

对提高PSC“命题说话”测试效度与信度的思考

陆妙琴

(咸阳职业技术学院, 陕西 咸阳 712000)

摘要: 多种原因导致了PSC“命题说话”测试效度与信度的降低。本文分析了目前影响PSC“命题说话”测试效度与信度的因素、产生的原因,并提出了改进的建议。

关键词: PSC; 命题说话; 效度; 信度; 影响因素; 原因; 建议

中图分类号: H195

文献标识码: A

文章编号: 94047-(2011)01-060-05

效度与信度是评估语言测试的两个重要标准。效度是指测试结果的有效性,即一套测试在多大程度上检测了所要检测的内容,多大程度上达到了测试的目的。信度是指测试结果的可靠性、稳定性,即一套测试测试结果在多大程度上具有一致性,也即测试结果的可信程度。普通话水平测试是对应试人运用普通话所能达到的规范程度、熟练程度的检测和评定,认定其普通话水平等级,属于标准参照性考试。“命题说话”是普通话水平测试(PSC)中测查应试人普通话运用能力的重点和关键。目前,绝大部分地方语言文字工作部门采用“读单音节字词”、“读多音节词语”、“朗读短文”、“命题说话”四项内容测试,免去了“选择判断”一项,“命题说话”的分值由30分调整为40分,这就使“命题说话”成为PSC中分值最重的一项内容,成了PSC的重头戏,这项测试的得分直接决定着应试人普通话水平的等级,这项测试的效度与信度直接影响着PSC的效度与信度。笔者认为,从提高测试效度的角度看,上述调整也是非常合理的,完全符合PSC的目的。因为PSC要测查的正是应试人运用普通话所能达到的标准程度,而说话是应试人普通话运用能力最直接、最有效的反映。正如孙修章先生在《普通话水平测试标准的研制与实践》一文中所指出的“说话是测试中难度较大的一项,

是测试的主要项目,说话测试难在于稍作准备后既要说得言之有物,又要标准、规范。对于普通话水平不高的人,这两项要求难以协调,照顾了标准,说不上几句话;顾了内容,顾不上标准。这样,说话测试就能真实地了解应试人的语言状况,这是其他几项测试不能代替的”。因此,“命题说话”本应是实现PSC效度的关键,但笔者在十多年的测试中发现,事实上“命题说话”却是PSC中测试效度与信度最低的一项内容。下面笔者将从影响PSC“命题说话”测试效度与信度的因素、产生的原因及改进建议几个方面谈谈自己的看法。

1 影响PSC“命题说话”测试效度与信度的因素

1.1 应试人背稿、内容雷同现象突出

测试中笔者经常发现应试人背稿现象,说话稿件有的来自网络、有的来自辅导书、有的来自《语言文字周报》。还有一些应试者,开始说话大体符合“命题说话”的要求,但说了没几句就想方设法利用现成的语料,甲用现成语料,乙也用现成预料,由此造成了说话内容的雷同。

笔者理解,目前普通话水平测试的内容可分为两大部分:“读单音节字词”、“读多音节词语”、“朗读短文”属于有文字凭借的“读”的

收稿日期: 2011-05-17

作者简介: 陆妙琴(1966—),女,副教授,全国汉语口语研究会会员,省级普通话测试员,咸阳市教学能手,主要从事汉语言教学及研究工作。E-mail:luckydog6666@163.com

测试；“命题说话”属于无文字凭借的“说”的测试，前三项内容重点测查的是应试人普通话语音的运用能力，而“命题说话”是对应试人运用普通话综合能力的全面测查。《普通话水平测试大纲》把“命题说话”测查的目的确定为“测查应试人在没有文字凭借的情况下说普通话的水平，重点测查语音标准程度、语汇语法规则程度和自然流畅度”。从语体色彩看，“命题说话”表达用的是口头语言，而背稿、运用现成语料表达，用的则是有文字凭借的书面语言。测试时背稿、运用现成语料的做法，使得本应是无文字凭借的说话变成了有文字凭借的说话，这完全背离了“命题说话”测试的目的，测出的不是应试人真实的水平，大大降低了测试的效度。

