



摘要

普通话水平测试是中华人民共和国国家级考试，是一种口语测试。普通话测试的结果会受到诸多因素的影响，本文考察心理因素、评分细则等方面对测试结果的影响，提出测试管理机构应加强对测试员思想、心理、业务水平、评分标准的把握和理解等方面改进办法，并对今后测试工作的发展提出新的设想。

关键词：普通话水平测试；影响；因素



Abstract

The standard checkout of Putonghua is the People's Republic of China state-level test, is one kinds of spoken language test, the test result influenced by many factors. This article analyzes to tests the result the influence from psychological factor, grading regulation aspects and so on , Proposed the management section should strengthen to the tester thought, psychological, the vocational level, the grading standard assurance with aspects and so on understanding improves the means, and propose the new tentative plan to the next test work development.

Key words: The standard checkout of Putonghua; influence; facts



华中师范大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所提交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

作者签名：吕洪维

日期：07年5月15日

学位论文授权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权华中师范大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。同时授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》，并通过网络向社会公众提供信息服务。

作者签名：吕洪维

日期：07年5月15日

导师签名：王同胜

日期：2007年5月15日

本人已经认真阅读“CALIS 高校学位论文全文数据库发布章程”，同意将本人的学位论文提交“CALIS 高校学位论文全文数据库”中全文发布，并可按“章程”中的规定享受相关权益。同意论文提交后滞后：半年；一年；二年发布。

作者签名：吕洪维

日期：07年5月15日

导师签名：王同胜

日期：2007年5月15日



1. 前言

美国未来学家耐斯比特在他的《大趋势》中展望前景时说：“要想真正取得的成功，你必须要懂得三种语言：流利的英语、西班牙语和电脑语言。”如果把他的话加以概括，那就是要懂得流利的母语、一门外国语和电脑语言。这可以说是人类进入 21 世纪的一个基本要求。

国家推广全国通用的普通话。普通话是以汉语文授课的各级各类学校的教学语言；是以汉语传送的各级各类广播电台、电视台的规范语言，也是汉语电影、电视剧、话剧必须使用的规范语言；是我国党政机关、团体、企事业单位干部在公务活动中必须使用的工作语言；是不同方言区以及国内不同民族之间人们的通用交际语言。

掌握和使用一定水平的普通话，是进行现代化建设的各行各业人员，特别是教师、播音员、节目主持人、演员等专业人员必备的职业素质。因此，国家规定在一定范围内对某些岗位的人员进行普通话水平测试，并逐步试行持等级证书上岗制度。

普通话是汉民族的共同语，是规范化的现代汉语。共同语和语言的规范化是不可分割的，没有一定的规范就不可能有真正的共同语。普通话的规范指的是现代汉语在语音、词汇、语法各方面的标准。普通话水平测试是推广普通话工作的重要组成部分，是使推广普通话工作逐步走向科学化、规范化、制度化的重要举措。推广普通话，促进语言规范化，是汉语发展的总趋势。普通话水平测试工作的健康开展必将对社会的语言生活产生深远的影响。

1.1 普通话水平测试的现状

《国家通用语言文字法》以法律形式确立了普通话和规范汉字作为国家通用语言文字的地位，这既标志着我国语言文字规范化、标准化工作正式步入法制轨道，进入一个新的发展时期，同时也意味着对普通话水平测试工作提出了更高的要求。早在 1997 年 12 月召开的全国语言文字工作会议上，许嘉璐主任就代表国家语委在工作报告中提出了推广普通话跨世纪奋斗目标：“2010 年以前，普通话在全国范围内初步普及”，“与口语表达密切相关行业的工作人员，其普通话水平达到相应的要求”。这一推普目标将新的历史使命摆在了我们面前：与口语表达密切相关行业的工作人员要在规定的期限里“达标”，普通话水平测试是普通话教学的一个重要



1. 前言

美国未来学家耐斯比特在他的《大趋势》中展望前景时说：“要想真正取得的成功，你必须要懂得三种语言：流利的英语、西班牙语和电脑语言。”如果把他的话加以概括，那就是要懂得流利的母语、一门外国语和电脑语言。这可以说是人类进入 21 世纪的一个基本要求。

国家推广全国通用的普通话。普通话是以汉语文授课的各级各类学校的教学语言；是以汉语传送的各级各类广播电台、电视台的规范语言，也是汉语电影、电视剧、话剧必须使用的规范语言；是我国党政机关、团体、企事业单位干部在公务活动中必须使用的工作语言；是不同方言区以及国内不同民族之间人们的通用交际语言。

掌握和使用一定水平的普通话，是进行现代化建设的各行各业人员，特别是教师、播音员、节目主持人、演员等专业人员必备的职业素质。因此，国家规定在一定范围内对某些岗位的人员进行普通话水平测试，并逐步试行持等级证书上岗制度。

普通话是汉民族的共同语，是规范化的现代汉语。共同语和语言的规范化是不可分割的，没有一定的规范就不可能有真正的共同语。普通话的规范指的是现代汉语在语音、词汇、语法各方面的标准。普通话水平测试是推广普通话工作的重要组成部分，是使推广普通话工作逐步走向科学化、规范化、制度化的重要举措。推广普通话，促进语言规范化，是汉语发展的总趋势。普通话水平测试工作的健康开展必将对社会的语言生活产生深远的影响。

1.1 普通话水平测试的现状

《国家通用语言文字法》以法律形式确立了普通话和规范汉字作为国家通用语言文字的地位，这既标志着我国语言文字规范化、标准化工作正式步入法制轨道，进入一个新的发展时期，同时也意味着对普通话水平测试工作提出了更高的要求。早在 1997 年 12 月召开的全国语言文字工作会议上，许嘉璐主任就代表国家语委在工作报告中提出了推广普通话跨世纪奋斗目标：“2010 年以前，普通话在全国范围内初步普及”，“与口语表达密切相关行业的工作人员，其普通话水平达到相应的要求”。这一推普目标将新的历史使命摆在了我们面前：与口语表达密切相关行业的工作人员要在规定的期限里“达标”，普通话水平测试是普通话教学的一个重要



部分。普通话水平测试(简称为 PSC)是我国为加快共同语普及进程、提高全社会普通话水平而设置的一种语言测试制度。是我国新时期推广普通话工作中采取的一项重大举措,它的诞生和推行,标志着我国推广、普及普通话工作开始向纵深发展,逐步进入制度化、规范化、科学化的新阶段。通过普通话水平测试,既能推动普通话的普及,又可以促进广大群众普通话水平的提高。

目前,全国已建起了较为完整的测试工作网络,拥有国家级和省级测试员 4.5 万人,有近 2 千万人次接受测试并取得相应的等级证书。有些省区已经积累了一些初步的实践经验。作为一项新的、前所未有的考试, PSC 在实践中取得明显成绩的同时,也暴露出了自身存在的一些问题。问题的直接表现是:对同样的测查内容和方法、甚至等级标准,不同地方、不同的测试员有着不同的理解;对应试人的普通话水平作出的等级评价与其实际能力之间有一定的距离;不同测试员对同一应试人的普通话水平所作出的等级评定,甚至是同一测试员在不同时间对同一应试人的普通话水平的评价结果也会不同,乃至错档。这一方面反映了 PSC 作为一种主观测试的明显特征,另一方面也表明 PSC 在测试的方法论上尚有缺陷,测试员的专业能力在整体上与完成测试的实际要求方面也存在一定的差距。因此加强 PSC 理论研究和测试员专业培养,是保证 PSC 的科学性,提高测试效度和信度的关键。

目前的当务之急,一是要提高对普通话水平测试重要性的认识,不要以为测试仅仅就是出张卷子,给个成绩,评个等级,发张证书;二是要结合测试中出现的重要问题,着力进行测试理论的研究,增强测试的科学性、规范性。普通话水平测试是一种专业性很强的大规模标准化考试,是一项十分复杂而又全新的工作,因此,必须精心组织实施,不断总结经验,建立和改进操作规程,特别是要正确把握测试标准,提高测试信度,确保普通话水平测试的质量。

普通话水平测试不是普通话系统知识的考试,不是文化水平的考核,也不是口才的评估,是应试人运用普通话所达到的标准程度的检测和评定。根据国家语言文字工作委员会、国家教育委员会、广播电影电视部《关于开展普通话水平测试工作的决定》,普通话水平测试工作先在一定范围内对某些岗位的人员实行。

为了便于操作和突出口头检测的特点,普通话水平测试一律采用口试。普通话有口语和书面语两种形式,测试也必须采取有文字凭借和没有文字凭借两种方式进行。有文字凭借的部分要包括适量的语音、词汇、语法的检测项,各类题目要有明确的目的、要求。要选取编制较好的试卷进行信度、区别度和难度分析,通过分析试卷可以在测试机构内作为标准试卷推广使用,并逐步建立测试题库。



1.2 普通话水平测试研究的现状

1994年10月，国家语言文字工作委员会、原国家教育委员会、原广播电影电视部联合发出《关于开展普通话水平测试工作的决定》，普通话水平测试（PSC）自此在全国正式开展。2000年10月颁布的《中华人民共和国国家通用语言文字法》又通过立法手段对普通话水平测试加以强调。近几年全国各地（包括港澳地区）测试工作发展十分迅速，形成了比较科学规范的管理和运行体系，产生了广泛的社会影响。普通话水平测试研究从论证到实施，从理论到实践，经过广大专家和业务人员的辛勤努力，也取得了很大成绩。1991年通过鉴定的由孙修章主持的国家社会科学基金项目“普通话水平测试标准”，奠定了普通话水平测试的基础。刘照雄主编的《普通话水平测试大纲》是目前为止普通话水平测试最重要的研究成果。据统计，1994年至2007年，全国发表关于普通话水平测试的论文438篇。可以说，普通话水平测试研究与工作实践目前正处于互动状态，形成了一个良性循环的发展态势。但是，普通话水平测试毕竟是我国新兴而且复杂的语言测试工程，与我国的汉语水平考试（HSK）及国内外其他一些著名的语言测试相比还显得不很成熟，研究成果与语言学其他领域相比，也显得不够丰硕。在发展过程中，形势和任务迫切要求全国有关领域的专家跨领域合作，共同攻关，采用新的研究手段和方法，解决一系列突出的理论问题与实践问题。

在理论研究方面，陈章太、仲哲明、刘照雄、孙修章、宋欣桥、毛世桢、王渝光、李宇明等进行了有益的探讨。但总体而言，某些具体问题研究较多，系统的基本理论研究仍显薄弱，普通话水平测试的理论框架体系尚未构建成熟。

为了进一步规范普通话水平测试的规程，修订考试的内容、评分方法、范围、测试题库建设等，都有待于进一步完善和提高，使普通话水平测试更具科学性、公平性、客观性、权威性。因此，此类规范标准的研究亦须加强。

1.3 本文研究目的和思路

普通话水平测试目前主要采用主观评测的方法，这使得测试成本很高，测试员劳动强度大，特别是测试信度受到一定局限。本文考察影响普通话水平测试的若干因素，主要是心理因素影响、普通话评分细则的影响及测试员素质的影响等，提出相应的解决办法，如测试管理架构要努力提高测试员思想素质、业务素质、心理素质，加强测评实践训练等管理。并指出今后测试的科学化方向，如人机测



1.2 普通话水平测试研究的现状

1994年10月,国家语言文字工作委员会、原国家教育委员会、原广播电影电视部联合发出《关于开展普通话水平测试工作的决定》,普通话水平测试(PSC)自此在全国正式开展。2000年10月颁布的《中华人民共和国国家通用语言文字法》又通过立法手段对普通话水平测试加以强调。近几年全国各地(包括港澳地区)测试工作发展十分迅速,形成了比较科学规范的管理和运行体系,产生了广泛的社会影响。普通话水平测试研究从论证到实施,从理论到实践,经过广大专家和业务人员的辛勤努力,也取得了很大成绩。1991年通过鉴定的由孙修章主持的国家社会科学基金项目“普通话水平测试标准”,奠定了普通话水平测试的基础。刘照雄主编的《普通话水平测试大纲》是目前为止普通话水平测试最重要的研究成果。据统计,1994年至2007年,全国发表关于普通话水平测试的论文438篇。可以说,普通话水平测试研究与工作实践目前正处于互动状态,形成了一个良性循环的发展态势。但是,普通话水平测试毕竟是我国新兴而且复杂的语言测试工程,与我国的汉语水平考试(HSK)及国内外其他一些著名的语言测试相比还显得不很成熟,研究成果与语言学其他领域相比,也显得不够丰硕。在发展过程中,形势和任务迫切要求全国有关领域的专家跨领域合作,共同攻关,采用新的研究手段和方法,解决一系列突出的理论问题与实践问题。

在理论研究方面,陈章太、仲哲明、刘照雄、孙修章、宋欣桥、毛世桢、王渝光、李宇明等进行了有益的探讨。但总体而言,某些具体问题研究较多,系统的基本理论研究仍显薄弱,普通话水平测试的理论框架体系尚未构建成熟。

为了进一步规范普通话水平测试的规程,修订考试的内容、评分方法、范围、测试题库建设等,都有待于进一步完善和提高,使普通话水平测试更具科学性、公平性、客观性、权威性。因此,此类规范标准的研究亦须加强。

1.3 本文研究目的和思路

普通话水平测试目前主要采用主观评测的方法,这使得测试成本很高,测试员劳动强度大,特别是测试信度受到一定局限。本文考察影响普通话水平测试的若干因素,主要是心理因素影响、普通话评分细则的影响及测试员素质的影响等,提出相应的解决办法,如测试管理架构要努力提高测试员思想素质、业务素质、心理素质,加强测评实践训练等管理。并指出今后测试的科学化方向,如人机测



试、网络测试等等。总之，要加强相关研究，采取相应措施，将影响普通话水平测试结果的种种因素降到最低，使普通话水平测试更具权威性和科学性。

本文通过对比实验法、调查法等分析出心理因素对普通话水平测试的影响，通过调查、对比等分析各省市评分细则对测试结果的影响，并通过某高校普通话测试站的测试实践，探讨相应的解决办法。



2. 影响普通话水平测试的心理因素

2.1 应试者的心理因素对普通话水平测试的影响

应试者的心理因素直接影响着普通话水平测试的结果。我们可以把应试者的心理因素分为两类：其一是主观心理因素；其二是客观心理因素。

2.1.1 主观心理因素

主观心理因素是指由于应试者自身的心理状态变化从而影响测试的诸多因素。这类因素主要包括应试者的动机、欲望、焦虑水平、情绪状态、意识水平等等。应试者参加普通话水平测试时，在动机、愿望、情绪、焦虑水平等方面与平时相比自然会强很多。在测试过程中，几乎所有的应试者都会出现程度不同的紧张情绪，在测试中常遇到这种情形：考生在开始测试时上来就是“第一题读单音节字……”漏报单位、姓名、编号。测试第一、二、三题时漏读、误读现象大量存在，如：将“拨”读为“拔”，“瘸”读为“拐”，“唱”读为“歌”，“醋”读为“酸”，“涮”读成“刷”，“俩”读成“两”，“贰”读成“两”，“儿女”读为“女儿”，“贿”读为“赂”，等等。眼里看的是“加把劲”，嘴里念的却是“加把油”，还有些简单的常见字也不认识了，测试结束后问：“一个单人旁加上一个牛念什么？”有的考生说话时内容准备得很充分，但报完说话题目，脑子一片空白，出现语塞。个别考生甚至呼吸急促手脚发抖，声音微弱，发音吐字不清晰，测试程序搞混，需要测试员多次提醒。这些现象都是心理紧张所致。

2.1.2 客观心理因素

客观心理因素是指客观环境对应试者产生心理影响的因素。如测试环境的光线、测试员的态度、仪表、言语、表情等等。

我们曾在某高校普通话测试站设计两个对比实验。

实验一：随机抽取该测试站 2006 年 11 月 2 日某场测试，现场测试时，考场现场使用录音笔进行了录音，录音效果比较好。两名测试员所给的分数如下：

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
87	80.8	90	87.3	81	79	84	87.8	84.	87	81	82	87	88.2	87.6	87.5	90	87	90.5	80.2

表 2-1 对 20 名学生现场测试的评分



2. 影响普通话水平测试的心理因素

2.1 应试者的心理因素对普通话水平测试的影响

应试者的心理因素直接影响着普通话水平测试的结果。我们可以把应试者的心理因素分为两类：其一是主观心理因素；其二是客观心理因素。

2.1.1 主观心理因素

主观心理因素是指由于应试者自身的心理状态变化从而影响测试的诸多因素。这类因素主要包括应试者的动机、欲望、焦虑水平、情绪状态、意识水平等等。应试者参加普通话水平测试时，在动机、愿望、情绪、焦虑水平等方面与平时相比自然会强很多。在测试过程中，几乎所有的应试者都会出现程度不同的紧张情绪，在测试中常遇到这种情形：考生在开始测试时上来就是“第一题读单音节字……”漏报单位、姓名、编号。测试第一、二、三题时漏读、误读现象大量存在，如：将“拨”读为“拔”，“瘸”读为“拐”，“唱”读为“歌”，“醋”读为“酸”，“涮”读成“刷”，“俩”读成“两”，“贰”读成“两”，“儿女”读为“女儿”，“贿”读为“赂”，等等。眼里看的是“加把劲”，嘴里念的却是“加把油”，还有些简单的常见字也不认识了，测试结束后问：“一个单人旁加上一个牛念什么？”有的考生说话时内容准备得很充分，但报完说话题目，脑子一片空白，出现语塞。个别考生甚至呼吸急促手脚发抖，声音微弱，发音吐字不清晰，测试程序搞混，需要测试员多次提醒。这些现象都是心理紧张所致。

2.1.2 客观心理因素

客观心理因素是指客观环境对应试者产生心理影响的因素。如测试环境的光线、测试员的态度、仪表、言语、表情等等。

我们曾在某高校普通话测试站设计两个对比实验。

实验一：随机抽取该测试站 2006 年 11 月 2 日某场测试，现场测试时，考场现场使用录音笔进行了录音，录音效果比较好。两名测试员所给的分数如下：

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
87	80.8	90	87.3	81	79	84	87.8	84.	87	81	82	87	88.2	87.6	87.5	90	87	90.5	80.2

表 2-1 对 20 名学生现场测试的评分



一周以后，我们请两名测试员对该场考试的考生进行录音评分。在进行录音评分时，排除了应试者的外貌、穿着打扮、长相、职业等因素的影响，同时排除测试场地环境等因素，主要表现为测试员心理因素对测试结果的影响。评分结果如下：

