

北京语言大学汉语水平考试中心学术文库

张凯自选集

张凯 著

从「试」偶拾

现状与对策

汉语作为第二语言的
教学研究

汉语水平考试建设和计算机辅助教育

问学录

汉语测试与话语分析探究

张凯自选集

HSK与语言问题

HSK考试管理信息化及基于统计的考试质量控制

汉语测试初学集

语言测试与信息研究

汉语作为第二语言的测试研究

汉语测试探微

汉语测试与句法研究探微

北京语言大学汉语水平考试中心学术文库

张凯自选集

张 凯 著



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

图书在版编目(CIP)数据

张凯自选集 / 张凯著. —北京: 北京语言大学出版社, 2011. 6

(北京语言大学汉语水平考试中心学术文库)

ISBN 978-7-5619-3057-1

I. ①张… II. ①张… III. ①汉语—测试—文集②汉语—语言学—文集 IV. ①H1-53

中国版本图书馆 CIP 数据核字 (2011) 第 129356 号

书 名: 张凯自选集

责任印制: 姜正周

出版发行: **北京语言大学出版社**

社 址: 北京市海淀区学院路 15 号

邮政编码: 100083

网 址: www.blcup.com

电 话: 发行部 82303648/3591/3651

编辑部 82303647/3592

读者服务部 82303653/3908

网上订购电话 82303668

客户服务信箱 service@blcup.net

印 刷: 北京联兴盛业印刷股份有限公司

经 销: 全国新华书店

版 次: 2011 年 6 月第 1 版 2011 年 6 月第 1 次印刷

开 本: 880 毫米 × 1230 毫米 1/32 印张: 4

字 数: 124 千字

书 号: ISBN 978-7-5619-3057-1/H·11097

定 价: 16.00 元

凡有印装质量问题, 本社负责调换。电话: 82303590

目 录

语言测验的测度和精度	1
关于结构效度	23
测量是理论的组成部分——再谈构想效度	36
Performance 是“运用”还是“表现”	56
Rasch 模型考辨	73
参数不变性研究之一: Lord 1952 及其他	85
决策草率导致作弊——作弊谈之一	114
两个用户,谁骗了谁——作弊谈之二	117
独木桥和罗马——作弊谈之三	120
后 记	122

语言测验的测度和精度^①

零 引言

语言测验，通常被看成是第二语言教学（对外汉语教学）的重要组成部分。无论是专门研究语言测验的人、语言测验的开发者（专职的和非专职的），还是语言测验的使用者，甚至接受语言测验的被试，恐怕都想知道这样两件事：第一，语言测验测的是什么？第二，它是怎样测的？

在这里，所谓“测的是什么”，不是泛泛而论的语言能力或交际能力，而是指测量程序（亦即测量工具）直接操作的那个（或那些）东西，也就是“测度”。因此，上述第一个问题，更准确地说就是：语言测验的测度是什么？相应地，第二个问题可以表述为：语言测验是怎样实现的？接着，我们可以提出第三个问题：语言测验有多高的精度？本文要讨论的就是这三个问题。

本文讨论的虽然是语言测验，但论及的理论和方法，同样适用于心理测量和教育测量。我们将用“测量领域”来指称语言测验、心理测量和教育测量。

① 原载《语言文字应用》，2004年第4期。

一 测量的基本要素

1.1 测量的定义

在测量领域，广为流传的测量定义是 Stevens（1951：22）提出的：测量就是按规则给客体或事件赋值。罗德和诺维克（1968：7）把这个定义补充为“测量就是给实验单位的特定性质赋值”，Crocker 和 Algina（1986：3）认为，这个补充使 Stevens 的定义更加精确。Nitko（1983：5）的定义与此类似：“测量就是给人的特定属性或特性赋值，以此揭示人在被测属性上表现出的实在的关系”。Bachman（1990：18～19）的定义是：测量就是按照严格的程序和规则对人的特性进行量化的过程。Bachman 虽然使用了“量化”一词，但他所谓的量化指的也仅仅是赋值。

Zeller 认为 Stevens 的定义只反映了信度的要求，为了纠正 Stevens 的片面性，他提出，“测量可以定义为抽象概念和经验指标的联系过程”（Zeller，1985：340）。这个定义可以认为是基本正确的，但它并不能完全弥补 Stevens 定义的缺陷。《简明不列颠百科全书》（1985：203）的定义十分简洁：“测量是指确定数和量的过程”。从 Stevens 到 Bachman，他们都只在定义中强调了数——赋值，而忽略了量。

