

# 基于信息熵理论的职业院校 诊改目标适切性测度模型构建研究

杨应崧<sup>1</sup>,王成方<sup>2</sup>,林娟<sup>2</sup>

(1.全国职业院校教学工作诊断与改进专家委员会,上海 200050;

2.浙江金融职业学院,浙江 杭州 310018)

**摘要:**职业院校在诊改工作中,质量生成主体事先设定的目标是其进行自我诊断的坐标,也是产生改进动力的源头,因此,实施诊改必须先有目标。受到内外部诸多因素影响,诊改目标不可避免存在适切性问题,且目标适切性的高低直接决定着诊改的实际成效和持久活力。以信息熵理论为基础,通过计算实践水平与预定目标水平差距的信息熵和相对熵,构建目标适切性测度模型,对目标适切性进行验证性的精准测度,为目标修正和后续(下一阶段)目标的制定提供量化参数和可靠依据,从而进一步激发质量生成主体解决问题、消除偏差的动力,在目标导向和问题导向的良性互动中促进学校办学水平的持续提升。

**关键词:**教学工作诊断与改进;目标适切性;信息熵;职业教育;高职院校

中图分类号:G710 文献标识码:A 文章编号:1004-9290(2018)0036-0030-05

以诊断与改进(以下简称“诊改”)为主要特征的内部质量保证体系建设已成为高职院校提升内部治理水平、保证人才培养工作质量的重要抓手。目标是诊改工作的起点,只有目标明确,才能制定清晰的标准,才有后续的设计、组织与实施,诊改工作才能正常运行。目标也是诊改工作的动力源泉,诊断通过现状和目标的比较来发现问题和偏差,从而产生解决问题、消除偏差的愿望,再转化为学习、创新、改进的动力。没有目标,就没有比较,诊改工作就成无源之水<sup>①</sup>。目标也是教育行政部门对职业

院校诊改工作开展复核的依据,鉴于此,制定发展方向正确、适应实际需求、发展张力适度、切合学校实际,能够切实激发学校内在动力和创新活力的适切的诊改目标,是每所职业院校科学有序开展诊改工作首要且必须解决的关键问题。

## 一、核心概念界定

### (一)基于最近发展区理论的诊改目标

20世纪30年代,苏联教育家Vygotsky提出了最近发展区概念,创造性地阐述了教学、学习与发展之间的辩证关系。Vygotsky认为,学生的发展有

收稿日期:2018-10-17

基金项目:浙江省哲学社会科学规划课题“公平与效率权衡:浙江省高等教育资源投入合理性综合测度研究”(项目编号:17NDJC338YBM,主持人:林娟)

**作者简介:**杨应崧(1944—),男,上海市金山区人,全国职业院校教学工作诊断与改进专家委员会主任,教授,主要研究方向为职业教育诊断与改进;王成方(1957—),男,浙江杭州人,浙江金融职业学院研究员,主要研究方向为职业教育诊断与改进;林娟(1979—),女,浙江平阳人,浙江金融职业学院副研究员,主要研究方向为高职教育资源投入、高职教育质量管理。

两种水平：一种是学生实际发展水平，指独立活动时所能达到的解决问题的水平；另一种是学生潜在发展水平，是通过成人的指导或与能力较强的同伴合作所获得的潜力。两者之间的差距就是最近发展区<sup>[2]</sup>。教育的目标应当着眼于学生的最近发展区，预测学生学习潜能，优化教学，从而促进学生的知识获得与发展。

最近发展区理论不仅为优化教学提供了理论基础，也为诊改目标制定工作提供了理论指导。根据最近发展区理论，学校的发展可有两种预期水平，一种是实际发展水平，指持续现有发展能力所能达到的水平，另一种是潜在发展水平，需要通过学校激发活力、努力创新，提高发展能力所能达到的更高的发展水平。二者的差距就是最近发展区。诊改目标应当着眼于潜在发展水平，以此调动学校内涵建设工作各环节中的部门和个体的积极性和创造性，激发他们的发展潜力，通过努力地“跳一跳”来接近或达到潜在发展水平，从而实现更高水平的发展目标。

