

国内对外汉语分级测试 设计及实施调查研究

A Survey on the Chinese Language Placement Test in China

罗莲*

luolian99@163.com

Luo Lian

中央民族大学国际教育学院
北京 中国, 100081

Minzu University of China
College of International Education
Beijing, China 100081

分级测试是语言教学的起点,它是“以学习者为中心的教学”的重要保证。然而国内对汉语作为第二语言分级测试的研究甚少,且人们对国内各教学机构的汉语分级测试状况缺乏了解。为此,本文进行了一次调查研究,意在抛砖引玉,发现对外汉语教学界分级测试存在的问题,以引起各教学机构的重视,并对之进行深入的研究。

[关键词] 分级测试; 调查; 设计; 实施

Placement test is the starting point of any language teaching programs, and it is the foundation of learner-centered teaching. However, very few researches on placement tests have been done in this field, and the general situation of these tests in China are rarely known to the researchers. Therefore, this paper investigates the problems in the placement test of teaching Chinese as foreign language by the approach of questionnaire, intending to draw attention from various language teaching institution and attract more researchers to conduct in-depth further researches in this field.

Key words: Placement test; survey; design; administration.

* [作者简介] 罗莲,北京语言大学博士,中央民族大学国际教育学院讲师。主要研究方向为语言测试与评估。

1. 问题的提出

作为第二语言教学的起点,分级测试对教学的顺利进行有着非常重要的意义。首先,分级测试不准确对教学单位不利。由于班级学习者水平不齐,有时不得不重设新班,增加了支出,教学单位也很难分配教学资源。其次,这样的测试对学习者的不公平,学习者可能因分错级别而不愿意继续学习汉语,或者导致他们进步缓慢、教育费用增加、学习时间延长、结业时间推迟等等,有时还会造成学习者之间关系紧张,打击学习者的自信心。第三,由于班级学习者水平不齐,给语言教师的课堂教学造成很大困难,同时给负责教学管理的工作人员增加了工作量。

目前,分级测试作为语言教学机构都要举行的测试,研究还非常少。这就需要我们首先通过调查了解国内汉语作为第二语言分级测试的设计与实施现状,据此展开相关研究。

2. 相关研究

在中国大陆对外汉语教学界,关于分级测试的研究较少,已有研究主要是关于口语测试题型以及口语测试组织等方面的研究(刘颂浩,2001;李海燕等2003);陈作宏、邓秀均(2005)从分级测试的性质与特点、等级与标准、测试方式以及具体原则等几个方面进行了深入的探讨,该研究对分级测试中的问题做了全面考察。任春艳(2006)利用汉语语法习得顺序,设计了一个简化的分班测试,施测后对测试分数的信度和效度进行了研究。辛平(2007)从一次分级测试的结果入手,对分级结果进行跟踪分析,强调了难度系数的重要作用,同时考察表现性评价的鉴别作用。赵秀娟(2012)考虑到考试时间的限制,针对不同水平的学习者共设计两套水平不同的试卷,尝试采用两套试卷区分不同级别的考生。不过,目前尚没有针对国内各教学机构汉语作为第二语言分级测试进行的调查研究。

3. 研究方法

为了更加全面和深入地了解情况,本文采用点面结合的方式,对几十所各对外汉语教学机构分级测试工作进行问卷调查,辅以对个别教学机构的访谈。为使问卷调查涵盖的问题更全面,并更贴近实际,在调查前,首先针对几所院校了解到一些基本情况,然后进行了问卷设计工作。除教学机构的基本信息外,问卷共分为五个部分,包括分级测试的一般情况、笔试、口试、信度和效度、分级测试的难点及其他,共50个问题。

为了扩大调查范围,我们做了大量工作,包括收集国际汉语教学以及对外汉语教学学术会议的通讯录、查阅各教学机构主页、联系相关研究者、对外汉语教师及学生进行问卷调查等。考虑到绝大部分院校是由院系主任、教研室主任和一部分普通教师来承担分级测试工作,因此,尽可能向承担一定教学安排

