

# 会计教学中试卷质量的定量评估

杨 绮\*

(厦门大学会计系 厦门大学会计发展研究中心 361005)

**【摘要】** 在会计教学中,试卷质量的定量评估对提高教学质量具有重要意义。本文以试卷的难度、区分度、信度和效度作为定量评估试卷质量的指标体系,分别给出数学计算方法和评价标准,并在此基础上,以《会计学原理》课程结业考试的试卷为实例,应用该定量评估指标对此试卷的质量进行分析与教学评议。

**【关键词】** 会计教学 试卷质量 定量评估 指标体系

## 一、会计教学中试卷质量定量评估之必要性

古今中外都以考试作为选拔和使用人才的重要途径。当今学校教学管理部门把考试成绩作为鉴别学生的学习水平及确定学生升级与否的依据。而教师则把考试作为检查教学效果的主要方式,通过考试可为教与学双方提供教学过程的反馈信息,使教师了解教学上的薄弱环节,使学生发现自己的学习缺陷。可以说,考试的重要性在任何学科的教育中均是不言自明的,在会计教学中亦是如此。但是,通过一次考试,能在多大程度上达到其既定目的,则完全取决于试卷的质量,即取决于试卷是否科学、准确,考试结果是否真实、可靠。要回答这些问题,单凭经验是不够科学的。毋庸置疑,在会计教学中,完全有必要对试卷的质量进行科学的定量评估,从而为筛选试题和改进试题,提高试卷的科学性提供可靠依据,为发现和改进教与学过程中的缺陷提供准确信息。

## 二、会计教学中试卷质量之定量评估指标体系

科学的、量化的指标体系是试卷评估工作的核心和实施的基础,我们设立难度、区分度、信度和效度4个项目,构成试卷质量的定量评估指标体系。其中,前两个指标所针对的是每个试题,而后两个指标则针对整份试卷。之所以将前两个指标列入试卷质量定量评估的指标体系,是因为试题是构成试卷的基本单位,每个试题的质量将直接关系到整份试卷的质量,对每个试题进行定量统计分析,必能为我们了解整份试卷的质量和日后选择或修改试题提供依据。

### (一) 难度

难度是用以描述每个试题的难易程度的指标,可用第*i*题学生得分的平均值 $\bar{X}_i$ 与该题满分 $Q_i$ 的比值表示之。该比值越大,说明难度越小,而比值越小,则难度越大。为了与一般习惯概念相一致,可改用下式计算第*i*题的难度 $H_i$

$$H_i = 1 - \bar{X}_i / Q_i \quad (1)$$

一般而言,  $0.3 < H_i \leq 0.7$ , 难度适中;  $H_i > 0.7$  为难题;  $H_i \leq 0.3$  为易题。

### (二) 区分度

区分度是描述试卷对学生能力大小的区分程度。区分度合适的试卷能把学生的不同能力水平按考试成绩区分开来。学习好、水平高的学生能得高分,学习差、水平低的学生只能得低分,从

\*杨绮(1976.12—),女,福建莆田人,厦门大学会计系讲师,厦门大学会计系博士生。

而反映出学生学习成绩的真实情况。一般可用两个高低分极端组中试题分数平均值的差异程度作为区分度的指标。首先把学生的试卷按总成绩的高低依次排列,然后取出27%高分数的试卷作为高分组,再取出27%低分数的试卷作为低分组。若以 $\overline{X}_i^{(H)}$ 和 $\overline{X}_i^{(L)}$ 分别表示高分组和低分组中第*i*题的平均分数,以 $Q_i$ 表示该题的满分值,则第*i*题的区分度 $D_i$ 为

$$D_i = \left( \overline{X}_i^{(H)} - \overline{X}_i^{(L)} \right) / Q_i \quad (2)$$

显然难度太低的题目,几乎所有学生都能正确解答;反之,难度太高的题目,可能所有学生都不能解答。这两种情况均造成高、低分组中该题平均分数之差值 $\left( \overline{X}_i^{(H)} - \overline{X}_i^{(L)} \right)$ 很小,因而 $D_i$ 值很小,区分度很低,不能区分出学习成绩好、中、差的学生。只有难度适中的题目,高分组的平均分数 $\overline{X}_i^{(H)}$ 高,低分组的平均分数 $\overline{X}_i^{(L)}$ 低,因而 $D_i$ 值大,区分度好。

