

GONGWUYUANLIYONGKAO SHIDEXINDUHEXIAODUYANJIU

陈社育 著

公务员录用考试的 信度和效度研究



南京大学出版社

陈杜育 著

公务员录用考试的 信度和效度研究

国家公务员录用考试命题研究暨命题指导专家权威解读与考前预测试卷



南京大学出版社

图书在版编目(CIP)数据

公务员录用考试的信度和效度研究 / 陈杜育著. — 南京: 南京大学出版社, 2012. 7

ISBN 978 - 7 - 305 - 10188 - 5

I. ①公… II. ①陈… III. ①公务员—考试制度—研究—中国 IV. ①D630. 3

中国版本图书馆 CIP 数据核字(2012)第 150272 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号 邮 编 210093

网 址 <http://www.NjupCo.com>

出 版 人 左 健

书 名 公务员录用考试的信度和效度研究

著 者 陈杜育

责任编辑 陈济平 编辑热线 025 - 83686531

照 排 南京紫藤制版印务中心

印 刷 盐城市华光印刷厂

开 本 880×1230 1/32 印张 6.125 字数 154 千

版 次 2012 年 7 月第 1 版 2012 年 7 月第 1 次印刷

ISBN 978 - 7 - 305 - 10188 - 5

定 价 22.00 元

发行热线 025 - 83594756

电子邮箱 Press@NjupCo.com

Sales@NjupCo.com(市场部)

* 版权所有, 侵权必究

* 凡购买南大版图书, 如有印装质量问题, 请与所购
图书销售部门联系调换

前 言

随着国家公务员制度的全面推行,近几年来,我国公务员录用考试发展迅速,考试规模与影响越来越大,不仅在公务员系统建立了“凡进必考”的进入机制,而且许多企事业单位也积极学习公务员录用考试经验,通过考试从社会环境吸取“新鲜血液”,实现了人员更新和素质更新双重功能,保持了组织的生机和活力。按照1994年国家人事部颁布的《国家公务员录用暂行规定》,考虑到近几年来各级党政机关正在进行机构改革,公务员队伍的补充主要从专业对口的应届毕业生中考试录用,因此目前我国大多数省市国家公务员录用考试的笔试科目主要是公共基础知识考试和行政职业能力倾向测验,面试方法主要是易于规范操作的结构化面试。这种由多种测评方法组合起来的人事选拔体系,在全面量才、择优录用方面起着重要的作用,目的在于确保新招公务员具有较高的整体素质。

但由于我国公务员考录制度起步较晚,发展速度又较快,缺乏足够的经验,所以在考试的某些方面存在着不足之处。诸如,考试体系上的模式化、某些试题的质量较差、面试评分中的主观随意性,再加上近几年社会应试力量的影响,致使考生平均成绩逐年提高等等一系列问题。这些不足严重阻碍了公务员制度的进一步推广,使其难以充分发挥应有的作用。为了促进国家公务员录用考试工作的健康发展,增强考录工作的后劲,提高考录工作的科学化水平,需要对目前国家公务员考试工作的可靠性和有效性有一个全面准确的了解。作为江苏省“国家公务员考试研究”课题组成部

分的“国家公务员录用考试的信度和效度研究”，将针对目前国家公务员录用考试中存在的实际问题，在调查研究的基础上，通过信度和效度分析，力求发现问题的结症所在，进而寻找解决问题的答案。此外，本项研究既是一种应用性研究，又是测评技术的基础理论性研究。作为基础理论研究，本项研究同时还具有两个方面的意义：①从人才资源开发效率上看，我国目前迫切需要建立一套科学、高效的人才选拔测评技术体系。作为多种测评技术组合使用的公务员录用考试，其信度和效度研究的结果，必将能够为现代人事选拔测评技术的改造、创新和合理使用提供理论上的支持。②从现代心理学科发展来看，心理测量要实现从客观测量向科学评价的转变，必须在心理测评技术上有所创新，有所突破。公务员录用考试的信度和效度研究不仅提供了一次应用现代心理测量理论研究实际测评问题的机会，同时也从实证层面上为心理测评方法的引用和创新提供了可资利用的信息。

