

20世纪心理学通览

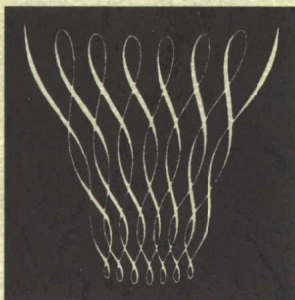
心理测验

Psychology

Psychological Testing

[美] 安妮·安娜斯塔西 苏珊娜·厄比纳 著

缪小春 竺培梁 译



浙江教育出版社

20世纪心理学通览

心理测验

Psychological Testing

[美] 安妮·安娜斯塔西 苏珊娜·厄比纳 著

缪小春 竺培梁 译

图书在版编目 (CIP) 数据

心理测验/ (美) 安娜斯塔西 (Anastasi, A.), (美) 厄比纳 (Urbina, S.) 著; 缪小春, 竺培梁译. —杭州: 浙江教育出版社, 2001. 8

(20 世纪心理学通览)

书名原名: Psychological Testing ISBN 7-5338-2580-2

I. 心... II. ①安...②厄...③缪...④竺...

III. 心理测验-研究 IV. B841.7

中国版本图书馆 CIP 数据核字 (2001) 第 041276 号

责任编辑 王晴波

封面设计 池长尧

责任出版 温劲风

责任校对 雷 坚

20 世纪心理学通览

心 理 测 验

[美] 安妮·安娜斯塔西 苏珊娜·厄比纳 著

缪小春 竺培梁 译

浙江教育出版社出版发行 (杭州市体育场路 347 号 邮编 310006)

杭州兴邦电子印务有限公司激光照排

杭州富春印务有限公司印刷

印数 1-5200

开本 850×1168 1/32 印张 25.625 插页 4 字数 580000

2001 年 8 月第 1 版 2001 年 8 月第 1 次印刷

ISBN 7-5338-2580-2/B·16 定价: 34.00 元

版权所有 翻印必究

“20 世纪心理学通览”序

心理学作为一门科学相当地年轻,但其源起,却可溯及人类思想的萌芽期。18 世纪之前的心理学思想,散见于古希腊和中国哲学家、伊斯兰教和基督教的官能心理学家、科学革命时期的机械论者等著述之中。到了 18~19 世纪,在前人研究的基础上,若干重大的理论和实践,包括联想主义、自然主义、实证主义、进化论、科学边缘的研究和实验生理学等,加速驱动着心理学从哲学中分化出来,为心理学成为一门独立的学科提供了基础。然而,心理科学的真正发展则是在 20 世纪。回顾一下心理科学在这百年间所取得的成果,不仅有助于人们认识心理科学自身的地位和作用,为今天的研究提供比较的范本,而且可以为 21 世纪心理科学的基本走向提供历史的经验和教训,促使心理科学更加健康地发展。我们出版“20 世纪心理学通览”的目的就在于此。

20 世纪的心理学出现了不少流派,成立了许多学术组织,出版了数量可观的研究文献,影响十分广泛。“通览”以 20 世纪心理学中具有重大影响的一派、一家、一人、一说为选题原则,涉及到的流派有元素主义、构造主义、机能主义、行为主义、策动学说、格式塔原理、精神分析、社会—文化—历史学派、认知理论、人本主义等。有些学派还可进一步分成若干家系。从每一流派或家系中有所侧重地选出一位创始人或代表人物的学说,从中

可以看到这些学说对 20 世纪心理学发展的影响,以及它们自身的地位和成就。当然,有些学说所属流派的划分并非十分严格,只能以思想观点大体相近为依据而暂时归入一个流派的名义之下。至于一些跨派别的学术著述,则根据其侧重的研究内容和方向,相应地归入人格、测量、人与社会等类别。

在“通览”中,每本译作都附有译序,旨在对倡导某一学说的代表人物及其学说在 20 世纪心理学中的地位、影响和价值做一基本的阐释,以说明为什么选择该书的理由。至于译序中对历史上同一具体问题的看法,译者之间可能存在一些差异,对此,难以强求一律。译序中的观点和结论由译者自负责任。此外,由于同一术语和概念在不同作者和不同时期的学说中表达的意思并非完全相同,所以,专业术语或概念的译名可能不尽一致,甚至同一人名译名,除约定俗成的外,也有可能不统一。对此,在主要术语和所有人名后都附有外文原文,以便读者辨析。

