

# 写作模糊综合评价研究<sup>\*</sup>

## ——以全国高校俄语专业八年级水平测试为例

上海外国语大学 许 宏 付 钰

**摘 要** 写作评价具有较强的主观性和模糊性,如何保证评分结果一致且可靠始终是高风险语言测试亟须解决的难题。本文基于对现行写作评分方法的分析,主张运用模糊综合评价法对写作评分标准进行量化处理,并以全国高校俄语专业八年级水平考试为例构建写作质量评估的模糊数学模型,分析该模型的优缺点,以期提供一套更加科学、客观、准确、全面的评分方法。

**关键词** 语言测试 写作 评分方法 模糊综合评价

**中图分类号** H319.3 **文献标识码** A **文章编号** 1001-5795(2021)03-0070-0006

外语测试与评估

### 1 引言

Alderson *et al.* (2000: 105) 认为“如果一个测试的评分是无效和不可靠的,那么之前为建造一个‘高质量’工具而进行的所有其他工作都将是浪费时间。”由此可见评阅工作的重要性。评阅环节的信用与测试的总体信用直接相关,如果评价不客观、不合理,不仅无法对教学起到推进作用,还可能会偏离预先设定的目标。制定科学的评分标准和评分方式,有助于提高测试结果的可靠性和客观性,为各高校推动教学大纲落实、促进教学改革等提供有效信息反馈。邹申(2012: 16)指出“写作测试是语言测试的重要组成部分,故当人们谈及考试效度时必提及写作测试效度。”写作评价具有较强的主观性和模糊性,其评判结果受多因素影响,模糊综合评价法可以为写作任务的评估提供一个实践性强且性能优良的评价模型,以保证评分结果一致与可靠,提高写作测试的信用、效度。模糊综合评价法是“在多级估量法上发展起来的一种综合评价法”,“实现了用‘亦此亦彼’的模糊评价来确定各因素的隶属规律,用 $[0, 1]$ 中的连续值取代了二值逻辑或多值逻辑,使评分者能根据自己的印象选择最合适的隶属度,充分表达自己对评判对象的印象”(陈希镇、梁飞豹,1993: 24)。本文以全国高校俄语专业八年级水平考试写作部分的评分方法为基础来建构模糊综合评价模型,以期提供一种更科学、客观、准确、全面的评价框架。

### 2 现行写作评分方法及其存在的问题

传统的写作评分有三种方法:整体评分法、主要特征评分法和分项评分法(Brown & Abeywickrama, 2013)。整体评分代表了评分者对写作任务完成情况的总体评价;主要特征评分是整体评分的一种变体,将一篇文章主要目的或特征的实现视为唯一评价要素,聚焦“学生在一个狭窄的定义范围内的写作能力”(Weigle, 2002: 110);分项评分指将考生的作文分成若干个子类别(组织结构、语法等),并为每个子类别分别进行评分。有学者指出,主要特征评分“是针对某个特定写作任务设计的,分数的意义不能外推到其他任务”,多适用于母语写作,外语写作中很少用到(李清华、孔文,2011: 18)。总的来说,“外语测试中写作任务的评判通常采用整体评分法或分项评分法”(孙海洋、韩宝成,2013: 48)。

按照内容是否切题、丰富,思想表达是否清晰,语言是否流畅、通顺,用词和句式是否恰当、多样等要求,整体评分法将评价结果分为几个等级,根据宏观层面划定等级分数线。在批阅时,评分者会参照学生的语言表现与上述条件的适切程度,根据主观印象和自身经验对学生的表现快速地进行整体判断,确定学生的成绩偏向于哪一个等级或分数阈值,再对比各个等级描述语的要求进行单独判断,核查细节错误,完成最终评定。

分项评分法将写作任务要求分解成不同维度,每个维度下都有反映考生语言表现的不同评分因子,以

★ 版权所有 文责自负 ★

及对应评价任务完成质量的不同等级水平,每个评分因子所属的各个等级水平应有表现的描述语,评分者按照作文表现与描述语的匹配情况给学生进行评分,最后取各维度评分的总值或均值作为最终成绩。

