

# 第一章 语言测验的 基本原理

本书的名字叫“语言测验理论与实践”，也就是说，它讲的是语言测验。“测验”，我们往往喜欢把它叫做“考试”。在日常生活和学习中；“期中考试”“期末考试”“汉语水平考试”“托福考试”，都是我们常常提到的。但在这里，我们为什么不叫“考试”而叫“测验”呢？这是我们首先要解释的一个问题。

## 第一节 语言测验的作用和目的

中国是考试的故乡，考试在中国已有 2000 年以上的历史。说到考试，我们就会想起科举考试。科举考试和我们现在说的测验是有区别的。科举考试的目的只有一个：选拔人才。现代的语言测验的作用与目的和传统科举考试是不一样的。语言测验的作用只有一个：科学地测量出学习者的语言能力。语言测验是一种测量工具，和任何测量工具一样，它有自己的专门的作用。比如说，秤的作用是

称出重量，尺子的作用是量出长度，手表的作用是量出时间。语言测验也一样，它的作用就是测出“语言能力”。

测量工具的作用和测量目的是有区别的。所谓测量目的，是我们使用测量工具时要达到的目的。秤的作用是称重，但为什么要称重呢？在市场上，称重是为了公平交易；在医院里（称体重）是为了了解健康状况；在（举重的）运动场上，是为了找出力气最大的人。由此可见，“称重量”在不同的情况下是有不同的目的的。语言测验的作用只有一个，即测量语言能力，那它的目的有哪些呢？换句话说，我们为什么要知道人的语言能力呢？在以下四种情况下，我们需要知道人的语言能力。

## 一、选拔

这里所说的“选拔”是广义的。中学生考大学时，大学需要知道他的外语或汉语水平；进了大学（或任何高一级学校）的学生可能是学过第二语言的，但他们的第二语言水平可能是不同的，为了有效地组织教学，我们可能要把水平不同的学生分到程度不同的班级，如高、中、低班；学生毕业的时候，我们可能要知道他的第二语言水平是否达到了毕业所要求的水平；学生走出学校找工作，用人单位一般要挑水平高的人。以上这些都是所谓的“选拔”，即在人群中选拔（第二语言）好的。在各种各样的选拔中，语言测验为我们提供了必要的信息：谁语言水平高，谁语言水平低。这是使用语言测验的目的之一。

## 二、诊断

在教学过程中，我们需要知道学生的语言水平是否有提高。经过一个学期（或一个月）的学习，学生的语言水平提高了吗？班里所有人提高的幅度都一样大吗？是否有人退步？这样的信息也要由

语言测验来提供。这个时候，我们的目的不是选拔：让水平高的学生赶快毕业、水平低的学生退学。不是的。这时我们的目的是通过测验来了解学生需要哪些帮助，就像医生给病人诊断一样，要知道他得了什么病，好决定给他吃什么药。这是使用语言测验的目的之二。

### 三、评估

一个教程学完，学生成绩如何？哪些合格，哪些不合格，都需要根据测验做出评估。评估不总是针对学生的。有的学生进步快，有的学生进步慢，难道都是学生的责任吗？不一定。也许教材的内容对一部分学生不合适，也许教师的教学方法不科学，也许教学内容和教学安排不合理。我们也可以根据语言测验提供的信息，对教材、教学法、教师的教学做出评估。我们也许会发现，学生的成绩不好，是因为有些测验内容是教学中没有教过的，而某些学生掌握得很好的东西，正是教师精心教学的结果。了解并评估学生成绩、教学或教材，是我们使用语言测验要达到的目的之三。

### 四、研究

为了研究语言本身，为了研究语言教学，为了研究学习过程，为了研究教材和教学法，我们都可能要用语言测验来为我们提供有益的信息，这是使用语言测验的目的之四。

总而言之，语言测验是一种测量工具，它能提供一些信息，我们（语言教师、语言学家）可以利用这些信息达到不同的目的，这正是现代语言测验和传统的语言考试之间的区别。换一个角度，我们也可以这样来解释。在传统的考试中，如果学生的成绩不好，那就都是他自己的责任；而从现代教育观点看，学生的成绩不好，可能有多种原因，可能有学生自己的原因，也可能有教师、教材、教

