



心理与教育的 测量和评价

(下)

[美] R. L. 桑代克 E. P. 哈根 合著
叶佩华 邹有华 刘蔚成 主译

人
民
教
育



外国教育丛书

心理与教育的测量和评价

下 册

[美] R. L. 桑代克 E. P. 哈根 合著
叶佩华 邹有华 刘蔚成 主译

人民教育出版社

**MEASUREMENT AND EVALUATION
IN PSYCHOLOGY AND EDUCATION**

FOURTH EDITION

ROBERT L. THORNDIKE

Richard March Hoe professor
of psychology and education
Teachers College

Columbia University

ELIZABETH P. HAGEN

professor of psychology and education
Teachers College

Columbia University

copyright © 1955, 1961, 1969, 1977, by John Wiley & Sons, Inc.

外国教育丛书

心理与教育的测量和评价

下 册

[美] R. L. 桑代克 合著
E. P. 哈 根

叶佩华 邹有华 刘蔚成 主译

*

人民教育出版社出版

新华书店北京发行所发行

顺义县永利印刷厂印装

*

开本 850×1168 1/32 印张 12 字数 210,000

1985年10月第1版 1987年12月第1次印刷

印数 1—6,200

ISBN 7-107-10169-2/G·1050

书号 7012·1001 定价 2.45 元

下 册 目 录

第九章 智力或学习能力倾向的标准测验	1
什么叫做学习能力倾向	1
作为心理构成物的一般智力	2
认知功能的团体测验	3
修订的斯坦福—比纳智力测验	8
韦克斯勒智力量表	13
简短的个人测验	17
不依靠语言或文化的测验	18
幼儿及学前测验	21
智力测量的团体测验和个人测验的对比	26
认知能力测验的可靠性和稳定性	29
认知能力测验测试出来的能力的实用意义	35
对文化背景有差异的人们进行的测验	40
预期的成就	44
认知能力测验与学校决策	56
概述	61
问题和练习	62
第十章 特殊才能的测量	65
职业能力的成套测验和单一测验	65
分辨与预测之分	88
预察测验	98
专业学校能力倾向成套测验	100
音乐能力倾向测量	101
艺术能力倾向测验	103
评价创造力	105

概述.....	107
问题和练习.....	108
第十一章 自我评价问卷与记录表.....	110
研究个性的方法.....	111
面谈.....	112
传记资料表.....	118
兴趣记录表.....	126
气质和适应记录表.....	139
态度问卷.....	157
概述与评价.....	164
问题和练习.....	165
第十二章 旁人眼光中的个性.....	167
介绍信.....	167
等级评定量表.....	170
在求得正确的等级评定过程中遇到的问题.....	172
提高等级评定的有效性.....	184
概述与评价.....	211
问题和练习.....	213
第十三章 行为的测验与观察.....	215
诚实的间接测验.....	216
其他类型的操作测验.....	220
投射技术.....	223
情境测验和评价程序.....	226
系统的观察.....	232
使用直接观察法的示范研究.....	238
系统观察的价值及其局限性.....	244
非正式观察——实例记录.....	248
概述.....	259
问题和练习.....	260
第十四章 设计学校的测验方案.....	262
马车和马的关系.....	262

测验方案的作用·····	263
测验方案应具备的特性·····	265
建议优先采用的测验方案·····	268
设计测验方案·····	278
学校测验方案的管理工作·····	287
向公众介绍测验的结果·····	295
校外测验方案·····	309
对校外测验方案的关注·····	311
概述·····	314
问题和练习·····	315
第十五章 分数与评分 ·····	319
分数的作用·····	320
正确评分法的技术问题·····	324
概述·····	338
问题和练习·····	339
第十六章 测验中的社会与政治问题 ·····	341
测验对个人权利的威胁·····	342
对某种人的权利的可能危害·····	350
测验所要取得的积极价值·····	351
最大限度地提高测试的积极效果·····	356
附录 专业词汇英汉对照表 ·····	359