现行的写作评分方法存在以下问题:第一,分数不易解释和校验。整体评分中,得分反映了评分者对多个评价标准的综合判断,即使分数相同也可能代表不同的意义。分项评分中的评分维度有限,经常出现多个评分因子合并评判的情况。例如,雅思考试使用“写作任务完成/回应情况”“连贯与衔接”“词汇丰富程度”“语法多样性及准确性”四个维度,其中,对“语法多样性及准确性”进行考查时会至少考虑语法多样性和语法准确性两个要素。如果等级划分过少,则无法有效区别学生在不同等级上的表现,所给分数很难进行校验。相反,等级划分过多会给评分者造成认知负荷。

第二,评价等级的界定比较含混。整体评分与分项评分都需要遵循采用多层次描述的原则,实际上很难区分“内容切题”“内容基本切题”“内容不切题”三者的界限(孙飞燕,2016:84)。标准模糊带来的影响在评分者进行评分的过程中会被进一步放大,加大评分误差。

第三,“宏观评级,微观调整”的阅卷方式容易导致评分者“习惯于找错并根据错误数量扣分,而全然不顾文章的思想内容和表达质量”(史铁强,2008:27),得分往往与作文的表面特征高度相关(李清华、孔文,2011:19),有的语篇的陈述毫无思想和内涵,甚至是套用文字模版,但语言错误不多,便能取得不错的高分;有的语篇逻辑结构性强,但语言错误较多,得分则会偏低。Heaton(2000:149)认为,这种计算错误的方法忽略了写作的根本目的——思想表达。它只集中在写作任务的消极方面,容易束缚考生的思维和创造力,迫使他们过分谨慎,难以提高写作水平。

第四,评分标准的设置简单、宽泛。国内很多高风险外语测试,如全国高校俄语专业八级水平测试(以下简称“俄语八级测试”),未提供评价写作文本结构、论据充分性以及语篇逻辑性和连贯性的依据,对语言错误数量和程度也没有明确说明。错误性质不同、程度不同,原则上应该扣不同的分数,但总体上错误的严重程度只是评分者的主观判断,错误之处的评判缺少统一标准。

总的来说,现行的印象式评分理念缺乏客观性,等级描述过于原则化,评分实践中难以兼顾语言和内容,分值的权重设计不够科学,无法从篇章层面审视学生语篇的优缺点,无法让评分者真正一致地使用并解释等级描述

语,也无法得知评分者是否严格依据等级标准进行评分。

### 3 写作评分方法的新设想

笔者认为,解决以上问题的根本方法是采取统一化、客观化的评分标准。这种评分标准可以通过模糊综合评价法来构建。模糊综合评价法是一种常见的基于模糊数学的综合评价方法,以“模糊集合”与“隶属度函数”两个概念为核心。“模糊集合”将评价指标与评价等级转化为“从0到1的不同程度的相对判定”(李金波,2010:117)。“隶属度函数”用于描述模糊集合,根据隶属度理论把定性评价在形式上转化为定量评价,将不确定性转化为确定性,“对受到多个因素制约的事物或对象作出一个总的评价”,“由于从多方面对事物进行评价难免带有模糊性和主观性,采用模糊数学的方法进行综合评判将使结果尽量客观从而取得更好的实际效果”(张铁男、李晶蕾,2002:132)。

目前的评分方法都属于精确评分,即给出具体的得分数字,但事实上,写作评分具有很强的主观性,评价过程具有很强的模糊性和不确定性。根据模糊的评分标准给出精确的评分数字的方法则不太合理,应该采用模糊评分的方法。基于此,指标体系结构多层次的评价系统适合运用模糊综合评价法,创建模糊综合评价模型。张文忠、郭晶晶(2002)指出,如果将评分标准看作测量学生某一语言能力的尺子,那么现行的精确评分法是将学生的得分(即学生的语言表现)与尺子上的某一刻度对应起来,而现实中,学生的语言表现与特定刻度(水平)的语言能力并不存在一一对应关系。在这种情况下,模糊综合评价法的优势就得以显现:评分者提供的是顺序数据,而不是数值数据,评价结果以向量的形式出现,而不是具体的数字,所考虑的评价因素对应评价标准的一个面,而不是某个点,从而能够较准确地反映语言能力的模糊状况。模糊综合评价法用数字量化评价标准的程度较深,需要用传统的数学方法进行分析 and 处理。评分者依据评分标准对学生的写作能力进行模糊综合评判,评判的结果表现为模糊评价矩阵,这种评分方法也可最终得出精确的分数,但是分数是由模糊合成运算形成,而不是评分者直接给出。