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
88.6	88.8	90.8	90.2	85.2	78.2	87.9	90.8	87.8	90.9	84.4	84.1	88.5	90.5	88.9	88.7	90.2	89	93.7	84.7

表 2-2 对 20 名学生测试录音的评分

Independent Samples Test

		Levene's Test for Equality of Variances		t-Test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
VAR00021	Equal variances assumed	.460	.502	2.204	38	.034	2.5000	1.13432	.20369	4.79631
	Equal variances not assumed			2.204	37.990	.034	2.5000	1.13432	.20367	4.79633

表 2-3 对 20 名学生现场测试评分与录音评分的统计检验

从两场测试的成绩来看，同样的被试，前后两次的评分有着非常显著的差异 ($t=2.204^*$, $p<0.05$)。如表 2-3 所示。这说明测试员的心理因素对测试结果有非常明显的影响，究其原因，除了测试员自身的业务水平、环境变化等原因外，测试员评分时受到情绪和心理定势的影响。

实验二：一周以后，我们组织参加上述测试 20 名考生到心理学实验室再次进行测试，在行为观察室中观察学生的行为表现。

首先将学生带入心理放松室中放松，然后在学校心理咨询老师的指引下，通过言语进行心理暗示，告诉他们等会每位同学可以在情景室中进行测试，情景室中没有测试员老师，仅只有录音笔录音。学生们依次进入情景室中参加测试。情景室中学生可以看见一面镜子，于是大家就面对镜子完成自己的测试。但实际上，在这种单向可视的镜子反面的观察室中，测试员们可以清楚地看见被试的一举一动，并通过情景室对考生进行隐秘评分，对测试员而言，这场测试与平常的考试没有很大的区别，但对考生而言，以为自己面对的是镜子的影像——即自己，排除了自己的一些紧张情绪。该 20 名考生这次的测试成绩如下：



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
89	85	92.8	89	85	84	86	89	87.5	91	84	86.5	88.5	90.2	89.6	89	89.8	91	93	86

表 2-4 对 20 名学生在行为观察室中测试的评分

	Levene's Test for Equality of Variances	t-test for Equality of Means								
						Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
		F	Sig.	t	df			Sig. (2-tailed)	Lower	Upper
VAR00001	Equal variances assumed	3.620	.065	2.820	38	.008	2.8500	1.01052	.80430	4.89570
	Equal variances not assumed			2.820	35.237	.008	2.8500	1.01052	.79902	4.90098

表 2-5 对 20 名学生现场测试评分与在行为观察室测试评分的统计检验

对 20 名学生现场测试评分与在行为观察室测试评分的统计检验结果，如表 2-4 所示，说明现场评分与隐秘评分之间存在着显著的差异 ($t=2.820^{**}$, $p>0.01$)。

我们可以看出应试者心理因素对普通话水平测试的影响。普通话测试通常由几名测试员共同测试一名应试者。测试员的一言一行对应试者都会产生很大影响。应试者心理素质的不稳定往往会影响到测试员的态度、情绪，测试员对考生的厌烦态度和不满情绪会造成评分时的不公正。测试员严肃的表情通常会使应试者产生紧张情绪，从而影响应试者发挥正常的水平，导致测试误差的出现。所以，测试员和蔼可亲的态度，哪怕一抹微笑，都会给应试者很大的鼓励与抚慰，能帮助他们克服测试过程中出现的紧张焦虑情绪，引导他们发挥出最佳水平。

2.2 测试员的心理因素对普通话水平测试的影响

由于普通话水平测试完全是口语测试，其主观心理的不稳定因素便导致了评价上的误差。在分析心理因素时，还必须明确误差问题具有两重性。一方面是误差的绝对性。任何测试都有一定误差，测试员之间的差别是客观存在的，而且就某一测试在同一时刻对某一测试员来说是不变的。另一方面是误差的相对性。任何误差都不是一成不变的，随着时间的推移和误差造成者心理活动的变化，同一问题的误差会不断变化，或扩大或缩小。不承认这两点，就失去了心理分析的理由和基础。



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
89	85	92.8	89	85	84	86	89	87.5	91	84	86.5	88.5	90.2	89.6	89	89.8	91	93	86

表 2-4 对 20 名学生在行为观察室中测试的评分

	Levene's Test for Equality of Variances	t-test for Equality of Means								
								95% Confidence Interval of the Difference		
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
VAR00001	Equal variances assumed	3.620	.065	2.820	38	.008	2.8500	1.01052	.80430	4.89570
	Equal variances not assumed			2.820	35.237	.008	2.8500	1.01052	.79902	4.90098

表 2-5 对 20 名学生现场测试评分与在行为观察室测试评分的统计检验

对 20 名学生现场测试评分与在行为观察室测试评分的统计检验结果，如表 2-4 所示，说明现场评分与隐秘评分之间存在着显著的差异 ($t=2.820^{**}$, $p>0.01$)。

我们可以看出应试者心理因素对普通话水平测试的影响。普通话测试通常由几名测试员共同测试一名应试者。测试员的一言一行对应试者都会产生很大影响。应试者心理素质的不稳定往往会影响到测试员的态度、情绪，测试员对考生的厌烦态度和不满情绪会造成评分时的不公正。测试员严肃的表情通常会使应试者产生紧张情绪，从而影响应试者发挥正常的水平，导致测试误差的出现。所以，测试员和蔼可亲的态度，哪怕一抹微笑，都会给应试者很大的鼓励与抚慰，能帮助他们克服测试过程中出现的紧张焦虑情绪，引导他们发挥出最佳水平。

2.2 测试员的心理因素对普通话水平测试的影响

由于普通话水平测试完全是口语测试，其主观心理的不稳定因素便导致了评价上的误差。在分析心理因素时，还必须明确误差问题具有两重性。一方面是误差的绝对性。任何测试都有一定误差，测试员之间的差别是客观存在的，而且就某一测试在同一时刻对某一测试员来说是不变的。另一方面是误差的相对性。任何误差都不是一成不变的，随着时间的推移和误差造成者心理活动的变化，同一问题的误差会不断变化，或扩大或缩小。不承认这两点，就失去了心理分析的理由和基础。



2.2.1 迁移的作用对测试结果的影响

迁移是已有的知识和经验对解决问题的影响，可以分为正迁移和负迁移，如果一种学习促进了另外一种学习就称为正迁移，一种学习干扰另外一种学习就称为负迁移。每位测试员的心理规则都是有差异的。心理规则具有双重含义：其一是《大纲》对测试员的要求，即个体的外在规则；其二是个体的自我认知，即个体的内在规则。一是外部《大纲》；二是个体在语言生活中的母语方言；三是个体的知识理论和认知能力。

2.2.2 心理定势对测试结果的影响

心理定势是指前面一个比较强烈的心理活动，对随后进行的心理活动的反应内容和反应趋势有明显的影响。势者，趋也，因利而制权。良好而健康向上的趋势是统一思想、振奋精神，树立良好的群体价值观，实现组织目标的有力保证，不可不察。定势是一种心理准备状态，其影响测试时的倾向性，有时有助于测试，有时又会妨碍测试。

2.2.3 情绪和动机对测试结果的影响

动机对测试的效率有明显的影响。简单测试动机越强，效率越高，太复杂的测试动机要小，有难度但难度不太大的测试，中等强度的动机最有好处，太强或太弱都不利于测试。情绪对人的影响也很大，积极情绪能激励人解决更复杂的任务，消极情绪则会使测试时智力活动受到障碍。

2.2.4 情感负担对测试结果的影响

测试是一个动态系统，在 15 分钟内完成对受测人五大项的心理评价，这是一项十分繁重而又紧张的脑力劳动，于是语音连贯的复杂性与心理评价高效率之间的矛盾成了测试员的情感负担。情感负担又由具体情境参数来确定。当受测人与测试员语音水平几乎没有什么差异时，一方面，测试员此时情感平衡受到破坏，一旦表现出态度上的“紊乱”，不仅将受测人作为潜在的对手，也使评价的误差率大幅度上升。另一方面，水平差异只有达到一定的程度才能引起测试员心理上的差异感觉，语音差异小，分化程度低，无法彰显心理规则，则难以得出比较客观的结论。不只是在水平相当时会造成评价误差指数上升，其实当双方差异达到一定水平的时候，也会出现糊涂评价。情感伴随认识过程而产生，全天候处于频繁的听、记、评等测试动作中，产生疲惫现象是很自然的。疲惫现象就是心理学所指体力及情感的疲乏症状。在这种情况下，心理规则的框架性未能自动进入系统，严重受我向性的干扰，依据先行经验，或凭借感觉作出评价，由此模糊性上升，准确性下降。



2.3 普通话水平测试的心理作用机制

语音是一种社会现象，普通话水平测试是基于社会认识的主观心理评价。从心理学的角度来看，这一主观评价就是对语音信息进行反应的心理过程。心理过程可以分解为听觉、记忆、控制和反应一系列阶段。语言能力是人的心理特征。在普通话水平测试中，这种特征是通过语音形式表现出来的，测试员评价语音是凭听觉感受进行的，听觉感受是人对语音信息的心理反应。心理与语音评价之所以可以联系起来，实际上是受测人语音信息与测试员心理规则和心理评价相互关联作用才得以实现的。测试员以自己的心理规则为依据，沟通语音和心理状态，产生心理评价的行为过程。这个过程包括：第一阶段为语音的“我化”。如在“朗读”“说话”评价中，测试员约以每秒4个音节的速度快速听辨。密集的语音在大部分情况下，测试员优先注意到其中与其心理规则相关的语音。例如，对于一个南方测试员，一般来讲他较容易忽视“儿化”“轻声”方面的语音，而对于一个北方测试员，对此却相对敏感。这是因为这两类测试员的心理规则有所不同。第二阶段为评价体验。该阶段测试员对我化的语音进行心理的分析辨认。上例北方测试员对“儿化”“轻声”关注后，根据提示听辨语音切分成份，产生对、错或缺陷的心理体验。第三阶段就是评分实施。心理构想形成后，测试员产生相应的行为。上例中北方测试员进一步用心理规则去处理“儿化”“轻声”语音问题。同时心理作用机制非常典型地反映在测试员评分的心态上，“评分心态实际暗含着测试员对不同应试对象所需要具备水平的心理预期，在这种不同心理的作用下，测试员在掌握评分标准时，也往往有不同的宽严尺度，这就很可能导致不同层次的应试人等分不等值，影响测试的客观公正与成绩的可比性。”^①

^① 申向阳：《试论普通话测试中心理因素的负效应基调控策略》，载《阿坝师范高等专科学校学报》，2005年底一期。



3. 影响普通话水平测试的评分细则因素

教育部、国家语言文字工作委员会 2003 年 10 月对普通话水平测试的大纲进行了修订,《普通话水平测试大纲》是普通话水平测试的指导性文件,也是纲领性文件。作为国家社会科学基金项目,《大纲》的研制历时四载,并经学术委员会严密论证后才予以公布,因此具有一定的科学依据。但因为是全国性文件,在测试大纲中的评分标准项出现了很多比较含混的术语,如朗读项评分中,语调偏误,视程度扣 0.5 分、1 分、2 分;停连不当,视程度扣 0.5 分、1 分、2 分;朗读不流畅,视程度扣 0.5 分、1 分、2 分;说话项评分中语音标准程度中“有方音但不明显”“有方音比较明显”“方音明显”“方音重”;词汇语法规程度中“词汇语法偶有不规范情况”“词汇语法屡有不规范情况”等,就给测试员特别是新测试员评分带来很大的不好操作性。测试员完全根据自己对《大纲》的理解评分,由于测试人员水平不一,个人判断理解能力不同,会造成评分标准带有极大的主观性,不可能准确的反映出受测人的普通话实际水平,使测试成绩有失公正。另一方面,受测人层次不一,要求不同,执行标准也要灵活掌握。因而测试员反映在工作中越来越感到困惑,不知道到底应该如何评分才能准确、公正、科学地反映出受测人普通话的实际水平。

有些地区的测试站在测试过程中,对评分标准放得过宽,受测人的总合格率达到 95% 以上,几乎做到人人过关,使普通话水平测试流于推普过关的低层次阶段。另有一些地区的测试站对评分标准掌握得过严,受测人的总合格率在 60% 左右,使受测人闻测色变,扼制了参加普测的热情,阻碍了普通话水平测试工作的健康开展。如果我们不很好地解决这一难题,就很可能使普通话水平测试这一国家性考试流于一种低层的推普过关形式而非强制性地检验推普成果的一种手段。与其他语言测试不同,普通话水平测试的成绩是挂在受测人的嘴上的,只要受测人开口讲普通话,他就已经向人们展示出他的普通话水平。我国 2010 年的工作目标是在全国大中城市普及普通话,实现这一目标需要全体社会成员的努力。作为推普工作的一个重要组成部分,普通话水平测试工作必不可少。因此各省测试中心都根据本省的实际情况,积极理顺、确定准确统一的可操作性强的评分细则,以便更加科学、客观、公正地反映出受测人普通话的语言面貌,并保证普通话水平测试体系的科学性、规范性、可操作性。



“国家大纲”在读字词部分中，将不准确的读音分为错误和缺陷两大类，并规定每字扣分多少，可以说有量化的标准，比较好掌握。语音错误指的是在普通话水平测试中，将普通话语音系统的一个音位发为另一个音位的现象。语音缺陷指的是普通话水平测试中，虽未将普通话语音系统的一个音位发为另一个音位，但尚未达到标准音位的现象。语音错误和语音缺陷是普通话水平测试语音评定中两个重要概念。普通话水平测试语音评定的一个重要特点是改变了以往语音评价“非此即彼”的二元价值判断的局限，使语音评定更符合实际和语感。但是对于语音错误和语音缺陷，人们尚存在不少认识的误区。有人认为方音成分是语音错误，地方普通话的语音成分是语音缺陷；有人简单地认为把甲读为乙就是语音错误，其他不到位的语音是语音缺陷。我们认为普通话水平测试一定要有音位学的观念，所以特别提出一个参照系，即普通话语音(音位)系统。把普通话语音系统中的一个音位发为另一个音位就是语音错误；虽未将普通话语音系统中的一个音位发为另一个音位，但尚未达到标准音位就是语音缺陷。“大纲”还明确规定缺陷的性质和内容，从声韵调等方面分别加以说明，也便于测试员掌握。但是还有以下问题难以很好把握：用量化分析的方法对应试者的普通话语音的规范程度作出判断，是普通话水平测试采用的基本方法。量化分析将音节正误情况分为三种类型：“正确”、“错误”、“缺陷”。但应试者发出的音节并非都与这三种类型一一对应。我们知道“学习普通话的过程是从方言母语向民族标准语转变的过程，普通话水平测试面对的是一种过渡性的语音，在测试中我们所接受到的语音信息是错综复杂的”^①。如复韵母的动程略大或略小、声带松紧调节过快致使阳平调值突然升高，音节长度不够、发单元音时舌部肌肉松弛，致使语言清晰度差等一系列的问题。由于测试大纲在这一系列的问题上界定模糊或无标准可依，加之普通话测试工作极强的专业性，使得不同的测试员在测试过程中对标准的把握宽严不一，测试结果也是五花八门。“语音本身也是一个连续的模糊变量，我们很难对某一个语音做出正确、错误、缺陷的精确切分。”^②另外语言能力毕竟是人的心理特征，语言本身具有模糊性，仅仅靠量化分析常常难以测出应试人的真实水平，所以《普通话水平测试大纲》中，除了一些量化的评分标准之外，还有一些模糊的评分项目，如朗读、说话项中的方言语调、语速、流畅程度、词汇、语法规程程度、语音面貌等。应该说《大纲》特意设立这些“模糊”评分项目是非常科学的。它可以弥补单纯量化的不足，使测试员对应试人的语音失误进行定量统计的同时，还可以在这些评分项中对那些一一对应的语言现象进行定性

^① 孙修章：《普通话水平测试标准》的研制与实践》载于《语言文字应用》，1994年第一期。

^② 孙修章：《普通话水平测试标准》的研制与实践》载于《语言文字应用》，1994年第一期。



的分析,从而使测试成绩更具科学性。但“模糊”评分并不是“糊涂”评分,定性分析也应该有一定方面的量的分析作基础,这样才使测试员便于操作。我们无法要求每位测试员都是语言学家,但我们可以让测试员知道一定量的分析基础上应该得出一个什么样的定性结论,缺陷才“比较突出”、“比较明显”、“略有反应”。由于《大纲》缺乏这此依据,致使测试中朗读项方言语调的扣分以及说话项中语音面貌的档次划分标准不一,出现“糊涂”评分的现象。所以模糊评价同样需要“量”的支撑,尽管这种“量”是约略的。但离开了建立在模糊评价基础上的约略的“量”,应试者的普通话测试总分及等级就难以做到精确,当然也就谈不上客观公正了。

3.1 各省市朗读项评分细则对测试结果影响

“PSC 评分细则”是对《普通话水平测试大纲》中有关评分标准内容的具体表述,是国家级、省级普通话水平测试员对普通话水平测试应试人的普通话水平等级进行测评的客观依据,应具有较强的可操作性,必须规范和统一。从目前全国各省市通行的“PSC 评分细则”来看,绝大部分与《大纲》的原则基本上是相符的,同时由于在个别问题上存在着认识上的分歧,在确定评分方法时就产生了一定的差别,这给测试员的测评工作带来了一定的麻烦,在一定程度上影响了应试人测试成绩的准确性,因此需要进行统一的修订与完善。从普通话水平测试的性质与效果看,各省市在制定“PSC 评分细则”时也应遵循统一的原则与方法。因此,从一定意义上说,有必要制定统一的与《大纲》相配套的、不违背《大纲》精神的、全国通用的“PSC 评分细则”。作为《大纲》的补充和细化,“评分细则”的制定必须严格按照《大纲》“总论”部分的阐述进行,对《大纲》的原则意见要进行正确的理解与阐述,不允许出现偏离《大纲》原则的评判标准,更不允许为了操作方便或尊重个别人的主观意愿而对《大纲》确定的原则进行随意修改。各省测试中心应根据本省情况,组织语言专家、骨干测试员等制定相应的评分细则,并在实施一段时间之后,召开相关人员会议,评价此评分细则是否科学并有操作性。我们比较了三个省的评分细则和国家大纲中的评分标准的区别,并提出相应的细则设想。