由于资料来源、外国语种、“通览”篇幅等方面的局限,加之我们经验的缺乏,所收的流派还不够全面;由于授权有先后,译作有快慢,很难依照流派活跃的年代、著述发表的年分、流派本身的结构、流派之间的关系等,有顺序、分类别地出版这部《通览》。对这些不足,我们将在具体的工作过程中适时地调整。对于这部“通览”为什么选定这位代表人物而不是那位代表人物,为什么收入某位代表人物的这部著述而不是那部著述等问题,读者可能会有不同的看法,对于来自各方的意见或建议,我们表示真诚的欢迎和感谢。

浙江教育出版社

中文版译序

一、关于书名和作者

本书的英文书名是 Psychological Testing, 我们译为“心理测验”。Test, 译为中文是“测验”, 而 Testing, 译为中文则是“编制测验”、“使用测验”。经常和“心理测验”一起提及的另一术语是“心理测量”, 那么, 这两者的关系如何呢? 心理测量是一个大概念, 心理测验则是一个小概念。心理测量包括心理测验, 心理测验是心理测量的方法之一。心理测量的方法, 除了心理测验之外, 当有其他许多方法, 例如观察、面谈、作品分析、个案研究等等。由于在心理测量中, 比较准确且比较可靠的一种方法, 非心理测验莫属, 在这种特定的意义上, 心理测验亦可以等同于心理测量。

和本书书名有关的另一个问题是“心理测量”和“教育测量”的关系。这两个术语的关系又如何呢? 一般来说, 在大学教育系开设的测量课程称为“教育测量”, 而在大学心理系开设的测量课程则称为“心理测量”或“心理和教育测量”, 在本系开设的测量课程之前冠以本系的名称, 这当然不无道理。然而, 我们认为, 在严格的、确切的测量学意义上, 心理测验和教育测验其实不是两个并列的概念。心理测验应该包括教育测验在内, 心理测验主要可以分为三类: 能力测验(包括智力测验、创造力测验

等),人格测验(包括兴趣测验、态度测验、价值观测验等)以及教育测验。

综上所述,本书的书名与内容倒是名副其实的。本书于1954年出版,其后,分别于1961年、1968年、1976年、1982年、1988年、1997年出版二、三、四、五、六、七版。其中,前6个版本都是安娜斯塔西(A. Anastasi)一人独立完成的,直至出版第七版时,安娜斯塔西已90高龄,才有厄比纳(S. Urbian)的加盟,与她共同完成最新版本的修订工作。

安娜斯塔西是美国著名心理学家,1908年出生于纽约,在巴纳德学院获文学士学位,21岁在哥伦比亚大学获哲学博士学位。安娜斯塔西先后执教于巴纳德学院、纽约市立大学昆斯学院、福德姆大学。1972年安娜斯塔西当选美国心理学会(APA)主席,这是50年来第一次由女性出任这个重要职务。1981年安娜斯塔西获得美国心理学会应用心理学杰出科学奖,1983年又获美国教育研究协会(AERA)教育研究杰出贡献奖。安娜斯塔西的主要论著,除了这部《心理测验》之外,还有《差异心理学》(*Differential Psychology*)、《应用心理学领域》(*Fields of Applied Psychology*)等。

二、关于测验的定义

本书的书名是《心理测验》,理所当然要对“测验”下一个科学的定义。早在1961年的第二版中,安娜斯塔西就给出了“测验”的完整定义:“心理测验实质上是对行为样本的客观的和标准化的测量。”到目前为止,这一定义仍为世界各国大多数心理学家所接受。安娜斯塔西从五个方面详细阐述了这个定义。

1. 行为样本(*behavior sample*)。心理测验是对小型的、经过仔细选择的个体行为样本进行观察。在这一点上,心理学家使

用的方法与其他学科的科学家用观察的方法是一样的。如果要检验病人的血液,医生只要分析一个或几个血液样本;如果要检验社区用水,只要分析一个或几个饮水样本。同样的,心理学家为了测验学生的词汇程度,只需使用一组有代表性的词汇;测验学生的算术计算能力,只需使用一组有代表性的算术问题;测验飞行员的手眼协调,则使用一组有代表性的运动测验即可。