### 4 模糊综合评价法——以全国高校俄语专业八级水平测试写作任务为例

#### 4.1 全国高校俄语专业八级水平测试写作任务的评阅现状

根据近几年阅卷教师的反馈,议论文采用整体评

分法,评分标准包括内容、结构、语言等维度,由阅卷教师在预定的评价等级内给予分值,评分量表不具备约束评分者达成一致性评价的功能,对写作题的评分质量产生较大影响,亟须制定一套新的评分量表与评分方式来详细划定评分教师的采分点,施行更为科学的分数计算方法,提高评分信度。

俄语八级测试采用计算机阅卷系统,标准设置过于简略,在降低误评率、避免评分标准宽严不一、减小评分者主观因素影响上的作用不大。虽然评分小组为了提高写作测试的评分质量,在阅卷环节设计了误差调控机制,对每份试卷的作文部分采取分立二评,甚至三评、四评,并通过后台监测程序“对给分过宽或过严的评阅人予以及时提示和纠正”(刘素梅,2007:54),但还是不能弥补评分标准本身存在的缺陷。这种评分模式很可能导致有些评分教师为了防止三评、四评的麻烦和风险,刻意避免给出极端分数,即使面对的作文质量很高或很低,也不会轻易给出很高或很低的分数,导致评分明显趋中,削弱了作文测试的区分功能,在一定程度上损害了考试的公正性与考生的利益。

#### 4.2 模糊综合评价法的评分量表

模糊综合评分法首先基于分项评分的理论,这与俄语八级测试的整体评分不同,它反映的是所测能力的构念,认为语言能力是可分的,“反映了一种分割式语言观”(李筱菊,1997:111)。通常,题量多的测试比题量少的测试信度高,但主观性测试很难增加题目数量,持这种观点的学者便将评分维度视为试题任务的因子,认为评分维度的增加会提高评分的信度(Carr,

2000;孙海洋、韩宝成,2013)。由此,大部分学者主张,分项评分更适用于外语写作测试,且信度高于整体评分。分项评分因子与所测能力模型中的各个子能力应保持一致性,每个分项都有对应的子能力(Bachman & Palmer,1996)，“这是保证量表效度,提高量表科学性的关键”(郑美玲,2019:20)。模糊综合评分与分项评分一脉相承,承认能力可分,使用多层次评价指标与总体写作能力模型中的各个子能力相对应。基于多因素能力模型的理念,结合史铁强(2008)提出的参考雅思、托福写作评分标准的建议,本文通过描述语之间的优化组合构建了评估作文质量的量表,见表1。

大部分评分量表都将语法和词汇作为评分标准中的关键类别进行单独提列,本文将二者合并为一个语言维度,力求改变当前俄语八级测试写作评分构念,将重心从细节挑错转向整体语言形式运用效果。

在评阅开始前,评分者要充分熟悉量表的分项等级描述语和判定方法,了解写作测试的任务特征和能力要求。写作能力属于表达意义的能力,各个评价指标与写作能力模型的一致性对应为:内容维度考查学生能否有效地针对写作主题和任务传递主题知识,能否克服写作内容贫乏、空泛等问题;语言维度旨在考查语法和词汇的准确性与丰富性;篇章维度旨在考查学生能否准确、有条理地表述思维过程,“运用多种衔接手段组句成篇”(郑美玲,2019:23),且保证连贯性,体现了对学生的分析概括能力、判断力、逻辑推理能力的重视;表达维度旨在考查学生的思想表达能力、思维能力与写作规范(学术写作要突出客观性,语态方面必