及管理工作的教师发出问卷。在整个调查过程中,共向 135 个单位发出附有问卷链接的电子邮件。为保证能收回更多的有效问卷,通过多次邮件往来、长途电话催促教师填写,最后收回的有效问卷为 60 份,这些问卷来自 26 个省市的 60 所教学机构,个别为孔子学院。根据回收的数据,本文介绍问卷调查得到的部分结果,之后介绍对个别教学机构访谈的情况,最后对我国对外汉语分级测试设计与实施情况做出总结。

4. 调查结果与讨论

4.1 分级测试总体情况

分级测试总体情况涉及多个方面。以下将从分级测试形式和分级依据两个方面进行说明。

4.1.1 关于分级测试的形式

关于国内的分级测试采用何种形式,问卷中设置了 2 个问题。即口试和笔试的分配情况。

表 1: 分级测试形式

选项	小计	比例
A 只有口试	6	10.00%
B 只有笔试	10	16.67%
C 口试+笔试	41	68.33%
D 其他,请注明	3	5.00%
本题有效填写人次	60	

如表 1 所示,绝大部分院校认为,口试与笔试测试的是不同技能,应分别进行。对只考口试和只考笔试的学校进行考查,发现大部分招生规模较小。如果口试和笔试都要进行测试将涉及到考试的先后次序,以及口试与笔试是否应该分别单独进行的问题。

表 2: 分级测试组织

选项	小计	比例
A 根据口试成绩分组笔试	4	9.09%
B 根据笔试成绩分组口试	6	13.64%
C 口、笔试单独进行	32	72.73%
D 其他	2	4.55%
本题有效填写人次	44	

在进行问卷调查之前，我们在对几所院校进行的访谈中发现，在实际操作中有几种形式客观存在：第一种方式是先口试，再根据口试情况将考生分到不同等级的小组中进行笔试。采用这种方式，口语测试教师可根据学生水平迅速调整问题，灵活且效率较高，而且笔试内容设计考虑到不同水平考生，针对性更强。但这种模式也有其不方便之处。首先，需要准备多个等级的笔试试卷，无法将不同水平的学生放到一张量表上进行区分。其次，增加了考试的组织工作量。教师需要首先确定考生大概的等级，如果学生没有通过某个等级试卷的考试，可能需要重新选择一个等级的试卷，再施测一次。第三，过于依赖口试教师的经验。学生的能力发展不均衡，来自欧美地区的考生听说能力强，但认读和写汉字能力比较差，来自汉字圈的学生听说能力差而阅读写作能力强，先进行口试很可能使得欧美学生分到了高级组，而亚洲地区的学生被分到初级组。为避免这种情况出现，可能需要更多地依赖口试教师的经验才能正确划分。

第二种方式，先笔试，再根据笔试成绩将考生分到不同等级中进行口试。采用这种方式，笔试可用同一份试卷，所有的学生都依据其笔试成绩被放到一张量表上，便于比较。在进行口语测试时，教师因为有笔试成绩作为依据，可以方便地挑选适合其水平的口语试题提问，提高了效率。但是，这种模式首先对笔试试卷的设计要求较高，题目水平需覆盖各个等级，且需要对试卷分数线进行精确划分。其次，虽然考试的组织工作量比第一种方式小一些，但对于笔试阅卷的要求较高，必须在非常短的时间内将所有成绩评阅完毕，并进行排列分组，工作强度大。第三，口语教师的评分也可能会受到笔试成绩的干扰。

绝大部分教学机构采用口试和笔试单独进行的方式，这样做可以保证口试和笔试成绩互不干扰。分班时可结合两者考虑，将考生以更客观的方式分配到不同级别。但是，这种方式对口试考官的压力比较大，对考官的经验要求比较高。如果口语试题设计有问题或口试考官经验不足，对中高级水平的学习者可能只能做到粗略区分，无法做到细分。此外，同第二种方式一样，这种方式对笔试试卷要求也较高。

