

## 第六章 学力评价与道德评价

### 一、学力的评定

学校教育的评定范围包括学习的评定,人格、操作的评定等方面,但这里只谈一下学力的评定。在学校教育中,学力的评定占有非常重要的位置。学力评定,虽说通过教育工作中的观察也是办得到的,但是作为有计划的方法使用的是学力测验。典型的学力测验是标准成就测验和教师编制测验。

#### (一) 标准成就测验

1. 标准成就测验的特点 标准成就测验 (Standardized achievement test) 是以各科教学和测验理论为背景,由专家编制的。标准成就测验由效度和信度高的问题组成,根据统计学原理制成全国性的标准。因此标准成就测验的一大特征就是借助全国学生成绩的成就测验水平为媒介而制订的标准,能够对学生所处的水平(例如用标准分数来表示)作出比较。然而现实情况是,未经充分标准化手续的测验却以“标准成就测验”的名义在市场上出售,测验的利用者很重要的一点是要识别出冒牌的标准成就测验与富有科学价值的标准成就测验。

2. 利用标准成就测验时的注意事项 要有效地利用标准成就测验,有一些注意事项。

(1) 提高测验利用者的修养 随着测验的诊断倾向的强化、

复杂和高深化要求测验的利用者具有使用与解释的能力。测验的利用者必需具备有关学力的教育和心理学的认识与测量理论的素养。

② 明确实施测验的目的 有效地利用标准成就测验的根本点是，教师自身要能够正确地认识并且做到了解学生，懂得教学要求与利用测量之间的相互关系。就是说，要阐明为什么要实施标准成就测验，合此便不能期待测验的有效利用。

③ 标准化的问题 标准成就测验的取样应当妥当，即按照标准化的手续抽取出来的样本（学生集体）应当是全国学生全域（Population）的一个缩影。在标准成就测验中有许多不过，是按照一定格式作了一番标准化手续罢了，并未真正符合科学的标准。此外，标准化的时间也是重要条件。在其它条件相同的条件下，标准化的时间愈新近愈好。

④ 效度问题 熟读测验介绍，要看看该测验是否具备应有的条件，即用标准成就测验可以测量什么？编制测验的理论基础是什么？在这些方面没有说明清楚的标准成就测验，最好避免使用。

## （二）教师编制测验

1. 教师编制测验的特点 在一个单元的中途和结束时，或是在期中、期末，就较短期间的学习内容进行的测验，它的特点就在于同教师的日常教学工作息息相关。教师编制测验的目的，从教师的角度看，是要了解学生是否达到了预期的指导目的，了解每个学生、全班学生掌握知识的趋势，以便有助于日后的教学工作。对学生来说，是要了解学习进行的程度，提高进一步学习的

动机作用。这种测验也称作形成性评定测验。

2. 制定教师编制测验时的注意事项 下面谈谈制定教师编制测验中应当注意的事项。

① 具体而明确地决定测验的评定目标 制定教师编制测验，特别是制定形成性评定测验的第一步，重要的是进行如下的教育目标编制工作：对应当学习的教育目标（内容）进行分析，并将分析出来的从属目标归纳在二、三节课或一周这样比较短期的教育活动中。关键是教育目标要规定得具体、可操作。

② 出题上的注意事项 教师编制测验主要着眼于诊断学生，所以更多地要使之具有诊断的功能。但是，毫无遗漏地拟出了解从属目标的达到程度所必需的所有问题，是很难的。因此，将从属目标同测验问题的目的的对应关系制定表5-3的PD表（问题范围维度表）（Problem-Dimension表）来表示是有效的。常用PD表有许多优点：可以促进教育目标与内容的分析工作，容易发现并排除测验内容不明确的问题，也可以核对必要种类的问题是否无遗漏地并且恰如其分地包括在内了。

表 5-3 小学六年级算术“分数”教材 P D 表 (织田守矢, 1974)

维度 (D)		1	2	3	4	5	6	7	8	9	10	11	12
		整数 × 分数	分数 × 分数	整数 ÷ 分数	分数 ÷ 分数	混 合	乘 法	除 法	加 法	减 法	假 分数	带 分数	真 分数
1	$2 \times \frac{2}{3}$	0					0				0	0	
2	$12 \times \frac{2}{3}$	0					0				0		
3	$\frac{2}{2} \times \frac{2}{4}$		0				0						0
4	$4 + \frac{2}{5}$			0				0			0		
5	$\frac{1}{2} + \frac{3}{4}$				0			0					0
...	...		...					...			...		

