

综合还是分立：写作试题评分标准与方法初探^{*}

南京大学 常晔

摘要：综合和分立分别指整体语言能力观和可分语言能力观指导下的两种对语言能力进行测试和评估的方法，鉴于这两种不同的语言能力观，语言测试理论中也形成了不同的评价标准。在语言测试理论发展到交际型测试理论的今天，虽然可分语言能力观略占上风，且新的测试理论与模式也层出不穷，评价机制不断翻新，但究其实质，也还是在综合和分立这两个根本出发点之间寻找平衡，寻求更切合实际的均衡点。本文从对两种语言能力观的理论探讨出发，依托高等学校德语专业四级考试大纲的规定，结合该测试写作阅卷的实际情况，对写作试题的评分标准问题做出新的探索，并提出了两步阅卷法的建议。

关键词：综合；分立；语言能力观；评分标准

作者简介：讲师，在读博士生，主要从事德语语言学及德语教学研究。电子邮箱：changxuan@nju.edu.cn

1 引言

综合和分立既是测试方法，又是评分方法，它们分别反映了语言测试理论中两种不同的语言能力观点。综合法源于将语言能力视作一个不可分割的整体，因此在评价它的时候只能采用综合性的整体评分方法；而分立法则认为语言能力是可以分割的，并依此对语言能力进行分项测试和分项评估。这两种语言能力观和语言测试评价方法虽然本身都只强调了语言能力与语言测试的一个方面，但在实践中只有把两者有机地结合起来，才能对写作试题的评分有实际的指导意义。

2 理论回顾

整体语言能力观的代表人物奥勒 (Oller, 1979) 认为，语言能力更像一种黏性物质而不是像机器一样可以被拆分成若干部分。这就是“不可分假说” (Indivisibility Hypothesis) 或者称之为“单一能力假说” (Unitary Competence Hypothesis, 简称为

UCH)。他将语言能力归结为一种期待语法生成系统,而这种期待语法生成系统本身的不可分割性决定了语言能力也是一种更像“黏性物质”的不可分割的整体。奥勒(Oller, 1979)在这种语言能力观的指导下,只需要设计出一种能整体反映被试者语言水平的试题就可以了(刘润清、韩宝成, 2000)。例如,通过写一篇命题作文或者就一个专题口头发表一定的观点,就可以比较全面而准确地反映被试者在一定情境下的交际语言能力了。

而巴赫曼(Bachman)则明确提出了语言能力不是单一的,而是多维度的。它由语言组织能力和语用能力组成。其中,语言组织能力包含语法能力和语篇能力,前者指生成或辨认语法正确话语的能力,而后者则是指能够理解话语中的主题内容以及把话语组织成口头或者书面语篇的能力;而语用能力则是指在具体的语言环境中理解和表达的能力,它包含了语义能力、功能能力和社会语言能力等三个方面。按照这种观点,对于被试者所拥有的语言知识和语言技能可以进行分别测试。

两种语言能力观不相上下,各有千秋。前者更注重语言能力的综合运用以及各知识点和单项技能间的关联,后者则更具可操作性。因此,目前绝大多数标准化测试实践中,人们普遍采纳了可分语言能力观的观点,并对被试者在听、说、读、写等方面的语言知识和语言技能进行测试,最后再综合评估,确定语言能力的高低。虽然可分语言能力观表面上占了上风,但同时也必须指出:一方面,构成语言能力的知识和技能并不是完全孤立、互不相干的。在测试实践中往往可以发现被试者不同技能之间存在着很大的相关性,以至于我们不得不承认语言能力是一个有机的整体。另一方面,在对于语言技能,特别是说和写等产出型技能的测试中,人们普遍还是采用了整体语言观的测试题型,如命题作文。综上所述,在测试实践中,测试设计者们实际上是在综合法和分立法之间游走,找出合理的平衡点。