三省朗读项评分细则比较:



的分析,从而使测试成绩更具科学性。但“模糊”评分并不是“糊涂”评分,定性分析也应该有一定方面的量的分析作基础,这样才使测试员便于操作。我们无法要求每位测试员都是语言学家,但我们可以让测试员知道一定量的分析基础上应该得出一个什么样的定性结论,缺陷才“比较突出”、“比较明显”、“略有反应”。由于《大纲》缺乏这此依据,致使测试中朗读项方言语调的扣分以及说话项中语音面貌的档次划分标准不一,出现“糊涂”评分的现象。所以模糊评价同样需要“量”的支撑,尽管这种“量”是约略的。但离开了建立在模糊评价基础上的约略的“量”,应试者的普通话测试总分及等级就难以做到精确,当然也就谈不上客观公正了。

3.1 各省市朗读项评分细则对测试结果影响

“PSC 评分细则”是对《普通话水平测试大纲》中有关评分标准内容的具体表述,是国家级、省级普通话水平测试员对普通话水平测试应试人的普通话水平等级进行测评的客观依据,应具有较强的可操作性,必须规范和统一。从目前全国各省市通行的“PSC 评分细则”来看,绝大部分与《大纲》的原则基本上是相符的,同时由于在个别问题上存在着认识上的分歧,在确定评分方法时就产生了一定的差别,这给测试员的测评工作带来了一定的麻烦,在一定程度上影响了应试人测试成绩的准确性,因此需要进行统一的修订与完善。从普通话水平测试的性质与效果看,各省市在制定“PSC 评分细则”时也应遵循统一的原则与方法。因此,从一定意义上说,有必要制定统一的与《大纲》相配套的、不违背《大纲》精神的、全国通用的“PSC 评分细则”。作为《大纲》的补充和细化,“评分细则”的制定必须严格按照《大纲》“总论”部分的阐述进行,对《大纲》的原则意见要进行正确的理解与阐述,不允许出现偏离《大纲》原则的评判标准,更不允许为了操作方便或尊重个别人的主观意愿而对《大纲》确定的原则进行随意修改。各省测试中心应根据本省情况,组织语言专家、骨干测试员等制定相应的评分细则,并在实施一段时间之后,召开相关人员会议,评价此评分细则是否科学并有操作性。我们比较了三个省的评分细则和国家大纲中的评分标准的区别,并提出相应的细则设想。

三省朗读项评分细则比较:



国家大纲评分标准	江苏省评分细则	浙江省评分细则	湖北省评分细则
声母或韵母系统性缺陷, 视程度扣 0.5、1 分	声母或韵母的系统性语音缺陷指: 有 3 个及 3 个以上音节具有相同的声母或韵母缺陷。(声调的缺陷不在此列) 只出现 1 类声母或韵母的系统性语音缺陷的, 扣 0.5 分。出现 2 类及 2 类以上声母或韵母的系统性语音缺陷的, 扣 1 分	声母或韵母的系统性语音缺陷, 应视程度扣分: 系统性语音缺陷较少可扣 0.5 分; 系统性语音缺陷较多则扣 1 分。	声母或韵母系统性缺陷, 视程度扣 0.5 (1—2 类)、1 分 (3 类以上) 声母或韵母有系统性的语音缺陷 (某类声母或韵母至少出现 5 次以上语音缺陷)
语调偏误, 视程度扣 0.5、1、2 分	这里的语调主要指明读诸要素中的重音、句调、速度, 不包括停顿。 语调偏误指: 重音不当, 词的轻重音格式不对, 声调有系统性缺陷, 句调不自然, 语速不当等。评分时应综合考虑各因素。 语调偏误有所表现, 扣 0.5 分; 语调偏误明显, 扣 1 分; 语调偏误严重, 扣 2 分。	“语调偏误”主要指字调、句调、轻重音格式、音变的失误, 应视程度扣分: 略有反映, 可扣 0.5 分; 偏误较多可扣 1 分; 偏误多则扣 2 分。	语调偏误, 视程度扣 0.5、1、2 分 语调偏误主要指语流中声调有系统性缺陷; 句调抑扬的失当; 字词的轻重音失当, 或不自然, 且有系统性表现; 音长不规范, 或节奏忽快忽慢; 语气词带有明显的方言痕迹。
停连不当, 视程度扣 0.5、1、2 分	停连不当指: 由不恰当的停顿或连读造成的对词语的肢解或对语义的误解。因换气需要或个人理解造成的短暂停顿 (并未曲解原文意思), 不视为停连不当。 停连不当 1 次, 扣 0.5 分; 2 次, 扣 1 分; 3 次及 3 次以上, 扣 2 分	“停连不当”指朗读时肢解词语、或造成言语误解、形成歧义等情况, 也包括朗读时节奏不当, 当断不断, 当连不连, 或字化、词化等情况, 按照应试人反映出来的情况, 视程度扣分。	停连不当, 视程度扣 0.5、1、2 分 停连不当指的是停顿对一个双音节或多音节词语的肢解; 停顿造成对一句话、一段话的误解, 形成歧义。
朗读不流畅 (包括回读) 视程度扣 0.5、1、2 分	朗读不流畅指: 回读, 停顿过多, 按音节崩读等。评分时应综合考虑诸因素。 朗读不流畅有所表现, 扣 0.5 分; 朗读不流畅明显, 扣 1 分; 朗读不流畅严重, 扣 2 分。	“朗读不流畅 (包括回读)”主要反映应试人朗读短文的熟练程度。不太流畅可扣 0.5 分; 明显不流畅可扣 1 分; 结结巴巴, 频繁回读则扣 2 分。对回读部分的音节语音以最后一遍为准, 如有漏读、增读、读错的音节均以错误记评。	朗读不流畅 (包括回读) 视程度扣 0.5、1、2 分 朗读不流畅指的是: 在语句中出现一字一进、或者一词一进现象, 不连贯; 语速过慢, 或较长时间停顿; 朗读中出现回读情况。

表 3-1 朗读项评分细则的比较



在声韵缺陷项扣分中，国家大纲仅仅只有：“视程度”这个模糊术语，江苏和湖北省的评分细则将其细化，明确出什么是声韵缺陷，但在缺陷的概念上有所不同。江苏省认为，3个或3个以上的声韵语音缺陷称为系统性；而湖北省指某类声母或韵母至少出现5次以上语音缺陷；浙江省干脆也模糊处理，跟国家大纲一样视程度扣分；至于扣分时扣0.5还是1分，江苏省提出只出现1类声母或韵母的系统性语音缺陷的，扣0.5分；2类以上扣1分。湖北省提出1—2类扣0.5分；3类以上扣1分。而浙江省则为系统性语音缺陷较少可扣0.5分；系统性语音缺陷较多则扣1分。评分越量化测试员越好评分。湖北和江苏省的评分细则定量与定性结合得较好。而浙江省的评分细则造作性就差了一些。至于声韵缺陷系统性是3个还是5个，我们认为湖北省的评分细则更加科学，3个缺陷即为系统缺陷稍严。

“语调偏误”的扣分细则中，三个省评分细则对语调的描述不尽相同，浙江省认为语调指的就是字调、句调、轻重音格式、音变；江苏省认为语调指的是指朗读诸要素中的重音、句调、速度，不包括停顿；湖北省认为语调指的语流中声调、句调、字词的轻重音、音长、节奏、语气词等。语调指的是在具体进行语言表达时，语音的快慢、高低、疾徐、停顿等变化，这些因素贯穿于整个句子，我们称之为语调。语调偏误的评分测试员比较难把握，主要还是和评分标准的不统一和模糊性相关。应统一为：语调指的是语流中的字调、句调、轻重音、音长、节奏等，在扣分上可以增设一项1.5分，这样在评分时更加细化，有利于操作。

“停连不当”一项的扣分，各省评分基本一致。浙江省增加朗读时节律不当，当断不断，当连不连，或字化、词化等情况。我们认为出现此种情况应在“朗读不流畅”中扣分，而不应该出现在“停连不当”中，否则会造成概念的混淆。

“朗读不流畅”项的评分，各省都作了相应的说明。回读、按音节崩读、一字一进、结结巴巴等都是扣分因素。但回读一项的扣分不明。因回读涉及到两个问题，回读一句话或者回读一个词、短语。这两者应该区别扣分。具体可细分为：回读一句话扣0.5分，2—3句扣1分，4局以上扣2分；回读3—4个词语扣0.5分；5—7个词语扣1分，7个以上扣2分。朗读不流畅（包括回读），回读并不是唯一扣分的依据，而应该在不流畅的因素中综合考虑。

3.2 各省市说话项评分细则对测试结果影响

说话项评分细则比较：



在声韵缺陷项扣分中，国家大纲仅仅只有：“视程度”这个模糊术语，江苏和湖北省的评分细则将其细化，明确出什么是声韵缺陷，但在缺陷的概念上有所不同。江苏省认为，3个或3个以上的声韵语音缺陷称为系统性；而湖北省指某类声母或韵母至少出现5次以上语音缺陷；浙江省干脆也模糊处理，跟国家大纲一样视程度扣分；至于扣分时扣0.5还是1分，江苏省提出只出现1类声母或韵母的系统性语音缺陷的，扣0.5分；2类以上扣1分。湖北省提出1—2类扣0.5分；3类以上扣1分。而浙江省则为系统性语音缺陷较少可扣0.5分；系统性语音缺陷较多则扣1分。评分越量化测试员越好评分。湖北和江苏省的评分细则定量与定性结合得较好。而浙江省的评分细则造作性就差了一些。至于声韵缺陷系统性是3个还是5个，我们认为湖北省的评分细则更加科学，3个缺陷即为系统缺陷稍严。

“语调偏误”的扣分细则中，三个省评分细则对语调的描述不尽相同，浙江省认为语调指的就是字调、句调、轻重音格式、音变；江苏省认为语调指的是指朗读诸要素中的重音、句调、速度，不包括停顿；湖北省认为语调指的语流中声调、句调、字词的轻重音、音长、节奏、语气词等。语调指的是在具体进行语言表达时，语音的快慢、高低、疾徐、停顿等变化，这些因素贯穿于整个句子，我们称之为语调。语调偏误的评分测试员比较难把握，主要还是和评分标准的不统一和模糊性相关。应统一为：语调指的是语流中的字调、句调、轻重音、音长、节奏等，在扣分上可以增设一项1.5分，这样在评分时更加细化，有利于操作。

“停连不当”一项的扣分，各省评分基本一致。浙江省增加朗读时节律不当，当断不断，当连不连，或字化、词化等情况。我们认为出现此种情况应在“朗读不流畅”中扣分，而不应该出现在“停连不当”中，否则会造成概念的混淆。

“朗读不流畅”项的评分，各省都作了相应的说明。回读、按音节崩读、一字一进、结结巴巴等都是扣分因素。但回读一项的扣分不明。因回读涉及到两个问题，回读一句话或者回读一个词、短语。这两者应该区别扣分。具体可细分为：回读一句话扣0.5分，2—3句扣1分，4局以上扣2分；回读3—4个词语扣0.5分；5—7个词语扣1分，7个以上扣2分。朗读不流畅（包括回读），回读并不是唯一扣分的依据，而应该在不流畅的因素中综合考虑。

3.2 各省市说话项评分细则对测试结果影响

说话项评分细则比较：



江苏省评分细则	浙江省评分细则	湖北省评分细则
一档：没有语音错误，扣0分；失误1-2次，扣1分；失误3-4次，扣2分。	一档：语音标准，或极少有失误。扣0分、0.5分、1分、1.5分、2分。	一档：语音标准，或极少失误。扣0、0.5、1、1.5、2分
二档：语音错误在5-7次之间，扣3分；语音错误在8-10次之间，扣4分。如方音明显，降为三档，分别扣5分、6分。	二档：语音错误在10次以下，有方音但不明显。扣3分、4分。	二档：语音错误在10次以内，有方音不明显，扣3分、4分
三档：语音错误在11-15次之间，无明显方音，扣5分、6分	语音错误在10次以下，但方音比较明显；或语音错误在10次-15次之间，有方音但不明显。扣5分、6分。	语音错误在15次以内，方音比较明显，有一类型系统性错误，但能部分区别，或有三类以内系统性缺陷，扣5、6分
四档：语音错误在11-15次之间，有明显方音，扣7分、8分。	四档：语音错误在10次-15次之间，方音比较明显。扣7分、8分。	四档：语音错误超过15次，方音明显，有2-3类系统性错误且大多数不能区分，扣7、8分
五档：语音错误在16-30次之间，视方音程度扣9分、10分、11分	五档：语音错误超过15次（16-45次），方音明显。扣9分、10分、11分。	五档：语音错误多，方音明显，声母、韵母有3类以上系统性错误，且基本不能区分，扣9、10分
六档：语音错误超过30次，视方音程度扣12分、13分、14分。	六档：语音错误多（45次以上），方音重。扣12分、13分、14分	六档：语音错误很多，方音重，但尚能听出足普通话，且话语表达清楚，扣12、13、14分

表3-2 说话项评分细则比较

说话项的评分是定量和定性相结合的评分，既要数错误的个数，也要对整个语流的情况作出定性的判定，难度较大，而且档次之间虽然只有1、2分之差，但评分时档次定得不准就有可能差到3、4分了。

在一档的评分中，江苏省的评分具体由语音失误的次数，浙江和湖北仅仅只是规定极少有失误，江苏省无0.5、1.5分的扣分。湖北和浙江省均有0.5、1.5分的扣分，这样既符合国家大纲的标准，也有利于区分一级甲等的成绩。

二档的评分中，错误的次数对应的评分基本相同，均为10次以下，江苏省还区别出3、4分的错误个数，但江苏省加了一条：如方音明显，降为三档，分别扣



5分、6分，也就是错误次数在10次以内，但方音明显，则降一个档次，此种说法与三档的描述“语音错误在11-15次之间，无明显方音”有些混淆。5、6分的档次到底是方音明显还是无明显方音？因此湖北、浙江省的评分细则的描述更好一些。

三档的评分中，各省的描述也不尽相同，从错误的次数来说，江苏省是11—15次之间，浙江是10次以内、或者10—15次之间，湖北是15次以下。从方音的描述来说，江苏省是“有明显方音”；浙江省是10次以内，“方音比较明显”，10—15次之间，“有方音但不明显”；湖北省是“方音比较明显”。这样的描述对测试员来说显然想念绕口令一样，很容易绕昏。而且从中文的意思来说，“方音比较明显”比“方音明显”感觉要明显一些。因此，这一个档次的评分细则可以概括为：语音错误在15次以下，（当然也包括了10次以下），有方音但不明显，有一类系统性错误或三类系统性缺陷，扣5、6分。

第四档的评分中，对于错误的次数各省也不相同，江苏省是11--15次之间，浙江是10—15次之间，湖北省是15次以上。从方音的描述上，江苏省“有明显方音”，浙江省“方音比较明显”，湖北省是“方音明显”。错误的次数如果相差太小，很难给应考者评分。

第五档的评分中，浙江省的16--45次，次数跨度太大，也不利于评分，而且在测试实践中，为保证测试流程的顺利进行，45次错误很难计数，因此可将其确定为：语音错误多，方音明显，声母、韵母有3类以上系统性错误，且基本不能区分，视方音程度扣9、10分。

第六档的评分中，各地均写明“方音重”。浙江和江苏明确地指出了错误个数（30次或45次以上），湖北省的细则中只提出错误很多，次数已经难以很好地记录，但是还是应和湖北的评分细则一样，加上方音重，但尚能听出是普通话，且话语表达清楚，扣12、13、14分。因为我们测试的是普通话水平，如果应试者说的是方言，哪怕错误的次数不太多，也不能归到此档次中。

3.3 临界分数对测试结果影响

当应试者的测试成绩到达临界线时，特别需要测试员值得注意。我们考察了2003—2005年某校测试站的测试成绩。



5分、6分，也就是错误次数在10次以内，但方音明显，则降一个档次，此种说法与三档的描述“语音错误在11-15次之间，无明显方音”有些混淆。5、6分的档次到底是方音明显还是无明显方音？因此湖北、浙江省的评分细则的描述更好一些。

三档的评分中，各省的描述也不尽相同，从错误的次数来说，江苏省是11—15次之间，浙江是10次以内、或者10—15次之间，湖北是15次以下。从方音的描述来说，江苏省是“有明显方音”；浙江省是10次以内，“方音比较明显”，10—15次之间，“有方音但不明显”；湖北省是“方音比较明显”。这样的描述对测试员来说显然想念绕口令一样，很容易绕昏。而且从中文的意思来说，“方音比较明显”比“方音明显”感觉要明显一些。因此，这一个档次的评分细则可以概括为：语音错误在15次以下，（当然也包括了10次以下），有方音但不明显，有一类系统性错误或三类系统性缺陷，扣5、6分。

第四档的评分中，对于错误的次数各省也不相同，江苏省是11--15次之间，浙江是10—15次之间，湖北省是15次以上。从方音的描述上，江苏省“有明显方音”，浙江省“方音比较明显”，湖北省是“方音明显”。错误的次数如果相差太小，很难给应考者评分。

第五档的评分中，浙江省的16--45次，次数跨度太大，也不利于评分，而且在测试实践中，为保证测试流程的顺利进行，45次错误很难计数，因此可将其确定为：语音错误多，方音明显，声母、韵母有3类以上系统性错误，且基本不能区分，视方音程度扣9、10分。

第六档的评分中，各地均写明“方音重”。浙江和江苏明确地指出了错误个数（30次或45次以上），湖北省的细则中只提出错误很多，次数已经难以很好地记录，但是还是应和湖北的评分细则一样，加上方音重，但尚能听出是普通话，且话语表达清楚，扣12、13、14分。因为我们测试的是普通话水平，如果应试者说的是方言，哪怕错误的次数不太多，也不能归到此档次中。