## (二) 诊改目标的适切性

适切的诊改目标是目标制定主体瞄准经济社会发展和学生全面发展对学校办学的要求，基于办学环境、教学资源、现实状态和发展能力等的综合

准确预测。它能较好地指导学校办学工作，激发学校发展潜力，实践效果往往也较理想，实际达成水平和预定目标相近或相符。但在目标制定工作中，受到主体认知局限等因素的影响，诊改目标不可避免会出现不同程度的适切性不足问题，导致实际达成水平和预定目标存在某些差距。更甚者，目标出现或无视挑战机遇，无须努力奋斗，没有发展张力，或严重脱离实际，根本不可能实现，犹如画饼充饥等适切性严重不足的问题，不仅起不了目标应有的作用，实际达成水平与预定目标差距过大，而且会带来消磨意志、涣散人心的副作用。

## 二、基于信息熵理论的诊改目标适切性测度模型的构建

### (一) 信息熵理论

熵(Entropy)的概念来源于热力学。1850年，蒸汽机工程师Clausius发现热力学第二定律，提出用熵来判断一个过程的不可逆程度。物理学家 Boltzmann 把熵和系统的无序状态联系起来，微观

上，系统越无序，宏观上它就越趋于平衡。1948年，信息论创始人Shannon在他的著名论文《通信的数学原理》(A Mathematic Theory of Communication)中借鉴热力学引入熵的思路，提出信息熵的概念，用以度量信息系统的平均不确定。

定义1(信息熵)若事件A是由n个变量 $a_1, a_2, \dots, a_n$ 组成，第i个变量 $a_i$ 发生的概率为 $P(a_i)(i=1, 2, \dots, n)$ ，则 $I(a_i)=-\log_2 P(a_i)$ 是变量 $a_i$ 发生所含有的信息量，称为变量 $a_i$ 的自信息量。对于由n个变量组成的事件A，其不确定性与 $a_i$ 的不确定性相关，因此，确定事件A所需要平均信息量，也就是信息熵为：

$$H(A) = -\sum_{i=1}^n P(a_i) \log_2 P(a_i) \quad (1)$$

其中， $0 < P(a_i) < 1$ ，且 $\sum_{i=1}^n P(a_i) = 1$ 。当 $P(a_i) = 0$ 时，定义 $0 \log_2 0 \triangleq 0$ 。对数底数为2时，信息熵单位为比特(bit)。

定义2(条件熵)若事件A、C，以 $P(a_i|C_k)$ 记事件C出现结果 $C_k(k=1, 2, \dots, m)$ 条件下事件A出现结果 $a_i$ 的概率，则事件C出现结果 $C_k$ 条件下，事件A的熵为：

$$H_{C_k}(A) = -\sum_{i=1}^n P(a_i|C_k) \log_2 P(a_i|C_k) \quad (2)$$

则事件C实现的条件下事件A的条件熵为：  

$$H_C(A) = -\sum_{k=1}^m P(C_k) \sum_{i=1}^n P(a_i|C_k) \log_2 P(a_i|C_k) \quad (3)$$

### (二) 基于信息熵的诊改目标适切性测度模型

为了便于组织实施、衡量考核，诊改目标一般都由一系列量化的子目标或具体指标(标准或关键结果)组成，这为基于信息熵的诊改目标适切性测度提供了可能性。根据诊改目标适切性的内涵，对诊改目标适切与否和适切程度的判断和测度可通过目标预测值与实际值之间的关系来判断。如实际值接近或等于目标预测值，目标预测值被偏离率较小或未被偏离，说明目标被较好地实现，其适切性较好；如果实际值与目标预测值差距较大，说明目标实现效果不理想，存在较为严重的适切性不足问题。一般而言，每个子目标或具体指标都可能存在偏离问题，因此，可将诊改目标偏离事件视为每个子目标或具体指标偏离率分布的概率系统，具备归一性特点，就可利用信息熵对诊改目标偏离事件所包含的平均信息量进行度量，由此，诊改目标适切性测度就转换为诊改目标偏离事件信息熵的计算。