表1 俄语八级测试写作分项评分量表

等级 \ 分项	内容	语言	篇章	表达
优	完全围绕题目展开,论述充分,论证合理,例子生动、多样、切题。	没有语法错误,复合句和简单句交替使用;用词准确,注意使用同义词、近义词,以避免重复。	整篇结构清楚、流畅,段落之间、句子之间有丰富、连贯的衔接手段,逻辑性强。	语态得体、书写规范的基础上充分展示思想与个性,表达真情实感;自然运用修辞手段。
良	围绕题目展开,论述和论证平淡,例子单调。	有少量语法错误但不影响表达,句式多样;用词偶有轻微拼写、搭配错误,但总体表达手段丰富。	介于两者之间。	介于两者之间。
中	内容与题目有一定关联性,论述勉强,论证不足,无举例或例子不切题。	有较多语法错误,句式单一;词汇存在多处误用,用词简单。	结构基本合理,个别地方连贯性和逻辑性欠缺,衔接手段单一。	用语基本得体、规范,但有些地方思想表达不清楚;无任何修辞手段的运用。
差	通篇论述与题目毫无关联。	有严重语法和词汇错误,基本是简单句,或使用的复合句基本都有错误;用词过于简单、重复。	结构混乱,前言不搭后语,逻辑性差,未使用任何衔接手段。	文体风格不适宜,思想难以理解。

须得体),看重在写作中反映自己的个性化声音,力图对学生的语言措辞有更高阶的要求,强调形式运用能力与意义表述能力的整合(郑超、常春娥,2014:514)。

雅思和托福考试对文章的论证内容、篇章结构、逻辑衔接、语言表达方面都作出了要求,“全方位客观评价写作答题的总体质量,而不是注意小错误和单一弱点”,“不是按语言要素的错误数目扣分”(史铁强,2008:27)。本量表的内容及设计理念与雅思、托福考试一致。

#### 4.3 模糊综合评价法的基本原理

确定评价因素集。“评价因素”指对学生的作文进行评分时需要考虑的因素。根据上述评分量表,评价因素集由两个层次的集合组成:一级指标因素集  $U = \{U_1, U_2, U_3, U_4\} = \{\text{内容}, \text{语言}, \text{篇章}, \text{表达}\}$ ; 二级指标因素集  $U_1 = \{U_{11}, U_{12}, U_{13}\} = \{\text{切题}, \text{论述}, \text{举例}\}$ ,  $U_2 = \{U_{21}, U_{22}, U_{23}, U_{24}\} = \{\text{语法准确性}, \text{句式多样性}, \text{用词准确性}, \text{用词丰富性}\}$ ,  $U_3 = \{U_{31}, U_{32}, U_{33}\} = \{\text{结构有序}, \text{组织连贯性}, \text{衔接手段}\}$ ,  $U_4 = \{U_{41}, U_{42}, U_{43}, U_{44}\} = \{\text{思想}, \text{语态得体性}, \text{书写规范性}, \text{修辞}\}$ 。

确定评价等级集。鉴于各种质量评估基本采用序化指标,本文将评价等级集确定为  $V = \{V_1, V_2, V_3, V_4\} = \{\text{优}, \text{良}, \text{中}, \text{差}\}$ ,这与上述评分量表相统一,表示对学生在每个指标下的写作质量评价分为“优秀”“良好”“中等”和“差”四种水平。

确定权重集。由于  $U$  集中每个评价因素在评估作文质量时的受重视程度不同,因而对分数结果的影响程度理应不同,需要赋予每个因素不同的权重。根据隶属度原则,某一评价指标的权重越大,写作成绩对它的依赖程度越高。模糊权重集  $A$  是基于  $U$  建立的,由  $A = \{A_1, A_2, A_3, A_4\} = \{0.25, 0.3, 0.2, 0.25\}$  构成,而子权重集也以子评价因素集为基础,确定模型的权重系数是  $A_1 = \{0.4, 0.35, 0.25\}$ ,  $A_2 = \{0.25, 0.25, 0.25, 0.25\}$ ,  $A_3 = \{0.25, 0.45, 0.3\}$ ,  $A_4 = \{0.4, 0.2, 0.15, 0.25\}$ 。评价指标与权重分配的合理性决定了这种评分方法的科学性、准确性。权重系数的确定大多通过专家咨询法和层次分析法实现(李金波,2010)。层次分析法适用于指标层次和数量较多的评价体系,本文选用专家咨询法。

