

[瑞典] T. 胡森 [德] T. N. 波斯尔斯韦特 主编

教育 大百科全书

教育研究方法

9

INTERNATIONAL
ENCYCLOPEDIA OF
EDUCATION

西南师范大学出版社 海南出版社

[瑞典] T.胡森 [德] T.N.波斯尔斯韦特 主编

R
G461
11/9

教 育 大百科全书

INTERNATIONAL
ENCYCLOPEDIA OF

西南师范大学出版社 海南出版社

教育大百科全书

第 9 卷

● 教育研究方法

目 录

教育研究方法

• 性向测验

职业兴趣和性向的测量 (Vocational Interests and Aptitudes, Measures of)	3
--	---

• 态度与环境的测量

态度的测量 (Attitudes, Measurement of)	12
测量教育风气的描述性量表 (Descriptive Scales for Measuring Educational Climate)	21
家庭和学校环境的测量 (Family and School Environmental Measures)	27
评分者测量 (Judgments, Measurement of)	32
瑟斯通量表 (Thurstone Scales)	37

• 教育研究中的数据资料与政策

人口普查和全国性数据调查 (Census and National Survey Data)	40
数据库和数据档案 (Data Banks and Data Archives)	43
传记和自传中的教育史 (Educational History in Biographies and Autobiographies)	48
* 教育研究成果的传播 (Dissemination of Educational Research)	(1:399)
教育研究和政策制定 (Educational Research and Policy-making)	52
教育研究的历史 (Educational Research, History of)	59
政策分析 (Policy Analysis)	68
政策取向的研究 (Policy-oriented Research)	74
教育研究的政治学 (Politics of Educational Research)	81
教育研究:性质、需求和优先性 (Research in Education: Nature, Needs, and Priorities)	83

• 测量的模型和方法

适应性测验 (Adaptive Testing)	90
学习方法的本质和测量 (Approaches to Learning: Nature and Measurement of)	95
学校教育中的评价方法 (Assessment in Schools, Methods of)	99
标准参照性测验 (Criterion-referenced Measurement)	106
发展水平的测量 (Developmental Levels, Measurement of)	112
测验的项目样本 (Item Sampling in Testing)	118
项目编写技术 (Item Writing Techniques)	124

潜在特质测量模型(Latent Trait Measurement Models)	129
教育研究中的测量(Measurement in Educational Research)	135
社会背景的测量(Measurement of Social Background)	143
智力模型(Models of Intelligence)	149
部分分数模型(Partial Credit Model)	155
非度量数据的偏序量图分析(Partial Order Scalogram Analyses of Nonmetric Data)	161
案卷、剖析图、记录、报告和证明(Portfolios, Profiles, Records, Reports, and Certification)	169
Q 方法论(Q-Methodology)	175
问卷(Questionnaires)	177
量表法(Scaling Methods)	182
中等教育中的选拔机制(Selection Mechanisms in Secondary Education)	189
• 测量理论	
经典测验理论(Classical Test Theory)	193
概化理论(Generalizability Theory)	200
项目反应理论(Item Response Theory)	205
拉施测验理论(Rasch Measurement, Theory of)	209
教育和心理测量中的信度(Reliability in Educational and Psychological Measurement)	220
库格技术(Repertory Grid Technique)	233
效度(Validity)	236
• 研究方法论	
行动研究(Action Research)	242
传记式研究方法(Biographical Research Methods)	248
案例研究方法(Case Study Methods)	254
对变化的测量(Change, Measurement of)	259
横向研究方法(Cross-sectional Research Methods)	265
事件历史分析(Event History Analyses)	272
实验研究(Experimental Studies)	278
面理论(Facet Theory)	285
女性主义研究方法论(Feminist Research Methodology)	291
解释学(Hermeneutics)	297
教育研究的历史方法(Historical Methods in Educational Research)	304
* 人的发展:研究方法论(Human Development; Research Methodology)	(4:475)
为选拔人员和临床研究进行的面试(Interviewing for Selection and Clinical Research)	310
合法性研究(Legitimatory Research)	314
纵向研究方法(Longitudinal Research Methods)	318
教育研究中的微机(Microcomputers in Educational Research)	329
调查研究中缺失的数据和无反应的被试(Missing Data and Nonresponse in Survey Research)	334
流动表(Mobility Tables)	339
模型和模型建构(Models and Model Building)	342
叙事研究(Narrative Inquiry)	349
教育研究中的口述史(Oral History in Educational Research)	354
参与观察与鉴定(Participant Observation and Verification)	358

