

# 高校试卷质量评价的同异综合评判方法

郭瑞林 季保平

(安阳工学院 生物与食品工程学院, 安阳 455000)

摘要: 提出了高校试卷质量评价的同异综合评判方法,并以安阳工学院生物与食品工程学院农学专业 2005-2006 年度第二学期六门课程的考试成绩为实例,介绍了这种方法的应用.在此基础上,讨论了这种方法的实用价值及其应用范围。

关键词: 试卷;质量;同异综合评判;方法

中图分类号: G640

文献标识码: A

文章编号: 1673-2928(2007)01-0151-04

高校学生各科考试成绩是衡量学生学业水平和教学管理部门进行决策的重要依据。因此,准确把握测试水平,提高试卷质量就成为教学过程中一个不容忽视的问题。通常,人们多以各科学生成绩平均分作为教师教学效果的一个定量指标。由于试卷难易程度不一,致使不同学科之间的教学效果可比性差。为此,钱漫如等提出以试卷难易适合度、区分度、效度和信度等四个指标作为评价试卷质量的依据<sup>[1]</sup>。本文在此基础上,运用集对分析原理<sup>[2]</sup>,提出了试卷质量评价的一种新方法-同异综合评判法,供高校教学管理部门决策参考。

## 1 高校试卷质量评价的四个指标

多年教学研究和实践表明,以试卷的难易适合度、区分度、效度和信度四个指标衡量试卷质量的高低具有一定的可靠性。

令全部考生为  $n$  人,将考生的考试成绩由高到低排列,分成高分组和低分组,两组的人数各取考生总人数的 15%,即高分组、低分组分别为  $N$  人,  $N = \text{INT}(0.15n)$ 。

### 1.1 试卷的难易适合度 $P$

该指标反映试卷对考生的难易适合程度。具体计算公式为:

$$P = \frac{1}{100} \sum_{i=1}^k Q_i P_i \quad (1)$$

其中,  $Q_i$  为第  $i$  题的满分值,  $P_i$  为第  $i$  题的难易适合度,

$$P_i = \begin{cases} E_i & \text{当 } E_i < 0.5 \\ 1 - E_i & \text{当 } E_i \geq 0.5 \end{cases} \quad (i=1,2,\dots,k) \quad (2)$$

$$E_i = 1 - (H_i + L_i) / 2Q_i \quad (i=1,2,\dots,k) \quad (3)$$

式中,  $H_i$  为高分组第  $i$  题的平均分,  $L_i$  为低分组第  $i$  题的平均分,  $E_i$  为试题的难度,  $E_i$  越大,试题的难度越大,一般认为,  $E_i > 0.7$ , 试题偏难;  $0.3 \leq E_i \leq 0.7$ , 试题难度适中;  $E_i < 0.3$ , 试题偏易。显然,  $P_i$  的数值在  $[0, 0.5]$  之间,且  $P_i \geq 0.3$ , 试题难度合适;  $P_i < 0.3$ , 试题难度不合适。

### 1.2 试卷的区分度 $D$

该指标反映试卷区分优、差等生的程度,体现试卷对考生真实水平高低的鉴别能力。令第  $i$  题的区分度

\* 收稿日期: 2006-11-22

作者简介: 郭瑞林 (1960-), 河南林州人, 安阳工学院生物与食品工程学院研究员。研究方向: 小麦育种与教育教学研究。

$$D_i = (H_i - L_i) / Q_i \quad (i=1, 2, \dots, k) \quad (4)$$

显然  $D_i$  的数值在  $[-1, 1]$  之间, 若  $D_i > 0$ , 则其数值越大, 试题越好, 一般认为,  $D_i > 0.6$ , 试题区分度高度显著;  $0.3 < D_i < 0.6$ , 试题区分度基本显著;  $D_i < 0.3$ , 试题区分度不显著。若  $D_i < 0$ , 则该试题比区分度不显著的试题更差, 起着负作用, 该试题应予以取消。

由  $D_i$  可得试卷的整体区分度

$$D = \frac{1}{100} \sum_{i=1}^k Q_i D_i \quad (5)$$

### 1.3 试卷的效度

试题的效度反映的是试题与试卷总体的一致性或者相关性。试卷的整体效度则是各试题效度的加权平均。一般, 第  $i$  题的效度用  $T_i$  表示:

$$T_i = \frac{\sum_{j=1}^n (y_{ji} - y_i)(z_j - z)}{\sqrt{\sum_{j=1}^n (y_{ji} - y_i)^2 \sum_{j=1}^n (z_j - z)^2}} \quad (i=1, 2, \dots, k) \quad (6)$$

