维普资讯 http://www.cqvip.com

2002 年第 4 期
第 13 卷
总第 53 期

民族教育研究
Journal of Research on Education for Ethnic Minorities

No.4,2002
Vol.13
General No.53

语言测试中的项目分析方法

张　红

（中央民族大学 外语系，北京　100081）

[摘　要]　对每一次测试的试题进行项目分析的目的，是为了确定试题的科学性，把那些符合测试规则、能够体现测试功能的科学性的试题保存在试题库中。项目分析的方法不仅适合于英语测试，还可以扩大到任何一种选项式的测试。语言测试在教学研究中有着很重要的作用。如何使它更为科学、更为精确，并对学生起到积极的引导作用是一项艰巨且有意义的研究工作。

[关键词]　测试；　项目；　分析；　难度；　区分度；　信度

[中图分类号] H319.6　　[文献标识码] A　　[文章编号] 1001－7178(2002)04－0038－05

一、对测试的计划安排

设计和安排测试，目的是对学生的学习进行检测，帮助他们掌握知识、总结学习过程中的经验教训。教师也可以通过对试题的分析研究，及时发现教学中的问题，适时地对教学计划进行调整，使学生的英语在有限的时间内有最大程度的提高。与此同时，对每一次测试的试题所做的项目分析的目的是为了确定试题的科学性，把那些符合测试规则、能够体现测试功能的科学性的试题保存在试题库中。由此可见，语言测试中的项目分析这项工作是十分重要的。

在众多的语言测试形式中，多项选择（multiple-choice items）是最流行的一种形式。虽然在设计测试时需要策划者花时间、花精力，但是，它的评判却很客观，丝毫不受阅卷人的主观判断的影响。现在，许多学校都可以利用计算机进行评判工作，效率十分高。因此我们在安排测试分析时就选择了这一部分作为实验分析对象。我们以中央民族大学 2001 级 120 名非英语专业的各民族学生作为受试者，采用的教材是高等教育出版社和复旦大学出版社共同出版的《21 世纪大学英语》。这次考试是对本套教材第二册第一课至第五课的期中测试，测试的范围就是这五课的语法、单词和词组。测试的条件是受试者在 35 分钟内完成 40 道多项选择题，不允许查看任何参考书和字典。每一道题都有四个选项，每题只有一个答案，多选无效。受试者可以根据自己的英语知识进行猜测而不受惩罚，即答错的题不倒扣分。另外，我们为受试者准备好答题纸，他们只要把选好的答案用铅笔在答题纸上涂黑即可。测试项目总分为 40 分，分值分配是 1 项 1 分。受试者进行 35 分钟的测试后，我们把答题收上来进行评判，由此取得一些统计数据，然后对这些数据进行分析，最终得出实验结论。

[收稿日期]　2001－05－21

[作者简介]　张红（1967－），女，北京人，中央民族大学外语系讲师，从事英语教学与研究工作。

维普资讯 http://www.cqvip.com

二、项目分析过程

(一)确定项目的难度(Item Difficulty)

这项工作是确定测试项目是否值得要,即是否有可能归入试题库;其公式为:$Po = 0.5 + 0.5/n$(这里"n"表示每一个项目所具有的选项)。需要说明的是:如果多项选择测试的项目是值得要的,可归入试题库的"项目难度系数"应为 $Po = 0.5 + 0.5/4 = 0.625$。例如:如果测试者为 100 人,只有 50 人把题做对了,那么"难度标准"应为:D = 50/100 = 0.5。这一结论低于项目难度指数 0.625,这意味着测试项目对于测试者来说有些难。每一个项目我们都要计算它答对的人数,之后,把它除以测试者的总人数,所得出的百分比就是每一个项目的难度系数,公式为:D = 答对的人数/参加测试的总人数。以下表格即为这次测试的每一个项目的难度系数:

	项目 1	项目 2	项目 3	项目 4	项目 5	项目 6	项目 7	项目 8	项目 9	项目 10	项目 11
D	0.90	0.6083	0.3833	0.70	0.80	0.375	0.775	0.75	0.8333	0.5667	0.9667
	项目 12	项目 13	项目 14	项目 15	项目 16	项目 17	项目 18	项目 19	项目 20	项目 21	项目 22
D	0.8167	0.8583	0.9333	0.6833	0.9583	0.90	0.7083	0.775	0.7917	0.1417	0.8583
	项目 23	项目 24	项目 25	项目 26	项目 27	项目 28	项目 29	项目 30	项目 31	项目 32	项目 33
D	0.85	0.6833	0.8917	0.7917	0.9667	0.9083	0.6417	0.9083	0.9583	0.425	0.4167
	项目 34	项目 35	项目 36	项目 37	项目 38	项目 39	项目 40				
D	0.7833	0.0917	0.775	0.8167	0.8667	0.575	0.0417				