学法甚至测验本身的原因。要知道原因是什么，我们就可以让语言测验告诉我们。如果是学生的原因，我们想办法帮助学生提高成绩；如果是教师的原因，我们要提高教师的水平；如果是教材的问题，我们要编写出更好的教材；如果是教学法的原因，我们应该使用更有效的方法；如果测验本身有问题，那就要改进测验。正是在这样的意义上，我们说现代语言测验和传统的语言考试是不同的。所以，本书用“语言测验”这一术语，而不使用“考试”。

## 第二节 怎样用科学方法 测量语言能力

我们已经知道，语言测验的作用只有一个，即测量人的语言能力，而语言测验的这个作用，可以为各种目的服务。接下来的问题是：语言能力是什么？语言能力是一种抽象的心理能力，它看不见，摸不着，不像长度那样可以直接测量。本书不打算集中讨论语言能力问题。我们现在要弄清楚的是，怎样才能用科学方法去测量语言能力这种抽象的东西？

语言测验的作用是测量人的语言能力，那么语言测验的对象就是人的语言能力，人的语言能力是客观存在，但又是抽象的。如果要科学地测量出这种抽象的能力，我们就需要对测量对象的性质有所了解，也需要对测量的要素有所了解。

### 一、测量的定义

如果有人问：什么是测量？对这个问题，我们也许不能给出一

个满意的答案。那么好，我们就在日常生活中寻找这个问题的答案。

在日常生活中，我们几乎每天都接触到测量。到市场上买东西，我们要买一公斤肉、一公斤菜、一公斤水果，于是卖东西的人就给我们称出一公斤肉、一公斤菜、一公斤水果。卖东西的人所做的事，就是测量。我们要做一件衣服，就要到商场去买衣料，售货员会按我们的要求，量出两米衣料，他所做的，也是测量。我们盼望自己的孩子赶快长大，于是便经常给他量身高，这还是测量。

在学习、工作和科学研究中，我们也经常要用到测量。在体育课上，我们测量学生的100米短跑的速度。一个车工，每天要用卡尺测量工件的尺寸。在科学研究中，我们要测量很多东西，如太阳和地球之间的距离、汽车或飞机的速度等等。这些都是测量。

我们每个人都会在不同程度上使用一些测量工具（如尺子、弹簧秤、钟表），都会进行一些测量活动（如测量长度、重量、时间）。我们能把各种各样的测量概括为一句话吗？或者说，我们能给测量下一个定义吗？让我们试试。

我们来分析一个具体的测量，看看在这个过程中我们做了什么，最后又得到（知道）了什么。我们的例子是长度的测量。

假定有一根木头，要测量它的长度，我们做些什么呢？首先是找到一把尺子，如一只卷尺，卷尺上有刻度：米、厘米、毫米等。然后，我们把卷尺上标有“0”的一头和木头的一端对齐并固定住。接着，我们沿着这根木头把卷尺拉直，看木头的另一端和卷尺的什么位置是对齐的。最后，我们看到，卷尺上标着“3米”的地方是和木头的另一端对齐的，于是我们知道，这根木头的长度是3米。整个测量过程到此结束。

在这个过程的开头，我们不知道木头的长度，而在过程的结尾，我们知道了它的长度，因此，就量木头这件事来说，这是一个确定长度的过程。现在我们可以这样来定义测量：测量是确定长度的过程。

长度是用两个东西来表示的，一个是数字，如“3”，另一个是量，即“米”。我们可以用这两个东西来代替上边的“长度”，把测量定义为“确定数和量的过程”。但显然，这个定义还缺少某种东西。