下 册 表 目

9.1	出生第一年内的智力测验与后来的测验之间的相关系数 (逐次测验累计产生的相关系数).....	22
9.2	典型的一般能力测验的信度.....	30
9.3	斯坦福—比纳L级测验中 IQ 值均为 100 时, M 级测验中 IQ 值的预期分布情况.....	31
9.4	不同职业中一般智力的估计量.....	38
9.5	智力测验预测训练和熟练标准的效度.....	39
10.1	选拔飞行员能力倾向测验之间的相关.....	66
10.2	100 个高中生在 GATB 中的得分的相关系数.....	77
10.3	不同学科的等级与 DAT 得分之间的相关中数.....	79
10.4	关于特定职业的 GATB 因素的平均成绩.....	82
10.5	GATB 因素成绩的效度及某些职业能力模式.....	83
10.6	各种职业成就的各种测验的平均效度.....	87
10.7	法律学生、工科学生和汽车技工的能力与兴趣测度的中 数(百分位数)及组间的差异.....	90
10.8	12 种主要专业.....	91
10.9	预测工程学校和汽车技工学校 GATB 成绩的有效系 数.....	92
10.10	用合并因素的标准成绩来预测工程学校的学生成绩.....	93
10.11	1965—五年跟踪专业计划组的辨别函数曲线.....	95
10.12	比利·威廉斯个人能力测验图.....	97

10.13	典型的阅读准备测验的各种作业	99
11.1	苏珊·斯诺的斯特朗—坎贝尔职业测验分数图	125
11.2	苏珊·斯诺个人量表的标准分数	129
11.3	从事经过挑选的某些职业的人的库德职业兴趣的平均百分等级	130
11.4	KOIS 女性主修科测量的高分表	132
11.5	不同职业小组的库德平均标准分数	137
11.6	能力倾向和兴趣因素相关表	138
11.7	吉尔福德—齐默尔曼气质调查 10 量表的相关数和信度	144
11.8	精神病人的 MMPI 测验图并附解释性评注	147
11.9	对妇女解放进行态度测量的部分结果	161
12.1	文尼兰社会成熟量表项目举例	192
13.1	测量作弊情况的测验信度	218
13.2	两次独立的评价所获参数的相关系数	230
13.3	表明评价预测和工作条件之间密切关系的达到中层管理水平的评价的百分比	231
13.4	品质评定的全面评价与所达到的管理水平的关系	232
13.5	语言相互作用的分析类目表	241
14.1	测验方案的作用	264
14.2	建议学校采用的测验方案	278
14.3	罗伊·道金斯测验成绩表	307
15.1	课程三个部分的分数	336
15.2	三部分的等级分布	337

下 册 图 目

9.1	《认知能力测验》项目的样本	4
9.2	类似《认知能力测验》中的项目的语言测验样本	5
9.3	类似《认知能力测验》中的数量测验的样本	6
9.4	类似《认知能力测验》中的非文字项目的样本	7
9.5	模型题目	19
9.6	起始测验年龄和测验——重测间隔对从早期测验预测后来的《斯坦福—比纳》IQ的影响	32
9.7	测验——重测时间间隔对从早期小组测验预测中学末期的小组智力的影响	34
9.8	不同种族的中层(M)和下层(L)儿童的相对成绩	43
9.9	预期成绩的三种看法	45
9.10	根据五年级非文字的认知能力测验标准年龄分数,用衣阿华基本技能测验所得出的与年级平均分相当的百分位数分布	49
10.1	了解认知能力的方法	68
10.2	四种职业能力的简略图	86
10.3	五年级跟踪专业计划组的判别图	96
10.4	霍恩美术能力倾向记录表测验项目实例	104
11.1	自我描述技术的相互关系	113
11.2	雇用和评价问话人工作指南	117
11.3	态度量表一例	161