3.3 临界分数对测试结果影响

当应试者的测试成绩到达临界线时，特别需要测试员值得注意。我们考察了2003—2005年某校测试站的测试成绩。



	一级乙等		二级甲等		级乙等		三级甲等		三级乙等
2003年	55人		787人		901人		98人		1人
分数分布	96—96.9	92—96	91—91.9	87—90.9	86—86.9	80—86.9	79—79.9	70—79	60—69.9
分数段人数	0人	55人	101人	666人	90人	811人	9人	89人	1人
2004年	108人		774人		1002人		194人		5人
分数分布	96—96.9	92—96	91—91.9	87—90.9	86—86.9	80—86.9	79—79.9	70—79	60—69.9
分数段人数	0人	108人	65人	709人	83人	919人	5人	189人	5人
2005年	120人		616人		675人		93人		5人
分数分布	96—96.9	92—96	91—91.9	87—90.9	86—86.9	80—86.9	79—79.9	70—79	60—69.9
分数段人数	0人	120人	52人	564人	93人	582人	0人	93人	0人
2006年	83人		493人		609人		88人		0人
分数分布	96—96.9	92—96	91—91.9	87—90.9	86—86.9	80—86.9	79—79.9	70—79	60—69.9
分数段人数	0人	83人	37人	456人	67人	542人	0人	88人	0人
合计	366人		2670人		3187人		473人		6人
分数分布	96—96.9	92—96	91—91.9	87—90.9	86—86.9	80—86.9	79—79.9	70—79	60—69.9
分数段人数	0人	366人	154人	2516人	243人	2944人	14人	459人	6人

表 3-3 2003—2005 年某校测试站的测试成绩

从表 3-3 可以看出,2003--2006 年,某测试站共有 6696 人参加普通话水平测试,其中临界线的考生就有 411 人,占 6.14%。临界线的考生成绩到底是上还是下?多零点几分就是不同的级别,会影响考生的上岗、评职称、毕业等,而且测试结束后,容易引起争议的也是临界线分数的考生。经过测试员的评分表分析,可以看出,1 分之差主要是在测试的第三、四项的评分上,因为一、二项错误缺陷的评分比较明晰,而朗读和说话项的评分受测试员得影响非常大,因此评分细则的制订必须科学、客观、好操作。从表格中我们还可看出另外的问题,临界线的考生每年所占的比例是不同的,2003 年占 10.86%,2004 年占 4.52%,2005 年占 6.91%,2006 年占 8.1%。经过对审听、组长以及测试员的调查发现,有些测试员有怕麻烦、怕投诉的心理,考生一到临界线的时候,就紧张、慌神,有的全部扣得很松,少扣一分就上升一个等级;有的全部扣得很严,多扣一分就下降一个等级。还有的和测试员、审听、组长对全场成绩的整体把握相关。有的组长、复审在测试之前就说:“遇到临界线的情况,如果是大一的学生或者是初次参加测试的考生,就多扣一点,因为他今后还



可以再考；如果是毕业班的学生或者是考了很多次的就少扣一分让他上升一个档次。”等等，这就严重地影响考试的信度。在考试时应尽量避免出现此类情况，测试员应据实打分，而不能因人而异。制定“普通话水平测试评分细则”是一项严谨而又复杂的工作，它需要在《大纲》原则的指导下，根据测试实践，运用语言学、语音学、方言学、语言测试理论等知识，依据测试手段、测试方式、测试对象的特点，进行科学而又具体的论述，其目的是统一测试员的评分标准，使PSC的评分减少主观性差异，更具有客观性，从而使普通话水平测试向着更为科学、客观、健康的方向发展。



4. 影响普通话水平测试的测试员的素质因素

普通话水平测试工作已经开展了 12 年，成为推广普及普通话的一项重要举措。随着普通话水平测试等级证书含金量越来越大，普通话水平测试的研究越来越强，科学进行普通话水平测试的水平也越来越高，然而影响普通话水平测试的因素越来越难控制。除了应试者、测试组织者、测试试题等因素之外，测试员的素质高低直接决定了普通话水平测试的成败。

4.1 测试员的思想素质

测试是一项理性的工作，工作情绪影响测试成绩。有的测试员认为自己是主考官，说起话来也是盛气凌人。如有一位年轻的测试员在测试时，应试者年龄较大，方音也比较明显，测试员对应试者不屑一顾，说：“你这样的水平，凭什么教语文？”伤了应试者的心。还有的测试员态度高傲，与应试者发生冲突，影响考生考试。有的两个测试员之间有分歧，也不注意调整，互不相让，认为自己是正确的。其实这些完全都是可以避免的。有的测试员测试时不带钢笔、钟表等；有的迟到，因为一个人迟到，有时影响整个测试不能按时开始。有的测试员在测试时吃东西，接打电话，表现出不严肃。更有甚者，有的测试员中途跑出考场，只剩下另一个人测试，或者到别的考场打招呼。有的测试员随意减少测试项目或测试内容。朗读、说话项都不足时间。诸如此类，反映了测试员的思想素质。我们认为这些都是不允许的。因此在培训测试员时，应该将思想教育放在首位。普通话水平测试是严肃认真的。测试员的态度也应该是严肃认真的。你的一言一行，一举一动都代表着整个测试工作。当你坐在考场上，开始进行测试的时候，你已经不仅仅是你自己了。你代表着你的工作单位，代表着省市普通话培训测试中心，代表着国家语言文字工作机构，代表着这一项工作的形象。因此在测试时，测试员应该是严肃认真、和蔼可亲，紧张有序的把测试工作进行到底。

4.2 测试员的业务素质

普通话水平测试员应具有相应的专业素质。《普通话水平测试大纲》中对测试员的专业素质提出了要求，测试员的资格认定中对测试员的业务素质有如下要求：



4. 影响普通话水平测试的测试员的素质因素

普通话水平测试工作已经开展了 12 年，成为推广普及普通话的一项重要举措。随着普通话水平测试等级证书含金量越来越大，普通话水平测试的研究越来越强，科学进行普通话水平测试的水平也越来越高，然而影响普通话水平测试的因素越来越难控制。除了应试者、测试组织者、测试试题等因素之外，测试员的素质高低直接决定了普通话水平测试的成败。

4.1 测试员的思想素质

测试是一项理性的工作，工作情绪影响测试成绩。有的测试员认为自己是主考官，说起话来也是盛气凌人。如有一位年轻的测试员在测试时，应试者年龄较大，方音也比较明显，测试员对应试者不屑一顾，说：“你这样的水平，凭什么教语文？”伤了应试者的心。还有的测试员态度高傲，与应试者发生冲突，影响考生考试。有的两个测试员之间有分歧，也不注意调整，互不相让，认为自己是正确的。其实这些完全都是可以避免的。有的测试员测试时不带钢笔、钟表等；有的迟到，因为一个人迟到，有时影响整个测试不能按时开始。有的测试员在测试时吃东西，接打电话，表现出不严肃。更有甚者，有的测试员中途跑出考场，只剩下另一个人测试，或者到别的考场打招呼。有的测试员随意减少测试项目或测试内容。朗读、说话项都不足时间。诸如此类，反映了测试员的思想素质。我们认为这些都是不允许的。因此在培训测试员时，应该将思想教育放在首位。普通话水平测试是严肃认真的。测试员的态度也应该是严肃认真的。你的一言一行，一举一动都代表着整个测试工作。当你坐在考场上，开始进行测试的时候，你已经不仅仅是你自己了。你代表着你的工作单位，代表着省市普通话培训测试中心，代表着国家语言文字工作机构，代表着这一项工作的形象。因此在测试时，测试员应该是严肃认真、和蔼可亲，紧张有序的把测试工作进行到底。

4.2 测试员的业务素质

普通话水平测试员应具有相应的专业素质。《普通话水平测试大纲》中对测试员的专业素质提出了要求，测试员的资格认定中对测试员的业务素质有如下要求：



省级测试员：具有大专以上学历，3年以上工作的教师、播音员和语言文字工作干部，熟悉推广普通话工作方针政策和普通话语音理论，熟悉掌握《汉语拼音方案》和常用国际音标，熟悉本地区方言和普通话的对应规律，普通话口语在一级。有较强的测试能力和测试工作经验。

国家级测试员：具有大专以上学历、中级以上专业技术职称的教师、播音员和专职从事语言文字工作的5年以上的干部。熟悉推广普通话工作的方针政策和普通话语音理论，熟练掌握《汉语拼音方案》和常用国际音标，熟悉方言与普通话的一般对应规律，普通话口语水平达到一级，有较强的普通话水平测试能力和较丰富的测试工作经验。

但这仅仅是一个概括性的描述。总结十几年来的测试经验，作为PSC的基层组织和测试实践者，我们认为，普通话水平测试是一项非常复杂的智力劳动。测试员要能够较好地承担普通话水平测试任务，对应试人的普通话运用水平作出准确或者较为准确的评价，保持较高的信度，必须具备技术的和理论的两个方面的专业素质和能力，两者缺一不可。

(1) 技术素质

技术素质指在实际的测试过程中，测试员为了完成对应试人普通话水平等级的评定所需要的基本的经验能力，包括以下几个方面：

a、标准的普通话口语表达能力：测试员的普通话水平应当达到一级。

b、评分办法和等级标准的掌握理解和运用能力：测试员必须全面、熟练掌握每一个测试项目的评分方法和每一等级的合格标准，正确理解各项评分办法之间，以及它们和等级标准之间的关系，并将它们贯彻到测试过程中。

c、话语现象的感知、听辨能力：PSC是口语测试。口头语言是在时间的序列上流动的，转瞬即逝，测试员必须在其出现的一刹那准确感知、听辨。

d、话语现象的分析判断能力：是测试员对通过听觉器官感知到的话语现象作出定量、定性分析的能力。

e、实际操作的能力：是测试员在测试过程中随着测试过程的顺序延进，将听辨、判断结果转化为书面记录的能力。

(2) 理论知识素养

测试员除了具有上述技术素质外，还必须具有以下几个方面的理论素养：

f、普通话基础理论：掌握有关普通话的基本知识，了解普通话语音、语汇、语法方面的系统特征，具有对普通话语音、词汇、语法等各要素进行理论分析的能力。



g、国家有关语言文字政策、规范标准等方面的理论知识：普通话水平测试中有多个项目要借助书面进行，但汉字是非标音文字，其字形、字音和字义之间关系复杂，测试员必须掌握和理解这方面的政策和已经公布的规范标准。

h、有关的方言理论与知识：掌握普通话和方言之间的关系，了解方言、特别是所在地区各方言在语言各要素方面与普通话之间的对应规律。具有分析和识别方言与普通话的能力。

i、语言学理论：主要是音位学方面的理论知识，能够运用常见国际音标记录语音现象，具有运用音位学理论分析语音现象的能力。

j、普通话水平测试理论：即与普通话水平测试有关的应用语言学理论和知识。如普通话水平测试的性质、对象和方法论等方面的理论知识。

对测试员来说，技术素质是外在的、直接的、实用的，而理论素养则是内在的、间接的、根本的。两者虽有不同，却紧密相联。我们可以作出这样一个合乎逻辑的推论：一个普通话口语能力较好的人，如果缺少上述专业理论方面的素养，经过短期的培训，可能也会在一定程度上获得测试应有的一些技术能力，但是由于这些能力没有足够的理论方法上的支持，可能是不全面、不稳定的，其测试的效果在一定程度上也就是不完全可靠的。而一个测试员具有相应的技术能力，又掌握了这些理论知识和方法，那么他就能够对普通话水平测试本身、测试的评分方法、等级标准等作出更加全面清楚的理解，从而自觉地研究测试对象，分析测试对象的表现特征，不断总结测试经验，全面提高自己的测试技能。测试员只有多参加测试实践，才能够发现自己业务上存在的问题，找到自己努力的方向，不断提高自身的能力。

4.3 测试员的心理素质

(详见 2.2 测试员的心理因素对普通话水平测试的影响)

4.4 测试员的职业素质

普通话水平测试通过口试的方法进行，主要是由个体来操作，因此，测试员的职业素质在测试中也是影响测试结果非常重要的因素。个别测试员认为自己是主考官，主宰着应试者的“生杀大权”，因此测试时态度高傲，而且自觉不自觉地认为应试者的普通话水平不是很高。同时，测试员自身的普通话水平测试成绩一般处在 92—96 分之间，一甲的测试员可以说是凤毛麟角，因此人们通常觉得一甲非常难得。



g、国家有关语言文字政策、规范标准等方面的理论知识：普通话水平测试中有多个项目要借助书面进行，但汉字是非标音文字，其字形、字音和字义之间关系复杂，测试员必须掌握和理解这方面的政策和已经公布的规范标准。

h、有关的方言理论与知识：掌握普通话和方言之间的关系，了解方言、特别是所在地区各方言在语言各要素方面与普通话之间的对应规律。具有分析和识别方言与普通话的能力。

i、语言学理论：主要是音位学方面的理论知识，能够运用常见国际音标记录语音现象，具有运用音位学理论分析语音现象的能力。

j、普通话水平测试理论：即与普通话水平测试有关的应用语言学理论和知识。如普通话水平测试的性质、对象和方法论等方面的理论知识。

对测试员来说，技术素质是外在的、直接的、实用的，而理论素养则是内在的、间接的、根本的。两者虽有不同，却紧密相联。我们可以作出这样一个合乎逻辑的推论：一个普通话口语能力较好的人，如果缺少上述专业理论方面的素养，经过短期的培训，可能也会在一定程度上获得测试应有的一些技术能力，但是由于这些能力没有足够的理论方法上的支持，可能是不全面、不稳定的，其测试的效果在一定程度上也就是不完全可靠的。而一个测试员具有相应的技术能力，又掌握了这些理论知识和方法，那么他就能对普通话水平测试本身、测试的评分方法、等级标准等作出更加全面清楚的理解，从而自觉地研究测试对象，分析测试对象的表现特征，不断总结测试经验，全面提高自己的测试技能。测试员只有多参加测试实践，才能够发现自己业务上存在的问题，找到自己努力的方向，不断提高自身的能力。

4.3 测试员的心理素质

(详见 2.2 测试员的心理因素对普通话水平测试的影响)

4.4 测试员的职业素质

普通话水平测试通过口试的方法进行，主要是由个体来操作，因此，测试员的职业素质在测试中也是影响测试结果非常重要的因素。个别测试员认为自己是主考官，主宰着应试者的“生杀大权”，因此测试时态度高傲，而且自觉不自觉地认为应试者的普通话水平不是很高。同时，测试员自身的普通话水平测试成绩一般处在 92—96 分之间，一甲的测试员可以说是凤毛麟角，因此人们通常觉得一甲非常难得。



g、国家有关语言文字政策、规范标准等方面的理论知识：普通话水平测试中有多个项目要借助书面进行，但汉字是非标音文字，其字形、字音和字义之间关系复杂，测试员必须掌握和理解这方面的政策和已经公布的规范标准。

h、有关的方言理论与知识：掌握普通话和方言之间的关系，了解方言、特别是所在地区各方言在语言各要素方面与普通话之间的对应规律。具有分析和识别方言与普通话的能力。

i、语言学理论：主要是音位学方面的理论知识，能够运用常见国际音标记录语音现象，具有运用音位学理论分析语音现象的能力。

j、普通话水平测试理论：即与普通话水平测试有关的应用语言学理论和知识。如普通话水平测试的性质、对象和方法论等方面的理论知识。

对测试员来说，技术素质是外在的、直接的、实用的，而理论素养则是内在的、间接的、根本的。两者虽有不同，却紧密相联。我们可以作出这样一个合乎逻辑的推论：一个普通话口语能力较好的人，如果缺少上述专业理论方面的素养，经过短期的培训，可能也会在一定程度上获得测试应有的一些技术能力，但是由于这些能力没有足够的理论方法上的支持，可能是不全面、不稳定的，其测试的效果在一定程度上也就是不完全可靠的。而一个测试员具有相应的技术能力，又掌握了这些理论知识和方法，那么他就能对普通话水平测试本身、测试的评分方法、等级标准等作出更加全面清楚的理解，从而自觉地研究测试对象，分析测试对象的表现特征，不断总结测试经验，全面提高自己的测试技能。测试员只有多参加测试实践，才能够发现自己业务上存在的问题，找到自己努力的方向，不断提高自身的能力。

4.3 测试员的心理素质

(详见 2.2 测试员的心理因素对普通话水平测试的影响)

4.4 测试员的职业素质

普通话水平测试通过口试的方法进行，主要是由个体来操作，因此，测试员的职业素质在测试中也是影响测试结果非常重要的因素。个别测试员认为自己是主考官，主宰着应试者的“生杀大权”，因此测试时态度高傲，而且自觉不自觉地认为应试者的普通话水平不是很高。同时，测试员自身的普通话水平测试成绩一般处在 92—96 分之间，一甲的测试员可以说是凤毛麟角，因此人们通常觉得一甲非常难得。



在测试时，具有一甲实力的考生的语音缺陷被不断放大，在最后评分时，只得到一乙的成绩也就不足为奇了。而三乙的考生，一般都年龄较大，而且多半长期在偏远的方言区生活，一段时间内很难练好普通话。在测试时，测试员对这类应试者普遍都有一种怜悯之心。“这么大年纪了，也已参加过培训，甚至有的考生都考过好几次了，怎么好只得个三乙呢？”于是同情之心顿起，把分数稍稍放松，三乙就成了三甲。还有的测试员在测试时，正好碰到熟人或“打了招呼”的人，于是在打分时，不排除存在人情分的情况。测播音员时的评分和测教师时的评分不同，测教师和测学生时的不相同，测毕业生和低年级学生时的评分不相同，测城里学生和测农村（方言区）学生的评分也不同。这种看人给分，打招呼分，打人情分的做法实际上是缺乏职业素质的表现。

普通话水平测试有着一套标准化的评分程序，但普通话水平测试采用的是口试“直接测试法”，即通过主试者和应试者面对面进行口语测试，评定测试成绩时或多或少地会掺杂主试者即测试员的主观判断成分，测试员的业务素质乃至心理素质等不可避免地会对测试结果产生影响，使得普通话水平测试的信度受到一定影响。因此，提高普通话测试员的思想素质、业务素质、心理素质、职业素质等，减轻个人因素的影响，排除人情因素的干扰，是保障普通话水平测试质量的关键。



5 对策研究

影响普通话水平测试的因素很多，如何克服和控制消极因素，促进普通话水平测试工作的正常开展，是我们必须认真思考的问题。

5.1 重视管理与培训工作

国家、省、市各级语委和普通话培训测试中心作为管理者，应该从以下几个角度来考虑测试员的培训、考核、安排等，只有这样，才能保证测试工作顺利完成，使测试质量逐步提高，测试信度逐渐增强，并培养出一批高素质的测试员队伍。

5.1.1 有计划地培训新测试员

按照当地普通话普及程度和人口范围、经济发展、地区分布等指标，科学的预算本地区进行普通话培训测试所需要的测试员的数量，以适应普通话培训测试的需要。拟定测试员培训计划，在从事普通话推广工作的广大教师、干部中选拔人才，分期分批进行培训。