如果有人对我们说：你去测量吧。根据这种指令，我们能做什么呢？什么也不能做，因为我们不知道测量什么东西。假如又有人说：你去测量木头吧。这个指令好像具体了一些，但我们仍然是什么也不能做，因为我们不知道是测长度、重量，还是密度。只有当他说：你去测量木头的长度或重量时，我们才知道如何去做。“长度”或“重量”称为事物的“属性”。于是我们说，测量是给人或事物的某种属性确定数和量的过程。

我们可以用各种熟悉的测量过程去检验这个定义。我们会发现，测量长度、重量、面积、体积、时间、温度等等，都是确定某种属性的数和量的过程。在这些过程中，属性分别是长度、重量、面积、体积、时间、温度；通过测量，我们得到一些数字，还得到了一些量，如米、千克、平方米、立方米、小时、摄氏度等。

## 二、测量的四个要素

现在我们知道，测量是一个过程、一种活动，在这个过程中，我们给某种属性确定了数量。但我们是否想过，在测量活动中，我们需要哪些条件，才能给属性确定数量？换句话说，一个测量过程包含了那些要素？回想一下长度测量，我们立刻会想到，测量要有对象，即可测量的东西，还要有测量工具，比如卷尺。我们再看，测量工具（卷尺）上有什么呢？有毫米、厘米、米这样的单位，有1、2、3、4等数字，在卷尺的一头，还有0。我们可以把这些归纳为测量的四个要素：测度（测量的对象、可测量的属性）、测量单位、数字（实数）系统、零点（0）。

## (一) 测度

测度是测量的第一个要素，这是不言而喻的。刚才说过，如果有人让我们去测量，却不告诉我们测量什么东西的什么属性，那他的话等于没说。反过来说，一定先要知道存在一个测度，然后才谈得上测量。那么，什么是测度呢？我们还是来看最简单的例子：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。

## (二) 数字（实数）系统

测量是给属性确定数和量，这表明测量离不开数字。所谓“数字”指的是“实数”，包括自然数（正整数）、负数和小数。数字系统是现成的，不需要再去创造，只要能正确地使用就行了。有一种观点认为，测量就是给属性赋值，这个定义虽广为流传，但它是不完备的。测量固然需要数字，但仅有数字，或仅仅是用一种什么办法给某种东西赋值，不一定是测量。我们可以用一个简单的例子来说明。

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。

如果说仅仅赋了值（确定了数字）就算测量，那事情就简单了。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。前一种分法之所以不合理，是因为它忽略了测量中的另一个要素——单位。

### （三）测量单位

数和量是两个不同的概念，等数未必等量。仅有数字，只能保证数的相等，但无法保证量的相等。要保证量的相等，我们必须找到标准单位。一人分一个面包，人们会觉得不公平，因为面包有大有小，面包的大小是由它的重量决定的。在分面包这件事上，我们需要做的不是让数字相等，而是让重量相等——每人都得到等重量的面包。要想达到这个目的，我们就需要有一个单位（标准单位），这个单位是“克”。如果我们使每个人都得到100克面包，我们对面包的测量才是科学、合理的。

类似的例子我们还可以举出很多。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。

数和量是有区别的，定数未必能定量；要定量，就要有单位。单位是一个基本量，它的作用是充当比较的标准。所谓单位，就是“一份”，“一份”是多少，是由单位表示的。在分面包的例子中，一份是100克。单位不是在自然界中选择的，就是人为规定的；归根到底，单位都是人规定的。

#### (四) 计算的起点(零点)

测量一个东西的大小、长短,一定要有一个起点。例如,一个人身高是 170 厘米,是说他从脚跟到头顶之间的长度(距离)是 170 厘米,计算的起点是脚跟。把两个东西放在一起比较,也要有一个共同的起点。一个面包重 200 克,另一个面包重 50 克。如何知道第一个面包比第二个重呢?我们先规定,一个面包都没有为 0 克,再规定面包的重量以 1 克为基本单位,引入数字系统后,可以知道,第一个面包包含了 200 个基本单位,而第二个面包包含了 50 个基本单位,两个面包的计算起点都是 0,所以,第一个比第二个重。这就是零点的作用。

