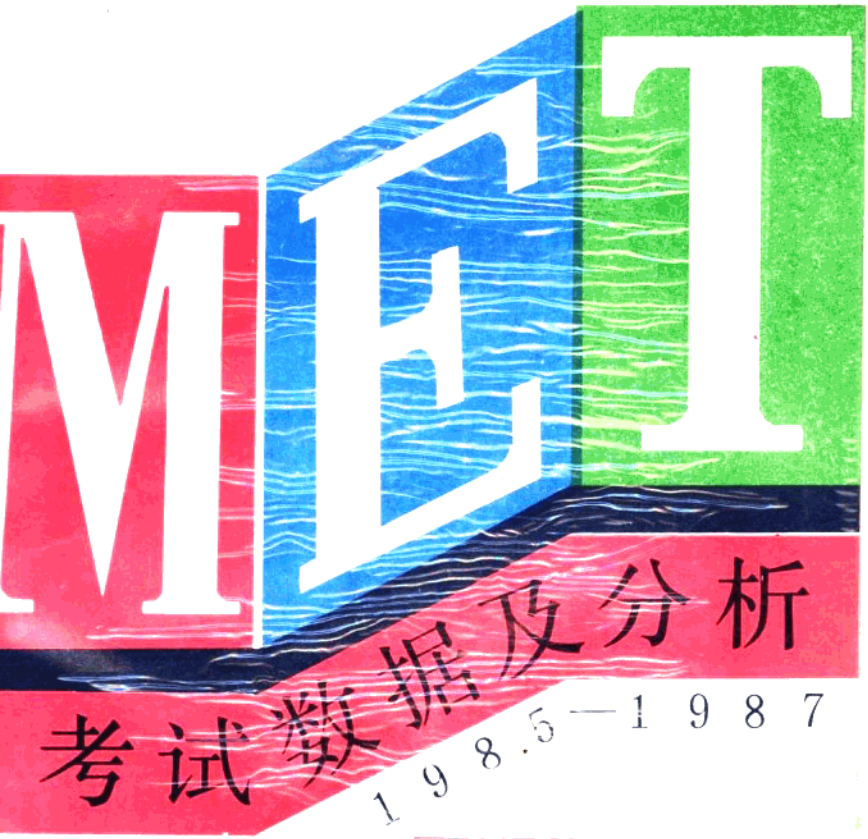


考试数据及分析

1985—1987

广东省高考英语改革小组

李 崴 执 笔

广东教育出版社

目　录

写在前面的话

一、本书的目的

高考英语标准化考试改革试验从1985年开始，至今已有三年多了。这几年来，试验组的同志们大胆地吸取国内外成功的经验，结合我国的国情，以有利于高校选拔人材和有利于促进中学英语教学改革为目的，正在摸索一条在我国实行标准化考试的路子。这两个“有利”中，更重要的是后者。因为选拔人材只是对当年的高考而言，而对中学英语教学的促进则更具有长远的意义。

按常规，MET的试题公诸于世，考生的成绩被充分使用，当年命题及考试的任务即告完成。但我们认为，通过MET考试，还应该让中学老师以及关心中学英语教学的人们从中得到更多的信息。如每一道大题的成绩如何？每一条题目考生的应答情况如何？考生在MET考试中哪些知识、能力、技巧掌握得较好，哪些方面还有不足？今后在中学英语教学中应如何改进？有了这些信息，对中学的英语教学将会起到更大的促进作用。本书正是为此目的而编写。书中所列的材料均是这几年高考MET中的原始数据。希望通过本书的出版，能更有力地促进标准化考试的试验和推广，有利于促进中学英语教学水平的提高；也希望能促进中学英语科考试的科学化，让教育测量学的理论在中学英语教学中进一步得到普及。

二、本书的内容

本书的内容主要是这几年来高考英语试题（MET）的

每一小题的项目分析表。将高考英语试题原始数据公开发表在我国尚属首次，为使老师们有效地利用这些来自高考考生的第一手资料，有必要在前面向大家具体讲述每一项数据的意义、来源（包括计算的公式与方法），以及就具体的数据举例谈谈我们对某些题目考生的应答情况的看法。书后还附上与项目分析表对应的当年的考题，以便查阅参考。