Stevens 式的定义可能造成了一个不良后果，即量化被等同于数字化。很多人以为，只要是用数字表示的，就是测量的结果。Zeller 对 Stevens 的批评含有这个意思，但他直接从效度入手考虑问题，所以并没有道出问题的实质。



1.2 测量的要素和量表

在测量领域，我们可能一直没有认真考虑这样一个问题：科学的测量由哪些要素组成？我们先来看一个简单的、直观的例子。假定有三个面包，一大俩小；大的那个重 200 克，小的两个重量相同，都是 50 克。现有三个人，年龄、饭量都一样大。问：这三个面包怎样分给这三个人才算公平？也许有人会说，这是小学二年级学生都会做的题。

别急，让我们按 Stevens 的测量定义来解这道题，看看会有什么结果。先给面包赋值，1、2、3，面包的连续体赋值为 3；再给需要面包的人赋值，也是 1、2、3，人的连续体赋值也为 3；结论：公平的分法是一人一个面包，因为每个人都得到 1。不行！即便是三个三岁的孩子，其中的两个也会提出抗议。真正公平的分法是：一个人得到两个小的，两个人各得半个大的。但这个做法和 Stevens 的定义是有冲突的。问题在于，Stevens 定义的测量中把数和量混为一谈了。数和量是两个不同的概念，等数未必等量，正如罗素所说，测量是“在某种量和某种数或所有的量和所有的数之间，建立一个唯一的相互符合的关系”（转引自 Torgerson, 1958: 13）。

1.2.1 测量的要素

严格意义的科学测量，可以定义为“估计或找出一定量单位的某个量值与同一属性的一个单位之间的关系”（Michell, 1997, 转引自 Kline, 1998: 41），这种测量的结果，“是数字和标准的乘积”，如“ 5×1 米”（哈瑞，1985: 79）。严格意义的科学测量中，包含了四个要素：实数、测度、零点、单位。

实数，这是任何测量都需要的，否则，我们拿什么去赋值呢？实数虽然是测量的要素，但它却是最不需要我们搞测量的人操心的东西，因为它是现成的（至少对非数学家来说是这样）。对于数字（即实数）我们只需要有一个清醒的认识即可：它是没有任何经验内容的。爱因斯坦（1949: 41）说：“数学仅仅涉及概念间的相互关系，而不考虑它们与经

验之间的关系。”Nunnally 和 Bernstein (1994: 8) 也说：“数学作为一个抽象的事业，不需要涉及实在。”数和数学是否涉及实在，数学家们有不同看法（克莱因，1980），我们只取其中的一种说法。Stevens 的测量在分面包时出了问题，正说明数是不涉及实在（面包）的。

测度 (measure)，在某种意义上说，是比实数更为重要的因素。在用英文写的测量文献中，measure 是一个常见的词。我们注意到，这个词是有歧义的 (Bachman, 1990: 51)，它很少用来指我们这里所说的“测度”。在缺乏词源学和数学史研究的情况下，我们暂且认为我们所用的“测度 (measure)”概念，是从测度论（数学的一个分支）借来的（很可能是测度论借测量之测度在先）。根据测度论，设 $\langle \cdot, F \rangle$ 为可测空间， μ 为定义在 F 上的非负函数，满足什么什么条件……则称 μ 为 F 上的测度（《现代应用数学手册》编委会，2000: 98）。这太难懂。幸好《简明不列颠百科全书》（1985: 201）有一个还算通俗的解释：

（测度是）数学中，对于不是由区间或矩形组成的某类点集而言的长度或面积的概念的推广。抽象地说，测度是使集合对应到数的任何一种规则，具有非负性和可加性，即两个不相重叠的集合的并集的测度等于它们各自的测度之和。

我们不用像上述定义那样作抽象的“推广”，只考虑未推广的概念之一：长度。任何视觉上可见的物体都有长度，如一条路、一根线等。长度是物体的一个属性，我们可以用测量工具对这个属性进行操作，也可以用数学运算对其进行操作。例如，我们可以把卷尺的一头和写字台的一头对齐，卷尺沿写字台的一个边展开，看写字台这个边上的另一头和卷尺的什么位置重合，然后读出卷尺上的数字，这是测量操作。我们还可以量出两个相同的长度，然后把这两个长度连接在一起，新的长度是原来一个长度的两倍，原来的一个长度是新长度的二分之一，这是用数学运算对长度进行操作。长度是一种测度。现在不妨这样来定义：测度就是可对其进行（施加）测量操作和数学运算的东西（属性、维度）。从更直观、更通俗的角度讲，要想测量，总得有东西可测，没有可测的东西，测量就谈不上；所谓可测的东西，