其中,  $y_{ji}$  为第  $j$  人第  $i$  题的得分,  $y_i$  为第  $i$  题的平均分,  $z_j$  为第  $j$  人的试卷总成绩,  $z$  为  $n$  位考生试卷总成绩的平均分。则试卷的整体效度为

$$T = \frac{1}{100} \sum_{i=1}^k Q_i T_i \quad (7)$$

### 1.4 试卷的信度

试卷的信度反映的是整个试卷与考生水平的接近程度, 一般用  $C$  表示:

表示:

$$C = 1 - \left[ \frac{\sum_{j=1}^n (z_j - z)^2}{nS^2} \right] \quad (8)$$

式中

$$S = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (z_j - z)^2} \quad (9)$$

若  $0.5 < C < 1$ , 试卷高度可信;  $0.35 < C < 0.5$ , 试卷基本可信;  $0 < C < 0.34$ , 试卷一般可信;  $C < 0$ , 试卷不可信。

## 2 试卷质量的同异综合评判方法与步骤

设有  $p$  类不同学科的试卷,  $m$  个试卷质量评价指标。则试卷质量的同异综合评判方法步骤如下:

1) 据式 (9), 计算各试卷质量评价指标值, 构成集合  $A$ 。

2) 确定各评价指标的理想值  $\alpha_k$ 。一般取各指标的最佳值为理想值。各指标的理想值便构成理想指标集合  $B$ 。

3) 计算待评价试卷中各指标  $a_{ik}$  与理想指标集中各对应指标值  $\alpha_k$  的同一度, 构成待评价指标与各理想指标的同一度矩阵  $R$ 。

$$a_k = a_{ik} / \alpha_k \quad (10)$$

与同一度相对的另一概念即为差异度, 令其为  $b_{ik}$ 。由式 (10), 不难看出,  $0 < a_k < 1$ , 因此, 同一度与差异度的关系为:

$$b_{ik} = 1 - a_k \quad (11)$$

4) 在试卷质量综合评价过程中, 各个性状有主次之分, 不能等量齐观。因此, 从专业的角度可以得到各性状的权重矩阵

$$W = (w_1, w_2, \dots, w_m) \quad (12)$$

5) 据此, 可求得待评价试卷与理想性状集的联系矩阵  $U$ 。

$$U=P \times W^T \quad (13)$$

U 中的元素  $A_l$  即第 l 类试卷与理想性状集的综合同一度。

$$A_l = \sum_{k=1}^m a_{lk} w_k \quad (l=1,2,\dots,n; k=1,2,\dots,m) \quad (14)$$

由此得到综合联系度

$$\mu_{\text{综}}(w) = A_l + B_l i \quad (15)$$

式中, 综合差异度  $B_l = 1 - A_l$ 。  $A_l$ 、  $B_l$  两者的大小差别反映了所论 2 个集合的同异联系程度。一般地,  $A_l$  越大, 表明第 l 类试卷接近理想试卷的程度越高, 反之亦然。规定 i 在区间[- 1, 1]内视不同情况不确定取值, 它是综合同一度与差异度联系的纽带和桥梁; 当 i 取正值时, 差异度  $B_l$  便朝着同一度的方向转化, 当 i 取负值时, 同一度  $A_l$  便朝着差异度的方向转化。一般情况下规定其取值 - 1, 以表示  $A_l$  与  $B_l$  是相反的两个方面。

6) 计算集合 A 与 B 的联系势。在联系度  $\mu_{\text{综}}(w)$  中, 当  $B_l = 0$  时, 定义同一度  $A_l$  与差异度  $B_l$  的比值为所论集对 H 中同异关系下的联系势为

$$S(H) = A_l / B_l, \text{ 当 } B_l = 0 \text{ 时} \quad (16)$$

则其联系同势为

$$S(H)_+ = A_l / B_l, \quad A_l / B_l > 1 \quad (17)$$

表明 2 个集合相同的趋势占优势。其联系异势则为

$$S(H)_- = A_l / B_l, \quad A_l / B_l < 1 \quad (18)$$

表明 2 个集合相异的趋势占优势。

7) 据黄金分割原理, 划分联系势。由黄金分割原理, 可将联系势进一步划分为强同势、同势、弱同势、弱异势、异势 5 级(表 1)。

表 1 同异关系下的联系势等级

等级划分	联系势	$A_l$ 、 $B_l$ 及其关系
一级	强同势	$A_l > B_l, 0.854 \leq A_l \leq 1$
二级	同势	$A_l > B_l, 0.618 \leq A_l < 0.854$
三级	弱同势	$A_l > B_l, 0.528 \leq A_l < 0.618$
四级	弱异势	$A_l < B_l$ 或 $A_l \geq B_l, 0.382 \leq A_l < 0.528$
五级	异势	$A_l < B_l, A_l < 0.382$

8) 据表 1 进行联系势测验。测验原则为: 若待评价试卷的联系势处于同一势级, 则表明不同试卷之间差异不明显; 若待评价试卷的联系势处于相同势态(如同势或异势)中的不同势级, 则表明不同试卷间差异明显; 若待评价试卷的联系势处于不同势态中, 则表明不同试卷间差异极明显。

