

高考英语听力口语考试的构想研究

谢利亚

(身份证号码: 150105198008170925)

摘要: 高考英语听力口语考试的设想。高校招生考试和英语教学改革旨在提高学生实际运用英语的能力, 侧重英语作为交际工具的地位。从课程规范得知, 语言综合能力强, 不仅会读写, 而且会说、懂; 语言学习能力高, 不仅会用, 而且能用。它注重语言在培养、发展学生方面的作用, 是为了突出其非工具目的性。因此, 在构思英语听说考试中, 将听说技能放在适当位置, 经肯定接纳听说技能后, 允许容忍生人、生词的存在。

关键词: 高考英语; 听力; 口语; 构想; 效度

一、引言

高考英语属于高校招生入学考试体系中的重中之重, 人数之多、风险之高、影响之远是极其独特的; 国内已积累了大量有价值的研究成果。2018年北京高考听说考试秉承了课程标准思想, 明确听说作为语言能力在语言体系中的地位和作用, 按照语言测试的理论模式设计观念; 课程标准关注发展, 关心人, 关于语言, 课程标准在语言上既有工具性, 又有人文性, 实现了高度一致的语言观; 掌握一门外语, 重在交际, 更在全面发展。

二、英语听力口语考试的考核概念

What is measured by a listening and a speaking test is the interpretation of a test, or what is measured by a language test. When we analyze the evaluation, we should comply with the difficulty of foreign language teaching and the exam reform, as well as incorporate pertinent language skills and theories into construction of listening and speaking syllabi. The examination should measure the standards in the syllabi and to cover all the aspects of the cloze procedure. In the meantime, understanding the real meaning of examination is to select exams and promote basic education. Validity arguments should be based upon the evidence provided by data collection on multiple aspects over time aimed at providing a basis for sound judgments across the life span of a test (Li Qinghua, 2010).

三、北京高考英语听说测试

本次开考的北京高考英语听说测试共包括听力和口语两个部分, 口语部分又包括短文转述、段落朗读和回答问题段, 听力

部分包括快速回答和听短文回答问题部分。此次测试的研究设计中采用交互式人机对话形式, 采用以当今社会发展现代化的电脑技术和头戴麦克风技术高度智能化的计算科学, 学生可通过计算机头戴麦克风设备的调节而录制他们的听力理解与口语内容, 随后使用计算机语音识别技术将其信息计算化, 最后利用人工智能系统进行准确打分。

1. 理论支撑及研究基础

效度是测试评价的核心问题, 也是测试质量的基本要求 (APA 等, 1999)。效度整体观认为效度是权威性解释的涵盖, 即效度是研究抓住证据和理性基础性成分来看待分数解释和使用的。这一理念的提出为语言测评带来了深远的影响, 然而该理性却遭到研究贫乏以及适用性不强的诟病。语言测评领域的学者们为此也建立起了一系列具体、细化的效度观。例如 Bachman (1990) 的语言交际能力模型, 以及 Bachman & Palmer (1996) 的评价工具实用性评价结构。后者充分阐释了 Messick 的效度整体观, 并使其于语言测评的实际中得以体现。

所谓考查英语口语测试的一致性或准确度, 可以从考查结果的可信度方面进行探讨, 可信度: 即英语口语测试可行性或信任度, 是指能否相信测试得到的结果。其常用的定距是同一班级重复考查同一项英语口语测试的 2 次分数背上, 如果是这两次分数背上高度一致, 说明它的可信度就很高, 反之可信度就有限。(秦明星, 2011)

