

大学英语四、六级考试的分数解释 PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/215/2021\\_2022\\_\\_E5\\_A4\\_A7\\_E5\\_AD\\_A6\\_E8\\_8B\\_B1\\_E8\\_c83\\_215931.htm](https://www.100test.com/kao_ti2020/215/2021_2022__E5_A4_A7_E5_AD_A6_E8_8B_B1_E8_c83_215931.htm) 大学英语四、六级考试现在考生数已达到每年900万，由于命题的科学性、评分的一致性、组织的严密性、成绩的可比性和可解释性，大学英语四、六级考试在社会上已有广泛影响，并且已经在国际上受到重视。由于种种原因，四、六级考试结果可能对学生有重要影响，因此是一项高风险考试，组织这样一项考试是极其严肃的心理测量学科学工作，又是一项艰苦细致的、政策性很强的工作。大学英语四、六级考试是一种大规模标准化考试。人们常常把标准化考试等同于多项选择题，又把多项选择题等同于模拟试题集，于是指责标准化考试考不出学生水平等等，这是一种极大的误解。英语标准化考试是指采用教育测量技术对学生的英语能力进行客观、准确、公正的测量，并提供一种公认的客观标准，通过对学生语言运用的抽样检查来确定学生的实际语言能力。大规模标准化考试为了保证考试的信度和效度，对考试有很高的质量要求。大学英语四、六级考试在设计时力求做到：1) 评分具有客观性、一致性 2) 对教学具有正确的指导性 3) 分数具有可比性、可解释性 4) 施考条件对所有考生以及不同考次的考生一律平等 5) 便于施考 为了保证考试的效度，大学英语四、六级考试通过调查研究制定了考试大纲、定义了所测量的英语语言能力、编制了详细的考试内容规范，使历年测试在形式和内容上都稳定不变，保持了评价标准的稳定性。考试的最终结果是以分数的形式报道考生成绩，这就涉及分数的可解释性问题。如

果分数任意性很大而且又不可解释，不同考次的成绩因为没有经过等值处理而不可比较，那么这样的考试就不符合标准化考试的规范要求，就很难说是科学的考试。为此，大规模标准化考试一般不采用原始分（即答对题数）来报道考生成绩，因为这样做分数的信息量太小，通常要进行一系列的分数转换过程，使最后报道的分数不但信息量大而且具有可解释性。大学英语四、六级考试每次考试结束后到发布考试成绩之前，要经过一系列的统计处理，分为数据输入、数据处理、成绩计算及统计分析三部分，其流程如图1所示。其中数据输入后的数据处理包括三个部分：1) 客观题部分首先要作加权处理，加权反映了语言测试设计者对语言能力不同部分的不同要求，反映了对不同教学目标的侧重。大学英语四、六级考试中阅读能力、听力、词汇等各个部分每答对一题的得分都是不同的。在经过加权处理后，还要对客观题部分用IRT（试题响应理论）模型作等值处理，以滤除不同考次间题目难易变化对得分的影响。尽管大学英语四、六级考试的全部题目在实际使用前都要经过“命题审题试题项目分析复审构卷”的漫长过程，在实考前整个试卷的难易度均已得到控制，但预测是在考生中抽样进行的，与实考结果难免会有所差异，这些差异须要通过IRT等值处理加以滤除，不能因题目难易而影响考生得分，这样做也保证了不同考次之间分数的可比性（见图2）；2) 主观题部分，如作文，要通过调整以滤除阅卷员主观因素的影响。为了保证阅卷员的阅卷信度（包括阅卷员本人的评分一致性、阅卷员之间的评分一致性、阅卷点之间的评分一致性），大学英语四、六级考试采取了一系列措施，包括制定明确的评分标准、确定评分参照

卷(range-finders)、严格的阅前培训、阅卷过程随机抽查等等。在采取以上这些有效措施后依然存在的误差可以称为系统误差，大学英语四、六级考试开发了一套软件系统，根据考生在主、客观题上得分的相关性进行调整，以滤除系统误差。事实上，四、六级考试每次要在全国范围内动员1500~2000名教师阅卷，阅卷工作是一种辛苦的付出，阅卷过程有严格的质量控制措施，并且用计算机滤除阅卷员的主观因素，使阅卷信度达到0.87，这在国内各种大规模考试中是很不容易做到的。