4.1.2 将学习者最后进行分级的依据

Hawaii (Manoa) 大学的 Brown、Hudson & Clark 三位研究者在 2004 年共同进行的分级测试调查中发现，169 家教学机构中，有 13 个教学项目完全依据分级测验分数对学生进行分级。中国的情况如何呢？

表 3：分级依据

选项	小计	比例
A 仅以分级测验分数为依据	9	15%
B 学过的教材	32	53.33%
C 学习时长	41	68.33%
D 是否在家中说普通话	5	8.33%

E 上学期期末成绩	12	20%
F HSK 成绩	41	68.33%
G 入学申请材料	24	40%
H 学习者的自评	15	25%
I 教师对学习者的评价(学习者努力程度等)	26	43.33%
J 其他信息, 详细	1	1.67%
本题有效填写人次		60

从表中可见, 分级依据有很多。如表 3 所示, 根据 HSK 成绩、学习时长和学过的教材进行分级是被调查院校中划分等级的三个主要依据。同时, 另有根据教师对学习者的评价(即学习者的努力程度等)、入学申请材料、学习者自评、上学期期末成绩、分级测验分数、是否在家中说普通话等项目作为分级依据。其中, 有 9 所院校仅以分级测验分数为依据, 其他各教学机构不同程度上参考了学习者的各种信息。我们前面也曾提到, 如果分级是以其他信息为主, 却把分级测试的成绩作为辅助参考的信息, 可能会失去设置分级测试的必要。在没有其他信息可以参考的情况下, 分级测试成绩应该是分级决策最重要和最可靠的依据。

4.2 关于笔试

问卷中, 笔试部分的问题涉及笔试的多个方面, 本文从设计依据与形式、客观题、主观题等几个方面来进行说明。问卷调查结果显示, 有 54 家设置笔试的教学机构回答了这部分问题。

4.2.1 笔试试卷的设计依据与形式

在美国, 有的教学机构选择商业化测试, 有的自己设计分级测试。由于国内目前没有出现商业化的汉语分级测试, 因此各教学机构使用的还是自己设计的分级测试。那么量身定制的分级测试, 其命题依据应该是怎样的? 应该采用何种题型? 调查中我们发现各教学机构的笔试试卷设计依据也是多种多样。

表 4: 笔试设计依据

选项	小计	比例
A 根据所用教材	36	66.67%
B 凭借教师经验	25	46.30%
C 依据相关部门公布的教学大纲	24	44.44%
D 参照其他学校的分班测试题	10	18.52%
E 其他, 请说明	1	1.85%
本题有效填写人次		54

选择其他一项的，只有一家教学机构，注明是以 HSK 试卷作为分级测试题。参考其他院校分级测试题目的院校接近五分之一，这一现象应引起我们注意。此外，我们认为新汉语水平考试（HSK）成绩可以作为最后分级决策的参考之一，但其可能并不适合直接拿来用作分级测试试卷。这是因为：

首先，虽然汉语水平考试是面向外国人的汉语作为外语的水平测试，但它分为多个等级，需要首先让考生选择参加某一级别的考试，或者告知考生考哪一级别，这一决定可能需要花费时间；其次，水平考试的目的与分级教学的目的并不相同，未考虑到具体的教学和教材情况。第三，新 HSK 多个等级的试卷并不在一张量表上，如果考生没有通过 HSK 五级考试且分数较低，我们要花费时间判断考生适合四级还是三级，从而增加了分级测试的组织工作量。第四，一般教学机构可以得到的新 HSK 的试卷均为公开发行的试卷，有可能一些国家的学习者已经做过这样的试卷而在分级测试中占优，造成最后分级结果不尽如人意（如果 HSK 试卷经过一定的改造并经过效度验证，则不属于直接使用范畴）。

表 5：笔试题型

选项	小计	比例
A 只有主观题（如作文、翻译、回答问题等需要教师主观判断的题目）	0	0.00%
B 只有客观题（如选择题）	6	11.11%
C 主观题加客观题	48	88.89%
本题有效填写人次		54

