

您当前位置: >> 语言文字 >> 学术交流 >> 浏览文章

试析影响PSC信度和效度的因素

郟远春

测试的信度指测试结果的可靠性、一致性和稳定性程度。一个高质量的测试,对同样一组被试反复多次测量,其结果应该保持不变。效度是指测试的有效性或准确性,即测试对它所要测的东西准确测试的程度。

普通话水平测试(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口语测试,全部测试内容均以口头方式进行。受试者在测试中所表现出的语音、词汇、语法规程程度,是评定其所达到的水平等级的重要依据。受试者的心理、测试手段和方式、测试内容等各种因素都会影响到受试者的普通话测试成绩。

一、受试者心理因素

心理因素包括人的感觉、知觉和情绪等,往往被称为事物发展变化的内因,是事物运动的源泉和动力,它规定着事物运动和发展的基本趋势。测试目的不同,使受试者在对待普通话水平测试时的感觉和情绪有着明显的不同,直接影响到其PSC的信度和效度。

心理因素包括心理过程与个性两个方面。心理过程是由认识过程、情绪过程和意志过程所构成。认识过程又包括感觉、知觉、记忆、思维与想象。感觉是人脑对客观事物个别属性的反映。知觉是人脑对客观事物的属性的整体反映或关系反映。记忆是对感知过或思维过的事物能够记住,并能在以后再再现,或在它重新呈现时能再认识的过程。思维则是人高级认识过程,它是人脑对客观事物间接的与概括性反映。想象是人脑对过去感知过的形象进行加工,产生一种新的形象。受试者通过外界宣传,感知普通话水平测试的价值,并在自己脑海中强化这样一种观念,并反映到受试过程中。

在人认识客观事物的时候,总会带有一定的态度,心理学把人们对客观事物态度的体验成为情绪。心理因素对人身心的作用主要通过人的情绪发挥作用。受试者通过感知,获得与之相关的情绪,主要表现在以下方面:

1. 严阵以待

受试将PSC测试看作一次非常严肃、正式和重要的考试,因而表现出慎重的态度。这直接会导致正反两种极端的结果:其一,普通话极其糟糕。由于将测试看得过于严重,受试者难免紧张。在这样的情绪影响下,有的受试者一进入考场便脸发红、手发抖、声发颤,本来很简单的一个字,他都有可能读错。如,把“牡丹”读成“杜丹”;错读、漏读现象严重,说话时思维混乱,毫无逻辑可言。总之,在整个测试过程中,其普通话表现得非常糟糕。然而,一旦考试结束,恢复到正常谈话时,你会发现受试的普通话并没有如其在测试时所表现的那么糟糕。笔者曾遇到一名受试者,在读前面的单音节字词和双音节词时,有着明显的方言口音,在后边说话时,由于紧张情绪的渐渐消失,他的方音明显变少。测试结束以后,在与笔者的对话中,竟找不到多少方音的影子。其二,普通话较好。有些测试者经过比较充分地准备,在测试时,放慢速度,尽量掩饰自己的语音缺陷,反而取得比较理想的成绩。然而,在实际的交际中,他的普通话水平却可能比较差。曾有一名考生,为了取得较好的测试成绩,他精心进行了准备。经过一段时间的训练,他的应试能力得到极大提高。最后,在普通话水平测试中获得了较好的成绩。然而,在实际与人交流中,他仍说着一口方音味很重的普通话。

2. 泰然处之

持这种心理的受试者有两种类型。其一,普通话水平比较好者。这一类人的普通话水平本身比较好,测试对他们来说易如反掌,只当一次随意的考试而已。其二,目的不明确者。持这种类型的受试者,本来没有非要考普通话的必要,只是见人考他也考,对于结果无甚要求。

上述两种表现中,第一种会影响到测试的准确和可信度,第二种能真实地反映受试者的普通话水平。

个性的不同也会影响到测试结果。个性包括个性倾向性与个性心理特征。个性倾向性指的是人与客观世界的相互作用中,形成了对事物的态度与趋向。一个人在生活实践中形成的需要、兴趣、信念、理想与世界观,反映出个性的倾向性。

个性倾向性不仅对改造客观世界有重要作用,而且对人的心理活动也会产生影响。需要是人对一定的客观事物的需求。动机则是激发

人去行动主观动因，动机是个体发动和维持行动的另一种心理状态。动机产生于需要。

PSC自1994年开始进行测试工作，每年有数以万计的人参加考试。参加考试人员的目的各不相同。大致有如下几种类型：

(1) 职业要求

根据《国家通用语言文字法》的规定，公务员、教师、广播电视节目主持人、以及银行、铁路、邮政等公共服务行业从业人员必须达到一定的普通话等级。普通话水平证书成为进入某一行业的必备条件之一。每年进入这些行业的人员数量惊人。以公务员为例，2011年，仅中央机关及其直属机构和参照公务员法管理的单位就招收1.6万余人；北京市计划招收4189人；上海市计划录用4048人；福建省计划招收2420人……。粗略统计下来，全国三十个省市、自治区、直辖市一年中共招收公务员不下10万人。如此庞大的从业人员，均需要进行普通话测试，他们也是普通话测试的主力军。在校大学生，尤其是准备进入这些行业的应届毕业生，都会参加普通话水平考试。