测试员培训属于较高层次的专业培训，学员必须具备大专以上学历，普通话水平达到一级以上，有一定的语言学理论基础。通过培训掌握《普通话测试大纲》和评分标准、测试方法，具有较强的听音、辨音、注音的能力，能够准确、公正地为应试者评定等级。培训班教员必须具备较高的语言测试理论水平和较丰富的语言测试实践经验。还可以委托高校。按照统一的大纲、教学计划开展培训工作，逐步形成培训网络。用几年的时间，培养一支测试员队伍。参看近几年来某省级测试员培训班的教学计划，从1996年普通话语音与测评训练各占50%，到2006年的语音部分只占20%，测评实践占80%。课程设置也越来越合理。在进行测评训练时，从原来的测试四项综合训练到分单项逐项训练等。但我们也应看到，PSC是一项长期的、艰巨的任务，需要有大量的测试员参与。因此有足够数量测试员是开展和推进测试工作的先决条件。但普通话普及的现实状况决定了社会上、特别是方言区既有较为标准的普通话口语能力，又有一定语言理论素养的人员相对较少，加上其他一些客观条件的制约，测试员的推荐选拔存在一定的困难。为了使这项工作能够顺利开展，在实际工作中只能降低要求。这样，由于参加培训的人员的水平参差不齐，加上培训时间较短，培训效果自然不是十分理想。另外，岗前培训的性质决定了它的主要目的是让受训者对PSC从不了解达到初步的了解，对普通话标准从只有经验



5 对策研究

影响普通话水平测试的因素很多，如何克服和控制消极因素，促进普通话水平测试工作的正常开展，是我们必须认真思考的问题。

5.1 重视管理与培训工作

国家、省、市各级语委和普通话培训测试中心作为管理者，应该从以下几个角度来考虑测试员的培训、考核、安排等，只有这样，才能保证测试工作顺利完成，使测试质量逐步提高，测试信度逐渐增强，并培养出一批高素质的测试员队伍。

5.1.1 有计划地培训新测试员

按照当地普通话普及程度和人口范围、经济发展、地区分布等指标，科学的预算本地区进行普通话培训测试所需要的测试员的数量，以适应普通话培训测试的需要。拟定测试员培训计划，在从事普通话推广工作的广大教师、干部中选拔人才，分期分批进行培训。

测试员培训属于较高层次的专业培训，学员必须具备大专以上学历，普通话水平达到一级以上，有一定的语言学理论基础。通过培训掌握《普通话测试大纲》和评分标准、测试方法，具有较强的听音、辨音、注音的能力，能够准确、公正地为应试者评定等级。培训班教员必须具备较高的语言测试理论水平和较丰富的语言测试实践经验。还可以委托高校。按照统一的大纲、教学计划开展培训工作，逐步形成培训网络。用几年的时间，培养一支测试员队伍。参看近几年来某省级测试员培训班的教学计划，从1996年普通话语音与测评训练各占50%，到2006年的语音部分只占20%，测评实践占80%。课程设置也越来越合理。在进行测评训练时，从原来的测试四项综合训练到分单项逐项训练等。但我们也应看到，PSC是一项长期的、艰巨的任务，需要有大量的测试员参与。因此有足够数量测试员是开展和推进测试工作的先决条件。但普通话普及的现实状况决定了社会上、特别是方言区既有较为标准的普通话口语能力，又有一定语言理论素养的人员相对较少，加上其他一些客观条件的制约，测试员的推荐选拔存在一定的困难。为了使这项工作能够顺利开展，在实际工作中只能降低要求。这样，由于参加培训的人员的水平参差不齐，加上培训时间较短，培训效果自然不是十分理想。另外，岗前培训的性质决定了它的主要目的是让受训者对PSC从不了解达到初步的了解，对普通话标准从只有经验



的表达能力达到基本的理论确认,初步掌握 PSC 的测试方式、测试内容和测试操作技能,在培训内容的安排上便侧重于普通话标准及能力和测试方式与技能的获得,而不能在一些重要的理论方面给予更加集中和深入的培养。因此测试员的岗前培训在相当大的程度上只能是测试员数量上的准备,而不能较好地兼顾 PSC 对测试员综合素质的更高要求。

5.1.2 加强后续培训,实行定期考核

现在国家语委培训测试中心及各省的培训中心对测试员的培训基本上是以培训测试能力为主,内容主要是熟悉《大纲》和测试实践训练,时间大多为两周,这种培训对于一部分有较高语音水平和较好语音听辨基础的测试员来讲,是符合“多快好省”原则的,但对于那些未受过正规语音学教育或口语水平有一定欠缺的测试员来讲,这种培训就显得很不够。如有些去参加省级测试员培训的教师连一般的汉语拼音能力都没掌握,两周之内,既要训练提高自己的口语水平,又要具备较高水平的语音听辨能力,就是勉强拿到省级测试员资格证书,若不进行后续培训,那么在具体操作中进行量化评判时,能否达到科学公正就可想而知了。测试管理机构应该对本地区的测试员有计划,有步骤地进行后续培训工作。在培训中注意加强教育测量学、心理测量学、数学、外语等的培训。作为一名优秀的测试员,普通话水平不能仅仅只要求自己 92 分,而应该到达 95 分以上。下面是某市 2006 年对全市省级测试员进行考核的成绩表。语音考核中有 132 人参加测试,126 人 92 分以上,占 95.46%; 43 人 95 分以上,占 32.6%; 6 人 92 分以下,占 4.45%。测试能力测考核中,60 分以上的 114 人,占 86.3%; 60 分以下的 18 人,占 13.63%; 80 分以上的 53 人,占 53%。

对于一名测试员来说,语音问题也不是一成不变的。有的测试员由于拿到资格证书后沾沾自喜,忽视了后续培训,导致普通话水平滑坡。而另外有一名测试员,在接受测试员资格培训考核之前,一直都是 91 分,二级甲等。在测试员培训班上,她刻苦学习,努力纠正自己的字音,终于以 92 分拿到了测试员资格证书。经过几年的锻炼,在市里的培训中,考到了 95 分,成为一名优秀测试员。但是她没有停住语言学习的脚步,仍然继续努力,参加省里的培训时,考出了 97.3 分的好成绩,并通过了国家语委的复审。

除了培训,还要定期认真开展考核工作。测试机构应定期对国家级测试员、省级测试员进行工作考核。根据考核结果,对表现出色的测试员进行表彰,对表现差和违反纪律的测试员进行批评,不称职的由聘任单位收回测试员聘书。将测试员的上岗资格和职务聘任分开,对测试员从思想素质、业务水平、工作量等各方面做好



的表达能力达到基本的理论确认,初步掌握 PSC 的测试方式、测试内容和测试操作技能,在培训内容的安排上便侧重于普通话标准及能力和测试方式与技能的获得,而不能在一些重要的理论方面给予更加集中和深入的培养。因此测试员的岗前培训在相当大的程度上只能是测试员数量上的准备,而不能较好地兼顾 PSC 对测试员综合素质的更高要求。

5.1.2 加强后续培训,实行定期考核

现在国家语委培训测试中心及各省的培训中心对测试员的培训基本上是以培训测试能力为主,内容主要是熟悉《大纲》和测试实践训练,时间大多为两周,这种培训对于一部分有较高语音水平和较好语音听辨基础的测试员来讲,是符合“多快好省”原则的,但对于那些未受过正规语音学教育或口语水平有一定欠缺的测试员来讲,这种培训就显得很不够。如有些去参加省级测试员培训的教师连一般的汉语拼音能力都没掌握,两周之内,既要训练提高自己的口语水平,又要具备较高水平的语音听辨能力,就是勉强拿到省级测试员资格证书,若不进行后续培训,那么在具体操作中进行量化评判时,能否达到科学公正就可想而知了。测试管理机构应该对本地区的测试员有计划,有步骤地进行后续培训工作。在培训中注意加强教育测量学、心理测量学、数学、外语等的培训。作为一名优秀的测试员,普通话水平不能仅仅只要求自己 92 分,而应该到达 95 分以上。下面是某市 2006 年对全市省级测试员进行考核的成绩表。语音考核中有 132 人参加测试,126 人 92 分以上,占 95.46%; 43 人 95 分以上,占 32.6%; 6 人 92 分以下,占 4.45%。测试能力测考核中,60 分以上的 114 人,占 86.3%; 60 分以下的 18 人,占 13.63%; 80 分以上的 53 人,占 53%。

对于一名测试员来说,语音问题也不是一成不变的。有的测试员由于拿到资格证书后沾沾自喜,忽视了后续培训,导致普通话水平滑坡。而另外有一名测试员,在接受测试员资格培训考核之前,一直都是 91 分,二级甲等。在测试员培训班上,她刻苦学习,努力纠正自己的字音,终于以 92 分拿到了测试员资格证书。经过几年的锻炼,在市里的培训中,考到了 95 分,成为一名优秀测试员。但是她没有停住语言学习的脚步,仍然继续努力,参加省里的培训时,考出了 97.3 分的好成绩,并通过了国家语委的复审。

除了培训,还要定期认真开展考核工作。测试机构应定期对国家级测试员、省级测试员进行工作考核。根据考核结果,对表现出色的测试员进行表彰,对表现差和违反纪律的测试员进行批评,不称职的由聘任单位收回测试员聘书。将测试员的上岗资格和职务聘任分开,对测试员从思想素质、业务水平、工作量等各方面做好



统计和考核工作。打破组长、复审终身制。组长、复审要能上能下，将优秀的省级测试员扩充到组长、复审的队伍中；并将长期不测试或态度不认真、业务水平下降的组长、复审予以解聘。组长、复审一年一聘，以利于此项工作的顺利开展。

5.1.3 建立测评监督体制

要保证测试质量，每场测试都要有业务组长，另外还须配备审听和复审，监督测试员的工作。组长、审听员和复审员由语言专家和优秀的国家级测试员担任。所谓审听，就是审听员到考场中和测试员一起测试应试者，根据评分情况，进行研讨；所谓复审，就是将已经测试过的应试者的录音由复审员测评，从中发现测试测试中可能出现的问题，有问题要进行讨论，以提高测试员的业务能力。而组长则应把每次测试中出现的各类情况作详细记录，报到省市测试中心，使每场测试在业务上都有负责人和责任人。

5.1.4 开展测试理论研究

目前，测试理论的研究在我国尚处于初期，而在国外的同行中，有很多都有了研究成果。测试管理机构应该加强测试员的理论研究，使测试员变埋头苦干型为科研型；变单一型为复合型；变单一型为配合型；变保守型为创新型。理论研究应立足于本行，向其他领域渗透。在测试实践的基础上进行定量统计和分析。摒弃传统的“没有功劳也有苦劳”，认为只要辛辛苦苦工作就行了的思想。如果不勤于思考，把自己的实践上升到理论层面，进一步提高就比较困难。

PSC只有短短十年的发展历史，作为应用性学科，其理论和方法涉及面较广，从事PSC理论与实践研究的专家学者相对较少。因此，省测试中心和各测试站大多没有能力或少有能力对测试员进行测试理论方面更加科学深入的综合培养。而国家测试中心虽然具有较多专家资源，但面对全国几十个省区测试员培养的需要，也没有能力按其所需给予及时全面、系统培养。而且目前测试员培养都是集中培训，方式上显得比较单一，虽然有它的优点，却不甚灵活，不能针对测试中出现的较为集中的问题，作出及时的研究性指导。另外在是培训经费上没有保证。“以测养测”是目前各测试站乃至测试中心在经费使用方面的基本情况，较少有其他来源，这在一定程度上制约了测试事业的发展，也制约了测试员队伍的建设。测试中心应及早意识到科研型测的力量，多争取各种科研课题，让测试员多参加科研活动，加强理论修养，促进测试实践。

除上述几个方面要加强之外，还要对测试员加强人文关怀。测试员都来自不同的工作岗位，为着相同的目标——推广普通话，维护祖国语言文字的纯洁性走到了



统计和考核工作。打破组长、复审终身制。组长、复审要能上能下，将优秀的省级测试员扩充到组长、复审的队伍中；并将长期不测试或态度不认真、业务水平下降的组长、复审予以解聘。组长、复审一年一聘，以利于此项工作的顺利开展。

5.1.3 建立测评监督体制

要保证测试质量，每场测试都要有业务组长，另外还须配备审听和复审，监督测试员的工作。组长、审听员和复审员由语言专家和优秀的国家级测试员担任。所谓审听，就是审听员到考场中和测试员一起测试应试者，根据评分情况，进行研讨；所谓复审，就是将已经测试过的应试者的录音由复审员测评，从中发现测试测试中可能出现的问题，有问题要进行讨论，以提高测试员的业务能力。而组长则应把每次测试中出现的各类情况作详细记录，报到省市测试中心，使每场测试在业务上都有负责人和责任人。

5.1.4 开展测试理论研究

目前，测试理论的研究在我国尚处于初期，而在国外的同行中，有很多都有了研究成果。测试管理机构应该加强测试员的理论研究，使测试员变埋头苦干型为科研型；变单一型为复合型；变单一型为配合型；变保守型为创新型。理论研究应立足于本行，向其他领域渗透。在测试实践的基础上进行定量统计和分析。摒弃传统的“没有功劳也有苦劳”，认为只要辛辛苦苦工作就行了的思想。如果不勤于思考，把自己的实践上升到理论层面，进一步提高就比较困难。

PSC只有短短十年的发展历史，作为应用性学科，其理论和方法涉及面较广，从事PSC理论与实践研究的专家学者相对较少。因此，省测试中心和各测试站大多没有能力或少有能力对测试员进行测试理论方面更加科学深入的综合培养。而国家测试中心虽然具有较多专家资源，但面对全国几十个省区测试员培养的需要，也没有能力按其所需给予及时全面、系统培养。而且目前测试员培养都是集中培训，方式上显得比较单一，虽然有它的优点，却不甚灵活，不能针对测试中出现的较为集中的问题，作出及时的研究性指导。另外在是培训经费上没有保证。“以测养测”是目前各测试站乃至测试中心在经费使用方面的基本情况，较少有其他来源，这在一定程度上制约了测试事业的发展，也制约了测试员队伍的建设。测试中心应及早意识到科研型测的力量，多争取各种科研课题，让测试员多参加科研活动，加强理论修养，促进测试实践。

除上述几个方面要加强之外，还要对测试员加强人文关怀。测试员都来自不同的工作岗位，为着相同的目标——推广普通话，维护祖国语言文字的纯洁性走到了



统计和考核工作。打破组长、复审终身制。组长、复审要能上能下，将优秀的省级测试员扩充到组长、复审的队伍中；并将长期不测试或态度不认真、业务水平下降的组长、复审予以解聘。组长、复审一年一聘，以利于此项工作的顺利开展。

5.1.3 建立测评监督体制

要保证测试质量，每场测试都要有业务组长，另外还须配备审听和复审，监督测试员的工作。组长、审听员和复审员由语言专家和优秀的国家级测试员担任。所谓审听，就是审听员到考场中和测试员一起测试应试者，根据评分情况，进行研讨；所谓复审，就是将已经测试过的应试者的录音由复审员测评，从中发现测试测试中可能出现的问题，有问题要进行讨论，以提高测试员的业务能力。而组长则应把每次测试中出现的各类情况作详细记录，报到省市测试中心，使每场测试在业务上都有负责人和责任人。

5.1.4 开展测试理论研究

目前，测试理论的研究在我国尚处于初期，而在国外的同行中，有很多都有了研究成果。测试管理机构应该加强测试员的理论研究，使测试员变埋头苦干型为科研型；变单一型为复合型；变单一型为配合型；变保守型为创新型。理论研究应立足于本行，向其他领域渗透。在测试实践的基础上进行定量统计和分析。摒弃传统的“没有功劳也有苦劳”，认为只要辛辛苦苦工作就行了的思想。如果不勤于思考，把自己的实践上升到理论层面，进一步提高就比较困难。

PSC只有短短十年的发展历史，作为应用性学科，其理论和方法涉及面较广，从事PSC理论与实践研究的专家学者相对较少。因此，省测试中心和各测试站大多没有能力或少有能力对测试员进行测试理论方面更加科学深入的综合培养。而国家测试中心虽然具有较多专家资源，但面对全国几十个省区测试员培养的需要，也没有能力按其所需给予及时全面、系统培养。而且目前测试员培养都是集中培训，方式上显得比较单一，虽然有它的优点，却不甚灵活，不能针对测试中出现的较为集中的问题，作出及时的研究性指导。另外在是培训经费上没有保证。“以测养测”是目前各测试站乃至测试中心在经费使用方面的基本情况，较少有其他来源，这在一定程度上制约了测试事业的发展，也制约了测试员队伍的建设。测试中心应及早意识到科研型测的力量，多争取各种科研课题，让测试员多参加科研活动，加强理论修养，促进测试实践。

除上述几个方面要加强之外，还要对测试员加强人文关怀。测试员都来自不同的工作岗位，为着相同的目标——推广普通话，维护祖国语言文字的纯洁性走到了



一起。省市语委、普通话培训测试中心作为管理机构，应该加强对测试员的人文关怀。在安排测试时，最好按照国家语委的要求，每个单位时间（半天）测试 15 人。有的地方一个单位时间内到达 20—30 人，这样超负荷的工作，会使测试员感到疲倦，或者为完成工作量只好缩短测试程序，影响测试质量。因此要为测试员提供必要的帮助。使测试员在测试工作中干得舒心，放心，有信心。营造出一种良好和谐的气氛，促进测试工作的规范化，提高测试质量。

只有通过测试管理机构对测试员严格的培养、关心和爱护，才能造就一批高素质的测试员队伍，保证测试工作有序、高质，推动全民普通话推广工作。

5.2 加强测试员测评能力训练

测试员的角色是多重的，但是在进行普通话水平测试时，其身份就是考官，要有良好的测评能力，准确无误地判定出应试者的普通话水平。

2006 年，某地区对其所在地的国家级测试员进行了后续培训和考核。在考核中，通过测评实践侧重于考核测试员的测评能力。测试中心选取了五名代表不同等级的应试者，由国家级测试员分别对这五名应试者进行测评，以下是测试员对五名应试者测试的结果：

考生	成绩	等级	评为一甲		评为一乙		评为二甲		评为二乙		评为三甲	
			人数	%	人数	%	人数	%	人数	%	人数	%
1	97.28	一甲	13	14	71	79	9	7				
2	93.95	一乙	4	4	77	83	12	13				
3	88.55	二甲			2	2	77	83	14	15		
4	84.65	二乙					12	13	80	86	1	1
5	76.41	三甲							2	2	91	98

