

改革后大学德语四级听力试题内容效度分析

郑 彧

(上海理工大学 中德国际学院, 上海 200093)

摘要: 以2010年大学德语四级听力测试大幅度改革后7套听力试题为研究对象, 结合《大学德语课程教学要求》和《大学德语四、六级考试大纲》对听力部分的要求, 以Bachman & Palmer的任务特征框架为依据, 对大学德语四级听力试题进行系统的内容效度分析, 旨在探讨改革后大学德语四级听力试题的命题质量。研究表明: 改革后大学德语四级听力试题的语篇输入和预期作答任务特征比较符合《教学要求》和《考试大纲》的相应要求。然而, 分析中也发现一些不足, 例如试题结构、任务类型单一, 部分试卷听力材料偏难, 语速波动大等。基于上述研究结果, 横向对比“欧洲语言共同参考框架”A2听力试题, 尝试为大学德语四级听力试题命制提供参考性建议。

关键词: 大学德语四级改革; 听力理解; 内容效度; 任务特征

中图分类号: H 314.2 **文献标志码:** A **文章编号:** 1009-895X(2022)02-0150-06

DOI: 10.13256/j.cnki.jusst.sse.2022.02.007

Evaluating the Content Validity of Post-Reform PHD 4 Listening Comprehension Session

ZHENG Yu

(Sino German College, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Based on corresponding teaching and testing syllabuses after reform in 2010, this paper analyses the content validity of listening comprehension session of PHD 4 within the framework of test task characteristics put forward by Bachman & Palmer (1996), aiming at examining tests' quality. The findings reveal that the characteristics of the listening tasks and expected response to the test items basically accord with the requirements of the corresponding teaching and testing syllabuses. However, some defects need noticing: test format and task types of some test papers are monotonous while other listening material and test items are too difficult in addition to the dramatically varied speech speed. Compared with GeR A2 listening comprehension, this paper puts forward some tentative suggestions for the future design of the PHD 4 listening comprehension.

Keywords: PHD 4 after reform; listening comprehension; content validity; task characteristics

2009年7月, 高等学校大学外语教学指导委员会德语组举行了全国大学德语教学与测试工作会议, 通过了《大学德语课程教学要求》(简称《教学要

求》), 提出了“培养学生不同层次的德语综合应用能力, 使他们在今后的学习、工作和社会交往中能够在一定程度上运用德语有效地进行交际, 完成

收稿日期: 2020-09-15

作者简介: 郑 彧(1985-), 女, 讲师。研究方向: 德语应用语言学、测试与教学。E-mail: joannadzeng@hotmail.com

各种任务，同时增强学生自主学习的能力，增强其积极参加中德交流的意识，提高其跨文化交际能力和综合文化素养，以适应我国社会发展和国际交流的需要”的教学目标^[1]。以此为依据，经过学习借鉴、反复讨论和论证，2010年4月新的《大学德语四、六级考试大纲》（简称《考试大纲》）正式公布执行，用以检查《大学德语课程教学要求》的执行情况，评估教学质量，检验学生的德语综合应用能力。

大学德语四级（以下简称 PHD 4）从 1995 年实施考试开始就有听力部分，在整套试卷中的权重为 18.5%^[2]。为达到《教学要求》确定的教学目标，从 2010 年开始 PHD 4 听力测试进行了大幅度的改革：1）试题比例从 18.5% 提高至 26%，分值从 15% 提高至 20%；2）答题时间从 15 分钟调整为 20 分钟；3）听力题型更加丰富，增加了正误判断题。鉴于听力在语言系统习得中的重要性^[3]和 PHD 4 听力测试大幅度改革的实施，对改革后的 PHD 4 听力试题的内容效度进行研究十分必要。目前，针对 PHD 4 考试的研究数量极其有限，研究内容多集中于测试介绍或应试技巧传授，例如刘贝贝^[4]，王轶、赵亘^[5]，缺乏对 2010 年改革以来 PHD 4 听力试题进行内容效度分析的研究。