一般 $D_i \geq 0.4$ 的试题,区分度好; $0.3 \leq D_i < 0.4$ 的试题,区分度良好; $0.2 \leq D_i < 0.3$ 的试题尚好,但须修改; $D_i < 0.2$ 的试题应该淘汰。

### (三) 信度

信度是描述考试结果的可靠性的指标。考试是教育测量的一种手段,任何一项测量都不可避免地存在误差,或者说存在有可靠性的问题。作为教育测量的考试自然也不例外。这就需要引入信度这个指标,用以描述考试结果的可靠性问题。教育测量误差的来源与一般测量误差的来源一样,有随机误差和系统误差。影响信度的误差是随机误差,它主要来源于试卷内部、考试过程、学生本人三个方面,如试题的代表性、题量、难度、评分客观性以及考场环境、学生心理状态等。

信度可用Cronbach在1951年提出的公式计算,即

$$b = \frac{n}{n-1} \left( 1 - \sum_{i=1}^n S_i^2 / S^2 \right) \quad (3)$$

式中: $n$ 为试卷的题数, $S_i^2$ 和 $S^2$ 分别为第*i*题得分和总分的方差,即

$$S_i^2 = \sum_{j=1}^N \left( X_{ji} - \overline{X}_i \right)^2 / N \quad (4)$$

$$S^2 = \sum_{j=1}^N \left( X_j - \overline{X} \right)^2 / N \quad (5)$$

式中: $X_{ji}$ 为第*j*个学生在第*i*题上的得分, $\overline{X}_i$ 为全体学生在第*i*题上得分的平均值, $X_j$

为第*j*个学生的总分, $\overline{X}$ 为全体学生总分的平均值, $N$ 为学生人数。

一般要求信度在0.6以上。

### (四) 效度

效度是描述试卷的准确性和有效性的指标。试卷是教育测量的工具,任何一种测量工具都存在有一个精度或准确度问题,它给测量带来系统误差。效度可分为内容效度和效标关联效度。一般来说,主要分析内容效度,要保证考试具有较高的内容效度,必须对有关的教学大纲、教学内容及教科书进行系统的分析,确定出各部分内容的相对重要性及其在试卷内容上应占的比例。关于试卷的内容效度目前尚难做到定量分析。

#### 一、实例分析

下文将应用以上试卷定量评估指标对《会计学原理》结业考试的试卷质量进行分析与教学评议。参加该结业考试的会计学专业学生共计180人左右，从中随机抽取座号自132号至161号的30名学生的试卷作为分析的样本。该结业考试的试卷全卷共有63个试题，满分为100分。首先计算出每个试题得分和总分的均值（见表1），并应用统计公式（4）、（5）计算出每个试题得分和总分的方差（见表1）。再把试卷按从高分到低分顺序排列，各取出27%的高分端和27%的低分端试卷组成高分组和低分组试卷，并分别算出这两组试卷中每题学生得分的平均值（见表1）。最后依次应用公式（1）、（2）、（3）分别算出每题的难度、区分度和整份试卷的信度（见表1）。

表1 试题的难度、区分度和信度

题号 $i$	满分 $Q_i$	均值 $\bar{X}_i$	方差 $S_i^2$	高分组均值 $\bar{X}_i^{(H)}$	低分组均值 $\bar{X}_i^{(L)}$	难度 $H_i$	区分度 $D_i$
1	1	0.80	0.058	0.906	0.781	0.20	0.13
2	1	0.92	0.068	1.000	0.844	0.08	0.16
3	1	0.72	0.144	0.938	0.594	0.28	0.34
4	1	0.86	0.090	0.875	0.750	0.14	0.13
5	1	0.73	0.202	0.875	0.875	0.27	0.00
6	1	0.79	0.078	0.906	0.688	0.21	0.22
7	1	0.43	0.240	0.625	0.250	0.57	0.38
8	1	0.88	0.040	0.938	0.781	0.12	0.16
9	1	0.83	0.122	1.000	0.688	0.17	0.31
10	1	0.82	0.116	0.938	0.750	0.18	0.19
11	4	3.3	0.722	3.625	3.000	0.18	0.16
12	4	3.0	1.00	3.750	2.438	0.25	0.33
13	4	3.2	0.130	3.125	3.000	0.20	0.03
14	4	2.45	1.254	3.063	1.750	0.39	0.33
15	1	0.70	0.221	0.875	0.500	0.30	0.38
16	1	0.87	0.119	0.875	1.000	0.13	-0.13
17	1	0.27	0.202	0.375	0.125	0.73	0.25
18	1	0.53	0.260	0.875	0.500	0.47	0.38
19	1	0.70	0.217	0.625	0.750	0.30	-0.13
20	1	0.30	0.217	0.750	0.125	0.70	0.63
21	1	0.70	0.217	0.750	0.500	0.30	0.25