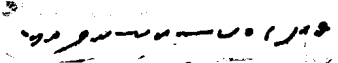
12.1	雇员成绩评价.....	189
12.2	作为等级评定基础的观察记录.....	204
12.3	四年级的社会关系图.....	207
13.1	击圈测验.....	217
13.2	罗夏型墨迹图.....	224
13.3	取自默里主题统觉测验样图.....	225
13.4	由“直接的”或“间接的”教师任教的班级的成绩和态度比较图.....	242
14.1	森杜威尔学校学科平均成绩和年级水平与系统平均分数的国家常模比较图.....	299
14.2	森杜威尔学校七年级学生的中位成绩与国家常模及与同一智商水平的学生常模的比较.....	299
14.3	伊斯顿学校七年级学生阅读成绩分数与智力测验成绩分数之比较.....	301
14.4	伊斯顿学校七年级成绩的中数和百分等级与国家常模之比较.....	302
14.5	同一组学生从四年级(1963~64)到七年级(1966~67)的拼写成绩图.....	303
14.6	衣阿华基本技能测验和洛奇—桑代克智力测验成绩报告表.....	306

第九章 智力或学习能力倾向 的标准测验

什么叫做学习能力倾向？

在前面几章里，我们讨论了作为检查学校教学最终效果的一些测验。这一章和下一章将讨论预示一个人今后的成就的一些测验。然而，无论是讨论过去学习的结果，还是将来的潜力，我们所能测试的只是现在的成绩。

现在的成绩和有组织的学校教学有不同程度的密切关系。因此，某个美国学生能够译出 *Arma virumque cano* 或者

、的意思，那几乎可以肯定是因为学生在学校的拉丁文课或者速记课中学过这些词句，而掌握这些东西，不过是表示从系统的教学中得到的进步而已。其他成绩，例如，了解字义、解决演算问题、领会一段复杂的散文的思路等，既要依赖于学生的校内经验，又要依赖于校外经验，这些知识在校内教过，但也需通过生活经验的日积月累才能完善。还有一些成绩和学校所教的东西关系极小。这方面的成绩包括有关解决空间的、形象的和其他非文字材料问题的能力，当然，也包括儿童在入学前就已经掌握的那些成绩。

应该强调：任何成绩如果不是依赖校内经验就是在某种程度

上依赖生活经验。任何语言测验都要求一个人必须会说那种语言；任何图形测验都要求一个人熟悉所描画出来的物象；无论什么种类的测验，都要求被测者必须懂得在测验时作出努力，尽力而为。由于同样的原因，任何成绩在某种程度上来说，都有赖于个人遗传上的潜能，即他成为人类一员的遗传潜能和他的个人的特殊基因。因此，这一章和下一章所讨论的能力倾向测验不同于第八章讨论的成绩测验，它是较少依赖学校的专门教学环节，而且它是用来推断将来，而不是用来取得关于过去的结论。这些测验不能看作纯粹的测量学习的固有潜力。

作为心理构成物的一般智力

在早期的心理测量史上，心理学家们就开始对人们使用概念和各种概念之间的关系的一般能力进行评价感到兴趣。虽然，引起这种兴趣的部分原因是出于科学的关注：要了解和描述智力的功能；但另一部分原因是出于一些实际的需要；他们要识别智力上有缺陷，在学校中很难获得正常进步的儿童。更严格地说，他们要识别连生活上起码的智力要求也达不到的一些人，——极端的案例是连下雨也不晓得进屋避雨的人。

设计出来用于对一般认识功能水平进行估价的测验，在心理学文献中称为智力测验，在通俗的讨论中称为 *IQ* 测验。然而，心理学家们对智力的定义从来没有明确的一致意见，而在智力包含什么，不包含什么的问题上也没有明确的一致意见。近几年来，出现一种越来越贬低智力这个专门名词的倾向，因为这个词的内容已经包涵过多的意思和感情，而不是指“一般的认识能力”或“学习的能力倾向”了。学习能力倾向这个术语特别清楚地表示出，这些测验的功能在于预测学校的成绩。

不管名称是什么,有充足的证据表明:各种各样的工作都依赖人们头脑中贮存的概念,和对概念关系的运用能力,这两方面总合起来,就形成了认识功能的一般能力。这种能力对测验成绩的影响范围广大,对学校内的和实际生活中很多方面的工作能力都有影响。在这一章里,我们将讨论这些测验所采用的形式,讨论它们通用于一系列测验方面和在时间上的稳定性的理由,讨论它们在学校和非学校标准方面的有效程度,以及它们在使用和解释中的问题。