一、研究设计

通过分析测试的内容和范畴，可以获得基于测试内容的凭证，这是重要的效度凭证之一。Bachman 认为，基于测试内容的效度证据研究，应包含两部分：内容相关性和内容覆盖性。前者指的是将试卷与权威理论规范相比较，从而得出两部分的相异同程度，二者越接近，则试卷的内容效度越高；后者则是考量该试卷是否涵盖了应该考查的技能范围^[6]。本研究对效度证据分析采用的是 Bachman & Palmer 提出的测试任务特征框架^[7]。笔者搜集了 2013 至 2019 年共 7 套 PHD 4 真题，主要从听力试题任务特征因素的角度开展分析。在进行上述分析的同时，对照《大学德语课程教学要求》（2010 年版）和《全新大学德语四、六级考试大纲》（2010 年版），分析 PHD 4 听力测试自 2010 年修订后的内容效度，明确该部分与相应《课程要求》和《考试大纲》的吻合程度。另一方面，以“欧洲语言共同参考框架”同级别语言水平测试（以下简称欧标

考试）为研究样本，参照上述框架分析听力试题，横向对比 PHD 4 听力试题，旨在为今后听力试题的命制提供一些参考性的建议。该研究选取欧标考试的参考依据为考试所面向考生的总学时数。大学德语教学分为 1~6 级，PHD 4 听力部分选材以《教学要求》规定的四级要求为原则，受试对象为累计学习达 240~280 学时的学生。欧标考试分为 A1、A2、B1、B2、C1、C2 等 6 个级别，其中 A2 德语考试面向累计学时达到 200~350 的学生，与 PHD 4 考试学时要求基本一致。因此，欧标 A2 考试作为横向对比的研究样本具有合理性。具体研究样本为 Goethe Zertifikat A2 Erwachsene 和 Telc A2，共 2 套欧标 A2 德语听力样题。

二、分析框架与理据

本文参照 Bachman & Palmer（1996）的任务特征框架，结合《教学要求》、《考试大纲》、PHD 4 及欧标 A2 听力试题各部分的特点，确定听力测试语篇输入、任务类型和预期作答任务特征的分析框架。语篇输入任务特征主要从体裁、题材、难度、话轮数、短文长度和语速方面开展分析；任务类型主要分析听力题型，预期作答则聚焦于所考查的听力技能（详见表 1）。

三、结果与讨论

笔者分别对 7 套 PHD 4 听力真题及 2 套欧标 A2 德语听力样题分析发现，就试题结构而言，PHD 4 由短对话（2~4 个话轮）和短文构成。在过去 7 年中，测试材料以短对话为主，但每年短对话和短文的配比均存在差异：从 2014 年对话和短文的比例为 5.5/4.5，到 2016 年的 7.5/2.5，再到 2019 年全部为短对话。这从一定程度上影响了试卷的标准化。欧标 A2 考试则包括短文、短对话及长对话，其中短文是主要的测试语篇类型。下面将从语篇输入、任务类型及预期作答任务特征方面具体分析与讨论上述考试的听力部分，探究改革后 PHD 4 听力试题的内容效度，并横向对比欧标 A2 考试。将分析结果先与《课程要求》和《考试大纲》的相关要求进行对比，分析其符合程度，再将其与欧标 A2 听力部分进行对比，开展进一步的讨论。

表1 任务特征分析框架

Tab. 1 The framework of test task characteristics

任务特征	项目	描述	涉及部分
语篇输入	体裁	记叙文、说明文、议论文、应用文	听力短文
	题材	日常活动、学校生活、工作求职、人物故事、科技等	短对话、长对话、听力短文
	难度	根据Flesch易读度公式计算得到的易读度值	听力短文
	话轮数	长对话的话轮数	长对话
	短文长度	听力短文的词数	听力短文
	语速	平均每分钟朗读每套试题所有题型听力材料(不含题目)的词数(wpm)	短对话、长对话、听力短文
任务类型	任务名称	单项选择题、多项选择题、填空题、配对题、正误选择题等	短对话、长对话、听力短文
预期作答	考查技能	理解中心意思、听懂重要的或特定的细节	短对话、长对话、听力短文