现象描述学(Phenomenography)	362
准实验(Quasi-experimentation)	367
教育研究:认识论问题(Research in Education:Epistemological Issues)	369
研究方法论:科学方法(Research Methodology;Scientific Methods)	379
教育研究范式(Research Paradigms in Education)	386
测验研究中的取样(Sampling in Survey Research)	391
* 教育研究中的选择偏好(Selection Bias in Educational Research)	(4:599)
教育研究中的自我报告法(Self-report in Educational Research)	402
教育研究中的符号学(Semiotics in Educational Research)	406
模拟研究技术(Simulation as a Research Technique)	412
单样本研究:测量变化(Single Case Research;Measuring Change)	416
社会测量法(Sociometric Methods)	420
调查研究法(Survey Research Methods)	424
教育研究工作者的培养(Training of Research Workers in Education)	431
教育研究中的三角测量法(Triangulation in Educational Research)	439
双生子研究(Twin Studies)	443
教育研究中的非目的效应(Unintended Effects in Educational Research)	447
教育研究中的录像和录音(Videotape and Audiotape Recording in Educational Research)	452
• 统计分析和数据管理	
数据分析中的贝叶斯统计(Bayesian Statistics in the Analysis of Data)	457
典型分析(Canonical Analysis)	462
聚类分析(Cluster Analysis)	466
概念图(Concept Mapping)	472
分类数据的列联频次分析(Configural Frequency Analysis of Categorized Data)	477
内容和文本分析(Content and Text Analysis)	483
列联表(Contingency Tables)	487
相关分析(Correlational Procedures in Data Analysis)	492
* 数据包分析(Data Envelopment Analysis)	(1:551)
教育中的数据库管理系统(Database Management Systems in Education)	501
教育测验中的决策理论(Decision Theory in Educational Testing)	508
描述性数据的分析(Descriptive Data, Analysis of)	512
判别分析(Discriminant Analysis)	522
教育研究中的调节影响和中介影响(Educational Research, Moderating and Mediating Effects in)	526
教育预测中的期望表(Expectancy Tables in Educational Prediction)	531
探索性数据分析(Exploratory Data Analysis)	533
因子分析(Factor Analysis)	541
因子模型(Factorial Modeling)	551
伽罗瓦格(Galois Lattices)	554
多层线性模型(Hierarchical Linear Models)	560
假设检验(Hypothesis Testing)	567
教育研究中的交互作用及其检测 (Interaction and Detection of its Effects in Educational Research)	572
对数线性模型(Log - Linear Models)	581

变异的测量(Measures of Variation)	589
元分析(Meta-analysis)	599
调查研究中的缺失值(Missing Scores in Survey Research)	608
多级分析(Multilevel Analysis)	611
多元分析(Multivariate Analysis)	619
非参数和自由分布统计(Nonparametric and Distribution-free Statistics)	627
路径分析和线性结构关系分析(Path Analysis and Linear Structural Relations Analysis)	633
潜变量路径分析(Path Analysis with Latent Variables)	644
教育研究中的预测(Prediction in Educational Research)	651
轮廓分析(Profile Analysis)	656
教育统计规划(Projections of Educational Statistics)	657
定量数据的回归分析(Regression Analysis of Quantified Data)	662
稳健统计过程(Robust Statistical Procedures)	671
调查研究中的抽样误差(Sampling Errors in Survey Research)	676
显著性检验(Significance Testing)	683
最小空间分析(Smallest Space Analysis)	690
社交网络分析(Social Network Analysis)	697
方差和协方差分析(Variance and Covariance, Analysis of)	703
• 测验	
成就测验量表(Achievement Test Scales)	708
* 计算机化的教育测试(Computerized Educational Testing)	(7:475)
* 教育成绩:比较研究(Educational Achievement:Comparative Studies)	(6:441)
论文:分数的均衡(Essays;Equating of Marks)	718
论文的计分(Essays,Scoring of)	724
统考(Examinations;Public)	731
外语测验(Foreign Language Testing)	738
课堂中的个别化测验(Individualized Testing in the Classroom)	744
测验和评价中的题库建设(Item Banking in Testing and Assessment)	749
项目偏见(Item Bias)	756
学习潜能及其测验(Learning Potential and Learning Potential Tests)	763
读写、计算和口语能力测验(Literacy,Numeracy,and Oracy Tests)	766
母语测验(Mother Tongue Language Tests)	772
选择题测验中的猜测(Multiple Choice Tests,Guessing in)	778
测量方法中的客观性测验(Objective Tests in Measurement)	782
人格问卷(Personality Inventories)	786
应用数学测验与科学测验(Practical Mathematics and Science Testing)	791
投射测验技术(Projective Testing Techniques)	796
评定量表(Rating Scales)	800
学生的学习进展图(Student Progress,Charting of)	807
测验管理(Test Administration)	814
测验偏差(Test Bias)	819
测验与课程(Testing and the Curriculum)	824
测验和态度量表的翻译(Tests and Attitude Scales,Translation of)	832

测验与考试的辅导 (Tests and Examinations, Coaching for)	837
不同类型的测验 (Tests; Different Types)	843
不同测验形式间的等值 (Tests, Equating of)	852
应试焦虑和成绩期望 (Test-taking Anxiety and Expectancy of Performance)	859
标准参照测验的标准设置 (Standard Setting in Criterion-referenced Testing)	864

* 英文原版以英文字母顺序排列分卷,中文版以专题归类分卷,对于同时可以归于几个专题的词条,用*在相关专题中标出其实际所在的卷数及页码,以便读者检索。

教育研究方法

石中英 审译

职业兴趣和性向的测量 (Vocational Interests and Aptitudes, Measures of)

在西方社会,大多数人的权力都来源于自己的职业,因为一般来说职业是人们金钱和身份的主要象征。在20世纪后期的民主社会里,有一种思潮变得越来越强烈:社会应该为经济落后地区的人和少数民族群体提供平等的机会,特别是获得教育和就业的机会。因此,为了免于让个体或团体在职业性向和兴趣测量方面处于不利的地位,关于职业兴趣和性向的测量及其结果有效和公平的使用引起了人们极大的关注。

虽然人们对如何更好地对那些已经处在不利地位的人拓宽教育和就业的机会还存在争论,但是至少大家认为这种拓宽是可取的,这个问题我们后面还会讨论。在本词条里,我们首先来定义一些关键词,然后讨论职业兴趣和性向(Aptitude)测量的目的、方法以及一些广泛使用的测量工具的特点。本词条对以上各个部分进行简单的回顾,然后介绍这个部分研究的重点,最后是对其进行一个公正的评论。

