

# 智力测验分数解释的发展趋势\*

张仲明 李 红

(西南师范大学 心理学院, 重庆 400715)

**摘 要:**智力测验分数的解释是智力心理学 一个重要的研究领域。智力测验分数解释最基本的一个目的就是为了理解。随着智力理论和智力测验的发展,智力测验分数的解释出现了一些新的特点与趋势即注重智力测验分数解释的理论基础,注重智力测验分数解释的效度丰富化与从有用信息的抽取来解释测验分数的趋向。

**关键词:**智力测验分数;解释;量表;趋势

中国分类号: B841.7

文献标识码: A

文章编号: 1003- 5184(2004)02- 0024- 04

## 1 问题的提出

解释智力测验分数、揭示出智力测验分数背后的深层含义一直是智力心理学研究的一个重要领域。测验分数的解释历来就非常困难。而对智力测验分数的解释是其中最困难的方面。有人认为对智力测验分数进行解释主要包括两个目的: 一是如何使分数有意义, 二是如何将有意义的信息传达给当事人<sup>[1]</sup>。Daniel<sup>[2]</sup>认为智力测验解释的主要目的是理解被试的认知状况和对未来发展进行预测, 如预测职业成功, 或者预测特殊干预措施的有效性。在这些观点中, 几乎都认为对测验分数的理解是测验解释最基本、最重要的方面, 即为理解而解释。但是, 在对智力测验分数的解释上, 传统的做法一般都集中在解释分数时的参照标准上, 如常模参照分数、内容参照分数和结果参照分数等。近年来, 智力测验发展出现了“由基于统计的测量转向基于理论的测量; 内容更加全面和深入; 评价的方式呈现出多元化; 一般智力测验与多重性向测验趋于融合; 运用了更多的测量理论和技术; 面向诊断、矫治和训练”等新特点<sup>[3]</sup>, 与智力测验的发展形势相呼应, 对智力测验分数的解释也出现了一些有别于传统的新趋势。

## 2 智力测验分数解释的发展趋向

### 2.1 注重智力测验分数解释的理论基础趋向

对智力测验分数的解释要从智力测验模式的理论基础出发。实际上, 最早的智力测验比纳- 西蒙量表的最早结构是根据经验编制的, 没有任何智力理论的构想。但是, 现在这种现象得到了明显的改变, 而且出现了进一步加强理论联系的趋向。目前智力测验的两个主要理论基础是层次心理分析能力模型和戴斯- 鲁利亚神经心理模型。

层次心理分析能力模型主要包括心理测验能力模型和三层次理论。心理测验能力模型和三层次理论差别不大, 因为它们都鉴别出 8 种相同的“一般能力”, 即流体智力、晶体智力、视觉加工、听觉加工、短时记忆、长时记忆的提取、认知速度和决策速度能力。但是心理测验能力把推理能力作为第九个一般能力, 而三层次理论则把推理能力看成是流体智力的一部分。心理测验能力模型和三层次理论之间最重要的区别是, 心理测验能力不包括在主要能力之上的第三级一般能力因素。而在三层次理论中, 一般能力因素对主要能力有不同程度的影响。戴斯- 鲁利亚神经心理模型<sup>[4]</sup>包括了三个主要的加工水平或者单元, 每一个水平都与大脑特定区域相联系。第一单元由唤醒和注意组成; 第二个单元与续时性或同时性信息编码形式有关。第三个单元由执行控制功能如计划与管理组成。这个模型独特之处是它

\* 基金项目: 全国教育规划“十五”重点课题( DBA010164) 和重庆市教育科研课题( 02- GY- 25)。

注意加工过程而不是能力,并且,在该模式中没有与一般能力结构相类似的成分。层次心理分析能力模型和戴斯-鲁利亚神经心理模型都强调能力的多维度。以此为基础,智力测验就给解释提供了大量的各种各样的分数,如,多重分数、合成分数、子测验分数等等。在这些测验分数上对智力测验进行解释,提高了智力测验解释的可靠性和准确性。正是如此,在智力测验分数的解释中,与理论的联系趋向还在加强。