(一) 语篇输入

1. 体裁

PHD 4 听力短文涉及体裁的分析。短文体裁多样,覆盖了《考试大纲》规定的“对话、广播节目、故事、叙述等”^[23]。表2显示,听力短文对应用文的考察最多,占总数的51.72%,其次为记叙文(31.03%),符合其中“增加了天气预报、交通路况报道、体育广播及其他各种新闻等短小的体裁类型”的描述,有助于实现《考试大纲》提出的测试“考生能否听懂接近正常语速、表达清晰、话题熟悉的口头表达并从中获取信息的能力”的目标^[23]。其中:2016年和2018年真题涉及记叙文、说明文等两种以上体裁,体裁类型较丰富;2013年和2014年涵盖记叙文和应用文两种体裁;2015年和2017年仅使用议论文这一种体裁。体裁数量分配上

表2 听力短文体裁分布

Tab. 2 Types of monologues

试题	记叙文/%	说明文/%	议论文/%	应用文/%
PHD 4 总体	31.03	6.90	10.35	51.72
2013 PHD 4	25	0	0	75
2014 PHD 4	44.4	0	0	55.6
2015 PHD 4	0	0	100	0
2016 PHD 4	40	0	20	40
2017 PHD 4	0	0	100	0
2018 PHD 4	20	40	0	40
2019 PHD 4	0	0	0	0
欧标A2	20	0	0	80

的变化从一定程度上反映出选材的不稳定。

对比两项考试的体裁分布,应用文和记叙文均为使用最多的体裁。应用文的特点是实用性强,多用于信息传递、解决实际问题等,是日常工作生活中常用的文体^[8]。记叙文是以描写人物的经历和事件的发展变化为主要内容的一种文体,是最基本、最常见、也是最主要的表达方式^{[8]61}。此外,PHD 4还对说明文进行考查(6.9%)。说明文通常用来扩充、增加知识等,信息量大,知识性强,理解难度要大于叙述性篇章^[8-9]。更值得注意的是,2015至2017年真题听力材料中还各出现了一篇议论文。这是一种论述事理、发表意见、提出主张的文体^[10]。特点是结构严谨、层次分明,与说明文一样属于理解难度较高的体裁^[11],进一步说明了听力文本选材难度差异较大。通过上述分析可以发现,PHD 4听力篇章难度要略大于欧标A2考试难度。

2. 题材

PHD 4听力题材涉及了日常生活、学校生活、工作求职、人物故事和科技类5种话题,符合《考试大纲》“日常生活范围内”“实用”等要求^[27]。日常生活是短对话和短文主要涉及的题材,分别占79.28%和72.41%;其次为学校生活,分别占9.01%和13.79%。此外,短文还涉及科技类题材,覆盖面较短对话更广。广泛的话题有助于更客观、全面地考查学生的听力理解能力。同时,所涉及的话题与学生日常和学习生活紧密相关,符合听力材料真实性的原则。

欧标A2考试涉及4类话题,分别为日常生活(78.26%)、工作求职(13.04%)、人物故事和学校生活(各4.35%)。短对话和短文以日常生活类

话题为主，分别占 80% 和 89.47%。长对话的题材为日常生活、工作求职和人物故事，分别约占 33.33%。对比两项考试的话题分布，发现题材分布相似度较高，涉及题材最多的是日常生活；PHD 4 学校生活的话题比例（10%）明显高于欧标 A2 级考试（4.3%），而后者工作求职的话题（13%）明显高于前者（9.3%），反映出两项考试面向考生受众上的差异。虽然对考生累计学时上的要求一致，但 PHD 4 面向在校非德语专业的大学生，而欧标 A2 考试是面向全社会的大规模语言考试。

3. 难度

听力短文涉及难度分析，难易度参考“Flesh”（易读度）指标。易读度通常用来衡量篇章难度，指语篇易于阅读和理解的程度或性质，主要应用于阅读测试和阅读教材的选材、命题及评估^[12]。目前，有两种常见的针对德语篇章易读度值的测算方式：一种是按原版 Flesh 易读度公式进行测算，但在音节数量统计时，排除非重读词尾“e”，参考按照德语篇章特点调整后的易读度量表^[13]；另一种是按 Amstad（1978）^[14]调整后的德语版公式进行测算，参考原版量表。为了衡量两种测算方式及量表的合理性，笔者选取两篇不同体裁的测试语篇，邀请两位具有 10 年大学德语教学经验的教师对语篇难度进行评价。综合易读度计算结果，最终确定仍采用原版公式和德语版“Flesh”参考量表，即 $FI = 206.835 - 84.6 \times WL - 1.015 \times SL$ 。其中，FI 指易读度指数，WL 指单词的平均音节数（不包含词尾 e），SL 指每个句子的平均词数。