我们希望，今后高考英语科项目分析的发表能成为标准化考试中的一个环节，并希望其在考试与教学中起到一种相互促进、相互检验的作用，进而为我国的英语教学改革尽一份力量。

第一部分　数据的意义

本节向大家介绍后面表中所列的各项数据的意义及其计算公式，但考虑到篇幅及普及的原因，没有提供公式的推导及来源。

1．答对率P与难度Δ

答对率P也即通过率，指选择正确答案的人数（Nr）占全体抽样人数（No）的比例，如抽样3000考生，答对该题的

$$P=Nr/No \qquad \text{公式（1）}$$

人数是1321人，则答对率P＝1321/3000＝0.44。由于通过率与难度的关系是非线性的，所以一般将通过率转为标准分Δ（希腊字母读作delte）作为难度值。Δ的计算公式为：

$$\Delta=13+4z \qquad \text{公式（2）}$$

其中$z=2.5071x+2.288x^3+33.5x^5-464x^7$

$$+2874.18x^9-4848.17x^{11} \qquad \text{公式（3）}$$

其中$x=\cdot 5-P$　　公式（4）

从计算公式（1）可知：通过率P值最大是1（表示所有考生都答对了），最小是0（表示所有考生都没答对），P值越大表示该题目越容易。Δ值则是越小表示题目越容易，Δ值为13时表示刚好有一半考生答对；Δ值为17（13加上一个标准差4的位置）时，大约只有16%的人答对；反之Δ值为9（13减去一个标准差）时，大约有84%的答对者。从这个指标可衡量考生对每一题以至整份试卷的掌握程度。

2．人数

表中人数这一栏中的“A，B，C，D，无效”，指的是

选A，B，C，D等各选择项的人数，如MET86笔试项目分析表的第一题中选A的有277人，选B的有74人，选C的31人，选D的23人，“无效”指的是漏选答案或重选答案的人数，五个人数加起来为总的抽样数。从人数这一栏可以看出考生选择干扰项的人数的分布，以便了解考生容易犯些什么错误。

3．平均分

这一栏中的“M_A，M_B，M_C，M_D，无效”指的是选A，B，C，D和无效选择的那些人的该大题的平均分，为了统一和便于比较，一律转换成13为平均分，4为标准差的标准分，计算公式如下：

$$M_A = 13 + \frac{4\ (x_A - M_t)}{S_t} \qquad \text{公式（5）}$$

$$M_B = 13 + \frac{4\ (x_B - M_t)}{S_t} \qquad \text{公式（6）}$$

$$M_C = 13 + \frac{4\ (x_C - M_t)}{S_t} \qquad \text{公式（7）}$$

$$M_D = 13 + \frac{4\ (x_D - M_t)}{S_t} \qquad \text{公式（8）}$$

$$M_{无效} = 13 + \frac{4\ (x_{无效} - M_t)}{S_t} \qquad \text{公式（9）}$$

其中$x_A x_B x_C x_D x_{无效}$分别为选A、B、C、D和无效选择的人的平均分（原始分）；$M_A M_B M_C M_D M_{无效}$分别为选A，B，C，D和无效选择的人的平均分（标准分）。从平均分这一栏我们可以观察到选各种选择项的考生到底是属于什么水平的考生，一般来说应该是选正确答案的考生成绩较高（如Δ大于13）而选干扰答案的考生成绩较低（如Δ小于13）。若不然则表示这道题目没出好，或是太难。从干扰项的平均分亦可了解到哪些是水平高的考生容易犯的错误，哪些是

只有水平低的考生才会犯的错误。

4．区分度

区分度是题目是否能够将水平高的考生与水平低的考生区分开来的指标。区分度的计算方法很多，我们采用的是计算该大题成绩与该选择的选答情况的双列相关系数 r_{bis} 来表示，计算公式如下：

$$r_{bis}=\frac{M_t-M_w}{S_t}\cdot\frac{P(1-P)}{y} \qquad \text{公式（10）}$$

或：

$$r_{bis}=\frac{M_r-M_t}{S_t}\cdot\frac{P}{y} \qquad \text{公式（11）}$$

其中M_r为选正确答案的考生的该大题的平均分（原始分）。

M_w为选其它答案的考生的该大题的平均分(原始分）。

M_t，S_t为全体抽样考生该大题的平均分和标准差。

$$S_t=\sqrt{\frac{\sum_{i=1}^{N_o}(x_i-M)^2}{N_o}} \qquad \text{公式（12）}$$

或

$$S_t=\sqrt{\frac{N_o\sum_{i=1}^{N_o}x_i^2-(\sum_{i=1}^{N_o}x_i)^2}{(N_o-1)N_o}} \qquad \text{公式（13）}$$