语言测试中写作试题的优点在于能够更全面地考查学生的语言综合运用能力,但却也给测试评分工作带来了难度,同时在写作试题的评分标准上,同样存在着综合和分立的矛盾,也就是说,作文应该被看成一个有机的整体还是按知识及技能考点分解成几个单项。李筱菊(1997)指出,测试中主观题的评分有着综合和分立两种评分方法。综合评分法把作文的内容和形式看作一个整体,注重总体印象和整体感觉;而分立评分法主张将作文分解为主题内容阐述、语言表达、修辞等方面进行分项的评分。

3 实践探索

德语专业四级考试是对德语专业学生完成四个学期基础语言学习后的一次全

面考查和总结,因此阅卷工作需要非常谨慎,力求做到公平合理。在该测试中,写作试题作为测试的一个单项占总分的15%,是一项非常重要的语言技能测试。《高等学校德语专业基础阶段考试大纲》(1997)(以下简称《考试大纲》)要求被试者在大约35分钟的时间内撰写一篇不少于120个单词的短文,做到内容切题、结构合理、表达连贯、用词得当、语法正确、书写规范。

在写作评分中,针对这种主观性较强的题型,阅卷组既要尽量克服评阅员之间的不一致,又要克服评阅员自身的不一致。评阅员之间的不一致是由于有的评阅员过于严格,而有的评阅员则相对宽松。同时,不同的评阅员由于各人喜好的差别,对于同一篇作文的总体印象又会有所不同。而评阅员自身不一致则往往由于评阅员在连续评阅了数篇佳作后,会对一篇略显逊色的文章格外严格,反之亦然。

针对这一现象,就要制定出既有可操作性,又能较好地消除评阅员之间以及评阅员自身不一致的评分标准来。对此,《考试大纲》对“书面表达”部分的评分标准做了“两分法”的分项规定。其中,内容方面8分,语言方面7分。具体如下。内容方面:主要看其是否达到交际目的,即内容是否切题,是否包括全部内容要点,是否符合语篇类型特征,脉络是否清楚,结构是否合理。

分数	评分标准
8—7	a. 内容切题,包含全部内容要点,完成交际任务。 b. 脉络清楚,结构合理。 c. 符合有关语篇类型特征。
6—5	a. 内容基本切题,包含大部分内容要点,基本达到交际目的。 b. 脉络基本清楚,结构基本合理。 c. 基本符合有关语篇类型特征。
4—3	a. 内容基本切题,但有错误,内容要点有疏漏,尚能达到交际目的。 b. 文章条理性差,结构有几处不合理。 c. 语篇类型特征尚属基本正确,但有明显错误。
2—1	a. 内容明显离题,只对少部分内容要点有所陈述,未能达到交际目的。 b. 结构明显有错误,条理性甚差。 c. 不符合有关语篇类型特征。

语言方面: 主要依据词汇、语法和正字法等项内容。

分数	评分标准
7	a. 词汇、语法、正字法合乎规范, 有个别错误, 不影响表达思想。 b. 语句通顺, 表达连贯。
6—5	a. 词汇、语法、正字法合乎规范, 有些错误, 但不影响表达思想。 b. 语句基本通顺, 表达基本连贯。
4—3	a. 词汇、语法、正字法方面多处出错, 比较明显影响表达思想。 b. 语句多处不通顺, 但尚能表达思想。
2—1	a. 词汇、语法、正字法方面的错误甚多, 严重影响表达思想。 b. 语句不通顺, 不知所云。

不难看出, 大纲规定的内容方面既包含了达到交际目的的内容要点, 也包含了文章结构和语篇类型特征这样的形式要素。而且在实际操作当中, 由于每次写作部分的题目不同, 文章从语篇类型到内容要点都千差万别, 所以制定更为细化的具体评分细则既可以更好地缩小评阅员自由裁量的空间, 同时也更利于评阅员把握, 从而减少评阅员在评分中出现大起大落的现象。因此, 2003年首次统一阅卷时, 便将语篇类型要素从考试大纲规定的评分标准的内容方面中单独分离出来, 迈出了“三分法”评分标准尝试的第一步。当时, 很多考生对于语篇类型的掌握较弱, 针对当时的现象, 开始制定的评分细则中内容方面占6分, 语言方面占6分, 语篇类型占到3分。在执行了一段时间后, 各高校加强了语篇类型知识的教学, 考生作文中语篇类型方面的错误大大减少, 却反而开始有重形式轻内容的倾向。因此, 评分细则又做出了相应的微调, 改为: 内容方面(7分), 语言方面(6分)和语篇类型方面(2分)。