1.2 说话缺时较为严重

说话时间，94版《大纲》规定说4分钟（不得少于3分钟），03版《大纲》规定说满3分钟。

测试中笔者发现“命题说话”一项能够说满规定时间的人不是很多，一部分应试者出现缺时的情况，且有的缺时还相当严重。这大大影响到测试的效度与信度。

1.3 测试员间评分差异较大

测试中笔者发现“读单音节字词”、“读多音节词语”两项，测试员的评分基本一致，“朗读短文”一项，评分有些微差异，评分差异最大的是“命题说话”一项。测试员间评分的较大差异直接影响到测试的信度。

1.4 测试中出现的当代汉语的新现象，由于无标准可依，测试员无从判断

测试中笔者曾碰到过运用网络语、流行语的应试人，还有的应试人表达时中英文夹杂，而《大纲》对这一语言现象没有制定明确的判定标准，测试员对此难以评定，这在一定程度上也影响到测试的效度与信度。

1.5 应试人普遍重视语音的学习，轻视甚至忽略了语汇、语法的学习和规范

重语音，轻语汇、语法的学习和规范，不仅影响到测试的效度与信度，也影响到汉语的规范化水平。

2 产生原因及改进建议

2.1 背稿、内容雷同产生原因及改进建议

笔者认为测试中出现背稿、内容雷同的根本原因是目前“命题说话”的命题方式还不够科学、严密，目前的命题方式给应试人提供了投机取巧、违规应试的可能和机会。由于“命题说话”话题是预设的，事先给定的，所以应试人有足够的时间在测试前自己动笔写说话稿，也可通过各种途径寻找现成的语料测试时运用，以便在测试中减少失误，获得高分。另一个原因是目前的评定标准对背稿、内容雷同太宽容，扣分太少，这就使应试人宁肯背稿、运用现成语料，也不愿自己老老实实地现场组织语言材料，即兴表达。因为依据目前说话项的评分标准，对于那些普通话水平不高的应试人来说，背稿、内容雷同扣除的分数与自己即兴表达产生失误扣除的分数相比，简直是芝麻比西瓜。

针对上述情况，笔者建议：首先，要尽快改变命题方式。如果能把事先给定话题变为当场命题，就会大大减少背稿、内容雷同的现象。笔者认为当场命题的方式更能体现“命题说话”测查的意图和目的。笔者理解“命题说话”测查的就是应试人在无文字凭借的情况下，随即组织语言，运用普通话即兴表达的能力，包括了把内部语言转化为外部语言的能力。背稿、内容雷同根本没有这一语言生成转化的过程，也就无从体现这一语言能力。其次，对背稿、内容雷同应加大扣分力度。笔者认为背稿、内容雷同至少应扣掉30分。因为背稿、内容雷同，违反了PSC测试规定，是一种不诚实的舞弊行为，只有加大扣分力度，才能维护国家立法通过的这一语言测试的严肃性和规范性。

2.2 缺时产生原因及改进建议

笔者认为测试中应试人缺时产生的原因主要有以下几点：一是测试前培训不到位，应试人不知“命题说话”的时限规定，以为随便说几句就可以了。二是所给某些话题距离应试人工作、学习、生活、兴趣有较大的距离，且话题缺少必要的话语情境，应试人一时难以打开话题，即使勉强打开话题，说话也难以继续。这主要是由于职业不同、生活内容不同、兴趣不同的人，大脑中所储存的信息类型也是不同的。通常情况下，应