1. 定义

当“职业”这个单词和其他的名词如“兴趣”、“性向”甚至和“教育”连在一起用时,就差不多是含蓄地指向那些不需要高深的文化知识的课程和工作。例如,在本百科全书的先前版本里,对在教育中使用职业性向测量的解释是“为了挑选学生或帮助学生选择某一个具体的职业,如售货员或经商”(1985 P. 5536)。在某些国家(如荷兰和加拿大的部分地区),职业教育或课程是面对那些在主流的学术教育中失败的那部分学生。《简明牛津词典》对“职业”有两种定义:第一个(来自拉丁语的 *vocare*,“去召唤”)是指“觉得自己适合或非常想从事某个工作或职业”,第二个是指一个人的“工作、行业、专业”。为了与本词条相符,“职业”这个名词被赋予了更为广泛的内涵,包括情感部分,如《简明牛津词典》关于职业的第二个解释,也包括所有工作技能范围的职业,而不仅仅是那些需要较少文化就能做的工作。

有一些作者将“职业”和“工作”作为本质上同义的词使用,但是实际上两者是有所分别的,“职业”是心理学里一个比较具有学术意义的术

语。例如,克赖茨(Crites)认为,职业行为就是对工作领域的刺激的反应,包括个体做出选择或者调整某个工作,制定退休计划等等(1969 P. 16)。舒佩尔(Super)认为,职业“对其心理学和经济学的意义进行区分,是指一个承担某种义务的工作,一种自我卷入,作为一种对个体有意义的活动,不仅仅是指其产品、分配或服务结果及其经济报酬”(1976 P. 10)。

从20世纪的早期起,“兴趣”成为对“工作进行明智选择”的关键理由,在过去很多年里人们用各种方式对兴趣进行了定义:从喜欢或不喜欢方面,从某种行为模式方面,从行为结果方面,从活动方面,从驱动力方面,从自我概念方面,或从人格的某些方面。《简明牛津词典》将兴趣定义为“比较兴奋(关注或好奇)或吸引某人的注意”,偏向给予兴趣以情感和动机方面的解释。最近这些年来认为兴趣是个体人格的深层次方面的东西,并且和个体的可解释的行为高度相关,这成为心理学界近年来对兴趣观点的普遍趋向。特曼(Terman)在对弗赖尔(Fryer 1931)关于兴趣测量的划时代观点的介绍里写道:“一个人的成就的数量和方向大部分是由兴趣的因素决定的。”(Fryer 1931 P. xvii)早期兴趣测量领域的杰出人物斯特朗(Strong)说道:“它们(兴趣)指的是人们愿意做的事情,以及人们对自己认为满意的事情的反应。”(1943 P. 19)霍兰(Holland)总结的一个比较著名的观点是:“(职业兴趣)代表了人们在工作、学校课程、业余爱好、娱乐活动和偏爱中体现出来的人格,简单地说,我们叫作‘职业兴趣’的东西仅仅是人格的另一个方面。”(1985 P. 7~8)

但是在界定“性向”时遇到了更多的问题。在《简明牛津词典》里,性向是指“天赋的倾向和才干”或“获得某种特定技能的能力”;能力则是“(做某事情)的能力”或“机智、才干、心理力量”。这种字典上的循环定义给心理学家的定义带来一定问题,“语义和概念上的不明确和混乱多于统一,使对性向的定义陷入一个困扰的境地”(Crites 1969 P. 26)。有些作者认为能力既包含“一个活动中已经掌握的部分”,也包括“推动学习某种任务的具体活动以及某种不随时间改变的东西”,因而比性向更为一般。在这种观点里,“能力”包括成就,而性向则更类似于某种特征。近期的观点则认为性向测验是关于能力发展的测验,从而在设计和使用目标上与成就测验区分开来(Anastasi 1982)。

虽然人们力图区分“能力”和“性向”,但在一些著作中它们是可以互换的,例如,就像弗莱斯曼和赖利(Fleishman and Reilly 1992)所说的,有些作者认为是一种能力的,另外的作者则更趋向于认为是一种性向。弗莱斯曼和赖利描述了52种假设的能力,分为认知、精神运动、物理和感知觉四类。相反,洪特尔(Hunter)认为“一般认知能力的测量是通过一些具体的性向的测量来进行的”(1986 P. 341),这些具体的性向有些是和弗莱斯曼和赖利列出来的能力相似。在本词条里,将“性向”定义为包括那些能够在预测工作成绩和工作前的学习安排中起到潜在作用的人类能力,不包括文化水平。

2. 测量职业兴趣和性向的目的

为什么心理学家热衷于测量人们的兴趣和能?教育者、顾问或招聘单位对这个问题的观点不同,从而测量职业兴趣和性向的目的也就不同。后者希望尽量扩大工作绩效和效率,而前者希望提供帮助人们选择获得个人满足感的职业信息。雇主倾向于将重点更多地放在适合的能力或性向,而顾问可能对兴趣和适合的能力或性向给予同样的重视而且有可能更加重视兴趣。这两种观点在理论上还是有所联系的——很久以来人们就假设如果工作的要求和环境与个体的兴趣和能相一致的话,个体会对工作更加满意并且做得更好。