表 5-1 某地区对其所在地的国家级测试员考核测试结果

从以上测试结果可以看出：测试员对一甲的评定正确率为 14%；一乙为 83%；二甲为 83%；二乙为 86%；三甲为 98%。也就是对一乙、二甲、二乙和三甲的评定正确率较高；对一甲的评定有较大的误差；对一乙、二甲的评定也不同程度地存在着偏严的现象。另外，在测试实践中，测试员对三乙的评定也存在不同程度的问题。



一起。省市语委、普通话培训测试中心作为管理机构，应该加强对测试员的人文关怀。在安排测试时，最好按照国家语委的要求，每个单位时间（半天）测试 15 人。有的地方一个单位时间内到达 20—30 人，这样超负荷的工作，会使测试员感到疲倦，或者为完成工作量只好缩短测试程序，影响测试质量。因此要为测试员提供必要的帮助。使测试员在测试工作中干得舒心，放心，有信心。营造出一种良好和谐的气氛，促进测试工作的规范化，提高测试质量。

只有通过测试管理机构对测试员严格的培养、关心和爱护，才能造就一批高素质的测试员队伍，保证测试工作有序、高质，推动全民普通话推广工作。

5.2 加强测试员测评能力训练

测试员的角色是多重的，但是在进行普通话水平测试时，其身份就是考官，要有良好的测评能力，准确无误地判定出应试者的普通话水平。

2006 年，某地区对其所在地的国家级测试员进行了后续培训和考核。在考核中，通过测评实践侧重于考核测试员的测评能力。测试中心选取了五名代表不同等级的应试者，由国家级测试员分别对这五名应试者进行测评，以下是测试员对五名应试者测试的结果：

考生	成绩	等级	评为一甲		评为一乙		评为二甲		评为二乙		评为三甲	
			人数	%	人数	%	人数	%	人数	%	人数	%
1	97.28	一甲	13	14	71	79	9	7				
2	93.95	一乙	4	4	77	83	12	13				
3	88.55	二甲			2	2	77	83	14	15		
4	84.65	二乙					12	13	80	86	1	1
5	76.41	三甲							2	2	91	98

表 5-1 某地区对其所在地的国家级测试员考核测试结果

从以上测试结果可以看出：测试员对一甲的评定正确率为 14%；一乙为 83%；二甲为 83%；二乙为 86%；三甲为 98%。也就是对一乙、二甲、二乙和三甲的评定正确率较高；对一甲的评定有较大的误差；对一乙、二甲的评定也不同程度地存在着偏严的现象。另外，在测试实践中，测试员对三乙的评定也存在不同程度的问题。



可见,在测试标准已经统一的情况下,不同的测试员按统一标准,对同一被试者仍可评出不同的等级。因此,加强测试员测评能力的训练非常有必要,这有利于克服诸多对测试员产生影响的因素。

5.2.1 克服应试者整体水平对测试员的影响

我国幅员辽阔,人口众多,加上历史的多种因素,形成我国汉语方言复杂、分歧严重的现实。方言的复杂对普通话的学习有很大影响。方言区的人学说普通话,除了方言语音影响外,往往还忽视了方言词语与普通话词语在音、形、义等方面的差异。他们往往将方言词语套上相应的普通话声调就认为是普通话词语了,因此,普通话水平测试的成绩较差。目前参加测试的人员,主要是师范生、教师、公务员等,他们大部分人的水平也就在一乙和三甲之间。如某地区 2006 年测试的考生共有 32819 人,其中一甲 2 人,一乙 3068 人,占 9.35%;二甲 11766 人,二乙 14070 人,占 78.72%;三甲 3235 人,三乙 144 人,占 10.30%;不晋级 3 人,占 0.01%。这符合测试中成绩正态分布的特点。由于平时测试员在测试中主要接触的就是这几个等级之间的考生,对于这几个等级的评定已经适应,其测试信度自然较高,而对于一甲、三乙的考生,由于平时接触得较少,适应性也比较差,因此测试的信度较低。

5.2.2 克服测试员自身语音水平的影响

省级或国家级测试员,其自身的普通话水平测试成绩要求是在 92 分,一级乙等以上。因为普通话水平测试的时间是很有限的,而且需要当场测定,这就需要测试员在短短的十分钟左右时间内,对应试者的语音情况做出准确判断。普通话水平测试的这种特点,要求测试员自身具有很高语音水平,如果仅仅只是刚够一级的低分水平,就很难成为一个优秀的测试员。2000 年,某市对 133 名省级测试员进行了后续培训和考核。参加测试的测试员,成绩在 92 分以上的有 126 人,占 95.46%,其中,92—93 分的 14 人,93—94 分的 39 人,94—95 分的 30 人,95 分以上的 43 人。但还有 6 人的成绩在 92 分以下,占 4.54%。除了测试成绩并不都是很高以外,有些测试员还存在着一些语音、语调错误或系统性的缺陷等,这就使得测试员在测评具有一甲实力的应试者时,觉得难以确定。另外,有些测试员有时可能听辨不出应试者存在的某些语音问题,也会导致测评的信度的降低。

5.2.3 克服测试员对评分标准理解的误差

测试是一种客观测量的方法,要客观地进行测量,必须做到在评分时采取客观的态度,排除种种干扰。参加评分的人员应该严格按照评分标准评分,自始至终用



可见,在测试标准已经统一的情况下,不同的测试员按统一标准,对同一被试者仍可评出不同的等级。因此,加强测试员测评能力的训练非常有必要,这有利于克服诸多对测试员产生影响的因素。

5.2.1 克服应试者整体水平对测试员的影响

我国幅员辽阔,人口众多,加上历史的多种因素,形成我国汉语方言复杂、分歧严重的现实。方言的复杂对普通话的学习有很大影响。方言区的人学说普通话,除了方言语音影响外,往往还忽视了方言词语与普通话词语在音、形、义等方面的差异。他们往往将方言词语套上相应的普通话声调就认为是普通话词语了,因此,普通话水平测试的成绩较差。目前参加测试的人员,主要是师范生、教师、公务员等,他们大部分人的水平也就在一乙和三甲之间。如某地区 2006 年测试的考生共有 32819 人,其中一甲 2 人,一乙 3068 人,占 9.35%;二甲 11766 人,二乙 14070 人,占 78.72%;三甲 3235 人,三乙 144 人,占 10.30%;不晋级 3 人,占 0.01%。这符合测试中成绩正态分布的特点。由于平时测试员在测试中主要接触的就是这几个等级之间的考生,对于这几个等级的评定已经适应,其测试信度自然较高,而对于一甲、三乙的考生,由于平时接触得较少,适应性也比较差,因此测试的信度较低。

5.2.2 克服测试员自身语音水平的影响

省级或国家级测试员,其自身的普通话水平测试成绩要求是在 92 分,一级乙等以上。因为普通话水平测试的时间是很有限的,而且需要当场测定,这就需要测试员在短短的十分钟左右时间内,对应试者的语音情况做出准确判断。普通话水平测试的这种特点,要求测试员自身具有很高语音水平,如果仅仅只是刚够一级的低分水平,就很难成为一个优秀的测试员。2000 年,某市对 133 名省级测试员进行了后续培训和考核。参加测试的测试员,成绩在 92 分以上的有 126 人,占 95.46%,其中,92—93 分的 14 人,93—94 分的 39 人,94—95 分的 30 人,95 分以上的 43 人。但还有 6 人的成绩在 92 分以下,占 4.54%。除了测试成绩并不都是很高以外,有些测试员还存在着一些语音、语调错误或系统性的缺陷等,这就使得测试员在测评具有一甲实力的应试者时,觉得难以确定。另外,有些测试员有时可能听辨不出应试者存在的某些语音问题,也会导致测评的信度的降低。

5.2.3 克服测试员对评分标准理解的误差

测试是一种客观测量的方法,要客观地进行测量,必须做到在评分时采取客观的态度,排除种种干扰。参加评分的人员应该严格按照评分标准评分,自始至终用



可见,在测试标准已经统一的情况下,不同的测试员按统一标准,对同一被试者仍可评出不同的等级。因此,加强测试员测评能力的训练非常有必要,这有利于克服诸多对测试员产生影响的因素。

5.2.1 克服应试者整体水平对测试员的影响

我国幅员辽阔,人口众多,加上历史的多种因素,形成我国汉语方言复杂、分歧严重的现实。方言的复杂对普通话的学习有很大影响。方言区的人学说普通话,除了方言语音影响外,往往还忽视了方言词语与普通话词语在音、形、义等方面的差异。他们往往将方言词语套上相应的普通话声调就认为是普通话词语了,因此,普通话水平测试的成绩较差。目前参加测试的人员,主要是师范生、教师、公务员等,他们大部分人的水平也就在一乙和三甲之间。如某地区 2006 年测试的考生共有 32819 人,其中一甲 2 人,一乙 3068 人,占 9.35%;二甲 11766 人,二乙 14070 人,占 78.72%;三甲 3235 人,三乙 144 人,占 10.30%;不晋级 3 人,占 0.01%。这符合测试中成绩正态分布的特点。由于平时测试员在测试中主要接触的就是这几个等级之间的考生,对于这几个等级的评定已经适应,其测试信度自然较高,而对于一甲、三乙的考生,由于平时接触得较少,适应性也比较差,因此测试的信度较低。

5.2.2 克服测试员自身语音水平的影响

省级或国家级测试员,其自身的普通话水平测试成绩要求是在 92 分,一级乙等以上。因为普通话水平测试的时间是很有限的,而且需要当场测定,这就需要测试员在短短的十分钟左右时间内,对应试者的语音情况做出准确判断。普通话水平测试的这种特点,要求测试员自身具有很高语音水平,如果仅仅只是刚够一级的低分水平,就很难成为一个优秀的测试员。2000 年,某市对 133 名省级测试员进行了后续培训和考核。参加测试的测试员,成绩在 92 分以上的有 126 人,占 95.46%,其中,92—93 分的 14 人,93—94 分的 39 人,94—95 分的 30 人,95 分以上的 43 人。但还有 6 人的成绩在 92 分以下,占 4.54%。除了测试成绩并不都是很高以外,有些测试员还存在着一些语音、语调错误或系统性的缺陷等,这就使得测试员在测评具有一甲实力的应试者时,觉得难以确定。另外,有些测试员有时可能听辨不出应试者存在的某些语音问题,也会导致测评的信度的降低。

5.2.3 克服测试员对评分标准理解的误差

测试是一种客观测量的方法,要客观地进行测量,必须做到在评分时采取客观的态度,排除种种干扰。参加评分的人员应该严格按照评分标准评分,自始至终用



“一把尺子”去测量,判断优劣。测试员参加测试的时间有长有短、次数有多有少,实践经验的积累不够和自身水平的限制,都可能会使某些测试员在评分标准的把握上产生误差,这也会影响到测试结果的正确判定,导致测试的信度降低。

上述因素对普通话水平测试产生的影响,有的体现在整体上,有的体现在具体的测试项目上。

在测试的第一、二项“读单音节字词和双音节词语”时,由于错误、缺陷是定量统计,因此测试员一般都容易把握,评分都很准确。第三、四项“朗读”和“说话”两项中,“朗读”的错漏、停顿、语速项扣分和“说话”的词汇语法、自然流畅、离题等几项的扣分,测试员把握得都很好,评分准确。但在“朗读”项的“声韵缺陷”和说话项“语调偏误”的掌握上,测试员之间评分有较大的差别,而这两项扣分的多少,将影响考生的成绩,直接关系到考生进入哪个级别。“语调”提出的时间较早,是一个被学术界研究多年的语言学概念。对于“语调”的研究到目前为止仍然是比较薄弱的环节,学术界历来说法不一,还没有形成一套公认的分析 and 描写语调的方法,至今也没有定性的概念表述。下面我们看一下在学术界有一定影响的工具书和代表性著作对“语调”的表述:

(1)说话的腔调,就是一句话里语音高低轻重的配置。(《现代汉语词典》)

(2)“语调”与“字调”相对而言,是指全句的高、低、抑、扬,和“字调”同是“音高”现象。……汉语的语调,特别显示在语句末尾的音节上。(徐世荣《普通话语音常识》)

(3)语调是一个句子中间高低、快慢、轻重、停顿的各种变化,同音高、音长、音强都有联系。(周殿福《艺术语言发声基础》)

(4)说话或朗读时,句子有停顿,声音有轻重快慢和高低长短的变化,这些总称语调。(黄伯荣、廖序东《现代汉语》)

(5)……能够帮助表达说话人的思想和感情态度,由此而产生的全句抑扬顿挫和其它方面的语音变化就是语调。语调主要由超音质成分,即音高、音强和音长组成。(林焘、王理嘉《语音学教程》)

语调的基本单位是句调,这是诸多研究者一致认同的。从上面各家的表述可以看出语言学界对“语调”研究的分歧主要集中于两点:一种观点认为语调就是指整个语句的声音高低变化(即音高变化);另一种观点则认为语调是由音高、音长和音强等多种因素构成的句子的节律模式。我们认为后一种观点的表述较为客观全面,某种程度上反映了汉语语句表达的基本特点:从句调的内部特征来看,音高特征明显,具体表现在句子声音的高低升降曲折变化上。从句调的外部特征看,音长、音高特



征都重要,有时音长的区别性特征反而更加明显。其中“节律模式”的包括范围不仅指句调的抑扬,同时也包含了语句停顿的长短、读音的轻重以及语速的快慢等。

综上所述,语调应该是汉语语句表达中快慢、高低、强弱、虚实等各种声音形式的总和,是由表达时语气的色彩和分量决定的。“语气的色彩、语气的分量,在有声语言中都具有社会性,都是可感的。但是,作为语句的声音形式,应该有自己的完整的运动过程。我们认为,语调就应该反映出这个运动过程。”^①

用周殿福先生的话来概括,“语调研究是语音研究的各部分的总和”^② 语调是字调、句调以及轻重格等因素的综合。对语调的要求,《普通话水平测试大纲》对每一个级别的标准为:

等级	基本要求
一级甲等	语调自然,表达流畅
一级乙等	语调自然,表达流畅,偶有字音、字调失误
二级甲等	语调自然,表达流畅,少数难点音有时出现失误,词汇语法极少有失误
二级乙等	方言语调不明显。有使用方言词汇、方言语法的情况
三级甲等	方言语调较明显,词汇、语法有失误。
三级乙等	方言语调明显,词汇、语法失误较多,外地人听其谈话有听不懂的现象

表 5-2 对语调的要求

在以上对每一个级别的要求中,特别是对“语调”的论述中,一级、二级甲等均为“语调自然”,即不能有方言语调;二级乙等为“方言语调不明显”,即只允许有不明显的方言语调;三级甲等为“方言语调较明显”,三级乙等为“方言语调明显”。方言语调的扣分往往是依据定性分析。按照国家测试大纲的标准,一级、二甲不应有方言语调;二乙的语调略有反映,扣1分;三甲的方言语调比较明显,扣1.5分;三乙的方言语调问题突出,扣2分。语调偏误普通话水平测试中,语调型式不符合普通话规范。语调研究是语音研究的难点,普通话水平测试对语调的评价,原《大纲》主要通过“方言语调”这一测试要素体现。但“方言语调”的提法从一开始便争议不断,一是概念本身内涵不清,二是不能对测试中的很多语调现象科学

^① 周殿福:《艺术语言发声基础》,中国社会科学出版社1980版。

^② 周殿福:《艺术语言发声基础》,中国社会科学出版社1980版。



评价(如,逻辑重音处理不当,很难用“方言语调”评价)。新《大纲》用“语调偏误”取代“方言语调”,内涵更明确,称说更科学。

从对每一级的描述中,可以看出,方言语调扣了分,就不能进入二甲以上的等级,也就是说,二甲、一级水平者在“语调偏误”一项不能失分。

由于语调主要由超音质成分,即音高、音强和音长组成,因此我们认为以上五个方面是“语调偏误”最为明显的特征,在进行普通话朗读项的测评时应该重点加以考察。

通过上述分析,我们在测试时不妨可以这样理解:普通话测试朗读项“方言语调”评判的依据主要是字调、句调、轻重格、节奏以及方言叹词等,字音错误和缺陷的类型可以作为“方言语调”评判的必要参考;前者为主,后者为辅。

由于国家测试大纲中“语调偏误”是定性分析,测试员只能凭感觉认定方言语调是“略有反映”“比较明显”还是“问题突出”,有时很不好把握。为了解决这个问题,便于测试员把握标准,某些地区在“语调偏误”中把定性的标准变为定量的,并予以分类,以利操作。语调偏误主要表现在:

(1)成系统的声调(字调)错误,使语调明显带有方言特征在普通话语调诸多因素中,声调(字调)是最为敏感的,“字调”是“语调”的基础或灵魂。这其中调形是首要因素,其次是在调形正确的前提下,调值是否到位(例如把阴平调值 55 念做 44 甚至 33,把上声调值 214 念做 211 或 212、213 等),也会影响到普通话语调。汉语声调和语调的关系非常复杂。赵元任先生曾经打过一个形象的比喻,他把声调比作小波浪,把语调比作大波浪。汉语是小波浪在大波浪上面,实际结果是两种代数之和。“由于语调和字调并存于语句之中,是一种叠加的关系,方言与普通话在声调(字调)调形上的差异越大(如一个‘升’一个‘降’的差异),表现在语调上的方言色彩越浓。”因此,方言声调(字调)的遗留,表现在语句中一定会影响普通话语调的准确,是构成方言语调的重要特征之一。

(2)句调基本调形错误,形成固定腔调句调调形是指句子内部的升降曲折变化和调势走向。根据贯穿整个句子语流波的升降曲折变化,汉语句调有升调、降调、平调、曲调四种基本调形:由于汉语是声调语言,相同调形的句子可以因为声调不同而产生不同的调形。一般的情况是以双音节音组(词)为语调的基础,调形顺势简化,前后音节调值的过渡尽可能平滑,并且有偏低的趋势。从整句话来看,音高变化的总体趋势仍然很明显,句尾升降调的音阶会高一些或低一些。应该说句调在句尾表现得最为明显。朗读时如果把语句的基本调形念错,就容易带有固定腔调,也就意味着带有方言语调。



