

# Analysis and Assessment of Physical Chemistry Examination Proposition with Educational Measurement

Qing-wei TIAN\*

College of Environmental and Chemical Engineering, Dalian Jiaotong University, Dalian, China

\*Corresponding author

**Keywords:** Educational measurement, Assessment, Examination proposition.

**Abstract.** This paper introduces the specific method of educational measurement by analyzing and evaluating final examination proposition of polymer chemistry. Long-term teaching practice shows that this method could guide and improve the teaching effects and should be widely applied and mastered among teachers.

## 用教育测量学分析评价物理化学考试命题

田庆伟\*

大连交通大学环境与化学工程学院，大连，中国

\*通讯作者

**关键词:** 教育测量学；评价；考试命题

**摘要:** 介绍了应用教育测量学原理对大连交通大学成型+软件专业2016年级，2017-2018学年第一学期学生的物理化学期末考试命题进行分析评价的具体方法和结果。经长期具体的教学工作实践表明，该方法具有广泛的普及和借鉴价值。值得提倡在教师中普及和掌握，以指导教学工作，提高教学效果，达到教学目的。

### 1. 引言

考试是学校教育教学过程中的一个重要环节，通过考试不仅可检验出学生对所学知识的掌握和理解程度，而且还可反馈出教师在教学过程中所存在的问题和不足，同时考试也是选择和发现人才的一个重要途径和有效方法。正由于上述原因，考试历来受到人们的高度重视，考试成为整个教育教学体系的一个重要组成部分和环节，在整个教育教学过程中具有极为重要和不可替代的作用。

### 2. 正文

考试通常是以考试命题的方式表现出来的，过去对考试命题的评价分析大都是从定性说明的角度去进行的，缺乏定量的分析和测算。因此得出的评估结论大都比较笼统，缺乏说服力。20世纪九十年代以来，国际上日益盛行一种新的考试命题质量分析评估方法，这就是教育测量学方法{1}。这种方法是采用数理统计对考试成绩进行有关考试信度、区分度和难度等项质量指标的测算，从而得出一系列的考试参数和数据，然后再将这些参数和数据同规定的标准参数指标相对照，看其是否吻合。如果所得出的有关考试质量参数均与规定的理论指标相符合或相近似，则说明这次考试的命题质量是比较好的，考试达到了预期的目的。所以，采用教育测量学{2}{3}{4}的方法对考试命题进行分析和测算，能够得出比较科学和公正客观

的评估结论。是一种比较先进和科学的方法，非常值得大力推广和应用。作者在二十世纪九十年代初，就已经将这一方法应用到自己的教学工作中，一直沿用至今。作者曾将自己的这一工作体会在本单位内部做过经验介绍和推广，引起了应有的重视。作者认为，作为高等学校的教师和长期从事教育教学管理工作的同志，熟悉和掌握教育测量学的科学方法，自觉地将这一科学方法应用到自己的教育教学工作中去，用其指导和规范自己的教学工作。提高教学质量和教学效果是很有必要和极为重要的。近年来，已有一些单位将这一方法付诸于他们的工作实践中，但是具体方法各异{5}{6}。考虑到这一科学方法的具体性和特殊性，为此，作者结合自己的实际教学工作。将作者用教育测量学方法对大连交通大学成型+软件专业16级2017-2018学年第一学期作者主讲的物理化学期末考试试题进行分析评价的具体做法和结果公布，此项工作未见文献报道，供有关方面参考，以期达到唤起广大教学同仁对此项工作的关注和抛砖引玉的目的。