2) 效度是指测试结果是否准确地反映了受测者的语言能力, 也是测试质量的重要指标。

证据包括内容效度、结构效度等。内容效度主要研究试题和目标能力之间的联系，而结构效度研究测试成绩如何去解释被测能力，使内容效度得以量化。这与韦宗嘉（2001）对内容效度的观点一脉相承。测试的内容效度是指测试材料在多大程度上对被测能力具有代表性和典型性。测试内容越具有效度，就越能准确地反映受测者的语言知识、技能和能力，符合考生的需求，并起到应有的良好效应。北京高考口语测试中 Reading 部分为朗读类任务，直接考察了考生的语音语调方面的知识；其余部分为交际类任务，间接考察了考生的其他语言知识。可见，目前中学英语教学并不重视考生的口语表达能力提高，学生害怕张嘴说英语。在测试的内容效度方面，Henning（1987）给出的定义为：当用来衡量被测能力的各项某人为材料所充分代表和涵盖时，就说这种测试具有内容效度。它体现在所用试卷材料是否具备很好的代表性，而试卷中题目是否足以代表受测者所拥有、掌握的同时具有一般性、典型性。若测试内容具有判别度，则意味着其可以客观明确、具体、精确地反映受测者在某项能力(名词、动词及词组搭配等)上所达到的水平，同时对受测者所测量的能力进行适当区分（如单项选择题、完型填空或改错题等）。对于当前高考口语测试中由 10 月份参与可行性测试得来的结果可以基本说明它是具有内容效度的，因此它还是比较可信的。由于本恩实在语言测试研究中首先指出了效度问题并给予了科学解释，故在编辑本领域的相关论文时我们坚持使用效度这一术语。整个 Reading 部分难度和词数定位适合高一学生，但选文比较单调且考察较多音节文中不真实音难度不算合适。

在谈论结构效度时，参照 Bachman 和 Palmer（1996）的论文，试卷品质的实质就是把考核分数作为能力或结构的评判标准，并对其内涵进行阐释的深度。它是一个永不停息的流变过程。记住，试卷分数揭示出来的解读并不能百分之百无误，而只是一种相对性的认知。北京高考英语口语测试评估了口头表达和听说综合能力。考生表达个人观点或回答问题，不必实时与他人交流。技术尚未完全实现人机互动，通常以录音形式进行。

2.公平和可执行性

大规模的重要考试大都没有设置口语考试的考核形式，主要是因为施测性和公平性的问题。当然也各自有自己的原因，比如涉及到大量的人力物力投入问题，如常规的面谈模拟形式的口语测试，将面临人力场地限制，另外虽可采用计算机与受试者对话的模式形式进行测试，但这将带来大量的计算机设备需求、大量可观经费支出、故障发生问题，后一所带来对公平性问题，因设备不存在能够同时命中所有受试者，需要分批迎接，若有因考试题库限制需要根据情况多备考题的需要，则引发不同版本间考试可比性问题，上面这些因素，导致许多重要考试并没有选择口语测试。认识到这些高风险测试可能对教育产生否定影响后，假如因技术问题的口语测评被拒绝鉴定高考语言口语能力将会继而延伸至普通教育阶段出现同样负面影响。学校教育很可能忽视对学生口语能力认知学习，数年以后学校的学生仍只能英语听说认知而不敢说。如此，有可能促使学生听说能力认知存在半数以上时间，并没有起到提高学生实际应用英文的能力。如今，人们更清晰认识到学习英文的优点与重要性，直接影响到国家层面、社会及个人层面的发展(郑方贤,2019)。高考实施口语测评具有很重要的现实意义。

3.试卷制作

据悉，为了保证考试的公平性，高考多套卷；但不同卷可比性的问题也随之出现。一般，前期测评的数据基础上，采用 IRT 模型将试题映射到同一个标度上，或考后进行结果等效性处理，均可完成多套卷数据的汇总工作。但是，高考保密的要求，使得无法提前就数据进行需要等效性处理，提前对数据结果中弥散的误差等因素进行控制，也不可直接报告结果。因此，多套卷只能够在出题过程中尽可能谨慎，但由于出题者和后续评分者不相同；而这也难以避免卷和结果等冗余不一致的情况。

4.批阅评分

智能评分系统采用定标的评分员的使用先天潜知识的方法学习评分规则，并经专家评估保障系统的评分质量。在评分员与系统在初评与终评的评分结果相差较大时，系统的评分结果可由专家对其判断做出更正。智能评分系统具有稳定性及一致性，但缺乏灵活性。人工评分灵活但不稳定，故二者应当相辅而不相互取代。合作排序和融合的最终效果应当取决于相辅相