绝大部分教学机构的笔试试卷采用了主观题加客观题的形式。这些数字表明，无论是否有语言测试理论作为指导，大部分院校的分级测试能够在一定程度上保证测试分数的信度及效度。限于问卷篇幅，问卷中没有调查主观题和客观题的比例。

4.2.2 笔试的客观题

既然所有教学机构在分级测试中都选用客观题这一形式，那么客观题考查哪些内容？客观题是如何进行评阅的？

表 6：笔试客观题考查内容

选项	小计	比例
A 听力	29	53.70%
B 语法	51	94.44%
C 词汇	52	96.30%

D 阅读	51	94.44%
E 汉字	48	88.89%
F 语音	29	53.70%
G 其他题型, 请说明	1	1.85%
本题有效填写人次		54

客观题考查的内容从高到低依次为: 词汇、语法和阅读、汉字、听力和语音。前四类知识或技能比例非常接近, 说明认可度较高。考查听力及语音的院校占了一半左右。

Brown 等 (2004) 对美国的第二语言分级测试进行调查时, 也有同样的现象, 但顺序有所不同。在 169 所院校中, 设置阅读测试的最多 (115 所, 占 68%), 其次是语法 (109 所, 占 64.5%), 第三位为词汇 (102 所, 60%), 这说明美国的语言教学项目可能更注重学习者的阅读技能训练。其他题型依次为写作 (80 所, 47.34%)、听力 (56 所, 33.14%)、口语 (24.85%)、完成真实任务的能力 (15 所, 8.88%)、文化知识 (1 所, 0.6%) 等。在其他题型中, 写作、口语一般是主观性试题, 听力是可以采用客观题形式考查的, 但是实施起来比较复杂。这可能是中美教师都不约而同地较少采用听力测试的原因之一。

笔试的客观题阅卷情况可见下表:

表 7: 客观题阅卷方式

选项	小计	比例
A 手工阅卷, 手工合分	49	90.74%
B 答题卡+光电阅读或扫描+ 评分软件阅卷	3	5.56%
C 直接在计算机上考试	2	3.70%
D 其他	0	0%
本题有效填写人次		54

如表 7 所示, 国内对外汉语分级测试在阅卷技术方面比较为落后。分级测试中采用客观题, 好处之一是试题抽样比较广泛, 有利于提高考试信度。好处之二在于, 使用计算机阅卷非常便捷, 效率高而错误少。当考生人数较少时, 手工阅卷尚可接受, 当考生人数较多时, 对客观题进行手工阅卷无疑是比较大的工作负担。而且, 手工阅卷意味着后续的一系列题目分析、信度分析都无法进行, 或需要花费大量时间。因此, 对分级测试客观题阅卷方式这一问题, 我们认为有条件的院校可以开发基于计算机的分级测试, 而在中小规模的院校中还需大力提倡和发展答题卡和光电阅读机扫卡方式, 或者答题卡扫描加评分识

别软件结合的方式。现阶段需要向这些院校积极推荐这些阅卷技术。

4.2.3 笔试的主观题

如果笔试试卷采用主观题加客观题的形式,那么各校主观题采用的是何种形式,又是如何评阅的?

表 8: 主观题题型

选项	小计	比例
A 作文	29	53.70%
B 回答问题	39	72.22%
C 完成对话	38	70.37%
D 完成句子	38	70.37%
E 造句	32	59.26%
F 翻译	12	22.22%
G 听写	7	12.96%
H 其他题型, 请注明	4	7.41%
本题有效填写人次	54	

如表 8 所示, 主观题型包括: 作文、回答问题、完成对话、完成句子、造句、翻译和听写等。其中回答问题、完成对话和完成句子是被调查学校主要采用的主观题形式。可见, 各教学机构采用的主观题题型很多。其中, 进行作文考试的教学机构占一半, 回答问题、完成对话和句子以及造句是最受青睐的主观题测试方式。