(2) 学习要求

中华民族是56个民族组成的大家庭，为增强国家和民族的认同感，实现各民族共同团结进步、共同繁荣发展，多年来，国家一直致力在各民族地区推广普通话。目前，普通话在民族地区的推广已经取得喜人的成绩。然而，目前民族地区普通话的整体普及水平还偏低，既不会说，也听不懂汉语的情况在有的地区还比较普遍，这严重影响和制约了民族地区加快经济社会发展的步伐。为改变此现状，少数有识之士努力学习普通话，这也成为其参加普通话水平考试的原动力。另一方面，一些考入高等院校的少数民族考生，由于学习和交际的需要，必须学好汉语汉字。通过普通话培训及测试，将大大有助于其普通话水平的提高。

(3) 无明确目的

也有部分人参加普通话测试，没有明确的目的，抱着身边的人都考，自己拿个证也没坏处的想法参考。这类人占极少数。

目的和要求不一样，其对待测试的态度就完全不一样。从目前来看，80%以上的受试是由于职业需求参加考试。比如：按照国家相关部门的规定，播音员、节目主持人、影视话剧演员为一级以上水平；教师为二级以上水平；公务员和社会公共服务行业从业人员为三级以上水平。同是教师，任职岗位不同，对其普通话的要求也不甚相同。如在中小学教语文或在大学教中文，则要求教师的普通话水平达到二级甲等以上，而在大学中担任理工科教学的教师，只需要达到二级乙等即可。所以，有相当一部分教语文（中文）的老师如果一次没有通过二级甲等（有的甚至要求达到一级乙等以上水平），他们进行第二次考试。相对而言，他们对待普通话水平测试的态度比较认真、严肃、慎重。

二、测试内容

普通话测试题共分四个部分：读单音节字词100个；读多音节词语50个；朗读短文一篇；说话。

1. 读单音节字词

该题主要考察受试者普通话声韵调的发音是否准确。题中有时会出现一些生僻字，由于受试者不认识这些字，故而读错。但这不能代表他的普通话水平低，只能说他的汉语水平比较低。另外，单音节字词的发音与日常语流的发音不太一样，需要读得比较饱满，到位。很多受试在读单音节字词时，由于缺乏对题目本身要求的了解，经常读成语流中的发音。这也不能代表他的普通话有问题，只能说他没有了解题意。

2. 读多音节词语

该题主要考察受试者的语流音变情况。在单读一个个音节的时候，普通话有四种基本声调，即阴平、阳平、上声、去声。然而，每个音节、每个字不是一个个孤立的单位，在词语、句子中会因相邻音节的影响，而产生变调。有些受试者接受了错误的观点，以为只要把每个字都读准就行，因此在读多音节词的时候会出现如下情况：（1）将词中的每个字分开读，导致声调不准确。例如：上声音节在阴平、阳平、去声（非上声音节）前，丢掉后半段上升的尾巴，调值由214变为半上声21。受试者如果接受了错误的观点，都会念成214调。

（2）儿化音的“儿”单独发音。如“小孩儿”，应念成“xiǎo hái r”，有些受试则读成“xiǎo hái ér”。这些现象在他们说话的时候不会出现。依据现在的评分标准，这样的读音都会扣分。但实际上这并未反映受试者真正的普通话水平。

3. 朗读短文

朗读是考察受试者在有文字凭借情况下的语音面貌、连读音变、语调语气等。现在所提供的50片作品中，有些作品缺少一些需要测试的项目，如：有的缺少“不”连读音变，有的缺少“一”变去声的，有的缺少“儿化韵”；而有些作品中考察这部分内容又太过频繁。再者，每篇朗读题目的难易程度有很大差异。这都会影响到测试的信度和效度。

4. 说话

说话是考察受试者在无文字凭借情况下用普通话进行表达的能力和规范程度。现在所提供的50个题目，其难度大小不同。如：“素质教育之我见”、“谈谈社会公德”、“谈谈科技发展与社会生活”等题目对不少受试者来说比较难，而“我最尊敬的人”、“我的业余生活”、“我的愿望”等题目相对来说比较好说。这与受试者的年龄阅历、好恶等因素都有关系。现在接受测试的主要对象是大中专学生，相对来说，他们阅历少，接触机会少，所以很多时候有的题目难以下手，这就势必会影响其测试成绩。