在人事选拔方面，各种测评技术方法的信度和效度一直是人事选拔研究工作所关心的根本问题。从20世纪40年代至今，尽管该方面研究的难度比较大，但世界各国在人事选拔工作中也断断续续地积累了一些研究结果，这些研究成果为人事选拔工作的改进和完善做出了不可磨灭的贡献。总结起来具有这几个方面的特点：①表面效度研究较多，各种测评技术方法内在机制的探讨不足。纵观国内外的各项同类研究，不难发现，虽然有各种测评技术的预测效度的陈述，但很少有对造成预测效度高低的内在因素的探讨。②国内外关于人事选拔测评技术信度和效度研究的文献较多，但结果存在很大的差异。如面谈法，信度在0.15~0.80，效度在0.25~0.80不等。研究结论不一致的主要原因可能有两个方面：一方面，面试研究多采用现场研究，变量的控制较难；另一方面，效标选取较为困难，各自所选的标准不太一样，因此结论也有很大的差异。③在人事选拔测评技术的信度研究上，各国研究

大多依据的测量理论是经典真分数理论。由于经典真分数理论只考虑单一笼统的误差,虽然设计了多种估计测验信度的方法,但实际上各种不同方法所估计的信度系数只能反映在不同测量条件下所形成的性质并不相同的测验误差,而不能将多种测试条件引起的信度变化反映出来。④我国在人事选拔测评技术研究工作的进展上,与西方相比差距很大。虽然国内的学术界已经意识到人事测评技术开发的重要性,但由于研究条件的限制,技术开发部门和技术使用部门相对独立,缺乏有效的合作,致使技术应用目前仍然处于较低层面上运作。可以说,人事测评技术研究上的不足,是造成我国人事选拔测评工作发展缓慢的主要原因。

针对当今人事选拔测评技术研究上的不足之处,本次研究借助现代心理测量理论,采用科学规范的实证研究方法,对公务员录用考试的信度和效度进行较为系统的研究,为公务员录用考试的深化改革提供较为全面的信息。具体表现在以下几个方面:

在研究的理论依据上,本次研究中各种考试技术的信度研究主要采用的是当今三大测量理论中的概化理论(Generalizability Theory)。与经典真分数理论相比较,概化理论具有多方面的优越性。尤其在测验误差的分析上,概化理论一方面利用方差分析技术,把测验变异分解成多个部分,每个部分对应于特定的误差来源,从而更便于测量误差的控制;另一方面又比较注重测验的整体设计,不仅能够对各种测量条件下引起的信度变化分别予以考察,而且能够将多种测量条件共同引起的信度变化反映出来,使研究结果更为全面,更加符合实际情况。

在研究题本选取方面,本次研究选用的题本是2000年江苏省国家公务员录用考试中使用的《公共基础知识考试》题本、《行政职业能力倾向测验》题本和《结构化面试》题本。这些题本在结构和内容上与中央国家机关招考公务员采用的考试题本基本一致,与全国大多数省市公务员录用考试中采用的题本也很相似。因此,

根据这些题本得到的研究结果不仅对江苏省公务员录用考试的改革具有明显的指导价值,对全国公务员录用考试改革也具有一定的借鉴意义。

在研究数据采集方面,本次研究主要采集两个方面的数据资料,一是用于信度分析、内容效度分析和构想效度分析的数据;二是用于效标关联效度中的预测效度分析的数据。为了得到有效的分析结论,本次研究在数据采集上具有三个方面的特点:① 真实性,即所有的样本数据均是来源于公务员录用考试的实际取样,并非模拟考试或实验考试结果;② 大样本,即笔试的两门考试答卷的随机抽样数目均在 1 000 份以上,抽取的面试考生也在 100 人以上,即使抽取新任公务员的考核人数也有 50 人,完全满足统计学的大样本的要求;③ 权威性,在预测效标获取方面,为了对抽取的 50 名新任公务员一年后的工作表现作出科学客观的评价,江苏省人事厅专门下发了《关于做好全省新录用公务员跟踪考核工作的通知》(苏人通[2000]206 号),由扬州市人事局牵头,成立了考核领导小组,通过多种途径从多个方面对新任公务员进行了深入细致的考核,确保考核结果科学、客观,符合考核对象的实际情况。由于本次研究在数据采集上严格把关,则本次研究结果更具代表性、客观性和可靠性。