表 9: 主观题评分人数

选项	小计	比例
A 一名教师评分	41	75.93%
B 双人评分取均分	9	16.67%
C 其他	4	7.41%
本题有效填写人次	54	

大部分教学机构分级测试主观题只由一人评分这种现象可能是由题型以及可用的阅卷时间决定的。回答问题、完成对话这样的题目有确定的答案范围或者回答要点, 虽然是主观题, 一人评分在一定程度上仍可以保证信度。但如果采用作文这种题型, 双人评分就更为公平, 也会减少误差。曾经有过这样的先例: 在分级测试的作文评分中, 对于同一篇满分为 20 分的作文, 两名教师给出的分数相差了 8 分, 相当于对学生能力的评价至少有两个等级的差异。

问卷统计数据还显示,有 40 所院校依据一定的成文标准进行主观题评分,而 10 所院校则没有成文的评分标准,仅凭教师的经验进行评分。

4.2.4 关于口试

问卷中口试部分涉及口试的设计、评分等多个方面。有些教学机构未设口试,因此填写的教学机构有 51 家。

4.2.5 口试的命题依据和形式

口试命题依据应该是怎样的?应该采用何种题型?

表 10: 口试题依据

选项	小计	比例
A 根据所用教材	33	64.71%
B 凭借教师经验	34	66.67%
C 依据相关部门公布的教学大纲	22	43.14%
D 参照其他学校的分班测试题	8	15.69%
E 其他,请说明	0	0%
本题有效填写人次		51

与笔试相比,口试试题依据教材的比例略有降低,凭借教师经验的比例则大幅增加,占到三分之二。口试命题依据教学大纲的学校数量与笔试相同,而参照其他学校分级测试题的院校却有所减少。现行的分级口试对口试考官经验的依赖是值得重视的问题。

各教学机构采用的口试的题型有多种,可见下表:

表 11: 口试题题型

选项	小计	比例
A 部分题型是笔试形式	5	9.80%
B 自我介绍	45	88.24%
C 朗读	39	76.47%
D 跟读句子	10	19.61%
E 回答问题	44	86.27%
F 认读不同等级的词汇	16	31.37%
G 看图说话	21	41.18%
H 叙述事情的经过	12	23.53%
I 话题(说明或议论)	17	33.33%
J 讨论	7	13.73%

K 听或读短文后复述	9	17.65%
L 角色扮演	3	5.88%
M 其他题型, 请说明	1	1.96%
本题有效填写人次	51	

口试使用的题型在一定程度上表明了实施分级测试的教学机构对口试题型的认可度。值得注意的是, 有几家教学机构采用间接测试的方式进行口试。

4.2.6 口试的评分

作为典型的主观性测试, 分级测试的口试一般都采用现场提问、现场评分的做法。我们比较关心口试的评分标准、考官对评分标准的把握、评分人数等问题。根据统计结果, 三分之二的教学机构有成文的口语评分标准。没有成文的标准, 可能使得口语测试的主观评分具有一定的主观随意性, 信度和效度不能得到很好的保证; 即使有成文的口语评分标准, 在实际施测时是否也不能够真正被采用: 我们从无经验口试考官是否需要经过培训这一角度来调查发现, 对无经验的口语考官不进行正式培训的教学机构占一半, 可能会增加评分中的误差。同样, 在实施口语测试时, 如果由一人评分可能误差会比较大, 尤其对中高级考生。现行各教学机构设置的考官人数见下表:

表 12: 口试评分人数

选项	小计	比例
A 1 人	20	39.22%
B 2 人	21	41.18%
C 3 人	8	15.69%
D 4 人及以上	2	3.92%
本题有效填写人次	51	

在实践操作中, 如果由一人提问且兼顾评分工作, 考官的负担可能会比较重, 评分的主观性可能过强。如果是两人评分, 分别或共同承担评分工作可能会减轻负担, 同时增加评分的客观性。若考官是 3 人或者以上, 虽然增加了评分客观性, 但口试成本也会增加。同时, 考生看到多位考官会比较紧张, 不利于发挥正常水平。因此, 两名考官进行口试更为合适。