3. 背景:测量方法和工具

3.1 兴趣

测量职业兴趣的传统方法是用纸笔工具。原先的兴趣测量是建立在“个体相似性”的基本原理上。即在同一种工作中的人们倾向于具有相同的爱好。根据这个基本原理,个人对那些自己的兴趣与工作中的他人的兴趣相似的工作更为满意。近年来,兴趣测量开始建立在“活动相似性”的基本原理上,这个原理认为假如人们喜欢的活动与工作中的活动相似,那么他们会对这个工作感到满意。因此使用最广泛的纸笔问卷不是包括一系列与工作相关的活动的描述就是一系列的职业名称。回答者根据这些题目来推测自己对工作中的人或工作活动的观点,并根据这些观点来完成问卷。有些工具则根据回答者对某种活动或职业的喜欢或不喜欢的强度来提供测量等级。那些依赖在两个或三个选项中进行强迫选择的工具有不同的使用方法,有的要求对大量的职业名称进行排序,有的使

用李克特量表,有的仅仅使用喜欢—不喜欢或喜欢—无所谓—不喜欢来作为反应项。

兴趣测量的“三巨头”——斯特朗在1927年出版了他的最初的《职业兴趣调查表》,库德(Kuder)在1934年出版他的最初的《兴趣爱好记录》,霍兰1958年首次公开了自己的《职业爱好问卷》,兴趣测量的工具都是建立在这三个先驱者的工作之上的。到现在为止,这三个人提出的理论在职业咨询中依旧占据着统治地位。从20世纪70年代开始这些工具开始持续发展并且出现了一些新的工具,但是后来的这些工具都没有霍兰在60年代的工作具有创新性。在这之前,斯特朗将指定的工作中的人与“一般的人”进行对比来倡导了“经验主义编码”方法。另一方面,库德虽然后来在库德《职业兴趣调查表(KOIS)》中加入了以经验为主的编码量表,但是他是最先开始制定等值性量表的人。斯特朗和库德的问卷都需要大量的调查和大样本来建立他们的量表。计分是非常复杂的,在90年代后基本上需要计算机来进行计分和从大量的职业量表中得到结果。

相反,霍兰的方法则比较简单——“他的爱好是用概念、理论或实践而不是心理测量”(Borgen 1986 P. 85)。霍兰提出这样的理论,“在我们的文化中大多数人(和环境)可以归为以下六类中的一类:现实型、研究型、艺术型、社会型、企业型和传统型”(1985 P. 2),并且联系这些类别制定了测量兴趣的工具。他认为“通过一个人在这六类模型的相似性的顺序,可以分出大约720种不同应对环境的人格类型”(P. 3)。这个理论对职业选择的作用是建立在人们会选择“允许他们锻炼自己的技能和能力……在问题和规则上具有统一意见的环境”(P. 4)的工作的假设之上的。因而,现实型会寻找现实的环境等等。霍兰的职业爱好问卷和自我指导问卷比较容易进行计分和自我计分,因而比较适合对被试给予及时的反馈。

霍兰的六种分类方案具有很高的效度,并且在1972年,坎贝尔将霍兰的6大职业领域引进了《斯特朗职业兴趣问卷》(因而这个问卷叫作斯特朗—坎贝尔兴趣问卷),并且这个量表现在还在使用。另外,许多新的问卷也具有与这个问卷相同或相类似的分类特征,如哈—欧生涯决策系统、约翰松(Johansson)的生涯评估量表、兰布(Lamb)和普雷迪格尔(Prediger)的男女通用美国大学测验兴趣问卷(UNIACT)以及美国职业服务机构使用的兴趣

问卷。霍兰的方案虽然不是被所有的研究数据支持(下面会进行讨论),但是鉴于它的重要作用,许多研究者都将其作为职业评价和咨询的基础(Lowman 1991)。如同博根(Borgen)写的那样,“……因为霍兰方法的简洁和精确,对中心概念的关注,以及这个方法允许有不确定性的因素存在,使通过这种方法做出的研究在职业行为中具有无与伦比的作用”(1986 P. 89)。对霍兰的研究的“正面的”和“负面的”的集中讨论感兴趣的读者可以参见文后的附录中的一些最近的期刊版本。

3.2 性向

在20世纪里,对人类心理能力或性向的分类理论从某种主要的单独能力或一般能力,发展到30年代瑟斯通(Thurstone)提出的几种基本能力因素,到60年代古尔福特(Guilford)的“智力结构模型”里的超过120种更为具体的能力,然后到邓尼特(Dunnette)的综述里总结埃克斯特龙(Ekstrom)的工作时又回到了大约10种能力(1976)。从这时开始,在一些著作里出现了更为深入的观点。从职业心理学的观点来看,90年代以后关于能力的争论不再关注其特征和结构,而是关注它们在预测工作绩效方面的作用。在本词条5.3中会对这些争论进行一些评价。

在20世纪60年代,工业心理学家如弗莱斯曼和他的同事开始对与能力或性向相关的工作成就进行集中研究。这些心理学家关注个体的身体、精神运动和感觉以及认知能力并且从非常具体的层次对这些方面进行分析。他们的主要目标就是首先定义能力,并通过各种能力对任务绩效所起的重要作用来使专家有效划分工作任务。彼得森和鲍恩斯(Peterson and Bownas 1982)认为在这些研究中证实了19种认知或知觉能力和18种精神运动或身体能力。弗莱斯曼和赖利(1992)则将这些能力扩展到52种。