根据德语“Flesh”参考量表，易读度分值越高的文档越容易理解（分值区间为非常容易 81~100，容易 71~80），易读度越低材料越难读（非常难 0~20，难 21~30），德语易读度 41 至 60 为“标准难度”。统计结果显示，PHD 4 短文的平均易读值为 56.3，为“标准”，不符合《考试大纲》中“句子结构比较简单”的要求^[23]。纵向对比近几年短文的平均易读值发现，2015 年和 2018 年的听力短文难度（49，47）远高于 2013 年、2014 年和 2016 年的短文难度（63.6，62.8，65.6），波动幅度很大，说明难度级别不稳定。欧标 A2 考试短文的平均易读度指数为“较容易”（67.4）。其中 Goethe A2 短文平均难度为“容易”（73.2）。

具体看各年份 PHD 4 短文易读度值分布，易读度指数以“41~60”为中心分布，以“标准”和

“较容易”两级为主。其中，2018 年的试题中有两篇短文的易读度值低于 30，难度级别属于“难”，达到高考（Abitur）难度^[12]。而 2014 年的试题中有一则短文易读度值为 91，难度级别为“非常容易”。难度级别从“难”到“非常容易”跨越了 6 个等级，进一步说明波动幅度很大，难度级别不稳定。而欧标 A2 考试短文的难度分布则较为均衡，跨度从“标准”至“非常容易”，且易读度指数以“较容易”为中心分布，以“标准”（33.3%）、“较容易”（20%）和“容易”（26.7%）此 3 级为主，难度适中，较容易理解。

对比两项考试后发现，欧标 A2 考试听力短文难度整体上低于 PHD 4。进一步对比两项考试短文难度级别的分布，发现 PHD 4 高难度听力短文（“难”和“有点难”）的数量为 13.1%，而欧标 A2 考试未涉及该难度的听力短文，所选用的听力材料均明显低于该难度。对比结果进一步印证了 PHD 4 在听力篇章选择上难度要略大于欧标 A2 考试。听力所用材料难度过高或过低都会影响考试的效度，进而可能影响对考生德语水平的准确测量。因此把握好听力测试材料的难易程度是测试设计者在选材时需要重点考虑的问题。

4. 话轮数及短文长度

长对话涉及话轮数的分析。PHD 4 听力部分不涉及长对话语篇输入，因此该部分仅对欧标 A2 考试听力理解部分进行分析。分析结果表明，Goethe A2 听力部分涉及两篇长对话，Telc A2 包含一篇长对话，均有 12 轮对话，话轮数稳定。对于初级水平的听力测试，过多的话轮数无疑会加大考生信息处理的负荷，使考生产生心理焦虑，进而影响其听力水平的正常发挥。因此，在设计长对话时话轮数的数量设定应把考生的语言水平纳入考量范围。

PHD 4 短文长度均在 30~40 词的范围。虽然考纲中对短文长度没有明确的定义，但符合《考试大纲》中对听力材料“短小的体裁类型”的描述^[27]。对比两项考试，发现短文的长度平均值相近。听力材料的长度是影响听力集中度的一个重要因素，因此，选材时适度把握听力短文的长度是必要的。

5. 语速

本文分析的语速是参照辜向东、李亚果^[15-16]提出的朗读每套试题所有题型听力材料（不含题目）的平均语速。计算公式为：每套试题听力理解部分

的平均语速(wpm)=不同题型听力材料(不含题目)的总词数(w)/朗读这些材料的总时间(m)。

结果显示(见图1),2013年的PHD 4听力理解部分的平均语速为42.04 wpm,明显低于每分钟约100词的考纲要求。德语口头表达平均语速在90~120 wpm区间内,也不符合考纲要求的“能听懂接近正常语速的内容”。研究表明,过慢的语速或者停顿时间太长也造成理解障碍^{[9][13]}。之后,PHD 4听力理解部分的平均语速有了显著的提高,除2016年外其他各年份平均语速都保持在每分钟100词以上,说明PHD 4听力部分的平均语速趋于适度。其中,2016年和2019年的平均语速分别为92.43 wpm和102.33 wpm,与考纲要求每分钟约100词相符。2014年、2015年和2017年略高于考纲要求,而2018年的平均语速则明显高于考纲要求,达到约125词/分钟。