(3) 成系统的轻重音错误

轻重音的方言模式是方言语调的另一个重要特点。轻重音主要表现为以下几种形式：第一，词的轻重音错误。这是轻重音模式中最基本的。一个词的几个音节，在实际呼读时，音量并不是均衡的。汉语的词，也和别的民族语言一样，一个词有一个重音。这个重音可以明显地表示这几个音节的密集性，同时可以依据重音来辨认是一个词，还是平行并立的两个词。一般来说，普通话的轻重音可以细分为四个等级，即重音、中音、次轻音、最轻音。普通话双音节词语的格式大多数是“中重”格式，但是在方言中却有把“中重”格式的词语读作“重中”格式或是“重轻(次轻、最轻)”，实际上也就是把第一个音节读得重而长，把第二个音节读得轻而短。这种念法与普通话的轻重音格式明显形成听感上的差异，因而就带有方言语调了。第二，语句的逻辑重音错误。句重音是在一句话或者一个音组中，一个或者几个音节发音在音强甚至音高、音长上突出而形成的重音。句重音往往与句子、词组的结构有关。

“逻辑重音”又叫做“强调重音”，为了强调句子的某一内容而把某个或几个音节说得用力一些，使它更加突出一些，这样形成的重音就叫逻辑重音。逻辑重音与表达的内容密切相关，提供的是说话者需要强调的意义信息，由不同的具体需要决定。当朗读者使用某种固定不变的声音形式，把词语纳入一种单一的格式，就会出现逻辑重音错误，普通话语句本身具有的重音模式发生了改变，从而形成方言语调。第三，轻声音节的重读。普通话的轻声音节大多是固定的，主要包括单音节中的一些结构助词、语气助词、方位词、趋向动词、词缀，双音节词后音节的轻声，三音节词的中缀和后缀，以及重叠式中的轻音。方言区的人在读上述轻声音节时，往往把这些轻声音节读得时值偏长，由于时长是区别汉语轻重音的主要特征，因此实际上已经失去轻声音节固有的语音特征了。

(4) 由语调整奏的变异而产生方言语调，包括语速、停顿、节拍群与普通话的差异。语句节奏是指语音的徐疾、高低、长短、轻重及音色的异同在一定时间内有规律地相间交替回环往复成周期性组合的结果。朗读文章时，如果不注意节奏的形式、层次，一字一顿，虽然意思能够听明白，不影响语义的表达，但听感上支离破碎；或者朗读时语速过快，造成滑音(吃字)现象，与普通话固有的节奏模式形成差异，这种差异就是方言语调。

(5) 大量方言语气词、感叹词的使用，包括普通话语气词“啊”的音变错误，也是方言语调的特征之一。汉语方言中的语气词、感叹词一般不会出现在朗读项的作品中，如果有，也是文体本身的问题，这与朗读者无关。普通话语气词“啊”有其特定的变化规则，此处不再赘述。但朗读时如果添加方言语气词、感叹词，或用方言语气



词、感叹词替换朗读文本中普通话的语气词、感叹词,该音变的“啊”没有音变,或音变错误,均应视为有语调偏误。“语调偏误”主要是扣“缺陷”分。每出现一类上述缺陷就扣 0.5 分,而要达到“一类”,也有一个量化的标准,那就是五次以上方可称为“一类”。尽管如此,在具体测试实践中,有的测试员仍然把被测者朗读时出现的一次两次缺陷视同出现了一类问题,于是就被扣去一个 0.5 分,甚至有的应试者达到了一级水平,在“语调偏误”一项中也被扣去 0.5 分。这样扣分的结果是降低了被测者的分数,造成评分偏严。

另外,在说话一项中,“语音面貌”的扣分也存在着不同的测试员松紧不同,或同一测试员时紧时松的情况,具体表现在程度较好的扣分较严,程度较差的扣分反而较松。这也是对大纲中的评分标准没有吃透的表现。还有些具体情况在此不一赘述了。

5.3 消除普通话水平测试误差的心理对策

心理因素尽管复杂,但仍可以认知与把握。可以通过心理训练,消除一些不良心理状态。从心理训练的角度,我们可以把认知技能分为听觉能力和评价能力。口语测试需要有关的认知技能,应加强对测试员认知技能的训练,即使已取得省或国家级资格者,也要有意识地在日常生活中训练和培养这方面的能力。训练和培养可从以下几个方面入手:

5.3.1 重视测试前的培训工作

测前培训包括两个方面,一个是测试员上岗培训,一个是应试者应试培训。很多测试员在参加测试员班学习时,不重视心理疏导,也不重视心理调节,认为自己普通话标准、测评能力强就是一个合格的测试员了。应试者在参加普通话培训时,有些人忽视测前培训,认为自己普通话还不错,不需要再花时间去听培训课,这实际上是一个误区。测试员和应试者都应该认真对待培训,努力调试自己的语音和心理,使自己在测试时心中有数,知己知彼。另外,培训教师在上课时也应该随时注意考生的心理疏导,在上课内容的设置上要加入心理学知识,并给予学生积极的暗示,并对考生的测试动机进行引导。

5.3.2 创设良好的环境和氛围

在现场测试时,测试员也应有良好的业务水平和心理素养,使考生进入考场之后如沐春风,对进入考场学生多一点微笑,多一点鼓励的话语,想到自己不仅是



词、感叹词替换朗读文本中普通话的语气词、感叹词,该音变的“啊”没有音变,或音变错误,均应视为有语调偏误。“语调偏误”主要是扣“缺陷”分。每出现一类上述缺陷就扣 0.5 分,而要达到“一类”,也有一个量化的标准,那就是五次以上方可称为“一类”。尽管如此,在具体测试实践中,有的测试员仍然把被测者朗读时出现的一次两次缺陷视同出现了一类问题,于是就被扣去一个 0.5 分,甚至有的应试者达到了一级水平,在“语调偏误”一项中也被扣去 0.5 分。这样扣分的结果是降低了被测者的分数,造成评分偏严。

另外,在说话一项中,“语音面貌”的扣分也存在着不同的测试员松紧不同,或同一测试员时紧时松的情况,具体表现在程度较好的扣分较严,程度较差的扣分反而较松。这也是对大纲中的评分标准没有吃透的表现。还有些具体情况在此不一赘述了。

5.3 消除普通话水平测试误差的心理对策

心理因素尽管复杂,但仍可以认知与把握。可以通过心理训练,消除一些不良心理状态。从心理训练的角度,我们可以把认知技能分为听觉能力和评价能力。口语测试需要有关的认知技能,应加强对测试员认知技能的训练,即使已取得省或国家级资格者,也要有意识地在日常生活中训练和培养这方面的能力。训练和培养可从以下几个方面入手:

5.3.1 重视测试前的培训工作

测前培训包括两个方面,一个是测试员上岗培训,一个是应试者应试培训。很多测试员在参加测试员班学习时,不重视心理疏导,也不重视心理调节,认为自己普通话标准、测评能力强就是一个合格的测试员了。应试者在参加普通话培训时,有些人忽视测前培训,认为自己普通话还不错,不需要再花时间去听培训课,这实际上是一个误区。测试员和应试者都应该认真对待培训,努力调试自己的语音和心理,使自己在测试时心中有数,知己知彼。另外,培训教师在上课时也应该随时注意考生的心理疏导,在上课内容的设置上要加入心理学知识,并给予学生积极的暗示,并对考生的测试动机进行引导。

5.3.2 创设良好的环境和氛围

在现场测试时,测试员也应有良好的业务水平和心理素养,使考生进入考场之后如沐春风,对进入考场学生多一点微笑,多一点鼓励的话语,想到自己不仅是



词、感叹词替换朗读文本中普通话的语气词、感叹词,该音变的“啊”没有音变,或音变错误,均应视为有语调偏误。“语调偏误”主要是扣“缺陷”分。每出现一类上述缺陷就扣 0.5 分,而要达到“一类”,也有一个量化的标准,那就是五次以上方可称为“一类”。尽管如此,在具体测试实践中,有的测试员仍然把被测者朗读时出现的一次两次缺陷视同出现了一类问题,于是就被扣去一个 0.5 分,甚至有的应试者达到了一级水平,在“语调偏误”一项中也被扣去 0.5 分。这样扣分的结果是降低了被测者的分数,造成评分偏严。

另外,在说话一项中,“语音面貌”的扣分也存在着不同的测试员松紧不同,或同一测试员时紧时松的情况,具体表现在程度较好的扣分较严,程度较差的扣分反而较松。这也是对大纲中的评分标准没有吃透的表现。还有些具体情况在此不一赘述了。

5.3 消除普通话水平测试误差的心理对策

心理因素尽管复杂,但仍可以认知与把握。可以通过心理训练,消除一些不良心理状态。从心理训练的角度,我们可以把认知技能分为听觉能力和评价能力。口语测试需要有关的认知技能,应加强对测试员认知技能的训练,即使已取得省或国家级资格者,也要有意识地在日常生活中训练和培养这方面的能力。训练和培养可从以下几个方面入手:

5.3.1 重视测试前的培训工作

测前培训包括两个方面,一个是测试员上岗培训,一个是应试者应试培训。很多测试员在参加测试员班学习时,不重视心理疏导,也不重视心理调节,认为自己普通话标准、测评能力强就是一个合格的测试员了。应试者在参加普通话培训时,有些人忽视测前培训,认为自己普通话还不错,不需要再花时间去听培训课,这实际上是一个误区。测试员和应试者都应该认真对待培训,努力调试自己的语音和心理,使自己在测试时心中有数,知己知彼。另外,培训教师在上课时也应该随时注意考生的心理疏导,在上课内容的设置上要加入心理学知识,并给予学生积极的暗示,并对考生的测试动机进行引导。

5.3.2 创设良好的环境和氛围

在现场测试时,测试员也应有良好的业务水平和心理素养,使考生进入考场之后如沐春风,对进入考场学生多一点微笑,多一点鼓励的话语,想到自己不仅是



一名语音测试员，还是一名心理咨询师。正确地引导考生，创设良好的测试环境和氛围，使考生以积极的心理面对测试，充分发挥自己的水平，取得好的成绩。

5.3.3 保持积极的心态

每分钟 400 多个音节，平均每秒钟 6-7 个音节。在这样快的语速下，测评员要综合视觉信息和听觉信息，通过视觉器官和听觉器官同时反映到大脑中，通过分析、判断，决定哪些发音正确，哪些发音错误或有缺陷，还有加字减字、语速语调等问题，都要迅速作出判断，并且在试卷上记录下来。在 6-7 分之一秒的时间里要完成这一系列动作，必须调动一切感官，手脑并动，眼耳并用，稍一疏忽，几个音节就滑过去了。如果没有高度集中的注意力、非常熟悉的试卷，测试中难免手忙脚乱，疲于应付。而人脑毕竟不是机器，能够集中注意力的时间有限。时间一长，就可能造成心理疲劳，听觉失灵，反映迟钝，对错误的语音习以为常，甚至对正确的语音产生怀疑，错判漏判明显增多，从而造成测试误差。所以测试员一定要调整好自己的心态。在业务水平不断进步的同时，提高自己的心理承受力以及对环境的抗干扰力，不受周围环境的影响以及测试搭档的影响。

5.3.4 克服测试员主观因素的影响

普通话水平测试作为一种测试，其评分标准是客观的。而很多时候由于测试员主观因素的影响会造成测评过宽或过严，导致测试误差出现。下面的一些情况通常会使评分偏宽或偏严：（1）第一印象。一般说来，如果应试者穿着整洁，行为举止彬彬有礼、测试时镇定自若、声音洪亮；或看上去朴实憨厚；或聪明伶俐、思维敏捷，给测试员一种美好的印象和愉悦的感受，评分时一般会偏宽。反之，若应试者衣冠不整、言行傲慢，或畏缩、害怕、说话声音太小，则会引起测试员的反感，打分会偏严。（2）前项印象。应试者在 1、2 两项测试中，如果声母、韵母、声调都读得比较准确，就会给测试员留下发音状况良好的印象，从而放宽对其它几项测试项的评分标准。反之，如果应试者在 1、2 两项中语音失误较多，就会被感知在后面的几项测试中也会出现失误较多的情况。例如湖北人前鼻音韵母 in 和后鼻音韵母 ing 不分，在 1、2 两项中若出现了这个问题，测试员会预测在后面的测试项中会出现同样问题。其次，语音转瞬即逝，导致测试员对语音信息的选择（语音听辨）带有一定的主观性，因而心理注意和心理规则就显得十分重要。当语流中相似错误的数量增加时，每个错误引起心理注意的几率就会减少，受注意程度就会下降，听辨音的误差就会增大；另一方面，听辨音能力的差异对心理注意和心理规则也有很大影响，会干扰对语音信息的搜索和判断。由于产生测试误差的主要原因是认知技



一名语音测试员，还是一名心理咨询师。正确地引导考生，创设良好的测试环境和氛围，使考生以积极的心理面对测试，充分发挥自己的水平，取得好的成绩。

5.3.3 保持积极的心态

每分钟 400 多个音节，平均每秒钟 6-7 个音节。在这样快的语速下，测评员要综合视觉信息和听觉信息，通过视觉器官和听觉器官同时反映到大脑中，通过分析、判断，决定哪些发音正确，哪些发音错误或有缺陷，还有加字减字、语速语调等问题，都要迅速作出判断，并且在试卷上记录下来。在 6-7 分之一秒的时间里要完成这一系列动作，必须调动一切感官，手脑并动，眼耳并用，稍一疏忽，几个音节就滑过去了。如果没有高度集中的注意力、非常熟悉的试卷，测试中难免手忙脚乱，疲于应付。而人脑毕竟不是机器，能够集中注意力的时间有限。时间一长，就可能造成心理疲劳，听觉失灵，反映迟钝，对错误的语音习以为常，甚至对正确的语音产生怀疑，错判漏判明显增多，从而造成测试误差。所以测试员一定要调整好自己的心态。在业务水平不断进步的同时，提高自己的心理承受力以及对环境的抗干扰力，不受周围环境的影响以及测试搭档的影响。

5.3.4 克服测试员主观因素的影响

普通话水平测试作为一种测试，其评分标准是客观的。而很多时候由于测试员主观因素的影响会造成测评过宽或过严，导致测试误差出现。下面的一些情况通常会使评分偏宽或偏严：(1) 第一印象。一般说来，如果应试者穿着整洁，行为举止彬彬有礼、测试时镇定自若、声音洪亮；或看上去朴实憨厚；或聪明伶俐、思维敏捷，给测试员一种美好的印象和愉悦的感受，评分时一般会偏宽。反之，若应试者衣冠不整、言行傲慢，或畏缩、害怕、说话声音太小，则会引起测试员的反感，打分会偏严。(2) 前项印象。应试者在 1、2 两项测试中，如果声母、韵母、声调都读得比较准确，就会给测试员留下发音状况良好的印象，从而放宽对其它几项测试项的评分标准。反之，如果应试者在 1、2 两项中语音失误较多，就会被感知在后面的几项测试中也会出现失误较多的情况。例如湖北人前鼻音韵母 in 和后鼻音韵母 ing 不分，在 1、2 两项中若出现了这个问题，测试员会预测在后面的测试项中会出现同样问题。其次，语音转瞬即逝，导致测试员对语音信息的选择（语音听辨）带有一定的主观性，因而心理注意和心理规则就显得十分重要。当语流中相似错误的数量增加时，每个错误引起心理注意的几率就会减少，受注意程度就会下降，听辨音的误差就会增大；另一方面，听辨音能力的差异对心理注意和心理规则也有很大影响，会干扰对语音信息的搜索和判断。由于产生测试误差的主要原因是认知技



一名语音测试员，还是一名心理咨询师。正确地引导考生，创设良好的测试环境和氛围，使考生以积极的心理面对测试，充分发挥自己的水平，取得好的成绩。

5.3.3 保持积极的心态

每分钟 400 多个音节，平均每秒钟 6-7 个音节。在这样快的语速下，测评员要综合视觉信息和听觉信息，通过视觉器官和听觉器官同时反映到大脑中，通过分析、判断，决定哪些发音正确，哪些发音错误或有缺陷，还有加字减字、语速语调等问题，都要迅速作出判断，并且在试卷上记录下来。在 6-7 分之一秒的时间里要完成这一系列动作，必须调动一切感官，手脑并动，眼耳并用，稍一疏忽，几个音节就滑过去了。如果没有高度集中的注意力、非常熟悉的试卷，测试中难免手忙脚乱，疲于应付。而人脑毕竟不是机器，能够集中注意力的时间有限。时间一长，就可能造成心理疲劳，听觉失灵，反映迟钝，对错误的语音习以为常，甚至对正确的语音产生怀疑，错判漏判明显增多，从而造成测试误差。所以测试员一定要调整好自己的心态。在业务水平不断进步的同时，提高自己的心理承受力以及对环境的抗干扰力，不受周围环境的影响以及测试搭档的影响。

5.3.4 克服测试员主观因素的影响

普通话水平测试作为一种测试，其评分标准是客观的。而很多时候由于测试员主观因素的影响会造成测评过宽或过严，导致测试误差出现。下面的一些情况通常会使评分偏宽或偏严：（1）第一印象。一般说来，如果应试者穿着整洁，行为举止彬彬有礼、测试时镇定自若、声音洪亮；或看上去朴实憨厚；或聪明伶俐、思维敏捷，给测试员一种美好的印象和愉悦的感受，评分时一般会偏宽。反之，若应试者衣冠不整、言行傲慢，或畏缩、害怕、说话声音太小，则会引起测试员的反感，打分会偏严。（2）前项印象。应试者在 1、2 两项测试中，如果声母、韵母、声调都读得比较准确，就会给测试员留下发音状况良好的印象，从而放宽对其它几项测试项的评分标准。反之，如果应试者在 1、2 两项中语音失误较多，就会被感知在后面的几项测试中也会出现失误较多的情况。例如湖北人前鼻音韵母 in 和后鼻音韵母 ing 不分，在 1、2 两项中若出现了这个问题，测试员会预测在后面的测试项中会出现同样问题。其次，语音转瞬即逝，导致测试员对语音信息的选择（语音听辨）带有一定的主观性，因而心理注意和心理规则就显得十分重要。当语流中相似错误的数量增加时，每个错误引起心理注意的几率就会减少，受注意程度就会下降，听辨音的误差就会增大；另一方面，听辨音能力的差异对心理注意和心理规则也有很大影响，会干扰对语音信息的搜索和判断。由于产生测试误差的主要原因是认知技