x_i为第i个考生的成绩。

M为全体抽样考生平均成绩。

M_o为抽样考生总人数。

P为选正确答案的人数占全体人数的比例。

y为正态分布中把P与1－P概率分开的纵线坐标值，可以由以下近似公式求得：

$$y=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \qquad \text{公式（14）}$$

z值求法同公式（3），但式中x＝P－0.5　公式（15）

考虑到命题的需要我们不仅算出了每一道题目的区分度，还计算出了每一选择项的区分度，所以区分度一栏中有A_r，B_r，C_r，D_r分别表示各个选择项的区分度，正确答案的那个选择项的区分度即本题的区分度，干扰项的区分度的计算同公式（11），所不同的是M_r为选择该干扰项的考生的该大题的平均分，P为选择干扰项的人数占总人数的百分比，并且为了方便使用，干扰项的区分度算出来后均乘上一个－1，这样区分度栏中凡是负数的，表示该选择项不但不能将考生区分开来，甚至是颠倒了。一般来说某一道题的区分度主要是看正确答案那个选择项的区分度，要求达到0.3以上，干扰项的区分度则只供参考。一道题的区分度越大，表示该题目答对的考生基本上是该大题成绩最好的学生，若区分度小甚至是负数，表示好的考生没答对，差的学生反而答对了。若某道题所有考生都答对了，则区分度为0。干扰项的区分度越高，则表示该选择项越能起到干扰作用——成绩差的学生选而成绩好的学生不会选。

大题的区分度和整个试卷的区分度的计算不能用算术平均数求取，其计算公式如下：

$$R_{bi_s}=\frac{e^{\frac{2Z}{N}}-1}{e^{\frac{2Z}{N}}+1}$$　　公式（16）

$$\text{其中 } z=\frac{1}{2}\sum_{i=1}^{N}\ln(1+r_i)-\ln(1-r_i)$$　　公式（17）

其中r_i是第i小题的区分度，N为总题目数。计算大题区

分度时为大题中的小题数。计算试卷区分度则是整份卷子的小题目数。大题的或试卷的区分度一般要求也应该在 0.3以上为好。本书中信度、偏态、峰值表中 整份试卷的 区分度 R_{bis}，就是以此公式算出来的。

5．信度

R_{11}和α信度是指考试的结果是否可靠的指标，其最大值为1，最小值为0。所谓可靠就是说这个考试的成绩可信，能反映出考生的水平。当然可靠只能是相对的，没有一个考试不存在误差，有误差就有不可靠的因素。因此考试的信度一般不可能达到1。求信度的分式有多种，我们采用的是Kuder-Richardson20公式：

$$R_{11}=\frac{N}{N-1}\left(\frac{S_t^2-\sum_{i=1}^{N}p_iq_i}{S_t^2}\right) \qquad \text{公式(18)}$$

其中N为整份试卷的小题目数；s_t是全体考生这次考试成绩(按一题一分计算)的标准差；p_i是第i小题的全体考生平均答对率，而$q_i=1-p_i$。从上式可看出当题目数越多或标准差越大，则信度R_{11}越大，也就是考试的成绩越可靠。

但R_{11}只能计算选择题的信度，对于有主观性试题的题目就要采用另一种办法计算信度。我们采用的是α，

$$\alpha=\frac{N}{N-1}\left(1-\frac{\sum_{i=1}^{N}st_i^2}{S_t^2}\right) \qquad \text{公式(19)}$$