认知功能的团体测验

在美国,大多数能力测验是以团体测验的形式进行的,即对整班学生同时进行测验。我们将介绍一系列团体测验,以说明这类测验方法的一些特点。其他一些广泛使用的测验系列及其主要特点,列举在附录四中。

《认知能力测验》(桑代克、哈根、洛吉,1968,1972)由四个不同的部分组成。用于小学低年级的测验是非阅读性的,所有测验指导都由口头给出。用于第四年级到第十三年级的测验包括:文字测验、数量测验和非文字测验。每个测验由几个子测验组成,子测验的题目形式相同,其进程是由易到难。每个测验有几套不同水平的题目,以便对智力成熟水平不同的人提供适当难度的测验。

小学的测验有四个子测验:口头词汇、关系概念、多重智力和数量概念。测验小册子里的材料全部是图形化的,图 9.1 显示了一些项目的样本。测验的要求以口头形式告诉学生。四个项目的说明基本上是以如下的方式进行口头指导:

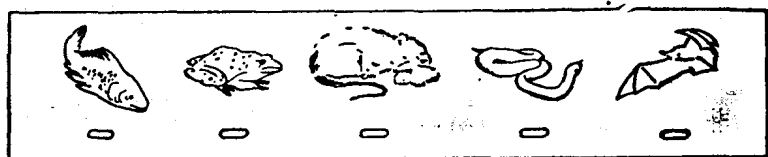
口头词汇:“在老鼠的图画下面的椭圆里填空……老鼠”。

关系概念:“在表示球在椅子上的图画下面的椭圆里填空……球在椅子

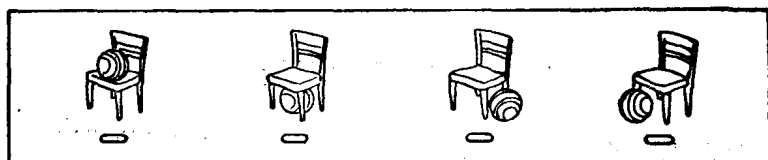
上”。

多重智力：“看这一排图画。有一张与其他的无关，在这张与其他不相同的画下面的椭圆里填空。……与其他的都不相同。”

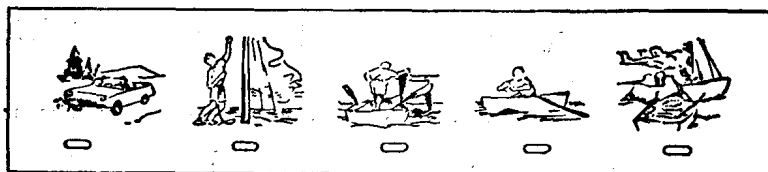
数量概念：“找出里面有两套两根棍子的盒子，在这个盒子下面的椭圆里填空……两套两根棍子。”



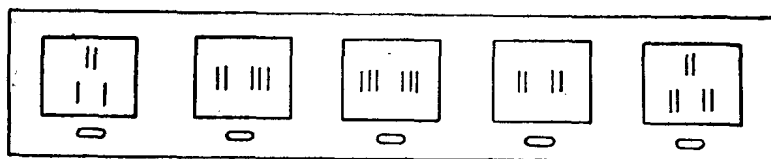
(a)



(b)



(c)



(d)

图 9.1 <认知能力测验>项目的样本。(a) 口头词汇；(b) 关系概念；(c) 多重智力；(d) 数量概念。

语言测验包含四个子测验，分别是词汇、句子填充、词语分类、词语类推。图 9.2 给出每类项目的例子。每个项目的正确答案下面划了线。为了提供合适的测验给低至四年级的儿童以及高至大

学的新生(意即对大多数成年人来说是够难的),这个测验以“多水平”的形式出现。每个子测验的项目都有从容易的四年级到难度大的十三年级的内容。但是,特定年龄或年级的学生只要做一部分测验题,从不同的地方开始,在不同的地方结束。这套测验有八个重迭的水平。因此,词汇测验的标准类型如下:

年 级	采用的测验的水平	开始项目	结束项目
4	A	1	25
5	B	6	30
6	C	11	35
7	D	16	40
8	E	21	45
9	F	26	50
10~11	G	31	55
12~13	H	36	60

词汇:

粗鲁的:

- A. 不快乐的 B. 忿怒的 C. 无信用的 D. 无礼的
E. 多咀的

句子:

填充:

马克斯很喜欢他的理科教员,但他不____他的数学教员。

- A. 服从 B. 争论 C. 尊重 D. 要求 E. 喜欢

词语分类:

鸽 鹰 鹁鹁 麻雀

- A. 蛾 B. 蝙蝠 C. 海鸥 D. 蜜蜂 E. 松鼠

词语类推:

豌豆对于蚕豆犹如桃子对_____。

- A. 煤坑 B. 树 C. 食物 D. 皮肤 E. 苹果

图 9.2 类似《认知能力测验》中的项目的语言测验样本。

每个学生测验 25 个词汇项目,所选的这些词汇项目最适合于他所在小组的能力水平。整个测验由 100 个项目组成,即四个类型,每类 25 个项目,见图 9.2。

因为运用数量性质的概念及其关系的能力稍不同于使用语言概念及其关系的能力,所以有一套单独的数量测验。图 9.3 是这套测验使用的项目的样本。在第一个子测验里,受试者必须确定两个数量中哪一个大些。在第二个子测验里,要选择一个答案,使数列合逻辑地延长。在第三个子测验里,调整数字和符号,使结果符合被选择的答案。

词的比较:

第 I 栏 如果 I 多于 II,作记号于 A

第 II 栏 如果 I 少于 II,作记号于 B

5×0 如果 I 等于 II,作记号于 C

数列:

18 16 14 12 10

A. 7 B. 8 C. 9 D. 10 E. 12

建立等式:

1 8 9 + -

A. 0 B. 3 C. 8 D. 9 E. 18

图 9.3 类似《认知能力测验》中的数量测验的样本。

为了测量不需要语言能力的特别是不需要阅读技能的推理能力,还有一系列非文字的测验。图 9.4 显示了图形推理、图形分类和图形综合这三类项目的样本。在前两类项目里,要求受试者选择一个正确的答案。在图形综合测验里,受试者必须确定把给出的一堆部分图形放置在每个图形上,能否使每个图形“变得完整”。非文字测验也制成多水平的形式,以便主试者为受试团体选择最适当的难度范围。

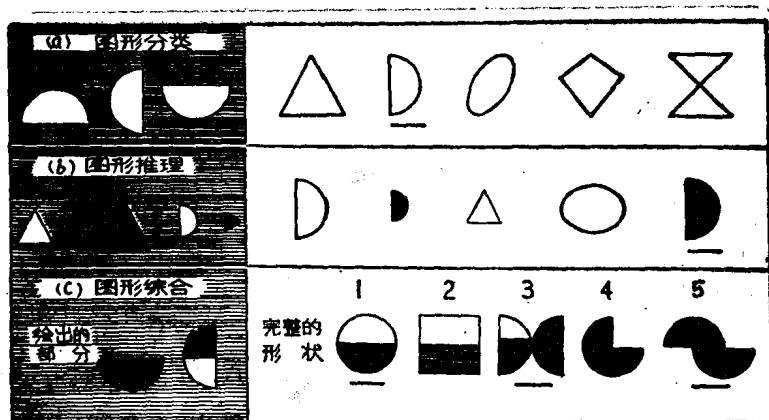


图 9.4 类似《认知能力测验》中的非文字项目的样本。

在同一年级团体的四套测验间的大量而又确实的相互关系中,可以找到证据说明,在这个范围相当广泛的测验里,自始至终存在着一种一般的认识能力。它们的关系大概如下表所示:

	语言的	数量的	非文字的
小学的*	.60	.58	.66
语言的		.75	.66
数量的			.72

*注: 小学测验和其他测验相隔两年时间。其他三个测验是在同一时间进行的。

显然,所有测验在较大程度上,都要求运用某种认知功能的普通技能。然而,正如我们将看到的,这些分开进行的测验的信度大约是0.90,因此,各种测验的功用也不相同。每个测验的独特性质足以使每个测验具有自己的意义。所有这些测验都表明与学校成绩的测验的良好相关,但就大多数科目来看,语言测验是一种更好的预测工具。不过,语言测验需要英语的阅读和运用能力。因此,非文