如果没有起点(零点),我们可能就没有办法进行比较或测量。假定我们走在一条公路上,公路的旁边有一条铁路和公路并行,两条路都是望不到头的。我们能知道这两条路哪条长吗?显然无法知道。如果我们知道铁路的起点和终点分别是北京和乌鲁木齐,而公路的起点和终点分别是吐鲁番和乌鲁木齐,那我们就知道铁路比公路长得多。

测度、数字、单位、零点是测量的四个要素,其中“测度”是任何测量都必须有的,否则我们就不知道测的是什么东西。其他三个要素可能出现可能不出现,这取决于测量水平。而且,在不同的测量中,后三个要素的性质也可能有所不同。

### 三、四种测量水平和四种量表

测量是确定数和量的过程,这个过程的结果是一个量表,或者说,这个过程的产物是一个量表。

测量的四个要素并不总是齐备的,特定的要素在不同的测量中也可能有不同的性质。根据某测量要素的有无和特定测量要素的不同性质,我们可以把测量分为四个水平:称名测量、顺序测量、等距测量、比率测量,相应地也有四种不同水平的量表:称名量表

(也叫名义量表)、顺序量表、等距量表(也叫等距区间量表)、比率量表。

### (一) 称名测量

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)、足球运动员身后的编号等,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。

### (二) 顺序测量

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。

### (三) 等距测量

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℃ 加 10℃ 等于 20℃(使 10℃ 的水的温度再提高 10℃,得到 20 的水),但 20 并不是 10 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)。

#### (四) 比率测量

比率量表一般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。

四种水平的量表性质见表 1-1。

表 1-1 四种量表及其性质

	称名量表	顺序量表	等距量表	比率量表
区别意义	+	+	+	+
有序性	-	+	+	+
等距区间	-	-	+	+
绝对零点	-	-	-	+

#### 四、测量对象及其性质

测量能够达到什么水平，不完全取决于我们的需要或愿望，而是取决于测量对象的性质。对于一些对象，我们能够以较高的水平来测量，从而得到水平较高的量表，对于另外一些对象，我们只能以较低的水平测量，而得到水平较低的量表。

对于人的感官而言，所有的对象（包括各种物质、人或物质的属性）可分为两类：一类是显性属性，另一类是隐性属性。显性属性是看得见、感觉得到的；隐性属性则是看不见、感觉不到的。

物体的长度、人的身高是看得见的；物体的重量、温度是感觉得到的。这些都是显性属性。电阻、人的智力和语言能力是看不见也感觉不到的，这些属于隐性属性。

对于测量而言，各种属性又可以分为两类：一类是可直接测量的，另一类是不可直接测量的。在整个自然界中，只有长度和时间是可直接测量的属性，其余属性都是不可直接测量的。

在测量中，我们把测量的对象称为“变量”。根据上面的划分，变量可以分成两类：直接变量和推断变量。只有当所测属性是显性的、直接可测的，它才构成一个直接变量。显性属性并不都是可以直接测量的，如温度，我们不能根据感觉说出一个物体的温度。不可测量的显性属性和所有的隐性属性，我们都称为推断变量。

## 五、测量工具

测量所涉及的变量分为两种——直接变量和推断变量，相应地，测量和测量工具也分为两种：对于直接变量，我们可以使用自我测量工具对其进行直接测量；对于推断变量，我们只能使用非自我测量工具对其进行间接测量。

直接变量只有两种——长度和时间，因此自我测量工具也就只有尺子和时钟两种。除了长度和时间两种变量是直接的以外，其他所有变量都是推断变量，因此除了尺子和时钟以外的测量工具都是非自我测量工具，语言测验自然也是非自我测量工具。