建立模糊评价矩阵。“模糊评价矩阵”是评分者从各个评价指标出发,对考生的作文根据不同的等级水平作出评判的结果组成的一个模糊评价模型。本文的模型由多级模糊评价构成,首先从低级层次开始,进行一

级模糊评价,然后向上层叠加完成二级模糊评价,得出最终写作成绩。模糊评价矩阵基于集合  $U, V$  建立,用于表示从  $U$  到  $V$  的模糊关系:  $R = (R_1, R_2, R_3, R_4)$ ,

$$\text{其中 } R_1 = \begin{bmatrix} r_{11} & \cdots & r_{14} \\ \vdots & \ddots & \vdots \\ r_{31} & \cdots & r_{34} \end{bmatrix}, R_2 = \begin{bmatrix} r_{41} & \cdots & r_{44} \\ \vdots & \ddots & \vdots \\ r_{71} & \cdots & r_{74} \end{bmatrix}, R_3 = \begin{bmatrix} r_{81} & \cdots & r_{84} \\ \vdots & \ddots & \vdots \\ r_{101} & \cdots & r_{104} \end{bmatrix}, R_4 = \begin{bmatrix} r_{111} & \cdots & r_{114} \\ \vdots & \ddots & \vdots \\ r_{141} & \cdots & r_{144} \end{bmatrix}, \text{以上算子}$$

表示所有参评人员对一篇作文的评价结果。

计算模糊综合评价值。利用普通矩阵点乘方法,计算公式为  $B = A \cdot R = (B_1, B_2, B_3, B_4)$ 。其中  $B$  表示将  $A$  转换为考生的写作表现在  $V$  上的模糊集合,  $B_1, B_2, B_3, B_4$  分别代表针对“内容”“语言”“篇章”“表达”等评价因素进行的一级模糊综合评价,由四个结果组成二级模糊评价矩阵  $R$ ,再完成二级模糊综合评价  $B$ 。符号“ $\cdot$ ”表示广义的合成运算,其中的模糊合成算子采用加权平均模型运算,并对计算结果进行归一化处理,根据最大隶属度原则,  $B$  取  $\text{Max}\{B_1, B_2, B_3, B_4\}$ ,由此得出学生综合写作能力的级别,为了进一步计算考生的最终写作得分,本文参照俄语八级测试已有评分标准等级对一级评价指标的评价等级进行赋值:  $V = \{V_1, V_2, V_3, V_4\} = V_T = \{15, 12, 9, 3\}$ ,并根据  $Y = B \cdot V_T$  计算考生的写作成绩(张国立等,2011;杨纶标等,2011;蒋泽军,2015)。

#### 4.4 模糊综合评价法的具体步骤

本文设定模糊综合评价法下的俄语八级测试沿用两位评分教师共同批阅一份试卷的机制,如针对一级指标“内容”中的“切题”这一项二级指标,两位教师分别对一份试卷做出“优”和“良”的判定,那么这项指标的行向量为  $(1/2, 1/2, 0, 0)$ ,即  $(0.5, 0.5, 0, 0)$ ,其他指标同理可得。

为详细说明模糊综合评价的操作步骤,笔者假设两位评分者对某一考生的作文进行判断,见表2。以一级指标“内容”为例,由表2可知“内容”模糊评价矩阵  $R_1$ ,与“内容”权重集  $A_1$  进行低级层次的模糊关系运算可得一级模糊综合评价  $B_1$ 。  $B_1 = A_1 \cdot R_1 = (0.4, 0.35,$

$$0.25) \cdot \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = (0.625, 0.375, 0, 0)。$$

同理可得  $B_2 = (0, 0.625, 0.375, 0)$ ;  $B_3 = (0.275, 0.725, 0, 0)$ ;  $B_4 = (0.1, 0.45, 0.325, 0.125)$ 。