4.3 关于分级测试的信度和效度

关于分级测试的信度, 我们从题目的统计指标(区分度、难易度)和试卷的信度来考察; 效度方面, 从分数线划分以及学生的调班率来考察。

4.3.1 关于分级测试题目区分度、难易度以及分数的信度

科学的分级测试至少应该依据经典测量理论计算题目难易度、区分度和测试分数的信度，并对效度进行研究。

表 13: 分级测试描述统计指标计算

选项	小计	比例
A 一般不计算	42	70%
B 有时计算	15	25%
C 每次都计算	3	5%
本题有效填写人次	60	

根据调查发现，有 15 家教学机构有时计算这些指标，占 25%。这个数字比我们预想的高出不少。但是，对这 15 家教学机构的回答数据再进一步调查，发现仅有 5 所有学习过语言测试课程的硕士或博士参与分级测试工作，且他们的客观题阅卷方式为手工阅卷。说明这 15 家教学机构未按照实际情况回答这个问题，或者是采用了手工阅卷并计算难度和区分度的方式。

以上调查结果说明，仅有 5%的教学机构对分级测试的科学性进行分析和调查。这个情况与美国 2004 年的调查相仿：在 169 家教学机构中，仅有 18 家声称对考试数据进行分析。经分析，我们认为造成这种情况的原因可能有以下几个方面：其一，分级测试风险并不是很高，其重要性尚未得到充分认识；其二，曾在硕士或博士阶段学习过语言测试课程的都是年轻教师，在分级测试工作中均尊重院校原有的约定俗成的做法；其三，专门从事语言测试的专业人士太少，因此对考试数据进行统计分析就成为教学机构的一项比较困难的工作。此外，汉语教学指导机构尚未对本校的分级测试和成绩测试做出指导性建议或规定。

4.3.2 关于分级测试分数线的划分

不管分级测试试卷如何评阅，也不管是否进行试题的分析，教学机构总要对分级测试的成绩划定分数线，并依据一定规则做出决策，将学习者分到合适的班级。那么分数线划分的方法各教学机构是如何确定的呢？

表 14: 分级测试分数线划分

选项	小计	比例
A 每次根据考试成绩百分比调整分数线 (如前 10%进 E 班，后 10%进 A 班)	27	45.00%
B 根据一定的统计方法 (如安哥夫方法、 边缘组法、对照组法、书签法等)	2	3.33%

C 每次考试分数线都相同, 个别微调	16	26.67%
D 先按照笔试成绩分组, 再根据口试成绩微调	18	30.00%
E 先按照口试成绩分组, 再按照笔试成绩微调	6	10.00%
F 笔试和口试各自有单独的分数线	14	23.33%
G 口试和笔试加起来确定分数线	13	21.67%
H 笔试的各个题型都有分数线	4	6.67%
I 其他, 请说明	4	6.67%
本题有效填写人次	60	

如果某家教学机构的学习者来源稳定, 每年的学员人数相似, 且每学期从各院校来的学生水平相差不大, 采用常模参照方法更为方便。但是, 一般来说各教学机构每学期的生源情况是不确定的。有时候学习者整体水平偏高, 而有时则初级学习者多一些。这时, 按照常模参照的解释来确定分数线就会产生问题。因此我们认为, 如果只有新生参加考试, 且试卷每年不变, 那么分级测试分数线应保持固定不变, 针对个别特殊情况可进行微调。

调出各教学机构的问卷数据查看时, 我们发现一些教学机构在新生老生一起参加考试、且使用同一套分级试卷的情况下, 每年的分数线维持不变。这种做法可能会造成老生在分级测试中占优势, 被分到等级较高的班级。经验告诉我们, 如果第一次到中国的新生和老生取得同样的分数, 那么很可能新生的能力要高于老生。

我们还注意到, 口试和笔试单独设立分数线的院校接近四分之一。对笔试各题型成绩都参考的仅有 4 家教学机构。这与 Weber (1985) 提出的要参考各种变量做出分级决策的建议相比, 还有较大的差距。