首先进行的是考试信度的测算，信度是指考试结果的可靠程度，通常用相关系数表示信度的高低。相关系数亦称为信度系数，常用 $r_{tt}$ 表示。在实际测量中，常用的信度系数有稳定性系数，等值性系数，等值稳定性系数和内部一致性系数等。具体采用哪种信度系数来表示考试测验的信度，应根据具体的考试测验情况来确定，不能千篇一律。这里，我们采用内部一致性系数来作为这次考试的信度，具体做法是：这次考试试卷的全部试题按题目番号分为两半，一半是奇数题，一半是偶数题。先把每个人的奇数题，偶数题得分分别相加，得一个奇数题总分，一个偶数题总分，然后求所有应试者奇数题总分与偶数题总分的相关系数，最后用斯皮尔曼——布朗公式校正，求整个测验的信度系数。如上所述，我们得到了如表1所示的全部测试情况：

表1 信度系数计算表

得分被试	奇数题总分X	偶数题总分Y	X <sup>2</sup>	Y <sup>2</sup>	X*Y
01	36	41.5	1296	1722.25	1494
02	27	48	729	2304	1296
03	42	52	1764	2704	2184
04	24	43.5	576	1892.25	1044
05	32	35	1024	1225	1120
06	33	38	1089	1444	1254
07	25	36	625	1296	900
08	30	44.5	900	1980.25	1335
09	32	44	1024	1936	1408
10	31	35	961	1225	1085
11	46	47.5	2116	2256.25	2185
12	37	43.5	1369	1892.25	1609.5
13	39.5	34.5	1560.25	1190.25	1362.75
14	30	38	900	1444	1140
15	30	25	900	625	750
16	42	41	1764	1681	1722
17	44	43.5	1936	1892.25	1914
18	30.5	23.5	930.25	552.25	716.75
19	40.5	49	1640.25	2401	1984.5
20	30	26	900	676	780
21	46	46.5	2116	2162.25	2139
22	30	32	900	1024	960
23	36	9	1296	81	324
24	30	40	900	1600	1200
25	27	27	729	729	729
26	33	45	1089	2025	1485

27	24	30.5	576	930.25	732
28	43	47.5	1849	2256.25	2042.5
29	16	21	256	441	336
30	27	30.5	729	930.25	823.5
	993.5	1118	34443.75	44518	38055.5
	$\sum X=993.5$	$\sum Y=1118$	$\sum X^2=34443.75$	$\sum Y^2=44518$	$\sum X*Y=38055.5$

根据公式(1):

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} \quad (1)$$

公式中 N 是被试者的数目, 这里 N=30, 将表中的数据带入公式 (1) 中有:

$$\begin{aligned} r_{xy} &= \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} \\ &= \frac{30 \times 38055.5 - 993.5 \times 1118}{\sqrt{[30 \times 34443.75 - (993.5)^2][30 \times 44518 - (1118)^2]}} \\ &= \frac{1141665 - 1110733}{\sqrt{[1033312.5 - 98704.25][1335540 - 1249924]}} \\ &= \frac{30932}{\sqrt{934608.25 \times 85616}} \\ &= \frac{30932}{\sqrt{80017419932}} = \frac{30932}{282873.51} = 0.11 \end{aligned}$$

根据斯皮尔曼——布朗公式, 见公式(2)

$$r_{tt} = \frac{2 \times r_{xy}}{1 + r_{xy}} \quad (2)$$

将得到的  $r_{xy}$  数值带入公式 (2), 即得到校正后的信度系数  $r_{tt}$

根据斯皮尔曼——布朗公式, 得到校正后的信度系数  $r_{tt}$

$$\begin{aligned} r_{tt} &= \frac{2 \times r_{xy}}{1 + r_{xy}} \\ &= \frac{2 \times 0.11}{1 + 0.11} = \frac{0.22}{1.11} = 0.20 \end{aligned}$$

即这次考试的信度值是 0.20。根据教育测量学的理论, 信度的取值范围在 0 与 1 之间:  $0 \leq r_{tt} \leq 1$ 。在 0~1 范围内, 信度系数  $r_{tt}$  越大, 考试的可信程度也越高。从测算得到的信度值可以看出, 这次考试的信度值虽然符合国际规定的标准范围, 但是相对较低, 从理论上讲是可信的但可信程度不是特别高。