当然,心理测验是一种间接测量。心理测验的过程如下:首先向被试呈现一组测题,然后引起被试的一组行为,最后推论被试的一种心理属性。这就是说,我们要作出准确的、可靠的推论,其前提是被试的一组行为要有代表性,或者说这组行为提供足够的有用信息,能够反映所要预测的心理属性,而这个前提之前提则是测题要有代表性,能够引起与所预测的心理属性有关的一组行为。

安娜斯塔西指出,测验题目不必与测验所预测的行为非常相似,但需要证实两者之间的经验性对应。测验样本与所预测的行为之间的相似性程度可以大不相同。在一个极端,测验样本与所预测的行为的一部分完全相合,例如外语词汇测验和驾驶员执照考试。大多数测验中所编入的任务与实际工作中所完成的任务之间只有适度的相似性,例如工作培训之前所实施的许多职业能力倾向测验。在另一个极端,测验样本的作用是引发被试的联想,我们试图根据被试的这种联想去预测他对日常生活中其他复杂性情境的反应,例如罗夏墨迹测验之类的投射人格测验。所有这些测验尽管表面上有所差异,但它们都由个体的行为样本所构成。一个心理测验要具有价值,就必须在经验上证实,被试在该测验上的表现与在其他情境中的行为相对应。

2. 标准化(standardization)。标准化是指测验的编制、实施、

评分以及分数解释的程序的一致性。为了使不同被试所得的分数可作比较,显然,测验的外部条件对所有被试都必须是相同的。在这种测验情境下,被试之间的测验分数的不同,才可单一地归因于测验个体本身的不同。

标准化的内容,具体涉及四个方面:(1)测验材料标准化。每个被试所使用的测验材料必须相同;测题文字表达确切,不会产生歧义;测题印刷质量清晰;操作器具的物理性能一致等。(2)测验实施标准化。每个被试的测验实施条件必须相同,包括测验指导语、时限、例题、处理被试提问的方法、测验环境等。(3)评分标准化。严格按照标准答案评分,不得加入评分者的主观意见。(4)测验分数解释的标准化。心理测验一般没有预先确定的及格或不及格的标准,每个测验的分数都是根据经验性资料加以评价。对于大多数测验目的来说,往往参照常模来解释个体的测验分数。

安娜斯塔西指出,建立人格测验常模的方法,本质上与能力测验毫无二致。能力测验的常模未必是满分即没有错误的分数,人格测验的常模也未必是最合乎要求的即“理想”的成绩。在这两类测验中,常模均与典型或一般个体的行为表现相对应。

3. 难度的客观测量(objective measurement of difficulty)。心理测验的特性是客观的和标准化的测量。讨论标准化时已经提到心理测验的客观性的某些方面。因此,如果测验的实施、评分、分数解释等与特定主试的主观判断无关,则测验就是客观的。理论上,不管主试是谁,任何一个测验被试应该获得相同的分数。虽然实际上尚不能达到完满的标准化和客观性,但是客观性至少是测验编制的目标,并且在大多数测验中已经达到相当高的程度。

心理测验的客观性也表现在其他一些主要方面。测题或整

个测验的难度水平的确定,需要根据客观的、经验性的程序。常模样本中答对人数最多的测题作为最容易的测题,答对人数相对较少的测题则作为较难的测题。这样,可以建立难度的经验性顺序,然后将测验的测题由易到难加以排列。

不仅测验测题的安排,而且测验测题的选择,都应该以常模样本中通过每个测题的人数比例为指导。因此,测验难度范围的各个部分,包括其易端或难端,测题的数量都要适量。

4. 信度(reliability)。这个测验好不好? 这个测验真的有用吗? 回答此类问题的唯一方法是通过经验性尝试。心理测验的客观评价就是确定测验在规定情境中的信度和效度。

心理测量学中所使用的“信度”一词表示测验结果的一致性。测验信度是指使用相同的测验或测验的等值型式,再次测验相同被试时,他们所得分数的一致性程度。测验的信度很重要,信度不高的测验是没有什么用处的。如果使用一个智力测验去测量一名儿童,第一次测得的智商为 110,而几天后再测时智商却为 80,那么两个分数显然都没有多少可靠性,因此可以说这个智力测验毫无用处。