试人对那些与自己工作、学习、生活、兴趣相关的话题较为敏感, 这些话题容易触动其心灵热点并引发其共鸣。所以, 如果抽到这样的话题, 应试人想说的话多, 会觉得说话时间不够, 不会说不满时间; 反之, 应试人则会觉得无话可说。三是一些应试人由于心理紧张, 思维受阻, 造成语流不畅或一时不知说什么好。四是个别普通话水平不高的应试人出于投机的考虑, 钻了评分标准的空子, 故意不把时间说满。与上述所说背稿、内容雷同情形相似, 依据目前说话的评分标准, 对个别普通话水平不高的应试人来说, 说话缺时扣除的分数, 远远不及说满时间, 因语音、语汇、语法失误扣除的分数, 个别应试人不把时间说满, 实际上是他们减少失误, 减少丢分的策略。以03《大纲》为准, “命题说话”语音标准程度一项评分, 语音错误扣分最多达14分, 而缺时只要不在2分30秒以上, 最多只扣6分。

针对缺时的种种情况, 笔者提出如下建议: 测试前加强对应试人的培训, 通过测试前扎实有效的培训, 让应试人充分了解说话项测查的目的, 明确测查的要求; 尽可能为应试人创设一个宽松的测试氛围; 设置话题情境; 增加应试人与测试员间的双向对话。这样, 应试人说话时就能消除紧张心理, 较容易地找到对象感和交流感, 发挥出他们的正常水平, 把规定的说话时间说满。另外, 为避免应试人的投机, 语音标准程度、语汇语法规范程度的评定, 应把失误次数的计算与说话时间长短、说话音节量的多少结合起来考查, 最后确定应扣除的分数。

2.3 测试员间评分差异较大产生原因及改进建议

笔者认为测试员在“命题说话”一项评分差异较大的原因有以下几个: 一是《大纲》规定的说话项的评分标准较为模糊, 定性的描述较多, 定量的成分不够, 不好把握, 难以精准地操作。比如说话项对语音标准程度的评定, 出现“方音不明显”、“方音比较明显”、“方音明显”、“方音重”等描述, 但具体出现几类系统性语音错误、系统性语音缺陷分别判定为“方音不明显”、“方音比较明显”、“方音明显”、“方音重”, 没有具体的说明。再如说话项语汇语法规范程度的评定, 有“语汇、语法偶有不规范”

和“语汇、语法屡有不规范”的描述, 但语汇语法不规范的情形有哪些; 说话3分钟之内, 语汇语法出现几次失误就属“语汇、语法偶有不规范”, 语汇语法出现几次失误就属“语汇、语法屡有不规范”, 每次失误扣多少分, 都没有作具体说明。类似以上这些情况, 由于各个测试员对评定标准的把握很难统一, 所以测试员间评分产生较大差异也就是必然的了。二是测试员自身素质、语言学素养及测试经验的差异, 导致他们各自对《大纲》评分标准理解、把握、运用上的差异, 这一差异最终导致各测试员间评分的差异。三是近年来说话项评分标准似乎有些混乱, 使得测试员无所适从。03《大纲》说话项评定分为语音标准程度、语汇语法规范程度和自然流畅度, 还补充了说话缺时扣分的标准; 而语教用司函[2009]5号《计算机辅助普通话水平测试评分试行办法》说话项评定分为: 语音标准程度; 语汇语法规范程度; 自然流畅程度; 说话不足3分钟, 酌情扣分; 离题、内容雷同; 无效话语六个方面。显然两个文件说话项评分标准有较大出入。笔者认为, 无论是“机测”还是人工测试, 每一测试项的评分标准应该是一致的, 不应该是“机测”一个标准, 人工测试一个标准。目前国家语委对此并无说明性的文件, 测试究竟采用哪个标准, 让人感到疑惑。四是某些测试员由于缺少及时学习、培训, 对测试的新规定、新动向不了解, 导致测试时一些测试员还在用老标准测试, 一些测试员用新标准测试, 不同测试员采用不同的标准测试, 必然产生对同一测试对象分数评定上的差异。

针对上述情况, 笔者建议: 一是增加说话项评定的定量成分, 使得评分标准更细化、更具体、更明确、更易于理解和操作。二是加强测试员上岗后的培养, 不断提高测试员的素质。如通过学习掌握跟测试相关的语言学和应用语言学的理论知识, 使测试员能够站在更高的层次上全面认识PSC; 通过及时学习、培训、测试案例的分析、研讨, 使测试员及时了解PSC的新规定、新变化、新动向, 更准确地理解和把握测试的评分标准, 更熟练精准地运用评分标准。三是把测试实践与测试研究紧密结合起来。测试实践与测试研