设诊改目标由  $n$  项内涵建设要素指标目标构成,  $x'_i (i=1,2,\dots,n)$  为第  $i$  项指标的目标预测值,  $x_i$  为第  $i$  项指标的实际值,  $a_i$  为第  $i$  项内涵建设要素指标实际值与目标预测值偏离率。根据最近发展区理论, 可设定目标预测值  $x'_i$  的最近发展区比例值  $b_i$ ,  $b_i$  为非负数, 其赋值区间为  $[0,1)$ 。当内涵建设要素指标实际值  $x_i$  在  $[x'_i, (1+b_i)x'_i]$  区间内, 令其偏离率  $a_i=0$ , 说明指标目标预测值适切性较好; 当实际值  $x_i > (1+b_i)x'_i$  或  $x_i < x'_i$ , 其偏离率  $a_i \neq 0$ , 其中,

$$a_i = \frac{|x_i - (1+b_i)x'_i|}{(1+b_i)x'_i} \quad (3)$$

时, 令  $a_i = \frac{|x_i - x'_i|}{x'_i}$ , 说明目标预测值适切性有不足, 存在设立过低或过高的问题, 造成实际值与目标预测值的偏离。由于诊改目标是由  $n$  项内涵建设要素指标目标构成, 相应地就存在  $n$  个偏离率  $a_i$ , 将偏离率  $a_i$  做归一化处理, 可得到:

$$P(a_i) = \frac{a_i}{\sum_{i=1}^n a_i} \quad (4)$$

其中,  $0 \leq P(a_i) \leq 1$ , 且  $\sum_{i=1}^n P(a_i) = 1$ 。可把  $P(a_i)$  视为随机变量偏离率  $a_i$  分布的概率, 则诊改目标偏离事件  $A$  的信息熵为:

$$H(A) = -\sum_{i=1}^n P(a_i) \log_2 P(a_i) \quad (5)$$

信息熵  $H(A)$  描述了由  $n$  项内涵建设要素指标目标构成的诊改目标偏离事件的平均信息量, 也就是不确定程度。一般而言, 诊改目标偏离事件信息熵值较大, 不确定性也就较大, 其适切性不足, 确定诊改目标所需的信息量就较多; 反之, 信息熵值较小, 不确定性就较小, 适切性较好, 确定诊改目标所需的信息量就较少。

若各个内涵建设要素指标目标偏离率  $a_i$  为等概率均匀分布, 即  $P(a_i) = \frac{1}{n}$ , 由 (5) 式计算可得诊改目标偏离事件的信息熵  $H(A) = \log_2 n$ , 不确定性达到最大值<sup>[4]</sup>, 说明诊改目标中各内涵建设要素指标目标发生等概率偏离, 诊改目标的适切性明显不足, 所有的指标均需要大量的信息来解决适切性不足的问题。

### (三) 基于条件信息熵的诊改目标适切性测度模型

基于条件信息熵的诊改目标适切性测度模型

是对基于信息熵测度模型的改进。诊改目标是职业院校在教育行政主管部门对其办学规范要求基础上对人才培养质量、内涵建设要素的综合或专项规划和预测, 因此, 诊改目标中的各项内涵建设要素指标大体可分为规范性指标和发展性指标两大类。其中, 规范性指标是指教育行政主管部门设定的办学基本要求指标, 其指标目标预测值  $y_j (j=1,2,\dots,m)$  以及应遵循规范化管理原则、环境建设原则, 不得低于教育行政主管部门规定值  $y''_j$ 。如果第  $j$  项规范性指标实际值  $y_j$  在规定值  $y''_j$  允许范围内, 该项指标被认为达标或实现; 反之, 则被认为不达标。发展性指标是学校结合办学实际制定的体现学校发展特色和水平的指标, 其目标预测值  $x'_i (i=1,2,3,\dots,n)$  设定应遵循适切性原则。

#### 1. 规范性指标不达标条件下的诊改目标适切性测度

由于规范性指标是职业院校办学的底线, 当规范性指标实际值  $Y = \{y_j\}$  中任一项规范性指标实际值  $y_j$  不达标, 遵循一票否决原则,  $Y$  不达标, 记为  $c_1$ , 已没有必要做诊改目标适切性测度。