在数据处理方面,本次研究应用了数种大型统计软件和概化分析 GENOVA 专用软件,采用基础统计分析、相关统计分析、多元因子分析、差异统计检验(t 检验)、概化分析等多种研究方法,对数据进行了系统的处理,对公务员录用考试中常用的三种测评技术的信度和效度进行了较为深入的探讨。本次研究得到如下几点结论:

1. 信度研究方面

(1) 本项研究表明,在目前的江苏省国家公务员录用考试中,公共基础知识考试的概化系数大体为 0.85 左右,行政职业能力倾

向测验的概化系数大体为 0.84 左右(不包括第一部分),结构化面试的概化系数大体在 0.60~0.85 之间,说明目前江苏省国家公务员录用考试的信度,就整体而言,基本达到要求。结构化面试虽然存在较大的评分误差,但是由 7 位考官从 7 个方面对考生进行面试评分,面试结果的可靠性仍然是较高的。

(2) 在公共基础知识考试中,相比四门考试科目,行政管理科目的试题在相关知识水平的测量上,其测量精度高于政治、语文、法律三门科目的试题。造成测量精度低的原因主要是因为试卷中存在着部分质量较差的试题。

(3) 在行政职业能力倾向测验中,阅读理解、图形分析和表式分析等三个分测验的内部一致性较高,而数学运算、事件排序、类比推理、定义理解、短文理解等五个分测验的内部一致性较低,尤其短文理解部分还有一道试题与分测验之间存在负相关,说明这五个分测验中尚有部分试题在其测量目标上存在异质性,尚有待于改进。

(4) 研究发现,提高公共基础知识考试和行政职业能力倾向测验的信度的途径有二种:① 提高试题的质量,尤其是试题的区分度;② 增加具有一定区分能力的试题量。

(5) 本项研究表明,结构化面试的信度与面试要素评分标准的客观化程度成正比。在政策理论水平、计划组织能力、宣传说服能力等三个面试要素上,各考官的评分结果之间相对比较一致。而在“举止仪表和情绪稳定性”上,考官的意见分歧最大,这是因为举止仪表和情绪稳定性是属于本质不同的两个测评要素,不能合并在一起。在思维能力、应变能力、言语表达能力等三个要素的测量信度上,各面试组之间存在一定的差异,这可能与有的考官小组面试前进行的考官培训不到位有关。

(6) 从面试各效应方差分量的分析上来看,在江苏省国家公务员录用面试中,各面试组的考官在评分基点上差距不大,这与目

前该省面试评分表的设计和对面试考官统一培训有关。在考生能力的测量精度上,研究表明,有的面试小组基本符合要求,有的面试小组则距面试要求还存在一定的距离。研究发现,造成面试精度不高的原因有三个方面:① 考官在评分时受到“趋中效应”的影响,考官在各面试要素评分取值上普遍存在评分幅度偏小的现象,不能拉开分数档次,比较容易受随机误差的影响;② 有的考官在面试评分时在一定程度上仍然受到“晕轮效应”的影响,不能准确区分考生在不同能力要素上的表现;③ 有的面试小组内考官对面试的评分标准缺乏统一的认识,在对考生能力素质的评价上存在较大的差异。

(7) 本次研究还发现,不管用人部门的考官还是非用人部门的考官在面试评分上均不存在考生性别上的差异,社会上对用人单位考官评分时普遍存在性别歧视的说法是缺乏根据的。但在对某些考生能力素质的评分上,用人单位考官与非用人考官之间存在明显差异,虽然这种差异没有达到统计学上的显著性水平,但任其发展下去,将有损于面试在人们心目中的公正性,应引起面试主管部门的重视。