4.3.3 关于效度

分级测试完成后, 我们最关心的是学习者的所在班级是否合适, 这也就是分级测试分数使用的效度。统计结果表明, 51.67%的教学机构一般不调查, 而 33.33%的教学机构有时调查, 每次都调查的只有 15%。

表 15: 分级测试调班率

选项	小计	比例
A 5%以下	23	38.33%
B 5-10%	24	40.00%
C 11-20%	5	8.33%
D 21-30%	0	0.00%
E 31-40%	0	0.00%
F 40%以上	1	1.67%

G 不清楚	5	8.33%
H 其他, 请注明	2	0.03%
本题有效填写人次		60

关于分级测试的调班率, 仅有一所学校承认调班率在 40% 以上。经查, 该校留学生人数很少, 即使调班人数很少都会造成调班率的大幅波动。对调班率不清楚的有 5 家教学机构 (8.33%)。另外有一名被调查者反映低年级的学习者调班多, 而高年级的调班少。这说明了初级班的分级更应精确, 而高级班弹性相对较大一些。中高级班中, 即使分得比较粗, 对每位学习者来说, 都有需要学习提高的内容。

4.4 分级测试中的困难

分级测试工作较为复杂, 常常面临一些困难。这些困难主要是:

表 16: 分级测试中的主要困难

选项	小计	比例
A 缺乏语言测试专业人员协助	33	55.00%
B 题目分析 (难度、区分度或信度计算)	32	53.33%
C 制定内容大纲	11	18.33%
D 题目的难度及等级控制	36	60.00%
E 控制题目和试卷的质量	18	30.00%
F 制定主观题的评分标准	20	33.33%
G 分数和等级的划分	19	31.67%
H 调班管理	19	31.67%
I 其他, 请说明	4	6.67%
本题有效填写人次		60

选择了其他选项的教师中, 有一名被调查者注明分级测试“耗时耗力”, 道明了参加分级测试教师的心声。配合其他问题的回答情况, 我们可以了解到, 现在的对外汉语分级测试确实存在的很多困难。

4.5 教学机构访谈

笔者在发放问卷的同时, 分别对两所教学机构 A、B 进行了访谈。访谈为无结构式。

A 校的语言生分为 8 个级别。该校采用口试和笔试相结合的方式, 口试和笔试单独进行, 有各自的分数线。该校的分级测试工作比较正规, 但是也存在

一些问题。例如，笔试的试卷在两个学期使用同一份试卷的同时，鼓励老生可以参加分级测试，以便决定他们是否可以跳级。A校的招生规模不小，但是笔试阅卷时，仍然采用手工阅卷、手工合分的方式。笔试主观题和口试有成文的评分标准，且会对评分员进行一定的培训，但是不管笔试还是口试，评分人仅有一人。此外，该校对分级测试统计数据不作计算，也不对分级测试结果进行调查。后来经了解，A校的情况反映了很多高校的分级测试实际状况。

B校的语言生初级分快慢班，以学生的学习能力和水平确定。慢班与快班使用同样的教材，但进度相对慢一点。该校分级测试操作模式一般是这样的：每学期开学两天报到，新生报到后参加分级测试，笔试和口试同时进行。笔试有两套依据水平大纲设计的试卷，每套试卷各分四级，按水平由低到高有四份，学生可以根据自己的水平来决定做几份，每份试卷达到60%就可以升一级。一般一天用一套试卷。口试则没有试卷，教师根据与学生的谈话判断其水平。最后将笔试和口试成绩结合，确定学生的等级。考试的时候采取学生来一个考一个的方式。这种题量很大的分级测试模式可以在一定程度上保证分级的效率，但也可以想象，该校教师的工作量也是非常大的。

从其他教学机构了解到的情况看，各个学校在分级测试工作中都存在一定的问題，而且费时费力是普遍存在的现象。

5. 结论与启示

根据以上调查和访谈，我们发现国内对外汉语教学机构的分级测试方式呈现多样化的情况。大部分分级测试能够保证分级教学的顺利进行，还有的学校采用了基于计算机的考试方式。但其中也存在很多问題，主要有以下几个：