## 六、测量中心

由于所测变量性质不同，我们使用的测量工具的性质不同，测量活动本身也有所不同。测量活动的不同表现为不同的测量中心——以主体为中心、以刺激为中心、以反应为中心。

### （一）以主体为中心

以主体为中心的测量就是直接测量。当所测变量是一个直接变量时，即该变量既是可见或可感的，又是可测的时候，我们就可以把这个变量当作可测的测度。所测的东西，正是要测的东西，这也就是以主体为中心。所谓以主体为中心的测量，就是用一种性质去测量同一种性质，或者说是测量工具本身所显示出的性质和它所要测量的是同一种性质。凡是以主体为中心的测量，使用的测量工具

都是自我测量工具；反过来说，只有使用自我测量工具，才能实现以主体为中心的测量。既然以主体为中心的测量是和自我测量工具相对应的，那么所有测量中，便只有长度和时间的测量是以主体为中心。尺子本身是有长度的，所以它是用它自己所有的一种性质去测量其他物体的同一种性质，即长度。钟表指针的转动是需要时间的，所以它是用自己具有的一种性质去测量其他事物的同一种性质，即时间。除了这两种测量之外，我们再也找不到以主体为中心的测量了。

既然只有长度和时间的测量是以主体为中心的，那么其他所有测量就都不是以主体为中心的了。既然如此，其他测量以什么为中心？其他测量不是以刺激为中心的，就是以反应为中心的。

## （二）以刺激为中心

以刺激为中心的测量是一种间接测量。以刺激为中心，顾名思义，就是测量工具所测的东西并不是我们想测的东西，而是一种刺激。我们对一种事物施加刺激，然后对这个刺激进行测量；这个刺激会导致事物产生某种反应，这种反应反映了事物的某种属性，通过对刺激的测量，我们可以有效地推断出事物的这种属性。电阻的测量是典型的以刺激为中心。

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，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。

### （三）以反应为中心

以反应为中心的测量也是间接测量。当事物受到刺激时，它可能会产生某种反应，这种反应反映了事物的这个属性。当刺激不可测量时（即不能以刺激为中心时），我们可以通过对反应的测量来推断事物的那个属性。用水银温度计测温度是典型的以反应为中心的测量。

温度是物体的一种属性，我们可以感觉到，但无法直接测量。我们知道，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。在这个测量中，我们通过对水银柱长度的测量，去推断物体的温度。

### （四）测量水平和测量中心

测量水平就是测量达到的水平。测量能够达到什么水平，和所测的属性（就直接测量而言）或所要推断的属性（就间接测量而言）有关。直接测量可以达到比率水平，因为我们既可以明确地规定单位，也可以明确地规定测量的零点（测量的起点）。在间接测量中，由于所要推断的属性不是直接可测的，我们往往难以规定测量单位或零点，因此，间接测量有可能达不到比率水平。

水银温度计只能达到等距水平，因为用水银温度计测温度只能规定基本单位，但却找不到绝对的零点。只要是物体，就有温度，我们不知道“没有温度”是个什么概念，因此“温度”没有绝对零点。在这种情况下，我们可以规定水结冰的温度为零度，标准大气压下水沸腾的温度为 100 度，然后再把零度和 100 度之间的水银柱长度划分为 100 个等份。于是，用水银温度计测量温度，就是只有相对零点和相等单位的等距测量。

### (五) 语言测验以什么为中心？

现在我们来考虑语言测验。我们说过，除了长度和时间的测量外，所有测量都是间接测量，这也就是说，语言测验也是间接测量。既然语言测验是间接测量，那它是以什么（刺激还是反应）为中心呢？

也许有人会争辩说，语言测验当然得以主体为中心。说这个话的人对“主体”一词的理解是有问题的。

有人把“主体”理解为人，语言测验测的不是一个一个的人吗？或者说，语言测验针对的不是人吗？因此语言测验是以主体为中心的。这个说法是不对的。语言测验测的不是整个人，而是人的一种属性——语言能力。类似地，身高测量测的也不是人，而是人的一种属性——身高（身体的长度）。人是一个极为复杂的有机体，任何一种测量工具都不可能测量整个人，而最多只能测量人的某一种属性而已。