同一受试，面对不同的测试内容，其成绩可能不一样。如果所出现的测试内容正好是受试所熟悉的，他的成绩相对会比较理想，反之则比较糟糕。

三、测试手段和方式

2007年以前, 各省市均采用人工测试手段进行普通话水平测试, 所有测试均由测试员判断。测试员的测试水平以及测试时的各种因素都有可能影响到受试的测试成绩。尤其是测试员的水平参差不齐, 同一个受试者, 不同的测试员所给的分也各不相同。主观性比较强。测试的信度和效度有待进一步验证。

2007年开始在全国开始试点计算机辅助测试。计算机辅助普通话水平测试是指应试人采用上机模式参加测试, 第一题“读单音节字词”、第二题“读多音节词语”和第三题“朗读短文”由计算机辅助普通话水平测试系统自动评分, 第四题“命题说话”由省普通话培训测试中心或各市(州、地)普通话培训测试站调配测试员通过网络在线评分。计算机辅助测试, 看似增加了测试的公平性和可信度, 而事实上却有很多问题存在。由于计算机测试程序属于新产品, 很多受试人员不会使用。另外, 在测试过程中也有很多问题, 如从试音到整个考试过程音量应保持基本一致, 声音太大或太小都会导致话筒无法正常录音; 考试时应保持适当的语速, 若念得太快, 按评分规则, 每一个字都有缺陷, 这样得分就不会高。

四、问题与建议

上文中, 我们总结了影响普通话测试信度和效度的几个关键因素, 同时也发现了一些问题和疑惑。

问题一: 被试由于紧张而未能发挥出其正常的水平, 这种情况下, 我们应如何评判他真实的普通话水平?

问题二: 被试由于不了解考试规则而出现的失误, 这种情况下, 我们应如何评判其真实的普通话水平?

问题三: 某些测试题是否能真正有效地测出被试的普通话水平?

问题四: 测试员是否能完全不受自己母方言或情绪的影响, 做出绝对公平和真实的判断?

事实上, 根据现有的测试标准和方式, 我们很难做到能绝对真实地测出被试的普通话水平, 所得分数仅仅只是一个相对值而非绝对值。这大大影响了PSC的信度和效度。因此, 我们认为, 如果想提高PSC的信度和效度, 必须对现有的评价标准做出适当的调整。

首先, 测试时仍以现在的测评标准为基准, 同时, 根据学生的实际状况, 灵活地做出调整。例如: 如果考生因为紧张在测试时未能发挥出应有水平, 可在其考试结束的时候诱导其与测试员对话, 进一步了解其真实的语言面貌。如果考生在试后所表现出的普通话水平明显高于试中, 我们可适当对其测试分数作一些调整。

其次, 进一步调整普通话测试内容, 使其能有效反映被试的普通话水平。例如: 读单音节词和多音节词中尽量避免出现多音字和生僻字, 朗读的短文应尽可能通俗易懂, 避免出现很多书面化的语言。

再次, 尽可能提高测试员的测试水平。在PSC中, 多数测试员都来自本土, 日常生活中常使用方言, 对普通话中的某些方音有着本能的认同感。尤其是一些有着固化的方音的测试员来说, 要他们敏锐地分辨出普通话和方音比较困难。

最后, 要加大对普通话测试的宣传力度。尤为重要, 要让每一个被试了解普通话测试的形式和规则, 不至于犯一些低级错误而失分。

参考文献:

1. 孟晖. 关于普通话水平测试员资格有效性问题的思考 [J]. 语言文字应用. 2010. (4).
2. 韦素玲. 普通话水平“机辅测试”环境对被测者影响的调查研究 [J]. 贵州师范学院学报. 2010. 09.
3. 吕忠民. 职业资格制度的研究及对策 [J]. 中国考试, 2008, (3).
4. 陈茜. 国家级普通话水平测试员队伍的培训与发展 [J]. 语言文字应用, 2007, (4)
5. 中华人民共和国教育部. 普通话水平测试管理规定 [Z]. 2003.
6. 曹艳丽. 建立和完善普通话水平测试质量保证体系的初步实践 [A]. 首届全国普通话水平测试学术研讨会 [C]. 北京: 语文出版社, 2003.
7. 李筱菊. 语言测试的科学和艺术 [M]. 长沙: 湖南教育出版社, 1997.
8. 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 [Z]. 1994.

作者简介: 郗远春, 女, 讲师, 博士, 主要从事方言及社会语言学研究。

来源: 西安工业大学人文学院

时间: 2012-1-11 16:08:27