接着我们进行教育测验质量指标的第二项测算——区分度的测算。所谓区分度是指试题对不同学习程度的考生的鉴别能力, 区分度通常用 D 来表示, 它是衡量试题质量的一项重要指标。

这次考试的试题属论文式测验试题。即试题的评分不能象客观选择性试题那样简单地按对、错或通过、不通过来加以评判。而必须根据一定的过程, 推理或计算, 推导程序, 得出的结论来具体地, 分步骤地加以评判。所以必须采用论文式试题区分度的计算方法。具体做法是: 第一步: 按测验总分由高到低依次排列试卷; 第二步: 从得分最高的一份试卷开始依次向下选出全部试卷的 27%, 做为高分组, 从得分最低的一份试卷开始依次向上选出全部试卷的 27%, 做为低分组; 第三步: 列试题分析表, 如表 2; 第四步: 按下面公式(3)计算区分度 (D)。

$$D = \frac{X_H - X_L}{N(H-L)} \quad (3)$$

公式(3)中,  $X_H$ 为高分组得分总数, $X_L$ 为低分组得分总数,  $N$  为应试总人数的 25%, 这次考试中  $N=8$ 。  $H$  为这道试题的最高得分,  $L$  为这道试题的最低得分, 我们以第一道试题为例, 把表 2 内的数值代入公式 (3), 具体看一下这道题区分度 ( $D$ ) 的计算:

表 2 区分度计算表

	高分组			低分组		
第一题	得分 (x)	人次(f)	得分总和 (fx)	得分 (x)	人次(f)	得分总和 (fx)
	16	8	128	16	6	96
				14	2	28
	H=16	$\sum f=8$	$X_H=128$	L=14	$\sum f=8$	$X_L=124$
第二题	得分 (x)	人次(f)	得分总和 (fx)	得分 (x)	人次(f)	得分总和 (fx)
	20	1	20	13	1	13
	19.5	2	39	12	1	12
	19	2	38	11.5	1	11.5
	18.5	3	55.5	11	2	22
				10	1	10
				6	1	6
				4	1	4
	H=20	$\sum f=8$	$X_H=152.5$	L=4	$\sum f=8$	$X_L=78.5$
第三题	得分 (x)	人次(f)	得分总和 (fx)	得分 (x)	人次(f)	得分总和 (fx)
	15	2	30	3	1	3
	14	5	70	2	2	4
	12	1	12	0	5	0
	H=15	$\sum f=8$	$X_H=112$	L=0	$\sum f=8$	$X_L=7$
第四题	得分 (x)	人次(f)	得分总和 (fx)	得分 (x)	人次(f)	得分总和 (fx)
	15	6	90	9	1	9
	14	2	28	8	1	8
				7	1	7
				6	1	6
				5	1	5
				4	1	4
			0	2	0	
	H=15	$\sum f=8$	$X_H=118$	L=0	$\sum f=8$	$X_L=39$
第五题	得分 (x)	人次(f)	得分总和 (fx)	得分 (x)	人次(f)	得分总和 (fx)
	15	4	60	7	2	14
	14.5	3	43.5	6	1	6
	14	1	14	2	1	2
				0	4	0
	H=15	$\sum f=8$	$X_H=117.5$	L=0	$\sum f=8$	$X_L=22$
第六题	得分 (x)	人次(f)	得分总和 (fx)	得分 (x)	人次(f)	得分总和 (fx)
	18	2	36	7	1	7
	17	2	34	6	4	24
	16	1	16	4	1	4
	15	2	30	3	1	3
	14	1	14	0	1	0
	H=18	$\sum f=8$	$X_H=130$	L=0	$\sum f=8$	$X_L=38$

$$D = \frac{X_H - X_L}{N(H-L)}$$

$$N = 30 \times \frac{27}{100} = 8, X_H=128, X_L=124, H=16, L=14。$$

$$\therefore D = \frac{128-124}{8 \times (16-14)} = \frac{4}{16} = 0.25$$

以上例相同方法，我们根据表 2 的数据得到这次考试每道试题的区分度 (D) 值，如表 3 所示。

最后进行的一项是试题难度的测算。所谓难度是指试题的难易程度。难度常用 P 表示。具体的测算方法是：第一步，第二步，第三步与试题区分度的具体做法相同，得到如表 2 的试题分析表。第四步：按下面公式 (4) 计算难度。

