

# 目 录

第一章 绪论	1
1.1 研究背景	1
1.2 研究目的	4
1.3 研究问题	5
1.4 小结	5
第二章 写作测试中的提示研究	7
2.1 写作测试中的提示固有特征研究	7
2.1.1 提示固有特征与写作得分	9
2.1.2 提示固有特征与文本特征	12
2.2 写作测试中提示与考生的交互研究	15
2.3 写作测试中提示与评分员的交互研究	19
2.4 小结	22
第三章 写作测试的效度验证框架与方法	25
3.1 效度的概念	25
3.2 效度验证框架	28
3.3 综合写作测试的构念	40
3.4 综合写作测试的效度验证框架	42
3.5 效度验证方法	44
3.5.1 Coh-Metrix 文本分析工具	45

3.5.2	多层线性模型 .....	49
3.5.3	结构方程模型 .....	53
3.5.4	多层面 Rasch 模型 .....	56
3.6	小结 .....	59
<b>第四章</b>	<b>研究设计 .....</b>	<b>61</b>
4.1	综合写作任务简介 .....	61
4.2	数据收集 .....	62
4.2.1	提示、作文、参赛学生与评分员数据 .....	62
4.2.2	认知维度的提示特征测量 .....	66
4.3	研究变量 .....	67
4.3.1	提示固有特征 .....	67
4.3.2	基于考生认知的提示特征 .....	68
4.3.3	基于评分员认知的提示特征与评分标准 .....	69
4.3.4	综合写作表现 .....	70
4.3.5	语言水平 .....	71
4.3.6	评分行为 .....	71
4.4	数据分析 .....	72
4.5	小结 .....	74
<b>第五章</b>	<b>综合写作测试的效度验证：提示固有特征维度 .....</b>	<b>75</b>
5.1	提示固有特征 .....	75
5.2	提示固有特征对综合写作表现的影响估算 .....	76
5.2.1	提示话题域与文本特征 .....	76
5.2.2	提示任务说明与文本特征 .....	81
5.3	提示固有特征对综合写作表现的影响 .....	83
5.3.1	提示话题域对文本特征的影响 .....	83
5.3.2	提示任务说明对文本特征的影响 .....	89
5.4	基于“评估使用论证”框架的讨论 .....	92

5.5 小结 .....	95
<b>第六章 综合写作测试的效度验证：考生认知维度</b> .....	<b>97</b>
6.1 基于考生认知的提示特征 .....	97
6.2 基于考生认知的提示特征对综合写作表现的影响估算 ...	100
6.2.1 基于多层线性模型的提示认知影响估算 .....	100
6.2.2 基于结构方程模型的提示认知影响估算 .....	109
6.3 基于考生认知的提示特征对综合写作表现的影响 .....	114
6.3.1 基于多层线性模型的提示影响 .....	114
6.3.2 基于结构方程模型的提示影响 .....	116
6.4 基于“评估使用论证”框架的讨论 .....	118
6.5 小结 .....	120
<b>第七章 综合写作测试的效度验证：评分员认知维度</b> .....	<b>121</b>
7.1 基于评分员认知的提示特征与评分标准 .....	121
7.2 基于评分员认知的提示特征对综合写作评分的影响估算	122
7.2.1 基于评分员严厉度的提示影响估算 .....	122
7.2.2 基于评分难易度的提示影响估算 .....	128
7.2.3 基于评分员与考生偏差交互的提示影响估算 .....	134
7.3 基于评分员认知的提示特征对综合写作评分的影响 .....	139
7.3.1 基于评分员严厉度的提示影响 .....	139
7.3.2 基于评分难易度的提示影响 .....	141
7.3.3 基于评分员与考生偏差交互的提示影响 .....	143
7.4 基于“评估使用论证”框架的讨论 .....	145
7.5 小结 .....	146
<b>第八章 结语</b> .....	<b>149</b>
8.1 研究的主要发现 .....	149
8.2 研究的理论价值与实践意义 .....	153

4 | 基于提示特征影响的综合写作测试效度研究

8.3 未来研究方向 .....	155
参考文献 .....	157
附录 .....	187
附录 1 “外研社杯”全国英语写作大赛简介 .....	187
附录 2 综合写作评分标准 .....	190
附录 3 综合写作任务样题 .....	191
附录 4 考生问卷调查 .....	193
附录 5 评分员问卷调查 .....	197
附录 6 各地区评分员与考生显著偏差交互统计 .....	204