也有人会说，语言测验测的是语言，因此它是以主体为中心的。这就涉及到了“语言”这个词的复杂含义。“语言”至少有两个意思，一个是指说出和写出的话的总和，如小说、报纸、杂志、论文等文字材料，以及广播中的新闻、故事等，也就是书面语和口语的总和。传统语言学研究的正是这个东西。我们知道，说出的话和写出的话，是语言能力的产物，不是语言能力本身，所以，把书面语和口语当成“语言”，并认为它们是语言测验的中心，是不对的。

语言测验所说的“语言”完全是另一个意思，它指的是人的和语言有关的心理活动，也就是使人能够说出或写出一些话的那种心理能力，我们把这种能力叫做语言能力。所谓的书面语和口语，只是这种能力的反应或产物，而不是能力本身。就好像水银膨胀是对温度变化的反应一样，水银或水银的变化并不是温度。语言能力和人的任何其他心理能力一样，都是大脑的活动，是看不见、摸不着、不可直接测量的。这样说来，语言能力（语言）虽然是语言测

验的主体，但语言测验却不能直接以它为中心。

也许我们可以认为，语言测验是以刺激为中心的。题目是对被试的刺激。客观性语言测验，如 HSK，有若干（比如说 100 个）题目，考试手册说明，一个题目 1 分。既然题目是刺激，而题目又是可以标明分值的，这不就是以刺激为中心吗？其实不然，题目固然是刺激，题目也可以标明分值，但题目在测验前和测验后并没有发生变化。我们可以回想电阻测量，在测量过程中，电流在电阻的作用下是有变化的，我们正是根据这个变化推算出电阻的。

另外一个思路是，题目看起来是可以赋值的，比如说一题 1 分。于是我们可以说被试答对一题，我就给他 1 分。这有点儿类似于扛麻袋，每个麻袋都是一样大、一样重的，只要你能把一个麻袋扛上车，我就给你五块钱，谁扛得多，谁挣钱就多。但这个思路在语言测验中是行不通的，因为语言测验的题目乃至任何一级单位的语言材料（如汉字、词汇、句子、段落、篇章）都未必是难度（或分量）相等的，而且我们说不清楚两个看似相当的语言材料，谁轻谁重。举个最简单的例子：两个汉字，“一”和“二”，仅就书写而言，第二个字的难度比第一个大一倍吗？迄今为止我们无法回答这个问题。除非我们能用实验的办法弄清一大批语言材料及题目本身的分量（权重），否则我们就无法确认一题 1 分是合理的。

以上两方面的原因决定了语言测验不可能以刺激为中心，现在剩下的路只有一条：以反应为中心。

以反应为中心的思路是这样的：测验题目给被试一个一个的刺激，被试作出一个一个的反应，每一个正确的反应给予一定的分数（比如 1 分），每一个错误的反应不给分。

这实际上是和上述以刺激为中心的第二个想法是一回事，以刺激为中心不能成立，是因为我们无法知道哪些刺激是等量的。即便假定若干个刺激是等量的，我们也无法确认对这些等量的刺激的反应也是等量的。我们可以用一个每人都有过的经验来说明这个问题

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。看上去这好像题目权重相等、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，给我零分，我一点儿怨言也没有；但在我有可能答对而因马虎等原因没答对时，也给我零分，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。

因为反应中含有不确定因素，所以我们无法把被试的任何一个反应当作一个基本单位。

## 七、怎样使不可测的东西变为可测的

我们想测量人的语言能力，但是，在语言测验中，我们不能以主体为中心，因为语言能力本身不是直接的测度；不能以刺激为中心，因为刺激自始至终没有任何变化。现在，剩下的只有反应了，可它又不满足测量对基本单位的要求。如此说来，我们岂不是就没有办法了？事情没有这么悲观，我们还有一种办法，那就是“虚构”一个测度，在这个测度上，我们可以得到测量所需的单位。

我们来看一个直观的例子——身高。平常我们说张三个子高，