就是测度。所以，我们说，测度是测量最重要的要素。

零点，是测量的第三个要素。量写字台的长度，要从一端量起；量房屋的面积，要从一面墙的墙根量起；量一块麦田的大小，要从地头量起。写字台的一端、墙根、地头，都是测量的起点，也就是零点，测量的零点，对应于实数中的零。

单位，是测量的最后一个要素。一千克、一米、一秒，都是单位。单位是作为比较的标准的一个基本的量。迪昂（1914：132）说过，如果没有单位，测量所得的信息是很不完备的。比如说，有人告诉我，有两个长度，一个是5，另一个是10，那么，我也许有理由认为第二个比第一个长，至于长多少，就不知道了。如果你还告诉我，二者的单位都是“米”，那我就知道第二个比第一个长一倍；如果你说前者的单位是米，后者的单位是厘米，于是我知道，不是后者比前者长一倍，而是前者比后者长49倍。马赫（1926：328）说：“如果我们不得不把连续变化的物理条件或物理量化归为量度，那么我们首先必须选择比较的对象或测量单位，并拟定如何判断另一对象等同于标准。”其实在日常生活中，我们都知道单位的重要性，比如买菜的时候。

以上是测量所需的四个要素。当然，如果我们谈的是广义的测量的话，这四个要素不一定同时出现在一个测量中，这就和测量水平及量表有关了。

1.2.2 量表

“把数字和考生表现联系起来的过程，称为‘称量（量度）（scaling）’，这个过程的结果是一个分数量表（scale）”（Petersen, Kolen 和 Hoover, 1989：221）。根据测量所含上述要素的多少，以及测量结果所关联的数学系统的复杂程度（罗德、诺维克，1968：12），量表一般分为四个水平，测量水平由低到高分别是称名（名义）量表、顺序量表、等距量表和比率量表。

称名测量是最低水平的测量，它就是把对象分成不同的类别，如男人编号为1、女人为2（如身份证），足球运动员背后的编号等，都是称名测量的结果，都是称名量表。具有同一量表值的对象至少具有一个共同特征，而这个特征是有另一个量表值的对象所没有的。例如，

男人编号为1，就意味着编号为1的都是男人，不能是女人。只要保证所测子集是互斥的，称名测量的赋值完全可以任意的，如我们也可以给女人编号为1或其他数值（如3、5），男人为2或其他数值（如4、6），只要编号能把男人和女人正确地分开就行，也就是说，在称名量表中，任何保持子集互斥的变换，都是允许的。

顺序量表处在第二个水平上。顺序测量除了把对象分成互斥的子集外，还要在对象之间建立顺序关系，即大于、大于等于，或小于、小于等于这样的关系。假定有三个人，身高不同，我们可以用数字1、2、3给他们赋值，3表示最高，2居中，1最矮，这就是顺序测量。在顺序测量中，任何保序的变换都是允许的。我们可以把上例中的数字3、2、1分别改为3万、2千、1百，其意义不变。

等距量表要高级一些了。等距测量不仅使对象可以区分、可以排出顺序，而且还要表明对象之间的差距是相等的。要使测量达到等距水平，我们必须规定一个相对零点和测量单位。等距量表的值和实数有一一对应的关系（称名量表与顺序量表和实数没有这个关系）。量表值可以作加法运算，但不能作乘法运算。摄氏温标是典型的等距量表，10摄氏度加10摄氏度等于20摄氏度（使10度的水的温度再提高10度，得到20度的水），但20摄氏度并不是10摄氏度的两倍。在等距量表中，只有线性变换是允许的（如每一个数都加上或乘上同一个数）。

比率量表一般被认为是最高水平的量表。除了等距量表的所有特点外，比率量表还要求具有绝对零点。有了绝对零点，量表值就允许作乘法运算，亦即不同的量表值之间有比率（倍数）关系，因此，在比率量表中，只有乘法变换是允许的。长度、重量（质量）都可以表现在比率量表上。

罗德、诺维克（1968：13）还提出一种绝对量表，和比率量表不同的是，绝对量表的单位是绝对不变的（不能任意规定），因此，对于绝对量表来说，任何变换都是不允许的。绝对量表的例子如原子量。