$$P = \frac{X_H + X_L - (2NL)}{2N(H-L)} \quad (4)$$

公式 (4) 中除 P 代表难度值以外，其他各项符号所代表的意义与区分度计算公式 (3) 中各项符号的意义相同，我们仍以第一试题为例，根据表 2 的数据，计算其难度值 P。

如上法，我们得到这次考试每道试题的难度 (P) 值，如表 3 所示。

表 3 区分度和难度数值表

类别 \ 题号	一	二	三	四	五	六
区分度 (D)	0.25	0.58	0.88	0.66	0.80	0.64
难度 (P)	0.88	0.52	0.50	0.65	0.58	0.58

根据教育测量学的理论，对于一般的考试，区分度 (D) 在 0.40 以上，可认为是非常好的试题；D 在 0.30~0.39.可认为是比较好的试题；D 在 0.20~0.29 的试题，是尚可以，但需再作改进的试题；D 在 0.19 以下是比较差的试题，应再作修改或从试卷中删除。按照这个标准，从表 3 中可以看出：试题二、三、四、五、六属非常好的试题；试题一属尚可，但需作改进的试题。这次试卷中没有出现属于比较差，应再做修改删除的试题。也没有出现属于比较好的试题。这表明这些试题具有区分学生学习水平的能力，是积极区分。教育测量学理论同时还指出：试题的难度对试题的区分度是有影响的。难度值 P 在 0~1 之间，难度系数 P 值越大，表明试题越难，P 值越小，表明试题越容易。试题太难或太易，对提高区分度都是不利的。理论分析证明，难度 (P) 值为 0.5 的试题的区分度是最高的。在选择试题时，一般以取 P 值为 0.3~0.7 或 0.4~0.6 为宜。为提高区分度，整个试卷的平均难度系数  $\bar{P}$  应取 0.5 左右。从表 3 结果可看出：一试题属于难度过大的试题，二、三、四、五、六试题属于符合标准难度范围之内，属于比较适宜的试题，其中第三题的区分度是最高的。从表 3 可得出整个试卷的平均难度  $\bar{P}$  应为 0.62。这说明在理论规定的标准难度值范围之内，说明试题难易适中。综上所述，我们可以看出这次考试的信度值还是可信的，表明考试成绩是可以信赖的。试题的区分度也很好，但试题的难易适中。综合分析表明，此次考试试题质量在信度和区分度上符合教育测量学理论规定的要求，取得了比较好的结果。

### 3. 结束语

以上是我们最初的工作方法和结果的具体范例，这一方法我们一直沿用至今，应用到我们的教学工作中。实践表明，取得了很好的教学效果。总之，利用教育测量学对考试命题进行评价分析，可以为我们提供有重要参考价值的信息和依据。这对我们改进教学工作，提高教学质量有着积极的指导意义。

**Reference**

- [1] Minqiang Zhang, Retrospect and Status Analysis of the Development of Educational Measurement in 20th Century, *Education Research* 11(1999) 32-37.
- [2] Minqiang Zhang, Educational Measurement, The People's Education Publishing Company (1998).
- [3] Maolan Wang, Educational Measurement, HenanUniversity Press(1987).
- [4] MaozuXu, Guihua Zhang, Fundamentals of Educational Measurement, China Railway Publishing House (1995).
- [5] Dong Wang, Liming Chen, etc, Test Paper Quality Analysis and Understanding with Educational Measurement, *China Journal of Modern Medicine*, 19 (2003) 154-156.
- [6] Wei Li, Jin Gao, etc, Application Research of Medical Curriculum Assessment with Educational Measurement, *China Higher Medical Education* 11(2014) 1-2.