其中N为大题目数，st_i为第i大题的标准差，s_t为总分的标准差。信度R_{11}一般要求在0.9以上。α一般要求达到0.8。

从本书的表3，表6，表9，表12，表15，表18，表21上可以分别看到近几年MET试题的信度值。

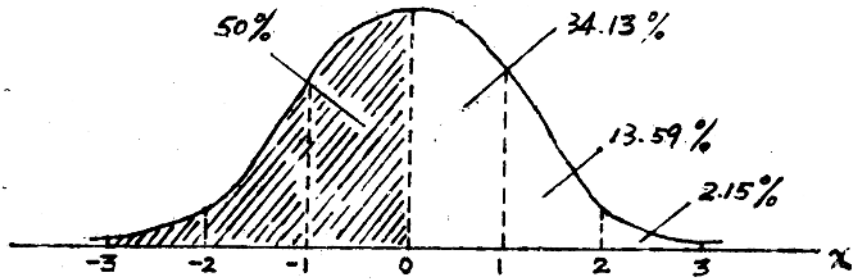
6．标准差

标准差S(Standard Deriation)是描述全体分数的离散程度的一个指标，标准差的公式见公式(12)和公式(13)。

从公式可见标准差越大表示这一批数据离平均分的“差距”越大，也就是分数分布较广，标准差越小则分数分布较窄或说“集中在平均分附近”，若所有考生得的都是同一个分数则标准差为0。

从标准差我们可以了解到考生成绩的分布情况。在与人的行为有关的社会现象中，大多数变量都呈现出或近似地呈现出正态分布的情形。所谓正态分布是指居中的数据量最多，两边的数据量逐渐对称地减少的数量分布情况（如图1所示）。而一个正态分布是由总体的平均数和标准差所决定的。若一组分数呈现正态分布，就是由全体考生的平均分和标准差所决定的。说得更具体一点就是：取得平均分附近的人最多，取得离平均分越远的两边的分数的人逐渐减少，并且减少的量与标准差的关系是衡定的。这就是：离平均分左右一个标准差以内的人数占全体人数的68.26%，离平均分左右两个标准差以内的人数占95.46%，三个标准差以内的人数占99.73%。或者也可以这样说，在平均分以下的人数占50%，到平均分以上一个标准差处的人数占84.13%，到平均分以上两个标准差处的人数是97.23%，三个标准差处则已经是99.87%了。

例如，MET85笔试中广东省全省的阅读理解平均分为22.45，标准差为8.67(见表23)，表示拿22.45分以下的约有50%的考生，(2735×50%≈1368人)，拿到平均分以上两个



正态分布曲线

标准差，即22.45+8.67×2=39.79分以上的约占100%－97.23%=2.77%(2735×2.77%≈76人)。

当然这种计算必须是在考试的成绩是呈正态分布的情况下才是如此。这几年来MET的成绩基本上是呈现正态分布的。

7．偏态值Skewness和峰值Kurtosis

前面所讲的许多指标都要求考试成绩的分布在正态分布下才有意义。什么为正态？图1的曲线反映的是理论正态曲线，而实际的曲线与理论曲线有一定的误差，这个误差可能有两个方面：一个是曲线的"峰"偏向了一边，若偏向高分一边称为负偏态，若偏向低分一边则称为正偏态，这可以从图上观察，也可以用偏态值(Sk)来表示，偏态值的计算公式如下：

$$Sk = \frac{\sum_{i=1}^{N_o} (x_i - M)^3 / N_o}{\sum_{i=1}^{N_o} (x_i - M)^2 / N_o \cdot \sqrt{\sum_{i=1}^{N_o} (x_i - M)^2 / N_o}}$$

公式(20)

其中x_i是第i个考生的成绩，M是全体考生的平均分，N_o是考生数。

另一方面，还要考虑分布曲线的“峰”是“高尖”还是“矮平”。因为“峰”的“高尖”或“矮平”都不是正态分布，“高尖”表示分数分布过于集中，“矮平”则表示分布过于分散。统计学中采用了峰值(Ku)来描述这一状况。峰值(Ku)的计算公式如下：

$$Ku = \frac{N_o \sum_{i=1}^{N_o} (x_i - M)^4}{\left(\sum_{i=1}^{N_o} (x_i - M)^2 \right)^2} - 3$$

公式(21)

式中变量意义同前。

当Sk＜0时表示负偏态，Sk＞0表示正偏态，Ku＞0表示峰“高尖”，Ku＜0表示峰“矮平”，当Sk和Ku都等于0或近似0时为完全正态或近似正态。本书中给出了这几年来MET每个考试的偏态值和峰值，以供参考。

8．常模(Norm)

与标准分(Standard Score)考试的常模指的是一种评价考生成绩优劣的参照标准或量表。在标准化考试中，为要对考生的水平作出较准确的估计，以及对考试做横向（如各

