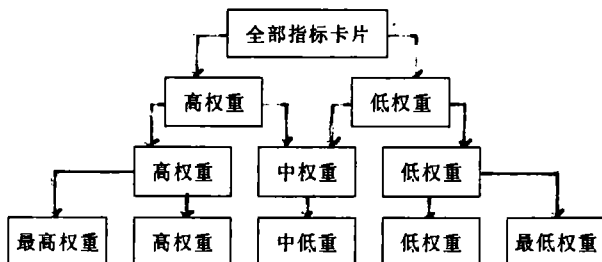


图2 Q分类示意图

权集合的构造方法除上述外,还有许多,例如关键要素归一定量法、专家协商确定法、已定指标权重环比定量法及指定指标权重对比定量法等。但是,最常用的还是上述四种,特别是前两种。

2 权集合品质的检验

无论采用何种方法构造出的权集合,都需对其科学性和合理性进行检验。检验方法通常有两种:相关法和符号秩次检验法。

2.1 相关法

相关法是建立在当参加评价的专家人数达到一定数量之后,专家在不分项目的综合评价中,其评定值的统计平均值能达到令人满意的程度假设之上的。设某一评价指标系统有 K 条指标,我们请一些专家对 N 所学校($N \geq 30$)或 N 个被评对象按照给定的指标系统进行评定,然后根据权集合进行计算,便可得到一组关于这些学校或被评对象的综合评定值,记为 $x_1, x_2, \dots, x_N$;与此同时,再请另一组专家对这 N 所学校或 N 个被评对象进行不分指标的综合评定,得到关于这些学校的综合评定值,记为 $Y_1, Y_2, \dots, Y_N$ 。这样每一所学校或被评对象都可得到两个综合评定值(X_i, Y_i)。根据皮尔逊相关公式(公式6)求相关。

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{N \cdot \sigma_X \cdot \sigma_Y} \quad (6)$$

如果求得的 r 值显著,则表明权集合构造合理,反之,则说明权集合不合理。但值得注意的是,用相关系数来衡量权集合的品质,其结果不能告诉我们哪一个权重不合理,它只告诉我们在指标与相应的权集合中确实存在着不合理、不科学的成份。因此,一旦 r 值不显著,便需对权重集合的部分或全部加以重新考虑。

2.2 符号秩次检验法

符号秩次检验法是由 *F. Wilcoxon* 提出的。利用该法检验权集合合理性的具体做法是请两组专家,其中一组按照给定的指标系统进行评定,另一组不按指标系统进行综合评定,这样每个被评对象都可得到两个评定值。将这两组相应的评定值求差数,然后按照差数绝对值的大小排序,并给每一个差数确定秩次,最后再给秩次添上原来的正负号。如果权集合构造合理,则正秩与负秩之和应当相等或接近相等。如果正秩与负秩相差较大,则说明权集合构造不合理,需重新加以考虑。

利用符号秩次检验法检验权集合品质分大小样本两种情况。在小样本($N < 30$)情况下,可首先计算正秩与负秩之和,记为 T_+ 和 T_- ,取 $T = \min(T_+, T_-)$,然后查符号秩次检验表。如果 T 大于表中临界值,说明权集合构造合理;如果 T 小于临界值,说明权集合存在问题,需重新考虑。在大样本($N \geq 30$)情况下,由于 T 的分布接近正态分布,故可以近似地采用 Z 检验。其步骤是:先求出 T 值的均值 $\bar{X}_T$ 和标准差 σ_T 。(式中 n 为不等于零的差值个数。)

$$\bar{X}_T = \frac{n(n+1)}{4}, \quad (7)$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}, \quad (8)$$

然后进行 Z 检验。

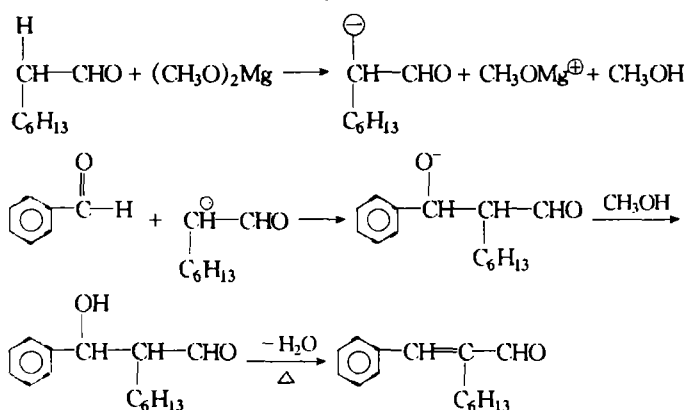
$$Z = \frac{|T - \bar{X}_T|}{\sigma_T}. \quad (9)$$

如果 $Z < 1.96$, 说明权集合构造合理; 如果 $Z \geq 1.96$, 说明权集合构造不合理, 则需重新加以考虑。

(上接第 8 页)

3.1 机理探讨

甲醇镁作为一个碱性试剂首先夺取辛醛中的 α -H, 形成的碳负离子(I)进攻苯甲酸羰基, 生成羟醛缩合中间体(II), II 经脱水后生成 α, β -不饱和醛, 即产物。



3.2 本法特点

与传统方法相比, 本法操作简单, 不耗用其它有机溶剂或萃取剂, 催化剂用量少而且易得。

参考文献

CA. 114: P61692w Payne Lawrence Sidney Ep392579(Cl. Co7c45/74) 17 oct 1990