编制或使用心理测验,都应该进行客观的测验信度的检验。测验信度有多种类型。比较相同的测验被试在不同的时间、使用不同组的测题、主试或评分者不同,或在任何其他有关的测验条件下所获得的分数,都可以检验信度。必须注意信度的类型和确定信度的方法,因为同一测验的不同信度可以有所不同。同样应该注意检验信度所使用的被试的性质和人数。

5. 效度(valiability)。效度表示测验实际上测量到它意欲测量的事物属性的程度。毫无疑问,测验效度是心理测验中最为重要的问题。效度直接检查测验实现其功用的程度。不管使用测验去测量什么东西,效度的确定通常需要独立的、外部的效

标(criteria)。例如,如果使用一个学业能力倾向测验来选拔高考学生,那么录取学生在大学的学业成就就是一种效标;如果使用一个职业能力倾向测验来选拔应聘人员,那么录取员工在单位的作业成就就是一种效标。测验与效标的关系可以使用相关系数即效度系数(validity coefficient)来表示。如果测验得高分的个体在效标中相对成功,而测验得低分的个体在效标中做得不好,那就说明测验分数和效标度量之间对应性好,因此测验的效度较高。效度系数使我们能够确定根据测验分数预测效标成绩的程度。

安娜斯塔西回答了人们关于测验效度的两种疑问。其一,如果必须追踪测验被试或以其他方法获得测验试图预测的事物属性的独立度量,那么为什么还需要测验呢?安娜斯塔西认为,应该区分建立效度的团体和出于使用目的最终对之实施测验的各个团体。在测验使用之前,必须根据一个有代表性的被试样本建立测验效度。这些被试的分数本身不是用于使用目的,而仅仅在检验测验效度的过程中起作用。如果使用这种方法证明测验确实有效,那么在缺乏效标度量的情况下,就能够将测验用于其他样本。

其二,为了获得测验试图预测的资料,我们只需要等待任何一个团体的效标高度的形成并加以利用,而不需要什么测验。安娜斯塔西认为,这种方法在理论上也许讲得通,但在实践上却根本行不通。它耗费大量的时间和精力,在大多数情况下都是绝不允许的。如果我们要确定哪些考生会顺利读完大学或哪些应聘人员工作会有所成就,那就要招收所有被试,并且等待他们后来的发展!我们编制测验,正是要把这种过程的极大耗费及其对个体的有害的情绪影响减到最小限度。使用测验,我们能够在一定的误差范围之内评定个体的相关属性的目前水平。测

验的效度和信度越高,测量误差范围也就越小。

三、关于训练对测验效度的影响

安娜斯塔西在 1981 年就指出,讨论训练对测验效度的影响时,一个根本的问题是,训练之后的提高是局限于测验所包括的具体测题,还是扩大到测验所要评价的广泛的行为领域。测验测题取样于有关行为的各个方面,显然,被试的测验成绩,应该反映他的正规的或非正规的、学校之内或之外的教育经历。这种广泛的影响决不会使测验没有效度,因为测验分数准确地表示被试在所评价的心理属性上的位置。从有效测验的观点来看,我们能够对影响的范围作出一种切实可行的区分。因此,只有当特定训练提高测验分数但没有明显影响测验所要测量的行为领域时,测验分数才没有效度可言。根据对测验效度的不同影响,安娜斯塔西把测验之前的训练分为三类:

1. 辅导(coaching)。辅导是指对类似于测验的测题做大量的强化练习。这种辅导完全可能提高测验成绩,提高的程度取决于多种因素,包括测验被试本身的能力和以前的教育经验、测验的性质,以及所提供的辅导的数量和类型。比起受教育较多、对测验有备而来的被试来,受教育较少、对测验准备不足的被试更有可能从特殊辅导中受益;辅导材料和测验内容之间越是类似,测验分数的提高也越大。

另一方面,辅导并没有相应地提高效标行为。辅导越是局限于具体的测验内容,提高就越是不可能扩大到效标成绩。辅导由此降低测验效度。在这种情况下,测验就不能有效地测量它所预测的广泛的行为能力,也不能准确地弄清被试是否具有在效标情境中取得成功所需的知识和技能。

编制良好的测验,应该选择受这种辅导的影响最少的测题。

如果通过短期针对性的辅导训练能够明显提高某些测题类型的成绩,则在施行的测验形式中便不再加以保留。例如有类测题要求简单的顿悟性解答,一旦得出解答,就能够直接用于所有类似的测题。将来再遇到这类测题时,它们测验的只是被试的记忆力而不是解决问题的能力。