职业性向测验从来都没有与能力的特征和结构的理论发展匹配起来,到目前为止在职业性向测量中占统治地位的仍旧是瑟斯通的“基本心理能力”模型。但也有例外(如在计算机领域,已经建立了一些特定的与性向相联系的测量方法;在文书领域,成套的测量方法已经建立了一段时间——一个最好的例子就是英国的国家教育研究基金会出版的现代职业技能测验),即使在一些作品中对这些问卷的价值有所争论,但是多重性向的成套问卷已经占支配地位并且到现在还得到广泛应用。到

目前为止最著名的成套问卷是美国劳工部建立的一般性向成套测验(GATB)和韦斯曼(Wesman)与贝内特(Bennett)建立的差异性向测验(DAT)。这两个成套问卷在美国之外的国家也都得到广泛应用。军队职业能力成套测验(ASVAB)每年在美国的大量学生中使用。另外还有一些建立在相似的基本原理上的应用稍微少一些的成套问卷,如美国出版的《员工性向问卷》和英国出版的《差异性向成套测验》。虽然上面引用的成套问卷都是为某个范围内的顾客设计的,但是其在主题上是存在差异的。

多重性向的成套问卷一般包括5~9个因素分数。例如,一般性向成套测验,包括言语性向、数字性向、空间性向、一般学习能力(由前面三个测验分数相加得到)、形状知觉、运动协调、文书知觉、手指灵巧以及手的敏捷的因素分数。区分性向测验偏向于在中学中使用而不是在就业选择中使用,这个问卷测量了言语性向、数字性向、一般学习能力(前两个测验分数的和)、抽象推理、文书性向、机械推理以及拼写和语言使用,虽然仍旧包括9个分数但是不包括感觉或运动维度。大多数具体的性向测验和组成这些成套问卷的测验都是纸笔形式的,但是一般性向成套测验有几个测验需要简单的用具,比如心理运动能力或性向的测验显然包括对一些材料如小钉板的操作。洛曼(Lowman 1991 P. 305~306)列出了一些有用的具体的性向测量(他称其为能力测量)。

虽然每个国家,以及国家内每个地区的情况都不一样,对高等水平教育和专业训练的职业兴趣和性向仍旧建立在过去一个或更多领域的成就以及一般的“学术性向”的基础上。出于这种目的,学术性向通常通过言语和定量的测验项目来进行评价,并且测验项目的内容应该尽可能地脱离具体的教育课程。多重性向成套测验认识到了一般认知性向或能力测验在评价某些领域的成就,如“工作”成就的重要性,因而这些成套测验经常将两到三个认知量表整合在一起来测量一般认知性向。

4. 兴趣和性向测量的创新

4.1 兴趣

沃尔什和奥西波夫(Walsh and Osipow 1986)详细地讨论了20世纪80年代中期以来的兴趣测量上的创新,包括通过解释一些与性别以及文化团体、子群体有关的项目的偏见和公平性来修订那些

使用广泛的测量工具。新的测量工具如伦林伯格职业兴趣问卷和兰布与普雷迪格尔的男女通用美国大学测验兴趣问卷都通过仔细的选择项目来缩小性别差异。约翰松对生涯评估量表的修改则突出体现在把重点放在了没有受过大学教育的职业量表上。霍兰则为了适应没有受过大学教育的回答者发布了一个低文化水平的自我指导探测系统。雅克松(Jackson)在70年代末期为发展雅克松职业行为问卷做了非常复杂的心里测量学方面的修改,通过一些具有创造性的过程,雅克松引用了斯特朗著作中的大量的数据库并将其用在了自己的量表中。

另外一些将问卷具体化为图示的和“卡片分类”的创新性的设计是为了提高职业兴趣测量的可接受性和适用性从而为更多的人使用。包括由美国的弗雷博格(Freeberg)和韦特拉(Vitella)在20世纪70年代制定的PAYES职业兴趣问卷的男女共用的图式问卷,PAYES的职业兴趣问卷中的量表模仿罗(Roe 1956)的职业分类设计理论,使用李克特式的回答方法,这是为成绩较差的少年设计的。另外一个80年代中期由贝克尔(Becker)制定的阅读自由职业兴趣问卷,该问卷目的与PAYES职业兴趣问卷相似,这个量表包括一系列由三个工作组成的项目,使用三个紧密联系在一起以工业为基础的职业群(如开汽车,做家务,洗熨衣服),要求必须选择其中的一个。也许这方面最有意义的工作由加拿大泰特罗和特拉汗(Tétreau and Trahan 1988)所制定的视觉兴趣测验(TTVIT),这个测量是建立在霍兰的人格六种分类的基础上的,用的是李克特式的反应方法。这个问卷是通过用幻灯片以一定的顺序和一定的时间放映约100个男女皆宜的工作来执行的。到了90年代,就可以用录像带的形式进行放映。泰特罗和特拉汗提供了这个量表的大量的国家和文化团体的结构效度。

从20世纪80年代早期开始,兴趣测量领域的重要发展趋势就是为咨询者和客户准备越来越容易理解的材料和手册,并不断把重点放在测验使用的适当性方面。职业兴趣测量逐渐的发展不仅为使用者选择和决定工作提供信息,还发展了使用者的自我了解(Hansen 1992)。尽管大量的专门技术在讨论和研究量表“清楚性”方面做出了很大努力,但是有证据表明,相同的客户所做的不同的问卷之间的区分度很小(Gottfredson 1986a)。如同汉

森(Hansen 1992)所说的那样,研究者和工具制定者注意了科莱和莫斯(Cole and Moss 1989)关于测验对个体和社会使用的边缘效应的警告,开始认识到最好将自己的努力放到保证测量工具适当的和公平的使用上。由于使用计算机计分,问卷的选项开始增多,结果报告也变得越来越详细,以及为手工计分问卷准备的解释材料越来越简略,问卷使用者确定测量是否用在适当的和辅助解释结果的责任开始变得越来越重要。