以上三个方面, 内容方面指作文为了达到相应的交际目的而必须涉及的话题的各个内容点, 视作文具体情况而定夺分数。语言和语篇所涵盖的基本点是一致的。在语言方面要求正字法符合规范、时态正确、用词得当、语法正确、语言连贯通顺。在语篇方面, 要求文章要有标题, 要有开头、结尾以及合理的段落划分, 语体还要符合语篇类型, 如在图表描述型语篇中就不可以加入个人的观点, 即要注意语言表达中的“去主观化”。

德语专业四级考试写作部分的评阅工作与一般课堂教学中的写作评阅不同, 应该采取更为谨慎的态度, 其分数对于被试者也有着重大的意义。因此不断完善评分标准是个很重要的课题, 综合法与分立法究竟孰优孰劣, 目前在理论界尚无定

论。笔者却认为，将综合与分立两种评分标准结合起来运用，互相取长补短，合理把握分寸才能更好、更客观地评价一篇作文。综上所述，笔者认为理想的评阅方法应该是含有先综合后分立两道程序的两步阅卷法，具体如下：

第一道程序：一审评阅员通读全文，根据第一印象给出总体评估意见，不给出具体得分，仅打出优、良、中、差的不同等级。在这一道程序中评阅员不做细致的定量处理，不去具体计算得分点的多少和语法错误出现的频度。笔者建议可以采用如下的标准：

级别	标准
优 (15—13.5)	语言错误很少，在通读过程中完全不影响对文章内容的理解，语篇结构合理，符合逻辑。文章很好地达到了交际目的。
良 (13—11)	语言错误少，在通读过程中除个别地方外，基本不影响对文章整体内容的理解，语篇结构基本合理，符合逻辑。文章基本达到交际目的。
中 (10.5—7)	在通读过程中常常难于理解文章内容，语言有较多的语法和语用错误，语篇结构不合理，逻辑性差。文章部分达到交际目的。
差 (6.5以下)	有很多的语法和语用错误，在通读过程中基本不能理解文章内容，文章没有语篇特征和结构。文章没有达到交际目的。

第二道程序：二审评阅员细读全文，根据三分法的评分标准，分项打分，在一审等级分的基础上打出具体分数。如果两道程序的评阅员出现不一致的情况，则由另外的一名评阅员进行终审。

全国德语四级考试中心自2008年以来，采取了上述先综合再分立的两步阅卷法。虽然阅卷的速度较以往有所下降，但阅卷质量得到了提高。

4 分析和评价

从2003年起的阅卷实践中，作文评阅小组先做出了将考试大纲规定的两分法评分标准改为三分法的尝试，后又将阅卷步骤改为先综合再分立的两步阅卷法，期间还对三分法评分标准的分数权重做了微小的调整。所有这些探索都体现了评阅小组在阅卷实践中，不断验证理论，又不断从实践中获取新认知的过程。

分析一下考试大纲规定的两分法的评分标准与实际阅卷中执行的三分法的评分标准，就不难发现，《考试大纲》规定的两分法评分标准虽然区分了内容方面与

语言方面,但内容方面仍然包含有语篇类型的成分。这样的评分标准实际上体现了把文章的内容和形式结合在一起,进行综合考量的精神。此外,其评分标准过于大而化之,不便于量化。如果只根据这个标准来打分,评阅员就有了更大的自由裁量空间。所以说,《考试大纲》规定的标准带有非常明显的综合评分的倾向,体现了整体观的思想。而在阅卷实践中制定的三分法评分标准就带有更多的分立法评分的特征,把内容和形式更加明确地分开进行考量,更多地倾向于可分语言观的思想。另外,阅卷实践中执行的三分法基本上采取了量化给分的方法,减少了评分员主观因素的影响。如对文章的内容点做出比较明确的规定,无论是偏严还是偏宽的评分员,只要做到按点给分,就基本可以避免评分员之间出现较大的出入。

