

试论教育评价中权集合的构造

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。权集合的构造是否科学、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,之所以整体水平不高,其中一个重要原因就是教育评价工作者在制定评价方案时,不能科学地构建出合理的权集合。因此,本文拟对如何科学地构造权集合作一探讨,以期对广大教育评价工作者有所裨益。

一、教育评价中权集合的构造

教育评价中权集合的构造方法很多,如专家协商法、关键因素归一定量法、Q分类法、已定指标两两比较法、已定指标权重环比定量法、指定指标权重对比定量法等等。本文主要介绍两种科学性较强而又实用的特尔斐法和层次分析法。

(一)特尔斐法(Delphi technique)

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。这一方法的基本做法是:将初拟的评价指标系统或拟提请咨询的问题,以分发问题表的形式,征求、汇集并统计参与咨询专家们的意见,然后通过多次反馈,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。如有可能,选出的专家最好是与被评客体没有直接利害关系。

第二步,咨询工作组织者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、人数优势等因素的负面影响。

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*个专家对第*i*项指标权重系数的给出值。

二是求出第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 Δ_{ij} 作出新的判断。经过几轮反复咨询后,一般说来 Δ_{ij} 会渐趋于零,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而不受各种干扰,加之该法操作简便,因而,它是目前较受人们欢迎的一种方法。

(二)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。采用

该法构造权集合的步骤为:

第一步,列出待求权重系数的子指标系统。

第二步,对列出的指标系统中的各项指标按照相对重要性等级(见表1)进行两两比较。

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等
3	略为重要	根据经验,一个比另一个评价稍有利
5	基本重要	根据经验,一个比另一个评价更为有利
7	确实重要	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	可以断言重要的程度为最高
2,4,6,8	两个相邻程度的中间值	需要折衷时采用

第三步,将两两比较的结果写成 n 阶倒数矩阵(见图1)。

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

图1 n 阶倒数矩阵

上述矩阵中 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。

第四步,对上述 n 阶倒数矩阵进行相容性检验。检验方法是先用下列行列式(公式3)求出最大特征值 $\lambda_{\max}$, 并按公式4计算出一致性指标 $C.I.$, 然后再求 $C.I.$ 与同阶矩阵的随机指标 $R.I.$ (见表2)的比率 $C.R.$ 。 $C.R.$ 又叫一致性比率。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 $C.R. > 0.10$, 说明作出的权重系数不够合理或不相容, 应当重新修正。

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

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

表2 $R.I.$ 表

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

$$W_i = \frac{1}{n} \sum_{j=1}^n (X_{ij} / \sum_{i=1}^n X_{ij}) \quad (5)$$

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。层次分析法虽然较特尔斐法复杂一些,但其科学性更强一些,因而也是一种常用的构造权集合的方法。

二、权集合品质的检验

无论采用何种方法构造出的权集合,都需要对其科学性和合理性进行检验。检验方法通常有两种。一是相关法,二是符号秩次检验法。

(一)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,评定值的统计平均值能达到令人满意的程度的假设之上的。设某一评价指标系统有 K 条指标,我们请一些专家对 $N(N \geq 30)$ 个被评单位进行不分指标的综合评定,得到关于这些被评单位的综合评定值,记为 $X_1, X_2, \dots, X_n$ 。与此同时,请另一组专家按照给定的指标系统进行评定,然后根据权集合进行计算,便又可得到一组关于这些被评单位的综合评定值,记为 $Y_1, Y_2, \dots, Y_n$ 。这样,每一个被评单位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。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{N \cdot \sigma_X \cdot \sigma_Y} \quad (6)$$

将求出的相关查积差相关显著性临界值表检验其显著性程度。如果指标系统的权集合构造合理,则 r 应显著。换言之,若 r 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存在着不合理、不科学的因素。因此,一旦 r 不显著,便需要对权集合的部分甚至全部进行重新考虑。

(二)符号秩次检验法

符号秩次检验法是由 F. Wilcoxon 提出的。其具体做法是请两组专家,一组不按指标系统进行整体综合评定,另一组按照给定的指标系统和权集合进行评定,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,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。反之,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,若 T 大于表中临界值,说明权集合构造合理,若 T 小于临界值,说明权集合存在问题,需重新构造。

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。其步骤为:

先求出 T 值的均值 $\bar{T}$ 和标准差 σ_T 。

$$\bar{T} = \frac{n(n+1)}{4} \quad (7)$$

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

式中 n 为不等于零的差值个数。

然后进行 Z 检验。

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

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