③ 决定测验形式时的注意事项 测验形式基本上可分为论文式测验和客观测验。两者比较起来, 客观测验的特点是, 不适于测量综合性的智力机能, 出题可以广泛而琐细, 容易介入偶然的成功, 评分客观, 效率高, 倘若常用会使学生的思维容易陷于单纯肤浅的危险, 等等。两种形式的测验各有长处和短处, 不能单纯说哪一种形式优劣。希望使用时采取灵活的态度, 考虑到实施测验的目的、测量的目标、测验的对象、场所等等条件, 善于利用两者的优点和长处。

大桥正夫编 钟启泉译: 《教育心理学》

上海教育出版社 1980年2月

## 二、测验的种类和功用

### (一) 测验的种类

这里讲的测验是标准测验，而不是非标准测验。标准测验除了要有常模外，还要具备下列条件：(1)给所有受试者实施有代表性的相同的一组测验，为他们的作业取得直接比较的基础；(2)实施测验的手续（包括测验指导语、测验时间限制、测验情境等）要有详细的规定，以保证每一受试者有相同的测验条件；(3)记分方法要有详细的规定，使评分误差极低。上述三点与得分的步骤有关，常模（平均数）则给测验分数提供参照点。如此，一个测验就能作为测量人的某种心理特性（智力、能力或人格等）的工具。标准测验可以按不同的分类标准加以分类。

#### 1 根据测验的对象来分

① 智力测验：目的在于测量智力的高低。

② 能力倾向测验：目的在于发现学生的潜在才能，深入了解其长处和发展倾向。能力倾向测验一般可以分为：(1)一般能力倾向测验：测量一个人的多方面的特殊潜能。例如区分能力倾向测验，它可以预测一个人哪一方面潜在能力较高或较低。(2)特殊能力倾向测验：偏重测量个人的特殊潜在能力，如音乐能力倾向测验、机械能力倾向测验。

③ 教育测验：测量一个人经教育训练或学习后的学业成绩。又称成就测验。它可分为学科测验和综合测验。前者测量学生某

学科的知识、技能；后者测量学生各学科的知识、技能。

④ 人格测验：测量兴趣、态度、情绪、动机、性格……等方面的行为。

### 2 根据测验的人数来分

① 个别测验：它只能由同一个主试在同一时间内测量一个人。个别测验的优点是主试对试的言语、情绪状态有仔细的观察，并且有充分的机会与被试合作，激起被试最大努力，所以其结果正确可靠。个别测验的缺点在于，时间不经济，测验的手续复杂，需要训练有素者方能胜任。

② 团体测验：可由一位主试同时测量许多人。各种教育测验都是团体测验，一部分智力测验也是团体测验。它的优点是时间经济，主试不必接受严格的专业训练即可担任。它的缺点在于对被试的行为不能作切实的控制，所得的结果不及个别测验正确可靠。

### 3 根据测验材料来分

① 语言或文字测验：它可以测量人类高层次的心理功能，编制和实施都较容易；人类的心智能力不能完全以图形或实物测量出来，故语言或文字测验应用范围较广，团体测验多数采用语言或文字测验。然而它不能应用于语言有困难的人，而且无法比较语言文化背景不同的受试者。

② 非文字测验或操作性测验：以图画、仪器、模型、工具、实物为测验材料，受试者以动作表达。它的长处和短处，正好与语言或文字测验相反。

#### 4. 根据测验的功用来分

##### ① 预测测验和成就测验

预测测验用于推测某人在某方面未来成功的可能性，智力测验、能力测验就属于此。它多数根据作业分析的结果来选择测验材料。成就测验在于考查某人在某方面目前的成绩，一般教育测验就属于此，因为它所测量的是学生现在的成绩，它往往是根据作业样本来选择测验材料。

##### ② 难度测验与速度测验

难度测验的功用在于测量受试者程度高低，它通常的时间限制的标准是使95%的被试都有做完测题的机会。测题由易到难排列，以测量被试解决难题的高度能力。速度测验在于测量被试作业的快慢，它的测题难度相等，但严格限制时间，看规定时间内做对几题。