(8) 本项研究表明,在结构化面试中,提高结构化面试的精度有两种途径:一是增加考官人数;二是提高面试评分标准的客观化程度。在不同的面试要素上,为了达到一定的面试信度要求,所需的考官人数可能有所不同,甚至相差较大,这与各面试要素评分标准的客观化程度有关。况且,随着考官人数的增加,面试的概化系数的增长幅度却在减弱,而面试成本却在增加。因此,根据全息理论,在目前的结构化面试中,为了提高面试的信度,提高面试要素评分标准的客观化程度比增加考官人数更具有重要的现实意义。

2. 效度研究方面

(1) 研究表明,2000年江苏省录用国家公务员公共基础知识考试卷的内容效度不够理想。在考试内容方面,政治、法律两科试

题在测试内容方面比较符合考试大纲的要求,重点明确,具有一定的覆盖面。语文学科中实际题量分布与大纲的要求不符,重点倒置。在整份卷中行政管理科目的题量偏少,测试内容方面的重点不够突出。在考试目标方面,试卷中的大多数试题的测评目标是考生对概念、原理和事件的记忆和初步理解,测量目标较低,缺少对考生的较高层次的认知能力的考察。

(2) 行政职业能力倾向测验具有较好的构想效度,通过因子分析得出的三个因子:抽象关系能力、资料分析能力、言语理解能力正是国家公务员素质结构中最基本的三个能力因素。但是本研究也发现,行政职业能力倾向测验中各分量表均具有较好的区分效度,但在会聚效度方面不够理想。“数量关系”“资料分析”部分的会聚效度相对较好,而“判断推理”“言语理解”部分的会聚效度相对较差,会聚效度较差的原因可能是因为该部分测验中存在少量的异质性试题。

(3) 当前公务员录用面试中普遍采用的结构化面试具有较好的构想效度,尤其对政策理论水平、思维能力等认知成分较大的面试要素具有较强的鉴别能力。对于应变能力这种认知和情感成分均有的复合素质也能较为准确地作出评价,说明目前江苏省的面试考官培训工作在这一方面已经产生了良好的效果。但对于宣传说服能力、言语表达能力两要素评分标准的掌握,大多数考官过多地偏重于两个素质的认知方面,而对两个素质的情感成分则不够重视。有的考官在面试评分时,很容易受到考生言语表达的影响,产生“晕轮效应”,这是因为有的考官不能准确地掌握某些面试要素的内涵和相应的外显行为表现所致。但就总体而言,结构化面试评分的区分效度好于会聚效度。从各个面试要素上来看,政策理论水平、计划组织能力上评分的会聚效度相对较好,举止仪表和情绪稳定性的会聚效度最差,说明会聚效度与各面试要素评分标准的客观化程度存在密切的关系。

(4) 公共基础知识考试和行政职业能力倾向测验成绩均存在着考生学历层次上的差异,本科生的行政职业能力倾向测验成绩普遍高于专科生,这一结果基本符合实际情况,也为笔试成绩的同时效度提供了予以支持的证据。

(5) 研究表明,公共基础知识考试的预测效度不够理想,考试成绩仅对录用后实际工作中的政策理论水平和知识面起到一定的预测作用,这可能与试卷中试题测量的认知目标普遍偏低有关。行政职业能力倾向测验成绩对计划能力、理解贯彻能力、观察判断能力、应变能力、工作质量、工作效率等部分工作表现起到一定的预测作用,但对于组织协调能力等含非智力因素较高的复合素质的预测性不够明显。这也暴露了行政职业能力倾向测验作为一种认知测验的局限性。

(6) 结构化面试中政策理论水平、人际合作能力、思维能力、言语表达能力等面试要素的评分对考生未来工作能力和工作质量具有较高的预测效度。面试要素计划组织能力上的评分对考生录用后工作中的计划能力具有较高的预测效度,对组织协调能力则相对弱一点。面试要素应变能力上的评分比较偏重于言语应变能力,不能涵盖整个应变能力的范畴,对未来工作的应变能力预测效度不够高,尚有待于改进。