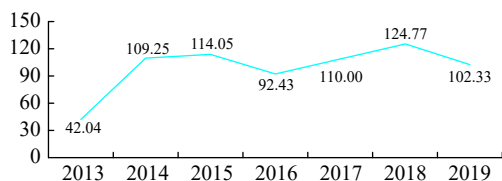


图1 PHD 4听力部分的平均语速

Fig. 1 Average speed of speech in PHD 4 listening part

欧标 A2 考试平均语速的平均值为 129.31 wpm,明显高于 PHD 4 近 7 年试题的平均值 99.27 wpm。排除 2013 年的极端情况后,欧标 A2 考试的平均语速依然明显高于 2014~2019 年的平均值 108 wpm。但欧标 A2 考试两套样题的平均语速有明显差异,Telc A2 平均语速为 112.36 wpm,Goethe A2 达到了 149.92 wpm。可见欧标 A2 考试的听力平均语速整体比 PHD 4 高,而 2015 年和 2017 年 PHD4 听力平均语速则与 Telc A2 接近。

(二) 任务类型

PHD 4 听力部分共分为 A、B 两节,每节各 10 道试题,题型均为提供四个备选答案的单项选择题,基本符合《考试大纲》中“A 节为单项选择题,B 节为单项选择题或正误判断题”的描述。考纲中虽提及“和改革前听力理解测试材料相比,现在增加了正误判断题型”,但在研究样本中依然只有单项选择题一种题型^[27]。而欧标 A2 考试中常见的测试任务有单项选择题、正误判断题、填空题及配对题,且听力部分均由两类及以上任务类型构成,各类任务使用较为平均。

(三) 预期作答

PHD 4 对所考核的听力技能有明确的定义,共包括 3 项微技能。分析结果显示,历年试卷全面覆盖了《考试大纲》所规定的全部考核技能,不但考核了“选择性获取信息”和“获取内容细节”等低层次听力理解技能,而且考核了高层次听力理解技能,即“获取听力材料主要信息”“能判断语言交际的环境”和“能判断或推测说话人的观点、意图和态度”,达到语篇层次的理解^[27]。对比各项微技能的分布可以发现,PHD 4 较注重于对听懂重要的或特定的细节的考查(69.3%),与《考试大纲》中“A 节测试考生获取听力材料的主要信息和关键词的能力”,“B 节测试考生有选择地获取听力材料中针对题目信息的能力”^[23],即两项听力技能各占 50% 的要求有所出入。对比两项考试发现,欧标 A2 考试着重考查的是学生能否听懂重要的或特定的细节信息,此类试题数量为 100%。由此可见,其对听力技能考查的比重不平衡,缺乏对理解主旨要义等高层次听力理解技能的考查。

四、结论与建议

以上分析表明,研究样本 PHD 4 听力试题基本符合《教学要求》和《考试大纲》的要求,具有较高的内容效度。语篇输入方面,体裁多样,话题广泛,短文长度适中;预期回答特征方面,听力技能全面覆盖《考试大纲》所规定的全部考核技能。然而,此次研究也发现了一些不足之处。比如,听力材料语篇类型、试题题型单一,短文整体偏难,语速波动较大,不稳定等。基于以上研究结果,笔者尝试为 PHD 4 听力试题的命制提供以下 4 点建议。

第一,听力测试材料的难度有待调整。本研究结果表明,PHD 4 听力测试有必要遵循易读度原则,从减少复杂的长单词数量及降低句子长度这两方面做些改进,从而降低听力材料的难度。具体而言,复杂的长单词可以用形式短小的近义词来替代,也可以将长单词拆解成短语;多使用动词,避免名词性短语;用主动语态替代被动语态;将复杂的长句改写成若干个简单短句。另一方面,按照《课程要求》中的教学评级,1~4 级为一般要求,5~6 级为较高要求,即 PHD 4 属于初级阶段的语言考试,因此其听力材料的难度不应该非常难,即易读度不应该包括 0~30 等级。