项目难度系数的范围是从 0 到 1。系数越大,这个项目就越容易。如果所有的项目都很容易,这个测试就不能区分学习好的学生和学习差的学生;如果所有的项目都很难,这个测试同样也不能对好的学习者和差的学习者加以区分,因此,测试项目应该处于一个适当的难度标准。从以上表格可以看出:40 个项目中有 30 个项目处于难度系数(0.625)以上,其余 10 个处于难度系数以下。戴维·P·哈里斯在他的《作为第二语言的英语测试》一书中写道:"那些非常简单的多选项目,比如说有 92% 的受试者答对了;或者那些非常难的多选项目,比如说有少于 30% 的受试者答对了,我们一般采取放弃不用的做法,因为它们没有尽到测试的职能,即:测试是用来衡量测试者的水平的一种方式。"[1] 因此,为了达到测试的最佳职能,那些难度指数低于 0.30 的项目由于太难的缘故应该丢弃不用,同样,那些难度指数高于 0.90 的项目由于过于简单的原因也应该丢弃不用。这样,我们就可以把那些难度指数处于 0.30 至 0.90 的测试项目作为进一步研究的目标。以上表格中符合标准的测试项目是:项目 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、17、18、19、20、22、23、24、25、26、29、32、33、34、36、37、38、39。

从上表中可以看出,有 3 个项目(项目 21、35 和 40)低于 0.30;而高于 0.90 的项目有 7 个(项目 11、14、16、21、27、28、30 和 31)。这说明容易的题目比难的题目多,测试题目偏易。

(二)确定项目的区分度(Determining Item Discrimination)

实验的第二步是测试的策划者评估测试项目,区分好的学习者和差的学习者的适合度。"项目的区分度"是指确定好的学习者和差的学习者的尺度。事实上,测试项目应该体现受试者的不同的语言熟练水平,这个语言熟练水平的差异应该在受试者所得的分数上得以体现。

首先,我们把受试者分成两组。对此,不同的学者有着不同的方法。一些学者认为我们应该在获得较高分数的受试者以及获得较低分数的受试者之中分别选出 25% 到 27% 的人数,然后统计他们答对的项目数。另一些学者认为我们应该把受试者分成两部分:一部分为"较高分

维普资讯 http://www.cqvip.com

数组";另一部分为"较低分数组",然后统计每一组所得的正确答题的项目数。如何分组取决于参加测试的人数:如果参加测试的人数很多,第一种方法比较好;如果参加测试的人数很少,第二种方法比较合适。由于参加此次测试的学生只有 120 人,所以我们选择了第二种方法。其次,使用公式:高分组答对的人数 - 低分组答对的人数/高分组答对的人数或低分组答对的人数。对于这个公式的具体操作方法是:把试卷按分数从高到低排列,然后一分为二,分别统计出高分组答对的人数和低分组答对的人数,之后,前者减去后者,所得数除以高分组人数或低分组人数。下面的表格表示此次测试中每一个项目的区分度:

	项目 1	项目 2	项目 3	项目 4	项目 5	项目 6	项目 7	项目 8	项目 9	项目 10	项目 11
DIS-	0.07	0.12	0.20	-0.07	0.17	0.22	0.25	0.17	0.27	-0.13	0.07
	项目 12	项目 13	项目 14	项目 15	项目 16	项目 17	项目 18	项目 19	项目 20	项目 21	项目 22
DIS-	0.13	0.18	0.03	0.07	0.08	0.13	0.12	0.25	0.28	0.02	0.08
	项目 23	项目 24	项目 25	项目 26	项目 27	项目 28	项目 29	项目 30	项目 31	项目 32	项目 33
DIS-	0.17	0.10	0.02	0.28	0	0.08	0.18	0.12	0.05	0.25	0.20
	项目 34	项目 35	项目 36	项目 37	项目 38	项目 39	项目 40				
DIS-	0.13	0.12	0.15	0.17	0.13	0.15	0.02				