成的效果,但智能系统综合排序之后的结果质量却仅取决于先前学习的质量。当2001年北京各高校联合组织高考大规模扩招的时候,北大海澜计算中心已经对两年之前两院联合举办的高考政治科目已经进行了试运作。中国政法大学法学研究院在与北大等高校配合开展联招考试选拔2000年《政治学概论》试运作期间,是死记硬背试用大量题型解答、迎合机械、敷衍塞责,在连续开展了两届同类型联招,大力支持北大与清华举办南北学术大联合的2001年春晚期,尖子生平均6次单位高校系部教授共计25次无差异问题不变引蛇出洞地上当受骗的成瘾;为了掌握智能阅卷系统的使用积极性,我们每次每考必答抽必评,且必须抽查随机选取和控制攸关性问题随机抽签进行设置抽题待考,及其所占比重重要权重要实际含义任务数据特征布局换位变换运动变量,并对6次30余所高校教学课验的第六次联招同密辑性类弹六联嘉密也举行下载会。如此连串行动铺垫了该系统后来无历史前照地显形显灵般顺利到位。(吕鸣 2015)

5. 考试场所和练习

北京市高中生英语听说测评模式中引入了人与计算机对话的方式。针对本市建设了许多考场,在考场里,每个考生的耳机上装有一个麦克风,用于考试对答用,并且经过法定标准化。所有这些考场装有作弊考场监控系统,可以充分保证测评过程的公平性和净度性。在测评过程中,开展了试卷A和B的考察交叉使用。规定:进行2场测评为宜。当进行第一轮的两场测评考完以后,A和B试题才会全市统一更换一次。使测评诚信性显著提升。本市考评学生需要模拟训练用的设备——常见的商业模拟系统是付费昂贵版的。

整个考卷预计用时22分钟,其中口语测考15分钟,怎样设计合理有效、使出题、考试、评分更加科学化,目前仍然值得探索。总而言之,将英语口语考试融入中高考英语考试体系,对于中学英语教育的健康成长,以及为高校选拔合适的人才,都是有益的尝试。然而,英语口语考试中依然还有不少问题亟待解决和目标需要实现,真正有意义的中高考英语教学前景还有待数据支撑和计划实现。

参考文献:

- [1] Bachman L & Palmer A. Language Assessment in Practice [M]. Oxford, Oxford University Press, 2010. 26. 外语测试与教学 2021 年第 1 期
- [2] 国务院. 国务院关于深化考试招生制度改革的实施意见 [EB/OL]. http://www.Gov.cn/zhengce/content/2014-09/04/content_9065.htm, 2014-09-03.
- [3] 侯艳萍. 外语高考听说测试改革的反拨作用研究[J]. 外语电化教学, 2018, (5): 23-29.
- [4] 教育部考试中心. 中国高考评价体系[M]. 人民教育出版社, 2019.
- [5] 吕鸣. 智能测评技术在大规模英语口语考试评卷中的探索与实践[J]. 中国考试, 2015, (10): 51-57.
- [6] 郑方贤, 徐雯. 新高考推动上海英语教育发展[J]. 中国考试, 2019, (9): 32-36.
- [7] 董连忠. 高考英语测试对高中英语教学的反拨效应探析——基于一所普通高校新生的问卷调查[J]. 外语测试与教学, 2014(2).
- [8] 董连忠. 高考英语听力考试的反拨效应研究[J]. 外语测试与教学, 2017(1).
- [9] 亓鲁霞. NMET 的反拨作用[J]. 外语教学与研究, 2004b(5).
- [10] 亓鲁霞. 高考英语的期望后效与实际后效——基于短文改错题的调查[J]. 课程·教材·教法, 2007(10).
- [11] 亓鲁霞. 语言测试的反拨效应理论与实证研究[J]. 外语教学理论与实践, 2011(4).
- [12] 秦秀白. 充分利用四、八级考试的反拨效应, 抑制教学质量滑坡[J]. 外语界, 2012(3).
- [13] 徐雯. 刍议新一轮高考改革背景下的高考英语(上海卷): 创新、思考与展望[J]. 外语测试与教学, 2016(4).
- [14] 郑方贤. 高考外语科目改革的效应[OL]. Retrieved from http://www.sohu.com/a/119952466_372526, 2016-11-25.
- [15] 邹中, 徐倩. 标准参照考试及其反拨效应——以 TEM 考试为例[J]. 外语教学理论与实践, 2014(1).