22	1	1.00	0.00	1.000	1.000	0.00	0.00
23	1	0.90	0.093	0.875	0.750	0.10	0.13
24	1	0.47	0.257	0.750	0.375	0.53	0.38
25	1	0.80	0.166	0.875	0.625	0.20	0.25
26	1	0.40	0.248	0.500	0.250	0.60	0.25
27	1	0.90	0.093	0.875	0.875	0.10	0.00
28	1	0.50	0.259	0.750	0.500	0.50	0.25
29	1	0.87	0.120	0.875	0.750	0.13	0.13
30	1	0.77	0.185	0.875	0.625	0.23	0.25
31	1	0.70	0.217	0.875	0.625	0.30	0.25
32	1	0.50	0.259	0.750	0.625	0.50	0.13
33	1	0.70	0.217	1.000	0.125	0.30	0.88
34	1	0.93	0.064	1.000	0.750	0.07	0.25
35	1	0.67	0.230	0.875	0.500	0.33	0.38
36	1	0.90	0.093	0.875	1.000	0.10	-0.13
37	1	0.43	0.254	0.500	0.500	0.57	0.00
38	1	0.43	0.254	0.500	0.375	0.57	0.13
39	1	0.97	0.033	1.000	0.875	0.03	0.13
40	1	0.70	0.217	0.750	0.500	0.30	0.25
41	1	1.00	0.00	1.000	1.000	0.00	0.00
42	1	0.70	0.217	0.875	0.500	0.30	0.38
43	1	0.87	0.120	1.000	0.750	0.13	0.25
44	1	0.90	0.093	1.000	0.625	0.10	0.38
45	5	4.18	0.905	4.750	3.500	0.16	0.25
46	5	3.83	2.213	4.625	2.812	0.23	0.36
47	2	1.58	0.260	1.875	1.188	0.21	0.22
48	2	1.95	0.040	2.000	1.938	0.03	0.03
49	2	1.05	0.144	1.000	0.812	0.48	0.09
50	2	1.32	0.405	1.938	0.938	0.34	0.50

51	2	1.97	0.033	2.000	1.875	0.02	0.06
52	2	1.85	0.089	1.875	1.812	0.08	0.03
53	2	1.97	0.016	1.938	1.938	0.02	0.00
54	2	1.68	0.267	1.938	1.188	0.16	0.38
55	2	1.42	0.088	1.438	1.250	0.30	0.09
56	2	1.33	0.126	1.438	0.875	0.34	0.28
57	2	1.83	0.109	1.750	1.875	0.08	-0.06
58	2	1.83	0.213	2.000	1.750	0.08	0.13
59	2	1.45	0.178	1.625	1.313	0.28	0.16
60	2	1.53	0.568	2.000	0.750	0.24	0.63
61	2	0.60	0.317	1.125	0.313	0.70	0.41
62	2	0.98	0.336	1.438	0.750	0.51	0.34
63	2	0.57	0.202	0.938	0.250	0.71	0.34
信度 $b=0.82$ , 均值 $\bar{X}=73.5$ , 方差 $S^2=83.485$ , 学生人数 $N=30$							