2. 测验的技巧(test sophistication)。测验的技巧是指纯粹的参加测验的练习,使用同一测验的复本,第二次的分数往往较高。不管实施复本的时间间隔的长短,第二次测验的平均数都显著提高;对于不同年龄的智力正常或超常的被试样本,情况同样如此。

另外,比起第一次参加测验的被试来,具有丰富的测验经验的被试,在测验成绩上占有优势。这种优势,一部分是由于克服了测验的陌生感,增强了自信心。另一部分是由于许多测验所包括的内容和机能的类型中有一定数量的重叠。熟悉通常的测题类型,成绩可能会稍稍提高一些。

训练测验技巧的目的,是使各个被试的测验技巧相等,消除或平衡他们以前参加测验经验的差异。这些个体差异是测验情境所特有的,也是影响测验分数本身的条件,但在测验所要评定的广泛的行为领域中未必反映出来。因此,这类训练方法将减少测验特有的因素的影响,可以更为有效地评定测验所要测量的广泛的行为领域,从而相应提高测验的效度。

3. 广泛的认知技能(broad cognitive skills)。这类训练的目的,是发展个体的广泛适用的智力技巧、工作习惯和解决问题的策略等。

训练广泛的认知技能是提供教育而不是提供辅导。其中所使用的许多训练方法,是为了发展个体有效地解决问题的行为,例如仔细地分析问题;得出答案时考虑各种不同方法、有关的细

节和含义;深思熟虑而不是凭一时冲动提出或选择一个答案;使用高标准来评价自己的成绩。

显然,如果对广泛适用的认知技能的训练是行之有效的,那么就能够提高受训者处理后来智力任务的能力。测验成绩中必然而且应该反映出这种提高,因为测验分数和效标成绩两者都得到了提高。这种训练使得测验效度保持不变,但重要的是,它大大增加了个体达到理想目标的可能性。

四、关于常模

在心理测验中,将被试的反应与标准答案相比较而得出的分数,称为原始分数(raw score)。原始分数本身是没有多大意义的,我们必须另有一种明确规定的、统一的参照框架。心理测验的分数,通常参照常模(norm)作出解释,常模表示标准化样本中被试的测验分数分布。有了常模,我们就可以表明被试在常模样本中的相对位置,从而参照其他个体评价他的成绩;我们也可以直接比较同一被试在许多不同测验上的相对成绩。安娜斯塔西把常模分为发展常模和团体内常模两大类型,并且着重指出各种常模存在的不足之处,以期引起人们的关注。

发展常模表示个体在正常发展线上处于怎样的水平。例如我们可以说一个8岁儿童具有10岁的智力年龄,也可以说一个四年级学生具有五年级算术水平或三年级阅读水平。常用的发展常模有以下三种:

1. 智力年龄(mental age)。法国比纳—西蒙智力量表首先使用智力年龄常模,被试的测验分数以他所能正确完成的测题所代表的年龄水平来表示。安娜斯塔西指出,智力年龄的明显缺点是单位不等。由于智力发展在年龄小时较为快速,而随着个体接近成熟年龄而逐渐减慢,因此智力年龄的单位并非随着

年龄保持不变,而是往往随着年龄的增加相应地缩小。

2. 年级当量(grade equivalents)。教育成就测验上的分数,经常按照年级当量来解释。年级常模表示每一年级学生所获得的平均原始分数。安娜斯塔西指出,年级常模存在着若干缺点:第一,各个年级的教学内容有所不同,因此年级常模仅仅适用于测验所包括的全部年级都教的共同学科。第二,即使同一门学科,各个年级的教学进度也不尽相同。换句话说,年级常模单位显然是不等距的。第三,年级常模容易引起误解,除非测验使用者牢牢记得导出年级常模的方法。第四,年级常模往往被误以为是成绩标准。

3. 顺序量表(ordinal scales)。这种常模源于儿童心理学的研究。通过对婴儿的行为发展的经验观察,人们描述各种机能随着年龄发展的典型行为。《格塞尔发展顺序量表》按月份表明,从4周到36个月的儿童在运动、适应性、语言、社会性等四种主要行为领域中所达到的大致发展水平。本世纪60年代瑞士儿童心理学家让·皮亚杰提出儿童四阶段认知发展理论,有些皮亚杰作业已被编入标准化智力量表。顺序量表用来鉴别儿童在具体行为机能的发展中所达到的阶段,它一般提供儿童实际上能做什么的信息。安娜斯塔西指出,皮亚杰方法有其贡献,但也有局限性。