# 表目录

表 2.1	写作评估中写作任务的各个维度 .....	8
表 3.1	效度渐进矩阵 .....	26
表 3.2	二语词汇能力测试中使用完形测试的理由分析 .....	28
表 3.3	测试公平性框架 .....	30
表 3.4	以证据为中心的评估论证模型 .....	35
表 4.1	“外研社杯”全国英语写作大赛省级复赛的综合 写作任务提示信息 .....	63
表 4.2	综合写作提示下作文的描述性统计 .....	64
表 4.3	参赛学生问卷的描述性统计 .....	65
表 4.4	评分员问卷的描述性统计 .....	65
表 4.5	考生问卷与评分员问卷的信度统计 .....	67
表 4.6	提示固有特征分类编码 .....	68
表 4.7	基于考生认知的提示特征描述 .....	68
表 4.8	基于评分员认知的提示特征描述 .....	69
表 4.9	基于评分员认知的评分标准描述 .....	70
表 5.1	提示固有特征分类 .....	76
表 5.2	社会话题作文中与写作得分显著相关的文本特征指标 .....	77
表 5.3	教育话题作文中与写作得分显著相关的文本特征指标 .....	77
表 5.4	商业话题作文中与写作得分显著相关的文本特征指标 .....	78
表 5.5	个人话题作文中与写作得分显著相关的文本特征指标 .....	78
表 5.6	社会话题提示作文中具有写作得分预测力的文本 特征指标 .....	80

表 5.7	教育话题提示作文中具有写作得分预测力的文本特征指标 .....	80
表 5.8	商业话题提示作文中具有写作得分预测力的文本特征指标 .....	80
表 5.9	个人话题提示作文中具有写作得分预测力的文本特征指标 .....	80
表 5.10	显性对立观点提示作文中与写作得分显著相关的文本特征指标 .....	81
表 5.11	隐性对立观点提示作文中与写作得分显著相关的文本特征指标 .....	82
表 5.12	显性对立观点提示作文中具有写作得分预测力的文本特征指标 .....	83
表 5.13	隐性对立观点提示作文中具有写作得分预测力的文本特征指标 .....	83
表 6.1	基于考生认知的提示特征的探索性因子分析结果 .....	98
表 6.2	基于考生认知的提示特征的探索性因子分析结果： 方差解释量 .....	98
表 6.3	基于考生认知的提示特征的因子描述表 .....	99
表 6.4	多层线性模型中各变量的描述性统计 .....	101
表 6.5	综合写作得分零模型的估计结果 .....	102
表 6.6	综合写作得分多层线性模型的统计结果 .....	103
表 6.7	内容得分零模型的统计结果 .....	104
表 6.8	内容得分多层线性模型的统计结果 .....	104
表 6.9	结构得分零模型的统计结果 .....	105
表 6.10	结构得分多层线性模型的统计结果 .....	106
表 6.11	语言得分零模型的统计结果 .....	107
表 6.12	语言得分多层线性模型的统计结果 .....	107
表 6.13	多层线性模型中提示因子变量的描述性统计 .....	108

表 6.14	综合写作得分多层线性模型的统计结果 (提示因子变量) .....	109
表 7.1	多层面 Rasch 模型下评分员严厉度的统计结果 .....	122
表 7.2	评分员严厉度分组结果 .....	125
表 7.3	二元逻辑回归分析的 Hosmer-Lemeshow 检验 .....	126
表 7.4	评分员严厉度与其提示认知的二元逻辑回归分析结果 .....	126
表 7.5	多层面 Rasch 模型下评分难易度的统计结果 .....	128
表 7.6	评分员对评分标准重要性认知的统计结果 .....	130
表 7.7	评分难度与评分标准认知的对比统计 .....	132
表 7.8	基于评分员认知的提示特征与评分难度的对比统计 .....	133
表 7.9	多层面 Rasch 模型下评分员与考生的偏差交互模式 .....	136

# 图目录

图 3.1	效度验证的社会认知框架 .....	29
图 3.2	解释性论证 .....	32
图 3.3	Toulmin 推断模型 .....	33
图 3.4	托福考试效度论证模型 .....	34
图 3.5	以论据为中心的评估论证模型 .....	36
图 3.6	“评估使用论证”框架的推理链 .....	38
图 3.7	Toulmin 论证模型 .....	39
图 3.8	“评估使用论证”框架的主张 .....	42
图 6.1	初始模型：考生对提示特征的认知与其综合写作表现之间的关系模型 .....	110
图 6.2	假设模型：考生对提示特征的认知与其综合写作表现之间的关系模型 .....	111
图 6.3	考生对提示特征的认知与其综合写作表现之间的关系模型 .....	112
图 6.4	考生对提示特征的认知与其综合写作表现之间的关系简图 .....	114