区分度的范围是从 -1 到 1。通常情况下,比较合适的区分度应该在 0.40 以上。因此,我们最好对那些区分度在 0.20 到 0.39 之间的项目加以修改。而对于那些区分度在 0.20 到 -1 之间的项目,我们只能弃置不用。事实上,分数之间差别越大,测试项目就越稳定。刘润清教授指出:"如果全体学生的分数相近,说明测试缺乏区分性,题目的难易范围不广,难易程度分得不细。题目一定要由易到难……如果突然加大难度,就会一下子难倒许多学生,出现分数集中的现象。逐渐提高难度,每道题只难住几个人,分数就会分散很广,显示出考生之间的细微差别。"[2]从上表可以看出:40 个测试项目所得出的区分度都处于 0.40 以下,其中有 9 个项目处于 0.20 到 0.39 之间,它们是项目 3、6、7、9、19、20、26、32、33。这个结论意味着受试者所得的分数没有很好地体现出他们的不同的语言熟练水平,未能很好地区分好的学习者和差的学习者。我们需要对以上 9 个测试项目进行修改,使它们的区分度达到 0.40 以上。

得到这个结果有各种原因,其中,项目的难度标准大多高于一般水平可能是原因之一。前面已经提到,项目的难度标准偏低就意味着这次测试对于受试者来说偏难;项目的难度标准偏高就意味着这次测试对于受试者来说偏易;项目的难度标准可能影响到项目的区分度。

(三)确定项目中各选项的效度(Determining the Effectiveness of Distracters)

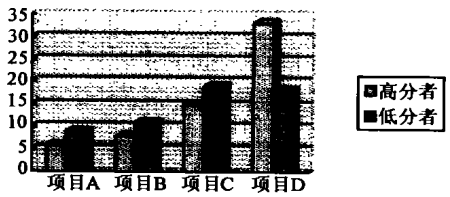
实验的第三步是判断各干扰项的作用。测试的每一个项目都有 4 个选项,受试者必须从中选出 1 个正确的答案。其余 3 个选项就是干扰项,其作用是干扰受试者做出正确的选择。干扰项如果干扰了所有的受试者或没有干扰所有的受试者,说明它没有起到干扰的作用。有时会出现这种现象:一个错误的答案使更多的高分受试者受到干扰,或者说选择这个错误答案的高分受试者要比低分受试者多。这个干扰项叫做"无用干扰项"或"故障干扰项",在项目分析中,应该去掉它们。下面的表格是对各个测试项目逐一进行统计分析后,筛选出的比较典型的数据统计表。

表中下画线的字母为正确答案。我们需要对各个项目逐一进行分析才能确定各干扰项的作用。以第 32 项、第 33 项为例:第 32 项的正确答案是 D,有 33 位高分受试者和 18 位低分受试者选择了它,高分者和低分者选择 A、B、C 三个干扰项的比例分别为 6:9、8:11、15:19;第 33

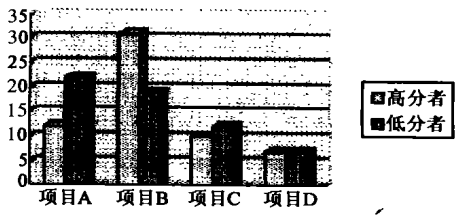
维普资讯 http://www.cqvip.com

项的正确答案是 B，有 31 位高分受试者和 19 位低分受试者选择了它，高分受试者和低分受试者选择 A、C、D 三个干扰项的比例分别为 12:22、10:12、7:7；我们可以通过下面的两个图表看出这两个项目的干扰项的设计是比较成功的，因为它们既干扰了高分受试者又干扰了低分受试者，而且干扰的人数相当。