4.2 性向

和兴趣相比,职业性向领域的测量要相对落后。其中一个进步就是在发展一些计算机适应性的性向成套测验和具体的性向测量的版本中引入项目反应模型(IRM)技术。倘若硬件设备已经具备并且已经有充分的研究结果用来确立测验的项目参数,这种类型的测验有可能在最短的时间内对人们的性向进行精确的估计。另外的比较可行的具有创新性的工作我们将在第五部分提到。

上面提到对伴随着兴趣测量应用中出现的详细的解释材料的评论同样适用于性向测验,特别是当性向测验作为全面的职业生涯探索系统的一部分时。这些系统可能会使用对性向的真实测量,但是在测量中要求有具有代表性的回答者的前测分数,例如在一般性向测验成套问卷或有关异性性向测验中,或者要求回答者对自己的能力或性向给一个笼统的自我评判,或者要求回答者评价自己对一系列与工作相联系的活动的熟练程度。上面两个系统比较著名的例子就是霍兰的自我指导探索系统和哈一欧生涯决策系统。这些工具都可以用计算机或者纸笔测验来完成。在这些例子里,虽然计算机的管理和计分可以对回答者的分数进行更为详细的报道,但是它们都需要在附带的解释材料中提出自己的要求。

4.3 计算机辅助指导系统

如果对20世纪80年代以来的兴趣和性向的测量的创新做一个综述的话,就不能不提到计算机辅助职业指导系统。在这里需要将对职业/教育和训练项目的信息的分类和选择的系统与更为灵活和全面的用于模拟一系列的指导性面谈的系统区分开来。前一个种类的系统已经存在不少,但是这里主要讨论的是第二种。因为制定这种系统花费较大,而且执行的时候需要比较复杂的程序,所以在现在的操作中很少有高质量的。最著名的莫过于加拿大的CHOICES系统、美国的DISCOVER和

互动式指导信息系统以及英国的 PROSPECT 系统。这些系统的一个共同特点就是包含一个“自我评价”或一个“自我探索”部分,这些系统的关键部分就是测量或自我评价兴趣、性向和技能。呈现在屏幕上的建议一般是告诉回答者如何使用这类信息,以及关于如何做出正确的职业选择策略和进行实际选择的信息。

虽然计算机辅助指导系统是很昂贵的(由于它的制作成本),但是个人电脑联网时代的到来意味着大量的使用者可以合理地分摊这些费用。它对大多数回答者来说是富有吸引力和激励性的,因而极有可能在人们未来职业生涯探索中起到越来越重要的作用,上面提到的美国的系统就是在其他国家也得到良好的使用效果。例如,许多欧洲国家使用了修改过的 CHOICES,澳大利亚也使用了互动式指导信息的修订版,新加坡也使用了国家教育协会制定的工作定位支持系统(JOBS),不可否认还有其他的国家和地方也使用了这些工具。

5. 研究重点

5.1 职业兴趣和性向的结构

因为前面已经讨论了能力或性向的结构,所以在这部分我们只分析兴趣的结构。兴趣和性向的分类系统可以和职业的分类联系起来,这是许多职业指导工作的基础。职业顾问用来进行讨论和帮助客户进行“明智”的职业选择的一种策略就是将客户的兴趣和性向特征与职业的分类联系在一起。在本词条的开始就提到过这个策略的基本原理,即在工作中,人们会对自己感兴趣的和更符合自己能力的工作任务更为满意和完成得更好。

因为计算机的使用使因素分析的应用更为广泛,许多研究开始探讨兴趣的结构。许多研究关注那些能够恰当地“包括”兴趣领域的潜在维度以及这些维度是否能够在平面空间里得到充分的解释。霍兰的理论著作提出一个六边形,根据六边形的顶点的顺序和距离来代表他对兴趣的六种分类之间的关系,这对兴趣结构的研究具有重大的影响。海兰(Hyland)和穆欣斯基(Muchinsky)总结了最近的 700 个研究,认为“这些研究在解释了霍兰兴趣结构理论的结构效度后开始关注验证这个六边形模型的正确性……大部分的研究结果都支持了这个结构”(1991 P. 75)。

虽然一般的研究结果都支持这个理论结构,但不是所有的研究都支持。一个典型的例子就是本

人通过澳大利亚全国性的青少年样本所做的自我指导探测系统结果就区分不出社会型和企业型的维度,而对联邦民用航空公司的 1 500 个成人工作申请者所做的职业爱好问卷的现实型和研究型的维度则混淆在一起(可能是由于受申请者样本范围的限制)。而因素分析时不止出现 6 个维度则很可能是因为最初的层面里存在不同种类的维度,例如雅克松的一部关于《雅克松职业兴趣问卷》的著作中,从 34 个雅克松称为“基本兴趣量表”中进行因素分析共得到 10 个维度——其中 8 个(类似于上面提到的罗的分类)来自与工作规则相联系的活动陈述,另外两个来自许多心理学家称作工作“价值”的陈述中。

大多数著作都承认兴趣的六种分类,霍兰的六边形分类理论会继续得到支持,但是各个顶点之间的距离会随着各个类别分数的经验性关系的改变而发生改变。霍兰(1992)声称他没有更多地留意实践者们是否能发现这个理论假设在他们对客户的工作中作用以及他们是否可以证实这种方法会给客户带来有利的作用。