3) 分数的正态化处理。大规模标准化考试的记分体制根据考试目的的不同通常采用正态分制或等级分制。这两种记分体制所表示的分数都带有大量信息，便于用户使用。大学英语四、六级考试的性质和目的，决定大学英语四、六级考试是一种尺度相关-常模参照考试(criterion-related norm-referenced test)，其记分体制采用以百分制为形式的正态分制。所谓常模参照考试是指以某一常模作为参照系来反映考生成绩的考试，首先要按照某一参照性群体建立常模，通常用均值和标准差来表示；某一考生的成绩是通过他的成绩在常模群体中与其他考生成绩比较的相对位置来表示的。常模参照考试所表示的分数是相对的，只有保持常模的稳定性才能保证分数解释的稳定性。尺度参照考试则要参照一定的尺度对学生进行考核。学生的成绩不是通过与其他考生的比较，而是通过与某种特定的尺度进行比较而得到解释的，学生的成绩必须达到某一尺度方能及格。这种特定尺度的选择或制定是尺度参照考试的关键，对于大学英语四、六级考试来说，这一尺度就是教学大纲中规定的教学要求。但不论哪一种性质的考试，分数的解释必须与某个参照系作比较，

才能表示出分数所代表的意义。而常模一经建立，其分数的意义和解释也是确定的，不会发生考生相互影响得分的问题。作为一种尺度相关常模参照考试，大学英语四、六级考试采用正态分制报导考生成绩，均值为72分，标准差为12分；为了适应我国传统，以百分制为形式，60分为及格，85分优秀。四级考试（CET-4）的报导分数有两方面的含义：1）判断考生是否达到了《大学英语教学大纲》所规定的四级教学要求；2）表示考生在标准样组即常模中的位置，常模由中国六所重点大学（北京大学、清华大学、上海交通大学、复旦大学、中国科技大学、西安交通大学）的近万名本科生组成。大学英语六级考试报导分的常模也由上述六所重点大学的本科生组成。大学英语四、六级考试的分数每次都分别经过等值处理，报导分的分数含义保持不变，四、六级考试分数相对于上述六校中一个学年的理论群体。这一记分体制为用户选拔人才提供了方便，例如若CET-4报导分百分位（优于...）CET-6报导分

60分	16%	62分	20%	66分	30%	69分	40%	72分	50%	75分	60%	60分	65%	64分	78分	70%	68分	78%	72分	82分	80%	85分	86%	76分	87分	90%	80分	93分	96%	85分	99分	99%	90分	100分	100%	100分
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------

某考生四级得85分，则相当于他的英语成绩在上述六所重点院校中优于86%的学生，而跟这位考生来自哪一所学校无关；又如若某考生六级得85分，则相当于在上述六所重点院校中优于96%的学生，也跟这位考生来自哪一所学校无关；同理，在选拔人才时，若要挑选相当于在上述六所重点院校中英语四级成绩最优秀的10%的学生，则可以定为四级87分；要选拔英语六级成绩最优秀的10%的学生，则可以定为六级80分。这样就为人才的选拔和合理

使用提供了方便。由于大学英语四、六级考试测量的科学性和分数的可解释性，目前境外和国外已不少大学采用大学英语四、六考试考试成绩作为录取大陆地区学生申请硕士、博士学位的英语成绩证明。大学英语四、六级考试口语考试的成绩报道则采用等级分制，其能力等级标准如图5所示。事实证明我国非英语专业的大学生有些已经具有很高的口头交际能力，凡获得A等和B等证书的学生都可以用流利的英语进行口头交际，达到相互沟通的目的；获得C等证书的学生也已经预备初步的英语口语交际能力。这些学生为我国广大的非英语专业的大学生学好英语树立了榜样。

等级 等级描述  
A等 能用英语就熟悉的题材进行口头交际，基本上没有困难  
B等 能用英语就熟悉的题材进行口头交际，虽有些困难，但不影响交际  
C等 能用英语就熟悉的题材进行简单的口头交际  
D等 尚不具有英语口语交际能力

以上这些措施不但保证了大学英语四、六级考试分数的科学性和可解释性，而且由于分数含有大量信息，也可以反馈到教学中帮助改进教学，同时也便于用人单位有效地选拔人才和合理地使用人才。（完）

全国大学英语四、六级考试委员会主任 杨惠中 2004.1. 100Test 下载频道开通，各类考试题目直接下载。详细请访问

[www.100test.com](http://www.100test.com)