基于一级模糊综合评价的计算结果,可以生成二级模糊评价矩阵  $R$ ,与已知的评价因素集  $U = (\text{内容}, \text{语言}, \text{篇章}, \text{表达})$  及其一级指标权重集  $A$  进行算子点乘,得到对考生写作任务的二级模糊综合评价  $B$ 。

$$\text{计算结果为 } B = A \cdot R = A \cdot \begin{bmatrix} B1 \\ B2 \\ B3 \\ B4 \end{bmatrix} = (0.25, 0.3, 0.2, 0.25)$$

$$0.25) \cdot \begin{bmatrix} 0.625 & 0.375 & 0 & 0 \\ 0 & 0.625 & 0.375 & 0 \\ 0.275 & 0.725 & 0 & 0 \\ 0.1 & 0.45 & 0.325 & 0.125 \end{bmatrix} = (0.23625, 0.53875, 0.19375, 0.03125)$$

表示该考生写作成绩是“优”的程度为 23.625%，“良”的程度为 53.875%，“中”的程度为 19.375%，“差”的程度为 3.125%。按照最大隶属度原则， $\text{Max}(0.23625, 0.53875, 0.19375, 0.03125) = 0.53875$ ，该考生写作任务的表现应评定为“良”，进一步计算其模糊综合评价值为： $Y = B \cdot V_T =$

$$(0.23625, 0.53875, 0.19375, 0.03125) \cdot \begin{bmatrix} 15 \\ 12 \\ 9 \\ 3 \end{bmatrix} = 11.84625$$

$\approx 12$ ，因此该考生的最终写作得分为 12 分。

表 2 写作部分的模糊综合评价结果

序号	一级指标	权重	二级指标	权重	等级水平			
					优	良	中	差
1	内容	0.25	切题	0.4	0.5	0.5	0	0
			论述	0.35	0.5	0.5	0	0
			举例	0.25	1	0	0	0
2	语言	0.3	语法准确性	0.25	0	0.5	0.5	0
			句式多样性	0.25	0	1	0	0
			用词准确性	0.25	0	1	0	0
			用词丰富性	0.25	0	0	1	0
3	篇章	0.2	结构有序	0.25	0.5	0.5	0	0
			组织连贯性	0.45	0	1	0	0
			衔接手段	0.3	0.5	0.5	0	0
4	表达	0.25	思想	0.4	0	0.5	0.5	0
			语态得体性	0.2	0.5	0.5	0	0
			书写规范性	0.15	0	1	0	0
			修辞	0.25	0	0	0.5	0.5

表 2 数据反映的是两位评分者的评价过程,是得出综合写作评价的依据,应报告于后台或计算机,而不是评分者。简言之,所有评分教师进行打分的量表,等级评价部分需要评分教师填写,允许设计成由评分教师自行选择在每一行的四个等级中打“√”来完成评估。

#### 4.5 模糊综合评价法的优点

第一,评分信度高。模糊综合评价法提供了客观化的评价程序,评分信度提高。已有测试对分数进行简单的加权处理,忽略了不同评分者及同一评分者在不同时间段的标准差异。模糊综合评分法利用模糊合成的原理,很大程度上降低极端因子的影响,保证不同评分者对不同应试者和题项在评分指标上达成一致,缩小个人评价之间的差距,减小测评误差。同时,它能避免主观臆断。

第二,操作较简单。评判时,评分者只需要依据学生的写作表现在特定指标中选择合适的等级水平即可,成绩的统计及计算工作均由计算机按照设定的算法执行,评分者并不知道考生的最终得分,极大程度地降低了人为因素的干扰。

第三,结果易解释。评判标准易掌握,分数易解释。相较之下,“优”“良”“中”“差”的等级更易把握,评定范围仅限于这四个级别,大大提高了评分的准确性和一致性。同时,这种序化指标更符合人脑的思维规律,减轻了评分老师的评阅心理压力(张文忠、张姮,2008)。

第四,权重多样化。评判时权重分配可多层次、多样化。写作任务考查的侧重点不同,应分配不同的权重加以评判。模糊综合评分法不仅能给不同评分维度加权,还能给各维度内部的子因素加权,使评分结果更具准确性。

第五,兼容性较强。模糊综合评分法可以将整体评分与分项评分结合起来,分别做整体与分项的模糊评价,甚至可以在评价因素集中增加“整体印象”维度,将各种评分方法的优势融合在一起,弥补当前评分标准的不足(金檀等,2008)。