团体内常模根据最可比较的标准化样本的成绩来评价被试的成绩。例如,把一名儿童的原始分数与相同年龄或相同学校年级的儿童相比较。常用的团体内容常模有以下三种:

1. 百分位数(percentiles)。它表示标准化样本中低于某一特定原始分数的人数百分比,也能够认为百分位数是100人团体中的等级。安娜斯塔西指出,百分位数便于计算,易于理解。主要缺点是单位明显不等,尤其在分布的两端。原始分数转换

成百分位数时,靠近分布中部的原始分数的差异被扩大,而靠近分布两端的原始分数的差异被缩小。另一缺点是百分位数只表示每一被试在标准化样本中的相对位置,而无法比较和说明不同分数之间的差异量。

2. 标准分数(standard score)。它按照分布的标准差来表示个体分数与平均数的距离。最初的原始分数,通过线性转换或非线性转换,都可得出标准分数。线性转换的标准分数,简称为“z 分数”,分布形态同最初原始分数完全一样。为了使不同分布形态的分数具有可比性,可以采用非线性转换,使分数大致符合正态分布,由此所得的标准分数则称为正态化标准分数,例如 T 分数、标准九分等。安娜斯塔西指出,只有当样本很大,又有代表性,并且有理由认为,造成偏态分布只是由于测验本身的缺陷,而不是由于样本的特性及影响所考虑的行为的其他因素,才能进行正态化转换。

3. 离差智商(deviation IQ)。早期的智力测验采用比率智商。一个重要的技术缺点是,不同年龄组智商分布的标准差不尽相同,因而不同年龄个体的智商就不可比较。主要由于这种原因,离差智商取代了比率智商,离差智商实际上是标准分数的另一种变式。韦克斯勒智力量表最早按照离差智商表示测验分数,其平均数为 100,标准差为 15。安娜斯塔西指出,只有当不同的测验采用相同数值的标准差时,我们才可比较它们的离差智商。离差智商的另一问题是,智力极低者得分偏高,而智力极高者则得分偏低。

五、关于“四度”

所谓“四度”,是指测验的信度和效度,以及项目的难度和区分度。信度是对测量结果的一致性程度的估计,而效度是对测

量结果的准确性程度的估计。项目就是组成测验的测题。项目难度表示被试反应与答案的吻合程度,而项目区分度表示在测验所要测量的特质上区分不同水平被试的程度。“四度”历来是心理测验理论的核心内容,也是本书的重头戏之一。我们认为,本书论述“四度”与同类论著相比,既有共同之处,亦有相异之处。

所谓共同之处可以“四度”的分类为例。例如,信度的类型可以分为:(1)再测信度。第二次重复实施完全相同的测验,同一组被试两次测验分数之间的相关,便是再测信度系数。(2)复本信度。第一次使用一种测验型式,第二次使用另一种测验型式,同一组被试两种型式测验分数的相关,便是复本信度系数。(3)分半信度。使用一种测验型式,实施一次测验,把测验分成两个等值的一半,每一被试得到两个分数。用通常方法计算两个半测验分数的相关,再使用斯皮尔曼—布朗公式校正,便是分半信度系数;也可以使用弗拉南根公式或鲁隆公式得出分半信度。(4)库德—理查逊信度和 α 系数。这种方法也对一种型式实施一次测验,但它所根据的是对所有测验项目反应的一致性。前者仅用于按照对错或有无等二值记分的测验项目,后者则适用于多重记分的测验项目。又如,效度可以分为三大类型:内容效度表示测验内容代表所要测量的属性的程度,效标效度表示有效预测个体在指定活动中行为表现的程度,构思效度表示测验能够测量到理论上的构想或特质的程度。再如,项目难度指标,既可以使用顺序量表的通过百分率,也可以把它转换成等距量表的 z 分数。最后,人格测验通常根据一个外部效标来评价项目区分度,而成就测验通常根据测验本身的总分来评价项目区分度。

至于相异之处,着重体现在三个方面:其一,安娜斯塔西用