二 测量和所测的变量

2.1 直接测量和间接测量

在测量领域里，人们经常提到一对概念：直接测量和间接测量，甚至有人常说，心理测量和语言测验基本都是间接测量。但是，说这话的人有时不太严谨，而听的人往往不认真对待。例如，一边有人说语言测验都是间接的（Baker, 1989; Bachman, 1990: 309），另一边就有人说有真实性的表现测验是直接测验（Davies, Brown, Elder, Hill, Lumley, McNamara, 1999: 47）。说语言测验（以及心理测验）都是间接的，人们无非以所测能力（如语言能力、智力）是潜在特质为由，并没有进一步论证；说表现测验是直接的，拿出的证据显得挺过硬：我（评分员）直接给他（被试）的口语或作文打分，这不就是直接测验吗？甚至有人会说，我根本不想推断什么作文能力，我只是给一篇作文打分，这总是直接测验了吧？

测量是直接的还是间接的，并不取决于我们的看法或认识。我们知道，测量是给某种属性赋值（尽管这个定义不严谨），这也就是说，测量操作完成后，一定有某个属性获得了量表值。只有当一种属性“可以作为一种可加的性质进行处理时，它才可以被描述成一种可测量的性质”（特拉斯特德，1979: 70）。那么，一个测量是直接的还是间接的，只取决于所测的属性。《简明不列颠百科全书》（1985: 204）说：“除非被测物理量可以直接被观测，否则一定要用其他可测量（liàng）来确定”，前者是直接测量，后者就是间接测量。所测属性是以变量的形式存在的，Stephens（1968: 118）把变量分成两种：直接变量和推断变量，前者是我们可以直接观察到的，后者的存在是我们推断出来的。根据这个区分，我们大致可以说，如果所测属性是直接

变量，测量就可能是直接的；如果所测属性是推断变量，那测量（如果可测的话）就一定是间接的。虽然变量（属性）可以分成两种，但属性还有可测和不可测两种可能。

2.2 两种变量四种状态

变量按是否直接分成两类，我们分别称为显性变量和隐性变量。两种变量又各有两种状态：可测和不可测，这可用图 1 表示出来。下面，我们一一说明这四种状态是否可以测量及理由。

		显性/隐性	
		显性、可测 (如长度、时间)	隐性、? (如语言能力?)
可测/不可测		显性、不可测 (如温度)	隐性、不可测 (如电阻)

图 1 属性的四种状态

直接可测的显性变量，在自然界中只有长度和时间两个，也就是说，只有长度和时间是可以直接测量的（哈瑞，1985）。这两种属性是我们可以直接感知的，也是可以用仪器（尺子和时钟）直接测量的，除此之外，整个自然界里再没有直接可测的了。

温度是显性属性，我们可以直接感觉到冷或热，但它却是不可直接测量的。温度不可直接测量，不是因为它不能被感知，而是因为我们不能根据感觉给它规定单位。由此可见，一种属性是否直接可测，还取决于是否可以规定出单位来。

电阻是隐性属性，而且不直接可测。电阻不可测的原因有两个，一是它不可直接感知，二是更不能规定单位。

现在，我们暂时可以这样说：显性的属性，如果能够规定单位，是直接可测的，如不能，则不直接可测；隐性的属性，都是不可直接测量的。但是我们也知道，温度和电阻现在是可以测量的。对不直接可测的属性的测量是怎样实现的呢？



三 测量中心或测度

Stephens (1968: 118) 把变量分成直接的和推断的两种, 相应地, 他又提出“直接测度 (measure)”和“指标 (index)”分别对应两种变量, 罗德、诺维克 (1968) 也使用了这一对概念。所谓直接测度, 就是由观测变量构成的连续体, 而指标是由推断变量构成的连续体。“如果被测量的是一个观测变量, 那我们就有了一个直接测度。如果你想‘测量’一个推断变量, 那你的测量工具只能给你显示一个指标” (Stephens, 1968: 118)。

如果不说明“指标”是相对于直接测度而言的推断变量的连续体, 这个有特定含义的概念是很难理解的, 尤其是翻译成中文以后。我们不妨用“推断测度”一词来代替“指标”, 这样, 观测变量—推断变量、直接测度—推断测度正好是两对对应的概念。推断测度也可以称为“间接测度”, 这也是名副其实的。