9) 对待评价试卷作出评价。规定评语集合  $E = \{ \text{优秀, 良好, 较好, 一般, 较差} \}$ 。按照黄金分割原理, 将综合同一度中介于 0 与 1 之间的值分割为相应于五种类型的评语值域, 即  $0.854 \leq A_l \leq 1$  为优良级;  $0.618 \leq A_l < 0.854$  为良好级;  $0.528 \leq A_l < 0.618$  为较好级;  $0.382 \leq A_l < 0.528$  为一般级;  $A_l < 0.382$  为较差级。并据上述评语值域和联系势, 对各个品种作出优劣评价。

### 3 应用实例

以安阳工学院生物与食品工程学院农学普招班 2005- 2006 年度 A、B、C、D、E 和 F 六门课程的考试成绩(略)为例, 说明这种方法的应用。

据公式 1) - (9) 计算出各试卷的指标值如表 2 所示。由表 2 可知, 各试卷难易适合度、区分度、效度和可信度各有优劣, 有的此指标好, 有的彼指标好, 数据杂乱无章, 很难作出评定。为此, 采用同异分析方法对其进行综合评判。

据各指标性质, 并征求教育专家意见, 可得其理想值及权重如表 3 所示。据公式 10) - (18), 得到各试卷的同异分析结果如表 4。

由此可分析各试卷质量, 在此不予赘述。

表 2 各试卷质量评价指标 \*

试卷	A	B	C	D	E	F
平均分	69.5	76.9	84.7	83.1	83.6	76.7
难易适合度	0.3075	0.2938	0.5375	0.1938	0.2200	0.2584
区分度	0.2100	0.4725	0.3350	0.3325	0.3750	0.3822
效度	0.6747	0.7370	0.6139	0.6946	0.6692	0.7265
信度	0.6726	-0.2403	0.0833	0.2375	0.1128	0.6516

\*A、B、C、D、E、F 分别表示不同课程试卷

表 3 试卷各指标理想值

试卷质量指标	试卷难易适合度	试卷区分度	试卷效度	试卷信度
理想值	0.5375	0.4725	0.7370	0.6726
权重	0.35	0.25	0.2	0.2

表 4 各试卷质量同异分析结果

考试科目	A	B	C	D	E	F
难易适合度	0.5721	0.5466	1	0.3606	0.4093	0.4807
区分度	0.4444	1	0.7090	0.7037	0.7937	0.8089
效度	0.9155	1	0.8330	0.9425	0.9080	0.9858
信度	1	-0.3573	0.1238	0.3531	0.1677	0.9688
综合同一度	0.6944	0.5699	0.7186	0.5613	0.5568	0.7614
联系势值	2.2651	1.3250	2.5537	1.2795	1.2563	3.1911
联系势	同势	弱同势	同势	弱同势	弱同势	同势
评语	良好	较好	良好	较好	较好	良好

#### 4 讨论

1) 试卷质量的高低不仅反映教师对所授课程的掌握程度,而且体现教师对学生学习能力的认知水平。因此,在教与学的过程中,教师既要加强对所授课程本身的全面了解,还要加强同学生之间的交流与沟通,使考试命题更具代表性和针对性,从而在教学互动过程中有效地提高教学效果。毫无疑问,开展试卷质量评价方法的研究,对于达此目的十分必要。

2) 试卷质量同异综合评判方法的提出,克服了以往教学效果评价过程中仅以各科成绩平均分数为依据的局限性,对于促进高校教学改革,提高高校教师命题能力,提升高校教学管理水平具有重要意义。

3) 试卷质量同异综合评判方法,不仅适用于高校,而且适用于中、小学校。十分明显,这种方法的应用,对于提高中、小学校教学质量和水平同样具有一定的借鉴作用。

#### 参考文献:

- [1] 钱漫如,蓝 荣,赵恩昌,等.试卷质量分析的可拓评判方法[C].蔡 文,从物元分析到可拓学[A],北京:科学技术文献出版社,1995: 343- 346.
- [2] 郭瑞林.同异分析的联系势测验及其在小麦品种区域试验中的应用[J].麦类作物学报,2004,24(1):63- 65.

## The Method of Similarity- difference Comprehensive Evaluation on Test Paper Quality in College

GUO Rui- lin

(Biological and Food Engineering Faculty, Anyang Institute of Technology, Anyang 455000, China)

**Abstract:** The paper advanced the method of similarity- difference comprehensive evaluation on test paper in college, and introduced the application of this kind of method with the actual example of agricultural major in Biological and Food Engineering Faculty, Anyang Institute of Technology. Based on this, the practical value and its application were discussed.

**Key Words:** Test paper; Quality; Similarity- difference comprehensive evaluation; Method