省之间)和纵向(如历年之间)的比较，一般把考生的原始分数按常模转换成标准分数(也称常模分数)。

常模一般用常模平均分和常模标准差来体现，并且是稳定的。这几年来MET考试定了两种常模，一种是各省的常模(常模平均分为60，标准差为15)，一种是联省的常模(常模平均分为100，标准差为20)。在表22中，可以查看这几年来各省考生MET成绩，在全体考生中所处的百分位置，表中各省的平均水平在60分处。例如：四川省(仅文科)考生1987年MET成绩在联省常模中是100分，在联省考生中恰好在50%的位置；山东省考生1987年MET成绩在联省(七省)考生中是110(约69.1%的位置)；陕西省(仅外语)考生1987年MET成绩在联省常模中是105(即59.9%的位置)。同样，用这个表也可以查看本校或某一考生在联省常模中的位置。如广东省某中学1986年高考MET成绩平均分为82分，则从表中可看出这个中学在四省中大约在96%的位置上。又如山东省1987年某考生MET成绩是90分，其在七省中的位置就是99.9%，而广东省1987年某考生MET得分100在七省中的位置只是在99.7%。

标准分是一个既能体现出学生的水平，又能体现出其所处的位置的一种分数制度。标准分的建立用的就是前面所提到的标准差。

其公式是：

$$Y = Y_o + S_o Z \qquad \text{公式(22)}$$

其中Y是标准分；Y_o和S_o分别是常模的平均分和标准差；Z为Z分数，表示考生的位置的一个量。

$$Z = \frac{x - M_t}{s_t} \qquad \text{公式(23)}$$

M_t 和 S_t 为全体考生的原始分的平均分和标准差。如MET87广东省的常模平均分Y。定为60，常模标准差S。定为15。又知道该年MET全省原始分平均分是73.43，标准差是23.53，某考生的原始分若是135，则其标准分

$$Y=60+15\times\frac{135-73.43}{23.53}=99$$

从公式(22)可见，标准分应该是无所谓“满分”的。但前几年考虑到其他科目没有实行标准分制度，故人为地将大于100分的分数都削平为100分。此外，以上的转换也是建立在分数分布正态的情况下，否则转换后标准分所表示的考生的位置是不准确的。

从1988年起广东省高考的各科分数和总分都实行标准分制度，为克服上面所说的不足，各科标准分转换采取了正态化转换，其步骤如下：

①将每个考生的原始分从高到低排序，我们可以得到每一分数上的考生人数及百分比(如某一科考生总数有90000人，低于61分(原始分)的有67086人，则61分的累计人数百分比为67086/90000=0.7454；又如，低于35分的有8670人，则35分的累计人数百分比为8670/90000=0.0963)。

②将每一原始分数对应的百分比转换成正态分布的z分数(见附录表一)。查0.7454对应得61分的z分为0.66，35分的z分数是−1.30。

③将z分数按公式(22)转换成标准分Y，广东省定出的常模平均分各科均为500，标准差为100，那么上两例中原始分35分的标准分为：

$$Y=500+100\times(-1.30)=370$$

原始分61分的标准分为：

$Y=500+100\times(0.66)=566$

有了这个分数我们也可以从附录表二中反查出其在全体考生中的百分位置。

总分的标准分的实现与各科标准分转换是一样的。步骤如下：

①将各科标准分乘上权重得出原始总分（语文、数学权重为1.2，生物权重为0.7，其余各科权重为1）如某理科考生数学、语文、政治、英语、物理、化学、生物的标准分数分别是610，555，520，560，578，612，550，则原始总分为：$610\times1.2+555\times1.2+520+560+578+612+550\times0.7=4053$

②将各考生总分原始分从高到低排序，得到每一分数的累计人数和百分比，如得到4053分以下的人数有46929人（理科总考生数是60000）则4053分的百分位是$40929/60000=0.7822$。

③将各分数的百分比查附录表一，如0.7822得z分数为0.78。

④将z分数按公式(22)转换为标准分Y，总分的常模平均分仍是500，标准差为100，所以上例中 $Y=500+100\times0.78=578$。该考生的总分标准分为578分。

标准分表示的是考生所处的位置，这可以从附录表二上查出。

如某考生总分得了578分，则从附录表二上可看出其大约处于全体考生的78.2%的位置上。同样某考生某一科标准分也反映了该考生在该科的百分位置，如某考生英语得了750分，数学得了550分，从附录表二可查出该考生英语科的百分位是99.38%，（即比99.38%的考生要好），而数学科的