究有着密切的联系,测试实践是测试研究的基础,测试研究对测试实践具有指导作用,测试研究有助于解决测试实践中碰到的难题。笔者建议国家成立一个专门机构,邀请有关专家对测试中出现的问题加以研究和解决,以保证PSC沿着健康方向发展。

2.4 运用网络语、流行语、中外文夹杂表达产生原因及改进建议

笔者认为PSC“命题说话”表达出现网络语、流行语、中外文夹杂的情况是汉语发展演变中的正常现象。语言是一种社会现象,它随着社会的发展在不断发展。随着社会的发展,一些渐渐丧失生命力的旧语汇会走向消亡;相反,随着新事物的不断涌现及社会语言生活的变化,新语汇会大量产生。对于新产生的语汇,哪些应当吸收到汉语语汇中来,如何规范,应该是我们语言文字工作者要思考和解决的问题,这实际上是语言发展和规范的问题。正如著名的辞典编撰专家李行建所说,语言的发展和规范必须是统一的,光要发展不要规范不行,光要规范不要发展也不行。只有发展,语言才有生命力,但不讲究规范,语言就无法健康发展。

笔者建议对网络语、流行语中那些使用频度高,大部分社会成员能够理解的,应该大胆吸收到汉语语汇中来,测试不应判为不规范,这也完全符合语言的约定俗成性。如“菜鸟”、“微博”、“桑拿”、“给力”等词就属上述情况。对于表达时中外文夹杂则应坚决扣分。因为表达时中外文夹杂影响到汉语的纯洁性,使得汉语成了语言的大杂烩。如“我在看TV”,“OK,今天训练先到这里。Bye!”就应视为表达不规范。

测试中这一新问题的产生启示我们,PSC必须深入社会语言生活的实际,分析口语表达中的实际问题,PSC必须与时俱进,关注社会语言生活,关注汉语发展中的新现象、新问题,并对其加以规范和引导。

2.5 重视语音的学习,轻视甚至忽略语汇、语法的学习和规范产生原因及改进建议

笔者认为,应试人重视语音的学习,轻视甚至忽略语汇、语法的学习和规范,主要有以下三个原因:一是目前绝大部分地方测试采用“读单

音节字词”、“读多音节词语”、“朗读短文”、“命题说话”四项内容测试,免去了“选择判断”一项,使得本来占分值不多的语汇语法在整个测试中所占分值比重更小,算下来大概只有10分。因此,不少应试人认为只要学好普通话语音,就能应付PSC,语汇语法分值太少,不值得下功夫去学。二是个别测试员受自身能力的限制,对语汇语法不规范的情形无清醒的认识,不好做判定就不去判定。三是个别测试员以“口语”为挡箭牌,认为口语表达中语汇语法不规范是合情合理的,可以原谅的,对应试人语汇语法不规范采取了宽容的态度。对此,笔者认为测试员首先要有正确的认识。测试员应该认识到语音、语汇、语法是语言三个不可割裂的要素。语音是语言的物质外壳,语汇是语言的建筑材料,语法是语言组合的规律、规则,相对于语音,语汇、语法是更深层次的两个问题。因此,只有全面学习普通话语音、语汇、语法,才能在更高层次上提高普通话水平,进而提高汉语规范化程度。

笔者建议,目前凡是采用四道题测试的地方,应把“选择判断”一项的10分增加到说话项语汇语法的分值中去,这样,语汇语法分值就由10分增加为20分。语汇语法分值增加了,应试人就会渐渐地重视语汇语法的学习和规范。二是测试员在测试前的辅导应全面,既要让应试人明白方言与普通话在语音上的差异,又要让他们懂得方言与普通话在语汇、语法方面的差异,尽快使应试人测试前的学习、培训完善起来。