#### 4.6 模糊综合评价的缺点

首先,评价指标及其权重体系的确定需要由相关测试专家完成,即使各个指标的权重通过基于数据和实验的层次分析法来决定,其中也有专家的人为干预,因此包含了一定的主观随意性。若评价指标考虑不全

面,有可能影响写作测试的区分度。此外,模糊综合评分法本身不能规避评价指标之间相关、交叉甚至重复所导致的评价结果混乱,因而科学、合理的评价指标预选环节尤为重要。

其次,评价指标的权重根据隶属度原则而定,理论上权重系数都在 0 到 1 之间,所有权重系数之和为 1,不存在单个权重是 0 或 1 的情况。因此,模糊综合评分法理论上排除了满分和零分的极端案例。此外,给评价等级集 V 赋值时,“优”和“差”的数值决定了两极情况下(两位评分员给一份试卷在所有指标项目上都评“优”或“差”)考生能得到的最高或最低分数,因此,赋值可根据具体情况和考纲要求来决定学生作文被允许的有效成绩范围。

## 5 结语

本文从现行写作评分方法出发,综合分析了整体评分与分项评分中存在的问题,主张运用模糊综合评价法对写作评分标准进行量化处理,并以全国高校俄语专业八级水平测试写作任务为例,构建一套写作模糊综合评价模型。研究发现,模糊综合评分法提供了一种“主观评分客观化”的评价思路,具有评分信度高、操作较简单、结果易解释、权重多样化、兼容性较强的特点,能有效减少评分过程中的随机误差,提高测试结果的准确性及可靠性,消减已有评分方法的弊端,值得加以广泛应用。□

## 参 考 文 献

- [1] Alderson, C., Clapham, C. & D. Wall. *Language Test Construction and Evaluation* [M]. Shanghai: Shanghai Foreign Language Education Press, 2000.
- [2] Bachman, L. & A. Palmer. *Language Assessment in Practice* [M]. Oxford: Oxford University Press, 1996.
- [3] Brown, H. & P. Abeywickrama. *Language Assessment: Principles and Classroom Practices* [M]. Beijing: Tsinghua University Press, 2013.
- [4] Carr, N. A comparison of the effects of analytic and holistic rating scale types in the context of composition tests [J]. *Issues in Applied Linguistics*, 2000(2).
- [5] Heaton, J. *Writing English Language Tests* [M]. Beijing: Foreign Language Teaching and Research Press, 2000.
- [6] Weigle, S. *Assessing Writing* [M]. Cambridge: Cambridge University Press, 2002.
- [7] 陈希镇,梁飞豹. 模糊综合评判中的模糊评分法[J]. 福州大学学报(自然科学版), 1993(6).
- [8] 蒋泽军. 模糊数学理论与方法[M]. 北京: 电子工业出版社, 2015.
- [9] 金檀等. 口语测试模糊评分方法设计及实验研究[J]. 现代外语, 2008(2).
- [10] 李金波. 命题过程中对试题能力考查的预估方法[J]. 现代教育科学, 2010(4).
- [11] 李清华,孔文. 二/外语写作测试评分研究综述[J]. 外语测试与教学, 2011(4).
- [12] 李筱菊. 语言测试科学与艺术[M]. 长沙: 湖南教育出版社, 1997.
- [13] 刘素梅. 俄语测试网上阅卷系统与主观题的误差控制[J]. 中国俄语教学, 2007(2).
- [14] 史铁强. 测试对教学的导向作用——俄语专业四、八级考试改革的思路[J]. 中国俄语教学, 2008(2).
- [15] 孙飞燕. 关于俄语专业四级考试写作评阅工作的几点思考[J]. 中国俄语教学, 2016(1).
- [16] 孙海洋,韩宝成. 英语写作分项评分和整体评分比较研究[J]. 解放军外国语学院学报, 2013(6).
- [17] 杨纶标,高英仪,凌卫新. 模糊数学原理及应用[M]. 广州: 华南理工大学出版社, 2011.
- [18] 张国立,张辉,孔倩. 模糊数学基础及应用[M]. 北京: 化学工业出版社, 2011.
- [19] 张铁男,李晶蕾. 对多级模糊综合评价方法的应用研究[J]. 哈尔滨工程大学学报, 2002(3).
- [20] 张文忠,郭晶晶. 模糊评分: 外语口语测试评分新思路[J]. 现代外语, 2002(1).
- [21] 张文忠,张姮. 基于模糊数学原理的口语测试模糊评估框架[J]. 外语教学与研究, 2008(6).
- [22] 郑超,常春娥. 是传递意义还是炫耀形式——外语写作过程中的形-意缠结现象[J]. 现代外语, 2014(4).
- [23] 郑美玲. 摘译能力分项评分量表的设计与研发[J]. 外语测试与教学, 2019(2).
- [24] 邹申. 基于语料库的写作测试效度研究——以《英语专业写作教学语料库》为例[J]. 外语电化教学, 2012(1).