我们知道, 要测量, 就一定要有测度, 而且一定要有直接测度。如果想测量一个推断变量, 我们的选择有两个: 要么放弃测量推断变量的想法, 要么找一个直接测度, 通过对这个直接测度的测量, 间接地去测量那个间接测度。Torgerson (1958) 提出“三个测量中心”, 这为我们理解和解决间接测量问题提供了一个很好的概念框架。

3.1 重新定义三个中心

Torgerson (1958) 认为, 测量有三个不同的中心, 这三个中心是: 以主体为中心、以刺激为中心、以反应为中心。“三个中心”的概念是很有意义的, 但他的定义和解释是有问题的。本文由于篇幅的限制, 不详细讨论 Torgerson 的定义, 而是以物理测量为例, 重新定义三个中心的概念。

以主体为中心是最简单的测量。主体对刺激所作反应的方差归因于主体的差异，被赋值的只有主体（Torgerson, 1958: 46）。在物理测量和日常生活中，长度测量是典型的以主体为中心的测量。长度测量的实质是用一种属性（长度）去测同一种属性，因为长度是显性的属性，而且可以规定1米为基本单位，因此它是直接测量。以主体为中心的测量并不都是直接测量，但长度测量却是以主体为中心的最典型、最容易理解的例子。

在以刺激为中心的测量中，被赋值的只有刺激（Torgerson, 1958: 46）。电阻的测量是以刺激为中心的测量。电阻是导体的一种隐性属性，不可直接测量。我们现在知道，当一个特定强度的电流通过一有电阻的导体后，电流会发生变化，根据欧姆定律，电流、电压、电阻三者之间有一种特定的数量关系，所以，我们可以利用这种关系间接地测出电阻。具体地说就是，把测得的电流值和已知的电压值带入欧姆定律的公式，然后推算出电阻的值。在这个过程中，欧姆计测到的根本不是电阻，而是电流，也就是说，被测量工具赋值的不是电阻，而是电流。电流的每一个变化都对应于一个特定的电阻值，事先把这些值换算好，标在欧姆计的刻度盘上。自始至终，电阻都是一个潜在的变量，只是因为它和电流之间有数量上的对应关系，所以我们可以很有把握地通过直接测量电流而达到间接测量电阻的目的。

在以反应为中心的测量中，被赋值的应该只有反应，用水银温度计测温度是典型的以反应为中心的测量。它利用的原理是，温度的均匀上升或下降，会导致水银柱均匀地变长或变短。在一定范围内，温度每上升一度，水银会增加一个单位的长度；反过来，温度每下降一度，水银则缩短一个单位。水银的伸缩，是对温度变化的反应，所以，这种测量是以反应为中心的。在这个过程中，被赋值的只有水银的长度，而温度的值是从长度值转换过去的。

以刺激为中心的测量和以反应为中心的测量能够成立，是因为刺激或反应与要测的属性之间有特定的数量关系，亦即观测变量和推断变量之间有特定的数量关系，这种关系允许我们通过给直接测度赋值达到刻画推断测度的目的。

由于测量中心不同，亦即测量并不都是直接的，测量所达到的水

平也是不同的。以主体为中心的测量可以达到比率水平，另外两种方式就不一定能达到比率水平了。例如，温度测量只能达到等距水平，因为温度不能定义绝对零点。

现在我们知道，科学测量并不都是直接的，这是因为所要测量的变量不都是直接的，当所测的是推断变量时，我们就必须要用间接的方式去测。在间接方式中，究竟是选择以刺激为中心还是选择以反应为中心，还要看刺激和反应的属性哪一个和观测变量有数量关系。以上说的是物理测量，现在我们可以根据物理测量考虑语言测验的问题了。

3.2 语言测验中的变量

要问语言测验测的究竟是什么，首先要看它涉及的变量有哪些。语言测验中的变量有三个：语言能力（主体）、刺激、反应。

语言能力，无论怎样定义，起码是很多语言测验声称要测的东西，它是语言测验的兴趣所在，是语言测验中的主体（这和 Torgerson 所说的“主体”意思不太一样）。但我们也知道，所谓的语言能力是一种潜在特质，是不可直接观察、不可直接测量的。因此，语言测验要想以主体为中心，是不可能的。

题目（包括客观题和主观题）是语言测验中的另一个变量，它是作为刺激物出现在测验中的。题目呈现给被试，就是给被试一个或若干个刺激，这种刺激要求被试作答。作为刺激物的题目，是一个观测变量。