#### 2. 规范性指标全部达标条件下的诊改目标适切性测度

当规范性指标实际值  $Y = \{y_j\}$  中任一项规范性指标实际值  $y_j$  达标,  $Y$  达标, 记为  $c_2$ , 则可进行诊改目标适切性测度。此时, 诊改目标偏离事件信息熵的运算就转化为规范性指标全部达标条件下的发展性指标目标偏离事件的信息熵运算。

由于规范性指标达标与否并不会影响发展性指标目标值是否被偏离, 偏离多少, 因此规范性指标达标与发展性指标目标值偏离这两个事件是相互独立事件, 故有规范性指标全部达标条件下的发展性指标目标预测值偏离率分布概率  $P(a_i|c_2) = P(a_i)$ 。则诊改目标偏离事件信息熵为:

$$H_{c_2}(A) = -\sum_{i=1}^n P(a_i|c_2) \log_2 P(a_i|c_2) = -\sum_{i=1}^n P(a_i) \log_2 P(a_i) \quad (6)$$

同理, 诊改目标偏离事件信息熵较大, 不确定性就较大, 发展性指标目标适切性不足, 确定发展性指标目标所需的信息量就较多; 反之, 诊改目标偏离事件信息熵较小, 不确定性就较小, 发展性指标目标适切性较好。



当各发展性指标目标偏离率 $a_i$ 为等概率均匀分布,即 $P(a_i)=\frac{1}{n}$ ,则诊改目标偏离事件信息熵, $H_c(A)=\log_2 n$ 不确定性达到最大值,说明诊改目标中各发展性指标目标发生等概率偏离,目标的适切性明显不足,确定各发展性指标目标均需要大量的信息。

### 三、诊改目标适切性测度模型的个案应用研究

#### (一) 个案职业学院诊改目标适切性的测度

学校发展五年规划是职业学院对内涵建设各项工作在未来五年要达到的境地或标准的规定和预测,属于学校诊改目标链中的总目标,关系到其他子目标的设定。表1为个案职业学院“十二五”规划中目标预测值与实际值存在差异的内涵建设要素指标。根据表1可见,五项指标目标预测值和实际值均存在较大差距,其中鉴定、考证和培训收入到款额度实际值要大于预测值,其余四个指标的实际值均小于预测值。

根据社会经济发展和学生全面发展对学校办学的要求,基于办学环境、教学资源、现实状态和发展能力等的综合考量,及达到激发学校办学潜力的目的,设定5项内涵建设指标的最近发展区预测比例值 $i$ 分别为: $b_1=0.08, b_2=0.50, b_3=0.08, b_4=0, b_5=0.17$ 。

根据表2可见,鉴定、考证和培训收入到款额度实际值仍高于最近发展区最高预测值,存在目标预测值设置过低的问题,其余四项指标因实际值低于(最低)预测值,存在目标预测值设置过高的问题,因此,五项指标目标均存在偏离问题。根据偏离率计算公式,五项指标目标的偏离率值分别为: $a_1=0.05, a_2=0.11, a_3=0.15, a_4=0.15, a_5=0.33$ 。

其中,在校生规模指标偏离率最低(0.05),省级教学名师指标偏离率最高(0.33)。根据公式(4)对五个指标目标的偏离率 $a_i$ 做归一化处理,得出相应的偏离率分布概率 $P(a_i)$ 分别为: $P(a_1)=0.06, P(a_2)=0.14, P(a_3)=0.19, P(a_4)=0.19, P(a_5)=0.42$ 。

根据公式(5)最后计算出5项指标规划目标偏离事件的信息熵为2.09bit。由于学校“十二五”规划的内涵建设其他要素指标实际值均已达到预测值或落在最近发展区预测值内,所以目标偏离率均为零,对信息熵没有“贡献”,因此,该“十二五”规划目标偏离事件的信息熵为2.09bit。