项目 1	A	B	C	D	项目 2	A	B	C	D	项目 3	A	B	C	D
高分者	2	56	0	2	高分者	5	40	0	17	高分者	0	32	29	0
低分者	3	52	0	5	低分者	5	33	3	17	低分者	6	32	17	4
项目 7	A	B	C	D	项目 8	A	B	C	D	项目 9	A	B	C	D
高分者	54	0	5	2	高分者	3	4	3	50	高分者	1	0	1	58
低分者	39	7	7	6	低分者	9	8	3	40	低分者	4	2	12	42
项目 10	A	B	C	D	项目 11	A	B	C	D	项目 12	A	B	C	D
高分者	3	30	2	25	高分者	0	0	60	0	高分者	53	0	0	7
低分者	1	38	5	16	低分者	0	1	56	3	低分者	45	0	1	14
项目 16	A	B	C	D	项目 17	A	B	C	D	项目 18	A	B	C	D
高分者	0	0	0	60	高分者	1	58	1	0	高分者	46	7	2	5
低分者	1	0	4	55	低分者	3	50	4	3	低分者	39	10	2	49
项目 19	A	B	C	D	项目 20	A	B	C	D	项目 21	A	B	C	D
高分者	1	5	54	0	高分者	1	3	0	56	高分者	9	41	2	9
低分者	7	14	39	0	低分者	13	5	3	39	低分者	10	38	3	8
项目 22	A	B	C	D	项目 23	A	B	C	D	项目 24	A	B	C	D
高分者	3	2	54	2	高分者	56	0	3	1	高分者	10	4	2	44
低分者	2	2	49	6	低分者	46	5	7	2	低分者	16	4	2	38
项目 31	A	B	C	D	项目 32	A	B	C	D	项目 33	A	B	C	D
高分者	0	1	0	59	高分者	6	8	15	33	高分者	12	31	10	7
低分者	1	2	1	56	低分者	9	11	19	18	低分者	22	19	12	7
项目 34	A	B	C	D	项目 35	A	B	C	D	项目 36	A	B	C	D
高分者	4	3	51	3	高分者	9	0	28	23	高分者	51	2	3	4
低分者	6	4	43	6	低分者	2	1	34	23	低分者	42	10	4	4
项目 37	A	B	C	D	项目 38	A	B	C	D	项目 39	A	B	C	D
高分者	2	54	2	2	高分者	4	0	0	56	高分者	5	36	1	18
低分者	3	44	3	10	低分者	2	1	9	48	低分者	14	33	0	13



图一



图二

下面我们来看一看第 16 项：第 16 项的正确答案是 D，其中有 60 个高分者和 55 个低分者选择了它，只有 5 个低分者分别选择了干扰项 A 和 C，而对于高分者来说，A、B、C 三项根本没有起到干扰的作用，因此，我们应该去掉这些无用干扰项，代之以更有效度的选项。项目 21

维普资讯 http://www.cqvip.com

中,大多数的受试者选择了错误选项,而且,选择错误的高分者比低分者多。项目11中的A选项是比较典型的无用干扰项,因为,无论是高分者还是低分者都没有选择它。如果试卷中有很多干扰项是无用的,就会影响到整个测试题的信度。

三、测试的信度(Test Reliability)

测试的信度用来衡量测试的精确程度。学生的分数是否稳定是从信度系数来看的。信度系数越高,测试就越可靠。如果一个测试的"信度系数"为1.0,那就说明这个测试相当可靠,精确程度相当高。我们有几种方式可以得到测试的信度系数。由于此测试为多项选一的客观测试,因此我们选用KR.20或KR.21。KR.20和KR.21是用于客观测试、特别是多项选择测试的最佳公式。在这两个公式中,KR.20要比KR.21更为精确,而后者比前者更加简单易行。由于我们是在进行测试的科学性研究,所以,我们选用KR.20公式求出此次测试的信度系数。这个公式是:$r = k/k - 1[1 - \sum pq/S^2]$("K"指测试项目的数量,这里为40;"p"指某一个测试项目的正确答题的比例;"q"指1-p;"S"指"标准偏差",此处为4.96)。

首先,把数据收集起来并且进行统计,然后,求出测试试题的信度系数为0.78。这说明学生在此次测试中得到的分数是比较稳定的。根据信度高则效度高、信度低则效度低的原理,我们可以得出结论:此次测试基本达到了测试的目的。

四、结语

这种项目分析方法不仅适用于英语测试,还适用于任何选项式的测试。只要各个项目的分值一致,进行分析时,把数据带入公式计算即可。语言测试在教学研究中有着很重要的作用,如何使它更为科学、更为精确,并对学生起到积极的引导作用是一项艰巨且有意义的研究工作。

[参考文献]

[1] Harris, David P. *Testing English as a Second Language*[M]. New York: MacGraw-Hill Book Company, 1968.105.

[2] 刘润清.语言测试和它的方法[M].北京:外语教学与研究出版社,1999.20.

Analyses of Testing Items

ZHANG Hong

(Foreign Languages Department, the CUN, Beijing 100081)

[**Abstract**] One of the main tasks for language teachers is to make language testing more scientific and more accurate in order to keep those items which accord with testing regulations and can embody the function of tests in further use. Language testing plays a very important role in language teaching and research. Analyses of testing items can be applied in any multiple-choice tests. It is also hard and significant research work in foreign language teaching.

[**Key words**] analyses of testing item; item difficulty; item discrimination; test reliability

〔责任编辑 宝玉柱〕