第二, 减小波动幅度, 使平均语速趋于平稳, 符合考纲每分钟约 100 词的要求。研究结果显示, 个别年份听力试题的平均语速或明显低于考纲要求, 或要求过高, 这两种情况都可能影响对学生德语水平的准确测量。

第三, 听力测试材料的语言输入类型还有待完善。通过与欧标 A2 考试对比, 可以发现 PHD 4 听力材料的语言输入类型仅有短对话和短文两种, 类型较为单一, 不能较全面地还原生活中常见的各种语言交际场景。测试设计者应提高听力材料的语言输入类型, 让听力测试尽量忠实于目标语使用的环境和要求。

第四, 丰富听力理解试题的任务类型。虽然改革后 PHD 4 的听力部分增加了正误判断题型, 但在本文研究样本中未见新任务类型出现。现今听力测试中常见的测试任务类型有开放性和封闭性两大类, 包括排序、单项选择题、多项选择题、填空题等^[17]。测试设计者可以根据测试目的选择合理、多样的任务类型考查学生的听力理解能力, 以提高测试结果的信度^[17]。

参考文献:

- [1] 教育部高等学校大学外语教学指导委员会德语组. 大学德语课程教学要求[M]. 北京: 高等教育出版社, 2010.
- [2] 全国大学德语考试设计组. 全新大学德语四、六级考试指南[M]. 上海: 上海外语教育出版社, 2010.
- [3] FEYTEN C M. The power of listening ability: An overlooked dimension in language acquisition[J]. *The Modern Language Journal*, 1991, 75(2): 173 - 180.
- [4] 刘贝贝. 大学德语四级考试新大纲详解[J]. *德语学习*, 2011 (3): 59 - 62.
- [5] 王轶, 赵巨. 解析新版大学德语四级考试阅读理解[J]. *德语学习*, 2011 (3): 63 - 71.
- [6] BACHMAN L F. *Fundamental Considerations in Language Testing*[M]. Oxford: Oxford University Press, 1990.
- [7] BACHMAN L F, PALMER A S. *Language Testing in Practice*[M]. Oxford: Oxford University Press, 1996.
- [8] BRINKER K. *Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden*[M]. Berlin: Erich Schmidt Verlag, 2010.
- [9] SOLMECKE G. *Texte Hören, Lesen und Verstehen* [M]. Berlin: Langenscheidt Schulbuch, 1993.
- [10] HAUSENDORF H. *Textlinguistik fürs Examen*[M]. Göttingen: Vandenhoeck & Ruprecht, 2008.
- [11] GROTJAHN R. *Postgraduiertenstudium Deutsch als Fremdsprache Studieneinheit Leistungsmessung und Leistungsbeurteilung Kapitel 1 Grundlagen Rüdiger Grotjahn Juli*[M]. Patras: Hellenic Open University, 2000.
- [12] 辜向东, 关晓仙. CET阅读测试与大学英语阅读教材易读度抽样研究[J]. *西安外国语学院学报*, 2003, 11 (3): 39 - 42.
- [13] BACHMANN C. Die flesch-formel[EB/OL]. [2020-03-25]. <http://www.leichtlesbar.ch/html/fleschformel.html>.
- [14] AMSTAD T. *Wie verständlich sind unsere Zeitungen?* [D]. Zürich: Universität Zürich, 1978.
- [15] 辜向东, 李亚果. 改革后CET听力测试语篇输入与预期回答任务特征分析[J]. *西安外国语大学学报*, 2010, 18 (4): 71 - 74, 79.
- [16] 辜向东, 李亚果. CET听力测试语篇输入和预期回答任务特征历时分析(1996—2007)[J]. *外语测试与教学*, 2012 (3): 17 - 26.
- [17] HINGER B, STADLER W. *Testen und Bewerten Fremdsprachlicher Kompetenzen: Eine Einführung*[M]. Tübingen: Narr Francke Attempto, 2018.

(编辑: 朱渭波)