##### ③ 普通测验与诊断测验

普通测验在于考查一个人或一个年级的学生在某方面的大概程度；诊断测验则进一步去诊断学生某方面的特殊优点和缺点。教育上的诊断测验偏重发现学生困难之处，作为改进教学方法或进行补救教学的依据。

#### (二) 测验的功用

1. 辨别智慧 适应班级教学，测出全年级学生智力高低，以此作为分班依据，因材施教。

2. 人材的选拔 在众多的候选人中确定有最大可能成功的人，以此选拔人才。

3. 指导就业 智力测验和能力倾向测验可以作为职业指导的工具，职业上成功的条件之一，是一个人从事的职业适合自己的能力。各种职业需要不同的智能，所以指导就业时，除了依照兴趣、志愿等条件外，还需要进行智力能力倾向测验，以决定被试适宜从事哪些职业。

4. 诊断 (在本节第一大点第四点中已有叙述)

5. 预测 (同上)

6. 评定学生的成绩 由于旧法考试 (论文式的考试) 存在着评分不统一，记分不准确且测题的取样代表性差的缺点，作为评定学生成绩的准则就很成问题，标准化测验克服了上述缺点，为评定学生成绩提供了客观依据。

7. 用来进行实验分组，使之等组化。

戴忠恒：《测验的性质·使用·种类》，

摘自《教育科研情况交流》

1984年第3期



### 三、正确使用测验，防止滥用

滥用测验的危害性甚大。不少人误以为任何人都可以使用心理测验，其实心理测验像其它科学一样，必须加以适当的运用才能发挥其功能。如果滥用或不够资格的人员实施、解释，则会引起不良后果。正确使用测验必须做到：

1. 测验要保密：受试者事先未曾获知内容的心理测验才有价值，所以测验内容不可随便泄漏，也不可借给不够资格的人员使用，免使测验失去控制。

2. 够资格的心理测验工作者才能使用心理测验。测验的选择、实施，记分以及测验分数的解释等只有训练有素的心理测验工作者才能胜任，这好比唯有一定技术水平的外科医生才能给病人开刀一样。

① 测验的选择：每一种测验都有其特殊功能。因此选用何种测验予受试者，应慎重考虑。选择测验时也应了解它的信度和效度。这就要求起码要懂得信度与效度的含义。

② 测验的实施：任何测验实施过程中的操作需要按照该测验严格规定的要求进行，并且善于安定受试者的情绪以及掌握其他有关事项，使受试者乐于把全部能力发挥出来，或把其他特征表现出来。

③ 测验的记分：主试者应该熟悉每一测题的记分方法，记分标准及其实例。记分必须避免“月晕效应”，对每一个反应，应该只就这个反应本身来评量。不能以成见影响记分。

④ 要懂得如何解释测验分数：一般而言，不要把测验分数告诉家长和受试，而只须告诉测验结果的解释。如IQ（智商）是100左右，就说他的智力一般；IQ是120或130，就说只要好好努力，他没有什么不能做，IQ在90与80之间，就可说对他的要求可以低一些，但多作一些鼓励也可以对某些工艺性的工作做得较好。

3. 测验的常模是某一标准化样组在一定时空中实现的平均成绩，地区不同，常模也就不同，经过相当年限以后，社会经济文化有了发展，常模也就要变动。为此不能把一个原来良好的测验，不分时间、地区到处乱用。

4. 主试者要对受试者在测验实施中的反应和行为做出切实而仔细的记录。如身体健康；有否心理上、情绪上的疲劳；恐惧的心理和决定命运的预感存在与否。施测的时间、地点都要注明。这些都是为了对受试者负责，并给最后决定者参考。

5. 要正确地使用测验，就要对测验本身有一个正确的认识。不能认为测验分数对于人们的能力和人格提供了非常完善的指标。这就是说，不能仅仅依据测验结果来评判，而应该参照其他评判的标准一起考核，而后再作出决定。