基金项目: 本文为全国教育科学规划重点项目“‘一带一路’沿线关键土著语言文化通识课程体系建设研究”(项目编号: AFA180013)的阶段性成果。

收稿日期: 2021-01

通讯地址: 200083 上海市 上海外国语大学俄罗斯东欧中亚学院

## Research on Fuzzy Comprehensive Evaluation of Writing Test —Take TRM-8 as an Example

XU Hong & FU Yu

( School of Russian and Eurasian Studies , Shanghai International Studies University , Shanghai 200083 , China)

**Abstract:** Since writing assessment is highly subjective and imprecise , ensuring the consistent and reliable assessment standard has been a difficult and urgent problem of high-stake language tests. The evaluation of writing tasks in foreign language tests usually adopts the overall scoring method or the sub-item scoring method. The challenge lies in the inability of the rater to use and interpret the grade descriptions consistently , and it is also impossible to know whether they are scoring candidates strictly according to the grade standard. The author believes that the fundamental method to solve the problems above is to adopt a unified and objective scoring standard , which can be constructed through a fuzzy comprehensive evaluation method. The writing part of TRM-8 now uses the overall scoring method. In order to improve its scoring reliability , this paper built a fuzzy comprehensive evaluation model based on the writing scoring method of TRM-8 , compiled a new rating scale and described the basic principles and specific steps of fuzzy comprehensive evaluation.

This paper found that fuzzy comprehensive evaluation had the following advantages: First , the scoring reliability is high. The fuzzy comprehensive scoring method uses the principle of fuzzy synthesis to greatly reduce the influence of extreme factors , narrow the gap between personal evaluations , and reduce scoring errors. At the same time , it can ensure that the composition score is given according to the scoring standard , avoid subjective assumptions , and it is convenient to score candidates with unbalanced writing ability. Second , the operation is relatively simple. When judging , the scorer only needs to select the appropriate level in the specific indicators based on the student's writing performance. The statistics and calculation

of the score are performed by the computer according to the preset algorithm , and the scorer does not know the final score of the candidates , which greatly reduces the interference of human factors. Third , the results are easy to interpret. The “excellent , good , medium , and poor” scale is easier to grasp , and the scope of evaluation is limited to these four levels , which greatly improves the accuracy and consistency of the scoring. Fourth , the weight is diversified. The fuzzy comprehensive scoring method can not only weight different scoring dimensions , but also weight the sub-factors within each dimension , so that the scoring results are more accurate. Fifth , the compatibility is strong. The fuzzy comprehensive scoring method can combine the overall score with the sub-item score and make the overall and sub-item fuzzy evaluation respectively.

The disadvantage of fuzzy comprehensive evaluation is that the determination of the evaluation index and its weighting system contains a certain degree of subjective arbitrariness; the fuzzy comprehensive scoring method itself cannot avoid the confusion of evaluation results caused by correlation , crossover or even duplication between evaluation indexes; fuzzy comprehensive evaluation method excludes the situation with full marks and zero marks.

The fuzzy comprehensive scoring method provides an evaluation idea of “objectification of subjective scoring” , which can effectively reduce random errors in the scoring process , improve the accuracy and reliability of test results , and reduce the drawbacks of existing scoring methods. It is worthy of wide application.

**Key words:** Language Testing; Writing; Assessment Method; Fuzzy Comprehensive Evaluation