# 第一章

## 绪论

### 1.1 研究背景

从语言测试与评估的发展阶段来看，当前语言测试的研究重点已从离散性测试转向语言运用测试（Grabe & Kaplan, 1996; Hamp-Lyons, 2001），综合写作任务则是语言运用测试的一个典范，日渐成为当前二语写作测试的发展趋势。较之传统的独立写作任务，综合写作任务提供了有意义的语言背景（Lee & Anderson, 2007），即写作提示，其形式一般为听力提示、阅读提示或两者兼有，因而综合写作测试有如下两大优势：1）可提高测试真实性，即写作测试常涉及对所读所听内容的理解与阐释（Esmacili, 2002; Weigle, 2004; Leki, Cumming & Silva, 2008），更符合现实生活中的听、读、写相结合的写作任务；2）可提高测试公平性，即写作提示中富含写作思路与语言素材，其作用类似于语言智库（Plakans & Gebril, 2012），提供均等的背景信息以供考生参考，弱化先验知识与单一片面内容的影响。国外许多大规模标准化测试项目已逐渐采用综合写作任务，如美国教育考试服务中心的新托福考试（TOEFL iBT）、培生教育集团的培生学术英语考试（PTE Academic）、加拿大学术英语水平考试（CAEL）、新西兰诊断性英语语言需求评估（DELNA）。在国内，综合写作任务在“外研社杯”全国英语写作大赛中得以成功应用。

纵观国内外二语综合写作评估研究，聚焦点主要体现在如下四大研究领域：1）二语综合写作任务中考生的写作过程与策略运用研究（Plakans, 2008, 2009a, 2009b; Plakans & Gebril, 2012; Yang & Plakans, 2012; Yang, 2014）；2）二语综合写作文本的语篇特征研究（Cumming et al., 2005; Knoch, 2009; Guo, Crossley & McNamara, 2013; Taguchi, Crawford & Wetze, 2013; Yu, 2013; Riazi, 2016; Plakans & Gebril, 2017）；3）二语综合写作任务中评分员的评分决策研究（Eckes, 2008; Wall & Horák, 2008; Knoch, 2009; Wiseman, 2012; Kuiken & Vedder, 2014; Goodwin, 2016）；4）二语综合写作任务的自动评分研究（Sawaki, Stricker & Oranje, 2009; Enright & Quinlan, 2010; Ramineni & Williamson, 2013; Sawaki, Quinlan & Lee, 2013; Bejar et al., 2014; Bridgeman & Ramineni, 2017）。此类研究为二语综合写作评估提供了理论依据与实证支持，但二语综合写作评估本身仍有其难以规避的局限性（Cumming, 2013），具体如下：

- 1）所测能力不清，即写作能力与对写作提示的理解能力的测量相混淆；
- 2）测试用途不明，即写作能力测试与写作能力诊断相混淆；
- 3）提示体裁模糊，即文本体裁不明确的写作提示会增加评分难度；
- 4）测试结果可比性不强，即对语言能力要求高，导致不同能力水平考生的测试结果不具有很强的可比性；
- 5）提示文本摘抄，即考生摘抄提示文本时将难以区分考生的语言产出与提示文本语言。

根据 Messick（1989, 1996）的效度整体观，效度被视为以构念效度为核心的整体概念。上述二语综合写作评估的局限性将成为“构念无关因素”，影响到二语写作能力评估的构念效度。但并非所有的构念无关因素都易于辨认，如何弱化或消除其影响是专家学者们需要解决的难题（Haladyna & Downing, 2004）。随着综合写作任务在二语测试中逐渐盛行，综合写作任务的效度研究有待进一步深入。

在语言测试中，许多因素（如施测过程遭干扰、考生个人特质、

考生身体或精神状态的临时变化等)会影响测试效果,这也让测试开发者力所不能及,但唯一力所能及的则是试题特征(Bachman & Palmer, 2010),因此可从控制试题质量入手以保证测量效度。在二语综合写作测试领域,已有较多研究发现试题特征对写作表现的影响,其中备受关注的是写作提示的系列文本是否会影响到考生在综合写作中的能力表现,并可对写作提示的影响细分为三类:1)写作提示对作文得分的影响(如Jennings et al., 1999; Esmaeili, 2002; Lee & Anderson, 2007; Kobrin, Deng & Shaw, 2011);2)写作提示对作文文本特征的影响(如Cumming et al., 2005; Ohkubo, 2009; Kormos, 2011; Plakans & Gebril, 2013);3)考生在综合写作中的写作过程与策略运用(如Ruiz-Funes, 1999; Cohen & Upton, 2007; Plakans, 2008, 2009b; Yang & Plakans, 2012)。