首先，各教学机构缺乏从事语言测试或学习过语言测试的专业人员协助；

第二，试卷的设计方面，各教学机构的分级测试中采用了各种题型，但何种题型应该受到重视，还缺乏这方面的实证研究或数据；

第三，绝大部分教学机构的分级测试题目不计算难度、区分度、试卷的信度，也不进行效度验证工作，导致分级测试的科学性难以得到保证；

第四，试卷的使用方面，很多教学机构连续两个学期使用同一份分级测试笔试或口试试卷，使在该校学习的老生占了优势；多家教学机构使用多个等级的试卷，等于使用多把标有不同单位刻度的量尺测量学生；此外，学生选择考哪一等级的试卷需要花费时间，同时也给分级测试增加了工作负担；

第五，分数线的划分方面，分级测试的分数解释应偏向标准参照而非非常模参照，但几乎一半的教学机构按照常模参照的做法来划分分数线，只有极少数的教学机构按照标准参照的分数解释方法划分分数线，结合分级测试各题型分数确定分数线的教学机构也非常少。另外，很多教学机构做分级决定时考虑了其他因素，但这些因素哪些更为重要可能还未达成一致共识；

第六，测试的阅卷方面，笔试试卷客观题阅卷方法陈旧，效率低；很多学

校的笔试主观题以及口试题的阅卷缺乏标准,不重视对阅卷教师的培训,缺乏客观性。

出现上述问题的原因是多方面的。其中最主要的原因是:首先,国内分级测试的研究刚刚起步,教学机构可参照的研究或相关资料非常少;其次,即使有相关的资料,无论在外语教学界还是在对外汉语教学界,一些语言教学院系是在90年代以后才渐渐开始开设语言测试课程,语言测试研究者和工作者为数不多,无法保证这些资料能够被正确使用;第三,由于语言测试的特殊性,即使教学机构有学习过语言测试的教师,也难以保证科学语言测试的观念和做法完全贯彻到分级测试工作中。

分级测试的现状反映出对外汉语教学界对语言测试与评估这一学科重视程度不够。因此,唯有对汉语教师、国际汉语教育硕士加强语言测试与评估知识培训,加大培养语言测试专业人才的力度,才能从根本上改进分级测试中存在的问题。

参考文献

- 刘润清、韩宝成, 2000, 语言测试和它的研究方法(修订版)[M]。北京: 外语教学与研究出版社。
- 陈作宏、邓秀均, 2005, 外国留学生汉语进修班分班测试初探[J]。云南师范大学学报(对外汉语教学与研究版), (5): 32-37。
- 李海燕、蔡云凌、刘颂浩, 2003, 口语分班测试题型研究[J]。世界汉语教学, (4): 79-90。
- 刘颂浩, 2001, 口语测试的组织与实施探索[J]。暨南大学华文学院学报, (3): 15-23, 86-92。
- 任春艳, 2007, 关于简化分班测试的实验研究[J]。语言教学与研究, (6): 45-50。
- 辛平, 2007, 安置性测试的跟踪研究[J]。汉语学习, (6): 76-81。
- 赵秀娟, 2012, 来华留学生分班测试的笔试试卷建构研究[J]。语言文字应用, (1): 117-124。
- Brown, J.D., T. Hudson & M. Clark, 2004. *Issues in Placement Survey*. Honolulu: University of Hawaii, National Foreign Language Resources Center. Retrived from <http://nflrc.Hawaii.edu/NetWorks/NW40.pdf>.
- Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley & T. McNamara, 2002. *Dictionary of Language Testing*. Beijing: Foreign language Teaching and Rearch Press & London: Cambridge University Press.

Hughes, A., 2000. *Testing for Language Teachers*. Beijing: Foreign language Teaching and Research Press & People Education Press (London: Cambridge University Press).

Weber, J., 1985. Assessment and Placement: A Review of the Research, *Community College Review*, 13, 21-32.