现在著作中比较有影响的观点是“从职业兴趣测量中得出的”(P. 6674)。然而在 20 世纪 70 年代末期和 80 年代,人们开始拥护以色列的加蒂(Gati)在著作中提出的另一个不同的趋向。加蒂的首要观点就是提出了一个职业或工作(他似乎认为这两个词是可以通用的)兴趣的等级模型,因为在他的观点里,“环形—六边形的模型面临着许多理论和经验主义的问题”(1979 P. 72)。他主要关注的一个方面就是在霍兰的六边形模型中企业型、常规型和现实型是相互接近的与罗的八种分类里组织型、商业型和技术型是互相接近的之间的矛盾。

加蒂在 1979 年的论文里提出的兴趣等级模型类似于弗农(Vernon 1961)的能力等级模型。在这个模型的顶端是一般兴趣因素,接下来是两个更为广泛的团体因素(“人文科学”和“自然科学”),接下来是更多的小因素(有些类似于霍兰或罗的分类),然后是职业领域,模型的底端是具体的工作。加蒂的理论是建立在工作是可以通过其属性或特征的集合来进行划分的基础上。这些属性或特征包括工作环境特征,如户内还是户外、与人们的接触、独立水平、报酬高低、工作价值等,不是以兴趣来划分的,他认为“用等级的树状结构可以表达工作之间的相似性联系,并且工作兴趣的

模式是和这个树状结构相一致的”(1979 P. 93)。然后他呈现了对一些支持霍兰的六边形模型的数据的再分析,它们也支持等级模型,但是不完全支持罗的连续体模型。

加蒂和他的同事在20世纪80年代发表了很多论文,既从概念上也从经验主义的角度上支持了职业兴趣的等级结构,这些论文后来做成了一个合集(Gati 1991)。对这些论文仔细阅读后发现这些作者继续关注工作的价值方面而不是工作的兴趣,许多的经验主义的资料都是来自小样本。但是就做问卷的兴趣来说,他们的著作实际上说明了环形理论比等级模型更具有代表性(Gati and Nathan 1986)。斯旺森(Swanson)认为加蒂忽略了个体对工作的知觉和兴趣之间的区别,因而等级模型更多描述的是前者而不是后者。她指出加蒂所宣称的关于等级模型的优越性好像“只有在通过由单独的工作名称作为项目组成的个人问卷(VPI)收集来的资料中是正确的”,并且“从其他由不同的项目形式组成的问卷中获得的资料……对霍兰的六边形模型和加蒂的等级模型具有同样的支持度”(1992 P. 237)。除非等级模型有非常明显的优势,否则研究者和顾问们可能更愿意使用霍兰的方案。

5.2 职业兴趣和性向的联系

洛曼的一本很有潜在影响力的书认为“综合的职业评价”是建立在“能力、兴趣和其他为达到对职业动态理解有关人格资料的必然联系”的前提上的(1991 P. 177)。这本书对能力、人格变量和兴趣的跨领域的交互关系的研究做了一个简短的回顾,认为应该更多地研究两者之间的互动以及这种交互作用对职业行为的作用。兴趣—能力或性向的关系是本词条关注的问题之一。20世纪40年代到60年代之间做了大量关于兴趣—能力或性向的研究,但是这些研究只促进了近代的职业兴趣理论的发展。洛曼解释了为什么早期的甚至是后来的研究在两个领域中只发现了“很少的共同变量”,然后他总结了现在的(包括他自己的)一些研究,认为“已经纠正了早期的一些不足”并“扩展了相应的著作”(P. 183)。一个比较好的例子就是兰达尔(Randahl)的文章,他对一个比较大的、同质性的样本使用多元统计如剖析图分析,结果表明“当从类型学角度进行测量时是可以测量出能力和兴趣的关系的”(1991 P. 348),而且这种联系是比较“牢固的”和“与理论预测相一致的”

(P. 346)。洛曼相信关于跨领域关系的研究论文仅仅是刚刚开始出现,并且希望能有更多非临床的、和两个性别都有的代表性的样本的研究出现。

5.3 预测职业选择或职业成绩的效力

人们大力研究职业兴趣和性向的一个主要原理就是兴趣和性向能够帮助个体选择职业和雇主选择最有生产力的员工。因此,这方面的研究会一直关注以下几个方面:将各个变量联系起来预测给定的结果;计分是否需要调整(如是通过标准分数,还是通过对不同的群体使用不同的标准图表,或者预测不同的课程或工作是否成功时对不同的变量使用不同的权重等等——这个问题在第六部分还会有详细的讨论);新的变量或评价方法是否会更好地进行预测。下面会对这方面的研究做个简要的回顾。

在20世纪50年代到70年代之间,人们对确定成功预测不同的课程或工作成绩的变量之间的不同联系和不同权重进行了大量研究(French 1966, Horst 1957)。但是结果表明,这种“区分性预测”的研究在方法学上是很难实现的,而且所得到的结果也不是很确定。这对学业和工作成就的测量以及用何种变量和测量预示着学业成功的标准化问题做出了贡献(Prediger 1989)。在教育领域,经常使用的是课程成员团体(Prediger and Brandt 1991)。目前虽然探索研究成功地表示了兴趣变量在预测方程中的作用高于性向变量(Lokan 1977),但是目前还很难付诸实践。