被试的答案（包括多项选择答案和作文等）是被试接受刺激后作出的反应，它在很大程度上反映了被试的语言能力。被试的答案也是一个观测变量。

3.3 语言测验的测量中心

语言测验中，这三种变量哪个能作为测量中心，是我们接着要考

虑的问题。语言测验以主体为中心的可能性，我们已经排除了。剩下的可能性还有两个：以刺激为中心或者以反应为中心。

语言测验也不能以刺激为中心，理由有二。第一，测量电阻时，导体的电阻使输入电流和输出电流之间产生了差异，也就是电阻使电流出现了变化，测量这个变化，就是测量了电阻。语言测验中类似的变化是不存在的。测验题目，作为对被试的刺激，事前是那个样子，事后它还是那个样子，它没有因被试对它作出了不同的反应而出现不同的变化。因此，即便语言测验的刺激是可以测量的，就是测了，也没有意义。第二，刺激虽然是一个观测变量，但它却是不可测量的。如果把一个题目看成一个个体，我们无法确定它的权重，若干题目权重是否相等也是无法确定的。如果把测验题目看成一个连续体，我们无法为其规定单位，因为权重可能不同，所以任何一个题目都没有资格做基本单位。是否可以把题目难度作为测量单位呢？题目难度是依赖被试的，不是绝对的量，仍是不能确定其权重的。语言测验的题目和最小感觉差异测量使用的刺激物是不同的，最小感觉差异是以刺激为中心的测量，它测量的是重量，而重量的测量是不依赖人的感觉的（车文博，1998）。

现在只剩下被试答案，即反应了。如果我们能够根据对反应的测量推断潜在的能力，也是可以满足的。但这暂时也是做不到的，因为反应的性质和刺激类似，它的权重也是不确定的，被试作出了若干反应，没有什么东西保证这些反应是相等的。被试的反应和题目难度是互相依赖的，这是其一；被试的反应不全是他能力的反映，其中会有其他因素掺杂进来，我们称为随机误差，这是其二。这两个因素决定了反应的权重是难以确定的。

我们可以用人们熟知的情况说明这个问题。一个由多项选择题构成的测验，每个题的分值是1分，每答对一题给1分。看上去好像题目权重相等，被试的正确反应的权重也相等。其实不然。我们每个做过被试的人都有这样的体会：一个把握不大的题，我有可能答对；而一个把握很大的题，我也有可能答错。通过仔细的分析，我们可以看到，答对和答错并不是简单的两种情况，其中包含的可能性一共有五种——三种情况下会答对，两种情况下会答错。如果我的能力正好可

以答对这个题，我应该得1分；如果我的能力答这个题有富余，给我1分就是给少了，因为这个1分不足以反映我的能力，应该给我比如说1.2分；如果我的能力不足以百分之百地答对这个题，但运气好，居然答对了，这时候给我1分，那是给多了，应该给我比如说0.8分。再看答错时的情况：如果我没有可能答对一个题，给我0分，我一点儿怨言没有；但在我有可能会答对而因马虎等原因没答对时，也给我0分，岂不是有点儿冤？这时应该给我比如说0.3分？

这种复杂的情况说明，语言测验的反应本身是很复杂的，因此，尽管在评分规则中只能规定一题1分，但如果因此认为1分恰好代表一个单位的反应，那就缺乏根据了。正如 Angoff (1984: 5) 所说，原始分量表是没有固定的意义的。

我们现在几乎被逼上绝路：语言测验既不能以不可见、不可测的主体为中心，又不能以可见却不可测也不反映主体属性的刺激为中心。反应倒是可见的，也部分地能反映主体，但它却又是难以量化、难以确定单位的。于是我们面临两种选择，要么彻底放弃测量语言能力的努力，要么针对反应，再想想办法。

四 虚构测度——奎特勒方式

反应是可见的，是一个观测变量，但是，它不能构成一个可测量的测度。反应构不成测度的原因是，每一个反应的量都可能是不确定的，因此没有一个量可以定义为标准量。这时候，如果希望把这个不可测量的变量变成可测量的测度，我们可以使用统计方法，这种方法我称为“奎特勒方式”。关于奎特勒和奎特勒的思想（参见 Stigler, 1986），限于篇幅，我们不作介绍。

所谓奎特勒方式，其实就是我们熟悉的常模方式，常模是搞测量的人都熟悉的。之所以重提奎特勒，是因为我们往往忘记了常模的本质意义。我们现在对常模的关注，几乎只在于用它导出测验分数，用它估计误差，甚至误以为常模是一种标准（如 Glaser, 1963 等），而