国内外有关写作提示影响的研究中,所涉及的写作提示特征主要是其固有特征,如话题、语篇类型、阅读文本/听力文本体裁、任务说明等,且研究发现各有不同,如部分研究发现综合写作测试中存在提示影响(如Cumming et al., 2005; Plakans & Gebril, 2013),而部分研究则证明不存在此类影响(如Jennings et al., 1999; Kobrin, Deng & Shaw, 2011)。写作提示的内在特征对综合写作表现的影响还有待论证。

此外,写作表现还受到很多试题因素的影响,如试题的话题熟悉度、难度、写作任务类型、提示的话题范围等,考生对这些试题因素也会有不同的理解(Lee, 2008)。而在综合写作测试领域,较少有研究关注到考生视角下的写作提示特征及其对考生写作表现的影响。在写作测试研究中,评分员与写作任务在评估中发挥着重要作用,而作文的评分决策往往会或多或少带有一些主观色彩(Bouwer et al., 2015),即评分员很可能会将其主观意志带入评分决策,而这种主观意志可能体现在评分员对不同写作提示有不同认知,而对写作提示的不同认知或将影响到他们的评分决策。但这种有关评分员对写作提示认知的推测还有待进一步的实证验证。

综上,目前国内外关于综合写作测试中写作提示因素的研究基本

停留在写作提示的固有特征层面，只有极少数学者提出了考生视角和评分员视角的写作提示特征，但相关实证研究极少。因此，我们将从三个维度的提示特征入手来探讨综合写作测试中的提示影响，即提示固有特征维度、考生认知的提示特征维度和评分员认知的提示特征维度。下面我们将简要介绍本研究的研究目的与研究问题。

## 1.2 研究目的

基于上述研究背景简介，我们将研究目的设定为两个，具体如下：第一，采用“评估使用论证”（assessment use argument，简称AUA）框架对国内首次使用的综合写作任务进行效度验证。具体而言，本部分包括三个方面：首先，采用深层文本特征测量工具 Coh-Metrix 探讨考生作文的文本特征，并通过回归分析考查提示固有特征是否会影响考生的写作表现，为综合写作测试的构念效度提供初步证据；其次，采用多层线性模型和结构方程模型分析方法探讨综合写作测试中考生对提示特征的认知对其写作表现的影响，为综合写作测试的构念效度提供证据；最后，采用多层面 Rasch 模型和二元逻辑回归等分析方法探讨综合写作测试中评分员对提示特征的认知与其评分决策之间的关系，为综合写作测试的评分效度提供证据。第二，通过效度验证为国内综合写作测试提供实证启示。具体而言，本部分包括三个方面：首先，为综合写作任务设计及评分提供启示，即通过将提示影响具体化，明确不同提示特征对写作表现的影响，有利于试题设计人员在编写试题和研制评分标准时规避考试偏颇，尽可能减小提示影响，同时也应加强评分员培训，避免评分偏颇；其次，为综合写作教学提供启示，即通过揭示提示特征与文本特征之间的关系，为教师选材与教学提供建议，公平公正地评估学生的二语写作水平；最后，为机器自动评分模型的设计提供启示，即基于不同文本特征差异的发现和认知影响模型的构建，为不同提示特征下的评分模型提供新的启示。现有机器自动评分研究主要以写作文本特征为参数（如 Attali, 2007; Knoch, 2009; Sawaki, Quinlan & Lee, 2013），而鲜有对其组合提示、考

生与评分员的认知活动加以考查，从而忽视了这三者在综合写作中的交互作用。

### 1.3 研究问题

针对上述研究目的，本研究提出以下三个研究问题：

1) 提示固有特征对综合写作表现的影响研究，即提示固有特征是否会对综合写作任务的文本特征产生影响？若存在影响，不同提示下的文本特征差异体现在哪些方面？

2) 考生对提示特征的认知对其综合写作表现的影响研究，即考生对提示特征的认知是否会对其写作表现产生影响？若存在影响，具体是哪些提示特征，影响程度如何？

3) 评分员对提示特征的认知对其综合写作评分的影响研究，即评分员对提示特征的认知是否会对其评分产生影响？若存在影响，具体是怎样的影响？

### 1.4 小结

本章首先简要介绍了综合写作测试的优点与研究焦点，然后通过梳理综合写作测试的研究现状，指出现有研究的局限性并以此为起点，提出本研究的研究目的和研究问题。