6. 一个好的测验工作者要有测验道德，不能利用测验作为压制人的工具，对人不平等的理由，也不能有私心。

赖忠恒，《测验的性质·作用·种类》，

摘自《教育科研情况交流》

1984年第2期

### 三、如何衡量考试效果

美国密执安州立大学教育学副教授菲茨杰拉德博士对如何衡量考试的效果，从学生、教师、学校、家长的角度，提出了以下若干值得研究的问题：

#### 一、学生方面

- 1 学生是否害怕考试？如果害怕，那考试在学习中有什么影响？
- 2 是根据考试的成绩来评定一个学生，还是根据老师或其他学生？
- 3 学生是否知道自己的考试成绩？他们是否应该知道？
- 4 考试对学生思想的发展有什么影响？
- 5 学生阅读、数学和其他科目的成绩和兴趣之间的关系如何？
- 6 考试是否和学生的能力有联系，和学生的不同种族或社会经济背景有关？
- 7 考试成绩不好的学生，他们的能力和其他特点发展得如何？
- 8 给在考试中获得低分数的学生开办补习班，是否加剧学校的问题？

#### 二、教师方面

- 1 要花多少时间出考题、进行考试、评分和解释有关问题？

2 学校课程中哪些重要方面没有考到或考到了？为什么？课程是否缩小到可以考的程度？考试对课程安排和教学内容的选择有何影响？

3 考试是用何种方式来促进或妨碍教学工作的？

4 教师是否理解出考题和评分的积极和消极的方面？

5 教师对考试的态度如何？他们的态度会不会影响到学生？

6 考试是否会影响师生关系？影响教师和领导的关系？怎样影响的？

7 教师工作的成绩是否可用学生考试的成绩来评定？如按照这一条例，有何意义？

8 考试成绩是否影响教师对学生的态度？

9 考试标准中确定的起码要求，老师是否都教了？

10 教师是否教考试的技巧？这些技巧是否重要？

11 考试是否提供很多在课堂上老师不了解的儿童的行为状况？

### 三、学校方面

1 考试是否能衡量出学校要达到的教育目标？

2 学校和学区的好坏，是否依据考试的结果来评定？

### 四、家长方面

1 考试对于社区是否信任学校有什么影响？

2 家长是否懂得考试的价值和局限性？

3 家长对自己孩子的潜力的看法和考试的成绩有什么

么影响。

原载《教育科研情况交流》1982年第7期

#### 四、国外中小学成绩评定的研究

##### (一) 美国高校教育评定

美国学校教育评定的基本要求是：第一、评定要有连续性和定期性；第二、要利用具有效度与信度的测量工具，作出客观而正确的评定；第三、评定要综合地全面地抓住学习者的成长与发展；第四、评定要为确定具体的教学目的和改进学生的行动模式，发挥重要作用；第五、综合性的评定要有助于社会广泛认识该学校的教育程序的价值。

美国多数中小学采取一律升级的制度。平时测验和考试较少。从小学升入初中不考试；从初中升入高中也不是通过考试淘汰学生，而是根据平时学业成绩的考查和学力适性测验（SAT）一类的资料，参照个人志愿，分别指定升入学术科、职业科或普通科。

为了促进教育质量的提高，佛罗里达州率先以州立法的形式规定全州中学生的升级或毕业的最低限度的学力标准。1975年，该州颁布法令，责成州和地方教育董事会协同制定测验标准，

实行全州统考。测验对象是三、五、八、十一年级，测验科目限于国语、数学。其中特别严格的是高中毕业班（十一年级）的统考。该测验大体以八年级程度的读写试题组成。由州统一命题，统一评分。凡不及格者，不发毕业文凭。对于这种做法，有毁有誉。但多数人认为，这是一个良好的开端。虽说测验在考试方法上、实施的强制性上存在问题，不过，统考的结果暴露了教育上的薄弱环节。该州在实施第一轮统考后，随即在翌年度的教育预算中，增拨了2600美元作为补偿教育费，让学校对学力差的学生采取提高学力的对策。佛州教育署长R·D·特林顿乐观地认为，1978年统考成绩比1977年有了提高。这是州采取了补偿教育的措施，浓厚了学术空气的成果。他相信，这个计划将会对教育质量的提高，产生积极的影响。

这个计划陆续得到了全国三十多个州的响应。测验实施各州有所不同，但有两点是一致的：①以考察读与算基本技能为主要内容；②为学力低下者提供补偿教育的机会。

### （二）西德学校教育评定

在西德学校，一年有3-5次填发“联系手册”，目的是让家长了解自己子女的发展与成绩状况。西德中小学的成绩评定，原则上是1953年以来实施的六级记分制。即：超、优、良、中、可、劣。