效度与信度是语言测试的两大基本要求,也一直是PSC的关键。以上笔者以自己十多年的测试实践为依据,在总结、思考的基础上,对如何提高PSC“命题说话”测试效度与信度提出了一些粗浅的看法。笔者相信,只要我们全体语言文字工作者、特别是全体测试员多实践、多总结、多探讨,PSC一定会更加客观、公平、公正,PSC的效度与信度一定会不断提高,PSC也一定会朝着制度化、科学化、规范化的轨道健康发展。

参考文献

- [1] 教育部,国家语委.普通话水平测试大纲[Z].教语用

[2003]2号文件.

[2] 教语用司.计算机辅助普通话水平测试评分试办法[Z].教语用司函[2009]5号.

[3] 徐泉.陈佑林.关于影响普通话水平测试信度和效度因素的分析及对策[A].见:中国应用语言学会.第4届全国语言文字应用学术研讨会论文集.

[4] 钱华.PSC“说话”项若干问题的思考分析及对策[EB/OL].<http://www.zsdhxx.com/new/jy/jx>ShowArticle.asp?ArticleID=5302>

011-5-29.

[5] 朱丽红. PSC中“说话”项的题型分析和改进建议[A].?见:国家语言文字工作委员会培训测试中心.第二届全国普通话水平测试学术论文集[C].北京:商务印书馆, 2006.

[6] 周小兵.论普通话水平测试的信度[EB/OL].[Http://www.sdyc.cn/qgy/zhaoshengjiuye/zhaosheng2006/hongxinwebs/wenzhai/wenzhai9.htm](http://www.sdyc.cn/qgy/zhaoshengjiuye/zhaosheng2006/hongxinwebs/wenzhai/wenzhai9.htm),2011-5-29.

A Tentative Study on the Enhancement of Reliability and Validity of the Assigned-topic-statement Test in PSC

LU Miao-qin

(Xianyang Vocational & Technical College, Xianyang Shaanxi 712000)

Abstract: there exist various factors which contribute to the decline of reliability and validity of the assigned-topic-statement test in PSC (Putonghua Shuiping Ceshi, also called Mandarin Proficiency Test). The thesis, conducts an overall analysis of the existing factors, effecting the reliability and validity of the assigned-topic-statement test in PSC, as well as the causes for those factors, and then offers a set of suggestions for improvement.

Key words: PSC, the assigned-topic-statement, reliability, validity, factors, causes, suggestions for improvement

(上接第40页)

间,取得了每667m²平均年收入达10000万元,纯收入达7000-8000元的高效益。

2004(11):9-10.

[2] 魏际峰.追肥枪,实用新型专利:200520124982.

[3] 金宏伟.葡萄地膜覆盖技术[J].吉林农业, 2007(6):18-19.

参 考 文 献

[1] 贺普超.葡萄新架型-“V”形架[J].农村新技术,

Study on Water-Saving fertilization and Cultivation Techniques for grapes in Shaanxi Weibei Semi-arid Plateau

ZHAO Wen-jie¹, YANG Xu-wu²

(1.Department of Biochemical Engineering, Xianyang Vocational Technical College, Xianyang, Shaanxi 712000;

2.College of Chemistry and Materials Science, Northwest University, Xi'an, 710069)

Abstract: This paper introduces water-saving fertilization and “V” form cultivation techniques for grapes in Shaanxi Weibei semi-arid plateau at 700-1000 m above sea level and without irrigation condition. Techniques of covering the black plastic film and fertilizing by use of the fertilization gun combined with the “V” form cultivation were employed for grape-planting. Compared with the plain area having good water and fertilizer conditions, the results in Shaanxi Weibei semi-arid plateau indicated that the use of these techniques can economizes water and fertilizer well and obtain remarkable economic benefit. It is worth generalizing and applying these techniques to grapes-planting in Weibei semi-arid plateau.

Key words: Weibei semi-arid plateau, Red globe grape, Water-saving fertilization, “V” form cultivation technique