(二) 个案职业学院提高后续诊改目标适切性的做法

“十二五”规划目标适切性的测度为个案职业学院制定“十三五”规划提供了量化参数和可靠依据。“十三五”规划目标的适切性成为学校高度关

表1 “十二五”规划中内涵建设要素指标情况

序号	内涵建设要素指标内涵	内涵建设要素指标名称	目标预测值	实际值
1	在校生规模(人)	$a_1$	6 500~7 000	6 159
2	鉴定、考证和培训收入到款额度(万元)	$a_2$	30~40	50
3	专业设置数(个)	$a_3$	26~28	22
4	专任教师数(位)	$a_4$	230	196
5	省级教学名师(位)	$a_5$	6	4

表2 “十二五”规划中部分内涵建设要素指标信息熵

序号	内涵建设要素指标	预期最低值	最近发展区预测值	最近发展区预测比例 $b_i$	实际值	偏离率 $a_i$	偏离率分布概率 $P(a_i)$	信息熵(bit)
1	在校生规模(人)	6 500	7 000	0.08	6159	0.05	0.06	0.24
2	鉴定、考证和培训收入到款额度(万元)	30	45	0.50	50	0.11	0.14	0.40
3	专业设置数(个)	26	28	0.08	22	0.15	0.19	0.46
4	专任教师数(位)	230	230	0.00	196	0.15	0.19	0.46
5	省级教学名师(位)	6	7	0.17	4	0.33	0.42	0.53
	合计	——	——	——	——	0.79	1	2.09

注:目标预测值设定为区间范围的,可将区间左边值视为目标预测最低值,区间右边值视为最近发展区值,也可设置更高的最近发展区值,但要满足比例要求。

注的问题，尤其是这五项指标目标偏离的问题。学校相关部门对这些指标目标预测水平与实际水平发生偏离的原因开展了大量的调查、分析与反思，发现如在校内规模目标被偏离主要是个别专业招生计划未完成或报到率过低导致在校内实际规模大大低于预测规模，进而学校发现这些专业设置或招生规模并不适应地方产业发展或学生发展的需求，需要学校对这些专业设置或招生规模进行相应的调整。因此，在制定“十三五”规划在校内规模目标的时候，学校在以往预测工作基础上，充分考虑这些专业的调整及对所有专业在未来五年适应社会经济和学生发展需求的预测。同时，通过定期观测各专业毕业生就业率、社会认可度、学生满意度、招生计划完成率和报到率等指标建立在校内规模变化预警机制，动态掌握“十三五”期间各专业在校内规模变化，及时对目标进行修正。

#### 四、总结

本研究以信息熵理论为基础，构建了基于信息熵和条件信息熵的诊改目标适切性测度模型，并应用于个案职业学院开展诊改目标的“轨迹分析”，通过精准判断学校“十二五”规划目标适切性，为“十三五”规划目标的制定提供了参数和依据，同时也激发学校解决问题、消除偏差的动力，建立内涵建设要素发展预警机制以提高规划目标的适切性，在

目标导向和问题导向的良性互动中促进了学校办学水平的提升，取得了较好的实践效果。

高职院校人才培养工作状态大数据的不断生成和更新，为基于信息熵理论的诊改目标适切性测度模型在内涵建设和内部质量保证体系建设中的应用提供了广阔的数据基础。基于人才培养工作状态大数据，该模型不仅能对学校、专业、课程、教师和学生层面的诊改目标进行适切性测度，实现“轨迹分析”，也能应用于学校、专业、课程、教师和学生五个层面诊改目标的“张力分析”和“比较分析”，帮助诊改主体把握“发展张力”的适当区间，发现发展的瓶颈。同时，也将成为教育行政部门对诊改开展复核的有力工具，最终促成诊改工作水平与成效的提升。

#### 参考文献：

- [1] 杨应崧. 关注职业院校教学诊改工作：打造“两链”，找准诊改的起点[N]. 中国教育报，2017-9-26 (11).
- [2] 麻彦坤. 最近发展区理论在当代美国教学模式中的应用[J]. 外国教育研究，2005，(9)：60-64.
- [3] 李贤平. 概率论基础[M]. 北京：高等教育出版社，2010：233-234.
- [4] 田振清，周越. 信息熵基本性质的研究[J]. 内蒙古师范大学学报自然科学（汉文）版，2002，(4)：348-351.