20世纪70年代关于预测的偏见以及是使用兴趣测量的原始分数还是标准分数作为对学生进行咨询,引发了80年代中期的一场激烈争论,主要问题是用一般能力因素之外的能力或性向来预测职业以及与之相关的成就是否真的有作用。《职业行为》期刊有一期专门讨论这个问题,大多数人认为一般能力因素在职业中是非常重要的。美国就业局(USES)从不同的职业领域中规定出来的66种“职业性向模型”,使通过结合性向测量的不同分数来选择员工成为常事。雷和厄尔斯(Ree and Earles 1991)对这个问题所作的研究就是其中一个例子。

普雷迪格尔(1987, 1989)指出了“成功预测模式”的困难并总结了通过跨工作的研究。他认为这种针对某一个单独工作和通过跨工作来确立不同的工作任务模式的研究,以及他自己的关于“职业领域的6种工作集群种的14个适当的认知和非

认知能力”的研究,都清楚地表明各个职业之间的能力差异不仅仅是一般能力因素的差异。普雷迪格尔(1987)认为在存在“相似剖析图模型”的教育咨询中测量能力/性向的种类是非常有价值的。

上面两段提到的研究和争论都关注能力或性向测量的使用。职业兴趣的测量是有一定预测性的,特别是与对职业向往的表达结合起来使用(Holland 1986)——这样也可以提高对中等教育的课程体系和学习领域的预测,这种方法已经在大范围的研究中得到强调。安利(Ainley)等表明“兴趣和性向在初等或中等教育早期的作用开始明显,并且对11和12年级的课程选择具有重要的影响”(1990 P. 100)。在他的一个被视为全国范围内几千个澳大利亚学生的纵向研究中,10岁儿童的计算成绩与12岁时选择物理科学课程具有显著的相关,而14岁时对研究活动的兴趣与12岁时参加科学研究活动的数量相关。对艺术的兴趣与参加人文和社会科学活动、创造性和表演艺术活动相关。基德和内勒(Kidd and Naylor 1991)的一个纵向研究的结果验证了安利等的关于研究兴趣的发现,虽然他们的研究样本比安利的小。

普雷迪格尔和勃兰特(Brandt)的一个以2000名高中学生为样本的研究结果表明:兴趣和能力/性向测量能够帮助学生选择让自己获得满足和成功的职业培训项目。通过一系列的判别分析后,研究者发现“不同的职业安排之间是存在实质差异的”,因而“每个兴趣和能力测量的安排上也是有显著差异的”(1991 P. 137)。我们前面所提到的研究支持了普雷迪格尔(1987)的观点,他认为在现在已知的教育课程或职业群体中对照学生的兴趣和能力分数概图是进行教育咨询的一个很好的基础。

5.4 其他的评价方法

除了在一些极少的场合里,如测量实际设备操作中的手指灵巧,职业兴趣和性向测验都是可以通过纸笔测验的形式来进行的。从20世纪80年代开始,由于计算机在计分和结果解释方面的快捷和迅速的优势,由计算机来管理测验和问卷变得非常普遍。另外,随着个人计算机图形包的出现和质量的不改善,计算机进行的测量中开始提供图表材料。不过在以上的场合里,只是评价的表达方式而不是评价方法本身的特点。

一个更为创新性的水平就是,计算机越来越能够以非常变通的方式来进行测验。可以根据回答

者先前回答的模式来决定测验项目的进度从而来推测回答者的能力水平。针对这种测验的统计技术可以保证用相对较少的项目得到对回答者能力的可靠估计,从而节省测验时间。因为根据难度来建立等级任务相对比较简单,所以这种适应性的测验比较适合用来进行能力或性向的测量。但是这种评价方法不适合用在兴趣测量上,即不适合用在瑟斯通类型的量表上,这种量表包含的是同质性的项目,回答者的进度很难统一,而且很难确立一个等级任务。

许多国家在20世纪80年代末期到90年代开始兴起的能力训练,比较倾向于在专业和非专业的环境下都使用职业性向测验。与以前的关于一般能力因素的争论相反,90年代兴起的性向测量更多地关注于成功地完成某个工作或某个训练课程所必需的任务上。现在能力评价的重点则是尽量少用纸笔的性向测验,而计算机模拟测试,经过一段时间的培训和实践,会成为未来通用的测试手段。

6. 公平性问题

霍兰在20世纪80年代中期提出兴趣测验新的方向时写道,“受当今的信仰、知识、测验标准和经济来源的限制,在兴趣问卷评价中最复杂、困难和牵涉到社会情感的问题就是如何保证公平性”(1986 P. 253)。他相信上面所提到的关于使用原始分数还是标准分数的分歧就是一般的职业咨询和单独的兴趣测量的目的不一致的结果。他认为兴趣问卷的目的至少有三个:探索(扩大人们选择工作的范围)、支持(为人们希望的或潜在的工作提供支持)、自我了解(为人们了解个体兴趣的特点提供结构化的信息)。霍兰和其他人(Hansen 1992)认为解决问卷的公平性问题不能仅仅靠收集更多的资料,而且要靠对不同群体的回答者做同一个问卷所产生效果和职业顾问要对回答者做问卷的目的了解,要靠对设计结构化的解释材料和职业顾问对解释测量的充分应用。

性向或能力测量中的公平性问题可能比兴趣测量中的更为重要。现在都把性向定义为“已得到发展的能力”,阅历丰富和接受良好教育的个体可能更好地发展自己的能力,因而就业选择系统和高等教育课程的某些限定领域就面临这样的尴尬:一方面要选择最有可能成功的候选人,但是另一方面又要尽量避免对弱势群体的不公平。我们知道