能和心理规则，因而测试员不能满足现有的听辨音水平，应加强认知技能训练，缩小心理规则的差异。在日常生活中，努力克服方言的影响，加强对语音听辨能力的监控，即使是省级和国家级测试员也要有意识地在日常生活中训练和培养这方面的能力，确保现有认知技能不退化。

5.4 改进技术手段

从数字化媒体发展的角度而言，今后的普通话水平测试应该努力改变目前这种现场测试的唯一方式，尽量排除人为因素的干扰，排除测试员和应试者心理因素的影响，走多样化的道路，可以参考以下几种模式运行。

5.4.1 人一机测试

所谓人机测试，主要指借用计算机的语音识别系统，通过语音心理学等科学理论支持，设计研制出能进行语音评分的计算机软件。计算机评分可以最大程度上排除人的干扰，而且不知疲倦，也不会受到测试员自身身体、心理、业务水平的影响。计算机通过对语音波形的分析，可以有效地判定语音错误以及语音缺陷。目前我国已经有几所大学开发出此种软件，并逐步投入到实际中。

5.4.2 机一人测试

所谓机人测试，主要是应试者通过自己录音的方式参加测试。随着科技的发展，录音笔的普及，应试者可以将自己的测试语音录入到录音笔中，再送到相应的普通话测试员那里进行评分。这种测试可以不受场地、时间、交通等因素的约束，应试者也可以避免出现收到测试员的干扰，并使自己的心理因素影响减小到最低。目前我国只有一级甲等的复审工作采取此种方式，其他的也仅仅只用于复审的工作中，我们建议今后可以大量使用此种方式，节约成本。

5.4.3 机一机测试

随着科技的发展，数字化已经进入到人民生活的方方面面，普通话的学习以及普通话水平测试应该紧密结合起来，相关部门可以在网络上开展普通话远程教学活动，并由应试者在网络平台上由相应的测试软件进行评分，并开出诊断的方子，使考生能够很快知晓自己的普通话的情况并有的放矢进行学习。

普通话水平测试是一个复杂的系统工程，对每一名测试员都提出了较高的要求，“在素质方面，所有参加考试实施的人员，都必须经过一定的业务培训，具有高度的责任感”，才能“保证考试实施的各项工作能严格按照科学程序进行，有效控制



能和心理规则，因而测试员不能满足现有的听辨音水平，应加强认知技能训练，缩小心理规则的差异。在日常生活中，努力克服方言的影响，加强对语音听辨能力的监控，即使是省级和国家级测试员也要有意识地在日常生活中训练和培养这方面的能力，确保现有认知技能不退化。

5.4 改进技术手段

从数字化媒体发展的角度而言，今后的普通话水平测试应该努力改变目前这种现场测试的唯一方式，尽量排除人为因素的干扰，排除测试员和应试者心理因素的影响，走多样化的道路，可以参考以下几种模式运行。

5.4.1 人一机测试

所谓人机测试，主要指借用计算机的语音识别系统，通过语音心理学等科学理论支持，设计研制出能进行语音评分的计算机软件。计算机评分可以最大程度上排除人的干扰，而且不知疲倦，也不会受到测试员自身身体、心理、业务水平的影响。计算机通过对语音波形的分析，可以有效地判定语音错误以及语音缺陷。目前我国已经有几所大学开发出此种软件，并逐步投入到实际中。

5.4.2 机一人测试

所谓机人测试，主要是应试者通过自己录音的方式参加测试。随着科技的发展，录音笔的普及，应试者可以将自己的测试语音录入到录音笔中，再送到相应的普通话测试员那里进行评分。这种测试可以不受场地、时间、交通等因素的约束，应试者也可以避免出现收到测试员的干扰，并使自己的心理因素影响减小到最低。目前我国只有一级甲等的复审工作采取此种方式，其他的也仅仅只用于复审的工作中，我们建议今后可以大量使用此种方式，节约成本。

5.4.3 机一机测试

随着科技的发展，数字化已经进入到人民生活的方方面面，普通话的学习以及普通话水平测试应该紧密结合起来，相关部门可以在网络上开展普通话远程教学活动，并由应试者在网络平台上由相应的测试软件进行评分，并开出诊断的方子，使考生能够很快知晓自己的普通话的情况并有的放矢进行学习。

普通话水平测试是一个复杂的系统工程，对每一名测试员都提出了较高的要求，“在素质方面，所有参加考试实施的人员，都必须经过一定的业务培训，具有高度的责任感”，才能“保证考试实施的各项工作能严格按照科学程序进行，有效控制



能和心理规则，因而测试员不能满足现有的听辨音水平，应加强认知技能训练，缩小心理规则的差异。在日常生活中，努力克服方言的影响，加强对语音听辨能力的监控，即使是省级和国家级测试员也要有意识地在日常生活中训练和培养这方面的能力，确保现有认知技能不退化。

5.4 改进技术手段

从数字化媒体发展的角度而言，今后的普通话水平测试应该努力改变目前这种现场测试的唯一方式，尽量排除人为因素的干扰，排除测试员和应试者心理因素的影响，走多样化的道路，可以参考以下几种模式运行。

5.4.1 人一机测试

所谓人机测试，主要指借用计算机的语音识别系统，通过语音心理学等科学理论支持，设计研制出能进行语音评分的计算机软件。计算机评分可以最大程度上排除人的干扰，而且不知疲倦，也不会受到测试员自身身体、心理、业务水平的影响。计算机通过对语音波形的分析，可以有效地判定语音错误以及语音缺陷。目前我国已经有几所大学开发出此种软件，并逐步投入到实际中。

5.4.2 机一人测试

所谓机人测试，主要是应试者通过自己录音的方式参加测试。随着科技的发展，录音笔的普及，应试者可以将自己的测试语音录入到录音笔中，再送到相应的普通话测试员那里进行评分。这种测试可以不受场地、时间、交通等因素的约束，应试者也可以避免出现收到测试员的干扰，并使自己的心理因素影响减小到最低。目前我国只有一级甲等的复审工作采取此种方式，其他的也仅仅只用于复审的工作中，我们建议今后可以大量使用此种方式，节约成本。

5.4.3 机一机测试

随着科技的发展，数字化已经进入到人民生活的方方面面，普通话的学习以及普通话水平测试应该紧密结合起来，相关部门可以在网络上开展普通话远程教学活动，并由应试者在网络平台上由相应的测试软件进行评分，并开出诊断的方子，使考生能够很快知晓自己的普通话的情况并有的放矢进行学习。

普通话水平测试是一个复杂的系统工程，对每一名测试员都提出了较高的要求，“在素质方面，所有参加考试实施的人员，都必须经过一定的业务培训，具有高度的责任感”，才能“保证考试实施的各项工作能严格按照科学程序进行，有效控制



各类无关因素的干扰，做到实施规范、判分客观、结果准确”。因此，普通话水平测试员就要认真找出自身的不足，增强责任感，不断提高自己的普通话水平和听音辨音的能力，加强对测试大纲中评分标准的学习，多参加测试实践，全面提高思想和业务乃至心理的综合素质，准确高效地完成测试工作，保证测试的信度，促进普通话的推广与普通话水平的提高。



6 结束语

6.1 本文的基本认识

普通话水平测试是中华人民共和国国家级考试，是一种口语测试，测试结果受到诸多因素的影响。本文从心理因素、评分细则等方面来分析对测试结果的影响，提出测试管理机构应加强对测试员思想、心理、业务水平、评分标准的把握和理解等方面改进办法，并对今后测试工作的发展提出新的设想。

6.2 本文存在的问题

本文通过分析心理因素、评分细则等对普通话水平测试结果的影响因素分析，提出测试机构应加强对测试员的管理以及测试员应提高自身思想素质、心理素质、业务水平、对评分的理解等方面改进办法，并对今后测试工作的发展提出新的设想。但还有些问题尚未分析清楚，如：

(1) 量化比例

我们知道用量化分析的方法对应试者普通话语音的规范程度作出判断，是普通话水平测试采用的基本方法，其重要性是不容忽视的。而《大纲》目前评分标准本身带有弹性，量化不够，如朗读部分，声韵缺陷、语调偏误、停连不当、朗读不流畅等的判断，出入很大，大纲中均为“视程度”扣相应的分数。老测试员或者骨干测试员知道如何“视程度”，大部分测试员不知道如何“视程度”，更多是依据个人经验评判。在说话部分，虽然“大纲”规定了语音面貌和词汇语法规程度的打分标准，但是方音明显还是不明显，规范还是不规范，由于无量化评分细则，测试员主观判断起的作用相对更大，这是困扰测试工作的一大难题。

(2) 字音与语流问题

以测试第一项单音节词为例，有应试者两个字两个字一读，把单字当成了词语读，测试第二项双音节词语中，有应试者读词完全割裂词的亲缘关系，只注意音节的准确，如“英雄、准备、教条”等，读为“英一雄，准一备；教一条”，语音正确，语感很差，读词语的方式完全不符合汉民族口语习惯，应该如何评分？《大纲》中没有可凭借的依据。第三项朗读中，有应试者追求一个字一个字蹦出来，完全不考虑语流，只是机械的把文章中的每一个字都读准确，特别是把声调读到位，听起



6 结束语

6.1 本文的基本认识

普通话水平测试是中华人民共和国国家级考试，是一种口语测试，测试结果受到诸多因素的影响。本文从心理因素、评分细则等方面来分析对测试结果的影响，提出测试管理机构应加强对测试员思想、心理、业务水平、评分标准的把握和理解等方面改进办法，并对今后测试工作的发展提出新的设想。

6.2 本文存在的问题

本文通过分析心理因素、评分细则等对普通话水平测试结果的影响因素分析，提出测试机构应加强对测试员的管理以及测试员应提高自身思想素质、心理素质、业务水平、对评分的理解等方面改进办法，并对今后测试工作的发展提出新的设想。但还有些问题尚未分析清楚，如：

(1) 量化比例

我们知道用量化分析的方法对应试者普通话语音的规范程度作出判断，是普通话水平测试采用的基本方法，其重要性是不容忽视的。而《大纲》目前评分标准本身带有弹性，量化不够，如朗读部分，声韵缺陷、语调偏误、停连不当、朗读不流畅等的判断，出入很大，大纲中均为“视程度”扣相应的分数。老测试员或者骨干测试员知道如何“视程度”，大部分测试员不知道如何“视程度”，更多是依据个人经验评判。在说话部分，虽然“大纲”规定了语音面貌和词汇语法规程度的打分标准，但是方音明显还是不明显，规范还是不规范，由于无量化评分细则，测试员主观判断起的作用相对更大，这是困扰测试工作的一大难题。

(2) 字音与语流问题

以测试第一项单音节词为例，有应试者两个字两个字一读，把单字当成了词语读，测试第二项双音节词语中，有应试者读词完全割裂词的亲缘关系，只注意音节的准确，如“英雄、准备、教条”等，读为“英一雄，准一备；教一条”，语音正确，语感很差，读词语的方式完全不符合汉民族口语习惯，应该如何评分？《大纲》中没有可凭借的依据。第三项朗读中，有应试者追求一个字一个字蹦出来，完全不考虑语流，只是机械的把文章中的每一个字都读准确，特别是把声调读到位，听起



来非常生硬。测试员在评分时有的完全忽略其语感，只要音节正确，其余忽略不计；还有的测试员注意到这个问题给以适当减分，这又造成了所执标准不一的混乱情况。这种状况涉及到一个重要的问题即语音（或者说“字音”）跟口语的语流之间是什么关系？目前完全侧重于语音而不考虑交际与语流的现象是很值得研究的。语音正确，语流极差，无法正常说话，肯定要降等，但在实际操作中又没有量化项作为扣分的依据，反之，语流好，语音不好，不会倒过来加分，这之间如何作技术性处理呢？

（3）“语言面貌”问题

说话项中语音面貌的描述不十分清晰，测试员在此项评分中的误差率较高。当然这个评分项作为《大纲》有意设立的定性分析项，应该让测试员有一定的自由度。但定性分析并不排除用定量分析的方法来提供依据，对于“方音明显”“方音较明显”“方音不明显”应该有一个接近量化的界定。比如可依据有没有出现成系统缺陷或错误，成系统缺陷或错误出现多少，及这种错误或缺陷的社会认知程度怎样等因素来判定，尽可能量化。

（4）非语言现象问题

评分标准在某些方面过分强调一些非语言性的侧面，以致在一定程度上压倒了必要的、基本的条件。如说话项中，要求测试的时间是3分钟，有些考生说外语、唱歌或长时间停顿等拖延时间，还有的考生说话很慢，这样错误次数就少一些，就会使评分受到影响。

（5）等第的临界问题

等第的临界线问题事关重大，如何把握也是一个客观存在的难题。一甲与一乙、一乙与二甲、二甲与二乙、二甲与三甲之间，语音上的差别到底在哪里？量化可比度应精确到什么程度？一甲要严到什么程度？三乙和三甲的低分比较，到底差在哪？这些也是普通话水平测试中在评分问题上有待解决的问题。我们知道普通话水平测试是目标参照性考试。“一级甲等”是测试员在测试每一位应试人员时所参照的目标。但“一级甲等”实际上是一个理论概念，它无法绝对对应在某一个特定的人身上，每一个测试员心中都有一个自己认可的“一级甲等”，因而测试员所参照的目标也是不确定的。

从中我们也可以看出影响、制约着普通话水平测试评分的还有一些非客观性因素，这些因素对普通话测试结果产生怎样的影响，本文有的涉及到了，有的还没有涉及，今后将更进一步考察分析。



参考文献

- [1] 宋欣桥. 普通话语音训练教程[M]. 北京:商务印书馆, 2004.
- [2] 马学良等. 普通语言学[M]. 北京:中央民族大学出版社, 1997.
- [3] 韩永昌等. 心理学[M]. 上海:华东师范大学出版社, 1993.
- [4] 刘照雄. 普通话水平测试大纲[M]. 吉林:吉林人民出版社, 2002.
- [5] 国家语委普通话培训测试中心. 普通话水平测试实施纲要[M]. 北京:商务印书馆, 2004.
- [6] [英]阿兰·戴维斯. 语言测试原理. 北京:经济科学出版社, 1997.
- [7] 宋欣桥. 普通话水平测试员实用手册. 北京:商务印书馆, 2000.
- [8] 刘照雄主编. 普通话水平测试大纲. 吉林:吉林人民出版社, 1994.
- [9] 语言文字报刊社. 语言文字应用. 北京:语言文字应用杂志社, 1997-3.
- [10] 何文征. 加强普通话水平测试的客观性. 北京:语文建设, 1998-9.
- [11] 刘镰力. 汉语水平测试研究. 北京:北京语言文化大学出版社, 1997.
- [12] 李筱菊. 语言测试科学与艺术. 湖南:湖南教育出版社, 1997.
- [13] 《中华人民共和国国家通用语言文字法》, 2000年10月31日第九届全国人民代表大会常务委员会第十八次会议通过.
- [14] 香港中国语文学会普通话测试研究小组, 普通话水平考试大纲(征求意见稿), 1984.
- [15] 中国对外汉语教学学会, 《汉语水平等级标准和等级大纲》, 1988.
- [16] 刘导生, 新时期的语言文字工作, 1986.
- [17] 孙修章, 普通话水平测试标准的研制与实践, 语言文字应用, 1992. (1).
- [18] 刘照雄, 推广普通话的重要举措——普通话水平测试简论, 语言文字应用, 1997.
- [19] 陈章太, 论普通话水平测试等级标准, 语言文字应用, 1997 (3).
- [20] 仲哲明, 普通话水平测试若干问题的讨论, 语言文字应用, 1997, (3).
- [21] 许嘉璐, 普通话水平测试研究序, 语文出版社, 1997.
- [22] 李字明, 关于普通话水平测试的思考, 普通话水平测试研究序, 上海教育出版社, 2002.
- [23] 王 晖, 1994年以来普通话水平测试研究概述, 首届全国普通话水平测试学术研讨会论文, 2002 .5 (扬州).
- [24] 刘照雄主编, 普通话水平测试大纲, 吉林人民出版社, 1994.



- [25] 于根元, 二十世纪的中国语言应用研究, 书海出版社, 1996.
- [26] 国家语委普通话培训测试中心、《语言文字应用》编辑部合编, 普通话水平测试的理论与实践, 商务印书馆, 1998 .
- [27] 宋欣桥编, 普通话水平测试员手册, 商务印书馆, 2000.
- [28] 刘镰力主编, 汉语水平测试研究, 北京语言文化大学出版社, 1997.
- [29] 谢小庆 鲁新民主编, 考试研究文集, 经济科学出版社, 2002.
- [30] 叶 军. 普通话水平测试的社会语言学思考[J]. 语文建设, 1997, 12: 24.
- [31] 王明东. 精确与模糊[J]. 语文建设, 1996, 5: 39.
- [32] 宋欣桥. 普通话水平测试中的评分差异[J]. 语文建设, 1998, 9: 12.
- [33] 晓 东. 亟需解决的一个问题[J]. 语文建设, 1998, 3: 13.
- [34] 李宇明. 理论语言学教程[M]. 华中师范大学出版社, 1998. 79、80.
- [35] 邵志芳. 思维心理学[M]. 华中师范大学出版社, 2001. 21、75.
- [36] 张传曾. 普通话等级测试标准的音位阐释[J]. 语文建设, 1998, 11: 28.
- [37] 桂诗春. 实验心理语言学纲要[M]. 湖南教育出版社, 1991. 99、215.



后记

本文是在语言所汪国胜教授精心指导和大力支持下完成的。汪老师严谨求实的治学态度、高度的敬业精神、兢兢业业、孜孜以求的工作作风和大胆创新的进取精神，对我产生重要影响。他渊博的知识、开阔的视野和敏锐的思维给了我深深的启迪。同时，在此次毕业论文中，通过对普通话水平测试结果的影响因素分析，让我在实际工作中更加如鱼得水，能够理论与实践相结合，更好地完成本职工作。

另外，我还要特别感谢各位任课老师对毕业论文写作的指导，感谢李向农教授、范新干教授、徐杰教授、曹海东教授、卢卓群教授、周光庆教授、储泽祥教授、吴振国教授给我们上课，他们治学严谨、学高身正。语言所姚双云老师和阮蓓、冯志华等同学对我的无私帮助，湖北省普通话测试中心、武汉语言文字培训测试中心和江汉大学普通话培训测试站、江汉大学语言文学研究所对我的大力支持，使我得以顺利完成论文。在此我也表示衷心的感谢。

最后，再次对关心、帮助我的老师和同学表示衷心的感谢！