此外,本次研究在总结以上几方面研究结果的基础上,针对目前国家公务员录用考试中的存在问题,有重点地从四个方面给出了提高录用考试科学性的可行性建议。

(1) 做好国家公务员录用考试的整体设计工作。综合本次研究的各项结论,我们不难发现,目前的国家公务员录用考试,在设计方面存在两个问题:一是,笔试阶段的两门考试——公共基础知识考试和行政职业能力倾向测验是单纯的认知能力测验,对预测考生在未来行政工作中取得成功的可能性来说,目前的笔试成绩只是一种必要条件而非充分条件,从这个意义上来说,目前的笔试

成绩在预测失败方面比预测成功更为准确。二是,目前的国家公务员录用考试在甄选体系上采用的是择优制,即根据录取名额,按考生的笔试成绩由高到低按 1:2 或 1:3 挑选考生参加面试。由于目前报考公务员的人数较多,而录取名额相对较少,笔试阶段淘汰率较高,如果笔试阶段过于重视知识的测验,就很容易使得一些综合素质高而没有很好复习应试的考生因笔试成绩不够高而不能参加面试,在新任公务员的选拔上产生“高分低能”的现象。针对目前公务员录用考试体系中存在的问题,我们认为有两种改进方式,一种方式是在笔试阶段增加综合素质测评,以帮助具有较高综合素质的考生更有希望通过笔试参加面试。另一种方法是对现有的笔试成绩设立最低分数线,实行公务员录用考试笔试成绩资格证书制度,用人单位可对取得资格证书的考生进行面试,然后根据面试成绩择优录用。

(2) 改进命题质量,提高试题的鉴别能力。在笔试阶段,目前影响笔试测量信度的主要因素是试题的质量,把好命题质量关是提高命题科学性的重要举措。对于公共基础知识考试,还要注意增加测评较高认知目标(如应用、分析和综合等)的试题的题量,以提高公共基础知识考试的预测效度。在目前面试的命题中,评分标准的主观性是影响面试信度的主要因素,加强面试评分标准的客观化和操作性的改进,对于提高目前面试工作的科学性具有重要的现实意义。

(3) 加强面试考官培训,提高面试工作的科学化程度。在目前的国家公务员录用考试中,除了后面的体检和考核工作,面试成绩几乎直接决定最终的录用人选,况且有的省、市党政机关从应届研究生中公开招考公务员没有笔试,只有面试,因此面试在目前的整个公务员考录体系中具有举足轻重的地位。面试考官的素质,面试工作操作的规范性,直接影响面试的结果。加强面试考官培训,有利于提高目前面试工作的科学化程度。根据本次研究的结

果,面试考官培训在面试评分方面需要强调这几个方面的内容:

① 在面试要素的理解和各要素外显行为观察等方面加强对考官的指导,可以通过试评、讨论等形式,加强考官对面试评分标准的统一认识,缩小不同考官在同一考生能力素质评价上的差异。

② 要求考官在面试评分时尽量拉开分数档次。在面试评分过程中可以设计一个面试评分平衡表,对所有的考生先在平衡表上打一个初始分,待所有考生面试结束时,通过比较,找出表现最好的考生和表现最差的考生,将两位考生在有关面试要素上的分值拉大,再决定其它表现较为一般的考生的分值。

③ 要求考官在面试评分时,克服“晕轮效应”对面试评分客观性的影响,“晕轮效应”是指考官对考生某一方面看得过重或有深刻的印象,从而误认为该考生其他方面素质的优劣,也都与这一方面的优劣相类似。因此,在面试考官培训时,要求考官在面试评分时,紧扣面试要素的内涵,认真分析和判断相应素质的外显行为,对考生的各项素质作出较为准确、客观的评价。

(4) 加强公务员录用考试研究,确保新招公务员具有较高的整体素质。在目前的公务员录用考试中存在两种情况:① 考试形式比较固定,多年来没有太大的变化,并逐步趋向模式化,这就很容易受到社会应试力量的影响,致使考生的平均成绩逐年提高,不能有效准确地鉴别出考生的能力素质。② 用于测评综合素质的方法比较单一,结构化面试在素质测评方面存在较大的局限性,无法承担起准确测评所有从事行政工作必备的综合素质的重任。这就需要研究和开发新的综合素质测评方法,以满足公务员录用考试择优选才的需要。