在西德，主要用学力成就测验来作出成绩评定。包括标准成就测验和教师编制测验。前者是以科教学和测验理论为背景，由专

家编制的。标准成就测验由效度和信度的问题组成，根据统计学原理制成全国性的标准。这种测验在期末和学年末进行。它的一大特点是借助全国学生的成就测验水平为媒介而制订的标准，能够对学生所处的水平作出比较。教师编制测验又叫非正式测验，在一个单元的中途或结束时，或是在期中期末就较短期间内的学习内容进行的。其特点是着重了解每个学生、全班学生掌握知识的趋势，以利改进教学工作。

这些测验的教育功能总起来说是：①用来调节教学过程，检查教学效果（包括诊断原有的学习基础，巩固学习的成续、判明难点所在，激发学习动机等）。②授于资格，决定升学和出路。③向学生和家长提供建议。④用于改革教学方法、课程乃至学校制度。

问题的关键在于成绩评定的客观性。那么，西德考虑了哪些提高成绩评定客观性的对策呢？归纳起来主要有：①要使学习目标明确化，不是抽象的教学目标，而是各门学科都应规定具体的可以测量的教学目标。在这方面，据说布鲁姆提出的教育目标分类（分为六类：知识、理解、应用、分析、综合、评价）很有用处。②要注意防止评定者（教师）的主观偏见。③要兼用性质不同的评定，凡是不能评分的领域，要利用评语之类的方法，记录每个学生的优缺点。除了学力评定之外，还要评定努力程度、进步状况。④要缩小家庭文化水平和语言水平上的差距。例如，设立每周二次的教师家长恳谈日，商讨指导家庭作业的问题。儿童的能力不是凝固不变的。因此，重要的是要着眼于促进儿童的人

格发展和智力发展。

### (三) 法国高校教育评定

法国中小学校的考试制度极其严格。学生的成绩全凭考试来评定。小学升中学本来也由考试选取。1959年以后改为在中学前二年设“观察期”，一般不必考试。中学毕业后的就业或升学，全凭各种各样的“证书考试”。

法国素来对考试有研究。比奈制订智力测验量表，也是用来诊断和预测学生成绩的。不过，真正对考试制度进行研究的，是皮埃隆。是他，研究了考试学的独立的学问体系。

考试学的早期研究批判成绩评定的方式，揭示传统考试缺乏效度和信度，特别是评分极不一致。例如，皮埃隆把同一篇作文，采用20分满分制，让七十六名教师分别作出评定，结果情形如下表：<sup>①</sup>

0—1分	1名	8—9分	10名
2—3分	6名	10—11分	3名
4—5分	20名	12—13分	2名
6—7分	34名		

传统的成绩评定之所以产生误差，主要有三个原因。

第一、评定者的看法不一。

第二、出题的偏差。这是考试评定中必然存在的问题。这是由于教育影响的心理学性质和标准测验的数理性质之间的矛盾。



生的。

第三、学生的能力本身每日都在发展变化，这一点是不能忽视的。本来，成绩评定的一个重要作用是预测学生的能力。学生的能力不是恒定的，是不断变化的。这里提出了一个问题；不能单凭一次考试的成绩来决定学生的未来出路。

教育常常需要反馈。无论教师还是学生，都需要判明自身的行为可靠性的情报。在这个意义上，成绩评定在教学过程中应占有位置。不过还得用光凭考试不能得到的信息——测验中不能测到的敏感性、独创性，以及其他人格因素，通过长期的继续观察来补充。在现代法国学校中，尤其在观察指导课程中，学校心理学家的的工作格外突出，其理由也就在于此。

#### (四) 苏联学校教育评定

在苏联学校中谈到教育评定时不光指学生学习结果的检查评价，也包括学习过程的评定。苏联1966年出版的一本《教育学辞典》把“检查”界说为：“教学过程的一个组成部分。旨在控制学生的学习活动，评定学习效果。”

根据俄罗斯联邦教育部的命令，学生成绩的评定要注意：①内容的正确性，②回答的完整性，③应用于实际的能力，④判断、结论的独创性，⑤回答的连贯性、表现力。有的教育学家还提出要把口头表达与书面表达能力、误答的错误及其性质、完成课题的速度，作为评定的指标。

苏联中小学的“检查”，根据目的与实施方法，大体可分三种。