然而,在实际的阅卷过程中,虽然分立评分法可以比较有效地限制评阅员的自由空间,克服主观性带来的差异,但在执行该标准的时候也发现,如果绝对严格地按点给分,未必就能合理地对被试者作文给出公正的评分。作文与其他项目相比有其特殊性,它不仅考查语言的形式是否正确,同时还考查被试者是否能够通过文章来表达思想、完成交际目的。所以说,文章究其本质还是内容与形式有机结合的一个整体,单纯的分立评分标准有割裂文章内容与形式的倾向,缺乏对文章的总体评价。例如,有的被试者语言能力较弱,仅仅搭起一篇文章的空架子,并罗列一些内容点,但总体结构混乱、文不成句。如果评阅员根据总体印象,进行综合评分,只会给出一个很低的分数;反之如果采用单纯的分立评分法,单项得分点逐个相加,却可以得到较高的分数。另外一种情况则是,一些总体印象很好的作文,却会因为单项方面的一些细小不足之处而被扣掉比较多的分数。

总的来说,综合法有助于对文章进行总体把握,是一种定性的分析;而分立的方法建立在具体的分步分析之上,更为精确,有助于对文章的好坏进行定量分析。总之,采用综合法评分,被试者的分数差异更大,且具有更多的主观性和随意性;而采用分立法,则会使得分数相对趋中。正是基于上述分析,为了避免单一综合法与单一分立法各自的缺点和不足,写作部分评阅组最终于2008年开始用“评分标准三分法”和“先综合再分立的两步阅卷”步骤相结合的阅卷方式代替了《考试大纲》中较为笼统的两分法评分标准。

作文内容方面和语篇类型分数权重的微调体现了德语专业四级考试对于德语专业基础教学的检验与反拨作用。《考试大纲》规定,专业四级考试要检验考生在基础阶段的语言能力掌握情况,并检验基础阶段教学大纲的实施情况。根据教学的实际情况以及德语专业基础阶段学生的总体情况,对测试评分标准的分数权重做细微调整也可以提高测试的反拨效度。事实证明,自2003年以来,被试者作文语篇类型知识的掌握水平明显提升。特别是自2008年以来,45所A类院校的作文平均成绩

逐年上升,分别为2008年9.2分,2009年9.83分,2010年10.46分(全国高校德语专业基础阶段测试组,2008—2010)。

多年来,特别是近三年来的阅卷实践还表明,该评阅方法和评分标准受到了作文评阅组中来自全国各高校德语专业评阅教师的一致认可;此外,它能够更加客观和全面地对学生的写作水平予以评价,给出的分数更具合理性和说服力。

5 结语

诚如李筱菊(1997)所说的那样,“语言测试既是科学,也是艺术”。综合和分立看似是两种截然不同的语言能力观和评分标准,但实际上任何一种观点和方法都不是绝对的。对于被试者的一篇作文,要能够做到定性正确,定量准确,必须将综合与分立两种评分标准和方法结合在一起,找到最佳的平衡点,使两种方法各自的优势相得益彰。

此外,写作部分的评阅工作时间紧、任务重,采用两步阅卷法后,随着考生的逐年增多,工作量势必还会增加。所以在今后的写作评阅工作中,集中商定具体评分标准,集中对评阅员进行培训,使用熟练的评阅员是在实践当中需要注意的问题。

注释

* 本文为教育部人文社会科学研究项目“中德跨文化语篇中他者与自我的互融建构”(项目编号:09YJA740061)的阶段成果。

参考文献

Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Oller, J. 1979. *Language Tests at School*. London: Longman.

李筱菊, 1997,《语言测试科学与艺术》,长沙:湖南教育出版社。

刘润清、韩宝成, 2000,《语言测试和它的方法(修订版)》,北京:外语教学与研究出版社。

全国高校德语专业基础阶段测试组, 1997,《高等学校德语专业基础阶段考试大纲》,上海:上海外语教育出版社。

——2008, 2009, 2010, 全国专业德语四级考试总结及试题分析。