为直观分析整份试卷的质量，我们以难度和区分度为坐标构建直角坐标图，如图1所示。

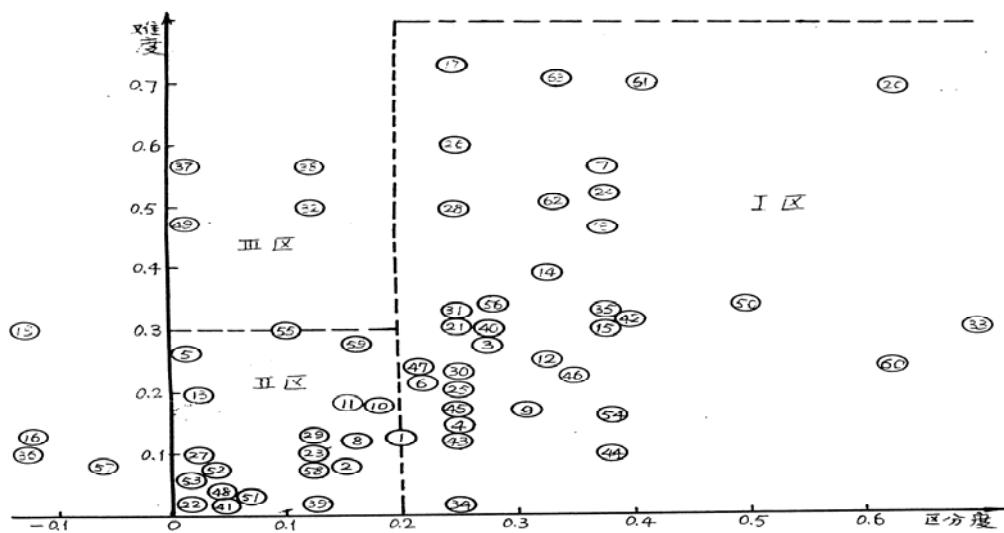


图1 难度——区分度坐标图

根据表1和图1可以得出如下研究结论：

(1) 从总体上看，整份试卷的信度高达0.82，说明本次考试结果的可信度是很高的。这是因为该试卷题量大，评卷前制订出详细的评分标准，且依据该评分标准认真阅卷评分，这些都是减少试卷随机误差的重要措施，从而提高了考试结果的可靠性。

(2) 落在I区 ( $D > 0.2$ ) 的35个试题，占全卷的55.6%，其区分度均大于0.2，作为结业考试，这些题目是合适的，它能够反映和区分出优劣生。其中难度大于0.3的有20题，占全卷的32%。

而难度小于 0.2 的只有 6 道题, 占全卷的 9.5%, 这些题目属于必须掌握的重要内容, 统计难度低反映出绝大多数学生能较好地理解和掌握这些重要内容, 只有少数差生尚未掌握。

(3) 落在 II 区 ( $H \leq 0.3, D < 0.2$ ) 的有 20 个试题, 占全卷的 31.7%。这些题目大都属于最基本的内容, 要求学生都应该能顺利通过, 这正如统计结果所表明的那样, 这些题目难度小, 区分度低。

(4) 落在 III 区 ( $H > 0.4, D < 0.2$ ) 的有 4 个试题, 占全卷的 6.3%。其中第 32 题和第 38 题属于简单的概念题, 但统计难度竟高达 0.5 以上, 而区分度只有 0.12。这种异常情况反映出优生对简单的概念问题有忽视倾向。第 37 题和第 49 题是属于会计学原理的应用问题, 本应区分出优劣生, 但结果却不然。虽然难度高而区分度却小于 0.1, 它提示我们在教学中要多注意培养学生的实际应用能力。

(5) 落在 IV 区 ( $D < 0$ ) 的有 4 个试题, 即第 16、19、36、57 题, 占全卷的 6.3%。这 4 个试题的区分度均小于零, 应予以淘汰。

#### 参考文献:

- 【1】 北京农业大学主编. 概率论与数理统计. 北京: 农业出版社, 1990: 165-166
- 【2】 邓方安等主编. 概率论与数理统计教程. 青岛: 青岛海洋大学出版社, 1999: 200-203

## Quantitative Assessment of Examination Paper's Quality in Accounting Teaching

**Abstract:** In accounting teaching, quantitative assessment of examination paper's quality is very important for improving teaching quality. This paper considers degree of difficulty, degree of differentiation, confidence and effectiveness of examination paper as index system for quantitative assessment of examination paper's quality, and provides mathematical calculating method and assessment criterion. Then this paper applies the above quantitative assessment method to the practical analysis and assessment of Accounting Principle's examination paper.

**Key words:** accounting teaching    quality of examination paper    quantitative assessment    index system