目 录

第一部分 导论	1
一、研究的对象、目的和意义	1
1. 研究对象的界定	1
2. 目前国家公务员录用考试中存在的问题	4
3. 研究工作的意义和目标	6
二、国内外相关研究成果综述	8
三、本次研究的理论依据	12
1. 信度分析理论	12
2. 效度分析理论	16
四、本次研究的方法	19
1. 研究题本的结构和内容	19
2. 研究数据资料的采集	22
3. 数据分析工具及其方法的选择	25
第二部分 “公共基础知识考试”的信度和效度分析	26
一、公共基础知识考试的信度分析	26
1. 单面交叉设计的数学模型及其计算公式	26
2. 公共基础知识考试的 G 研究	28
3. 公共基础知识考试的 D 研究	31
二、公共基础知识考试的效度分析	33
1. 内容效度分析	33

2. 效标关联效度分析	36
三、研究小结	38
第三部分 “行政职业能力倾向测验”的信度和效度分析	40
一、行政职业能力倾向测验的信度分析	40
1. 行政职业能力倾向测验的 G 研究	40
2. 行政职业能力倾向测验的 D 研究	46
二、行政职业能力倾向测验的效度分析	49
1. 构想效度分析	50
2. 效标关联效度分析	53
三、研究小结	57
第四部分 “结构化面试”的误差源研究	59
一、结构化面试的 G 研究	59
1. 测量目标的变异分析	61
2. “考官”侧面的变异分析	62
3. “面试要素”侧面的变异分析	63
4. “考生”与“考官”之间交互作用的变异分析	66
5. “考生”与“面试要素”之间交互作用的变异 分析	71
6. “考官”和“面试要素”之间的交互作用的变异 分析	72
7. “考生”“考官”和“面试要素”三者交互作用的 变异分析	74
二、结构化面试的 D 研究	79
1. 两面交叉设计的 D 研究	79
2. 单面(R)随机条件下的 D 研究	87
三、研究小结	92

第五部分 “结构化面试”的效度研究	95
一、结构化面试的构想效度分析	95
1. 因子分析	95
2. 多质-多法分析	98
二、结构化面试的预测效度分析	100
三、研究小结	106
第六部分 研究结论与建议	108
一、本项研究结论	108
1. 在信度研究方面	108
2. 在效度研究方面	110
二、公务员录用考试改革建议	112
附 录	115
附录一 公共基础知识试卷	115
附录二 行政职业能力倾向测验题本	132
附录三 结构化面试题本	157
附录四 关于对新录用公务员进行跟踪考核的实施方案	161
参考文献	171
后 记	178

第一部分 导论

一、研究的对象、目的和意义

1. 研究对象的界定

为政之要，首在择人。世界发达国家在选拔公务员时大多采用考试录用制度，公务员考试录用管理的科学化和法制化是现代政府人事管理的重要支柱。各国公务员制度的建立无一不是以考试录用制度的采用为其根本标志的。自从 19 世纪中期英国政府推行考试录用制度开始，各国竞相效仿，历时一个半世纪，考试录用制经过多次完善和检验，被证明是科学而行之有效的选人制度。从 20 世纪 80 年代初，我国党和政府开始吸收各国的考录经验，对干部人事制度进行改革。1982 年，原劳动人事部在《吸收录用干部问题的若干规定》中首次提出了“考试录用”要求，即国家机关、企事业单位吸收录用干部“实行公开招收，自愿报名，进行德、智、体全面考核，坚持考试，择优录用”的方法。1989 年 1 月，中共中央组织部、国家人事部联合下发了《关于国家行政机关补充工作人员实行考试办法的通知》，要求县级以上国家行政机关补充非领导职务的工作人员，要按照德才兼备的标准，公开考试，严格考核，择优录用。经过几年的实践探索，1993 年 8 月 14 日，国务院正式颁布了《国家公务员暂行条例》。根据暂行条例，人事部于 1994 年 6 月 7 日颁布了《国家公务员录用暂行规定》，它标志着考试录用工作开始进入法制化、规范化的轨道。1994 年 7 月，中央国家机关面