## 2.2 注重智力测验分数解释的效度丰富化趋向

### 2.2.1 效度的考察更加注重研究方法的多元化

在智力测验的解释中,对效度的考察已经出现了方法多元化的趋向。首先,在子测验水平,可以进行内容和任务分析。内容和任务分析是测验解释的起点。例如,卡罗尔<sup>[5]</sup>就对 400 多个能力数据进行了因素分析,总结了认知任务之间的相关强度。在对测验进行内容和任务分析时,狭义能力与每个主要能力的相关仍然有价值。以戴斯-鲁利亚神经心理模型为基础的智力测验,如考夫曼的儿童成套评价(K-ABC),因为清楚地描述了每一个功能单元的加工,所以内容和任务分析特别依赖于子测验的结构效度。

其次,对智力测验的效度分析的又一方法是对两个或多个成套测验进行相关分析和探索性因素分析。尤其,当一个标志性子测验或量表功能很强,对具体能力具有比较纯的测量时,相关分析就是一个很强的工具。但是要注意,当成套测验的交叉相关稳定地低于子测验或量表的信度时,它的结果就不能提供丰富的信息。而对探索性因素分析同样要注意,作为抽取大量的非确定性因素的解释工具,探索性因素分析的作用还是很弱;特别,当一个研究得出了几个不同的结果时,要形成清楚的理解就困难。而因素怎样旋转的不确定性,也妨碍做出清楚的结论。

对智力测验的效度分析非常有用的一个方法是验证性因素分析。它比探索性因素分析具有更多的优势。一方面,对测验进行分析时,子测验的内部相关可以与预期的相比较,并且预期的相关和观察到的相关的不适合程度能够数量化。由此,可以对子

测验或量表测量的内容的假设进行抉择。当数据符合模型时,这个负荷对解释才有意义。另一方面就是对两个或更多成套测验共同地进行分析。验证性因素分析能够检验在一个量表中的潜质与另一个量表中的潜质是否相同。运用这个技术已经表明 WISC-R 操作测验和 DAS 空间能力量表和 K-ABC 同时性加工量表都测到了相同的能力因素。因此,Keith<sup>[6]</sup>等认为验证性因素分析有助于结构效度和对智力测验的解释。

### 2.2.2 效度的考察更加注重理解因素与量表的关系

验证性因素分析研究的大量成果及发展和心理测验能力模式的出现,为智力测验结构效度方面开展丰富的研究提供了动力。它们提出了一个非常重要的问题:量表怎样才能很好地测量到隐含因素。在不同的测量工具中,用相同的能力维度解释两个量表,并不意味着这些量表的分数高度一致,也不能意味着对这个维度进行了有效的测量。量表效度即量表分数和因素之间一致性程度,依赖于几个方面。一是构成量表的子测验的效度。每一个子测验的方差对那个因素的贡献是多少?在子测验中,决定性因素的方差的比例提高时,子测验的公共方差对总量表分数的影响也逐渐提高。二是子测验取样的代表能力。选择表面特征和它们所测量的狭义能力不相似的子测验是提高有效总体的策略。但,还不清楚在总体中需要多少不同的狭义能力才能产生对主要能力的有效测量。Woodcock<sup>[7]</sup>认为最小数为 2,但是,对一些主要能力而言,可能需要的数量更大。第三,子测验的信度影响了总量表对隐含因素测量的效度。随着子测验数量的提高,误差方差对总体的影响会降低。因此,当子测验信度较低时,就需要更多的子测验来形成有效量表。另外,还可以用计算量表与隐含因素的相关的程序来考察因素与量表的关系。Gorsuch<sup>[8]</sup>描述了这个程序。在单个成套测验的分析中,了解量表分数与潜质怎样相关,很有用。Gustafsson 和 Undheim<sup>[9]</sup>就分析了在 WISC-全量表智商和一般能力因素之间的相关系数为 0.84。由此可见,考察因素与量表的关系不仅可以指导选择测验而且可以帮助解释在相似量表中的分数

差异。

### 2.2.3 注重从总体来解释测验分数

每一个智力测验,都有一个总体分数。每个智力测验的总体在测验的解释中占据了重要地位。总体分数是一种强有力的预测,至少是解释认知评价的一个有用起点。从这个观点看,对智力测验分数的解释再没有比总体测量到了什么内容更重要的问题了。一个观察到的事实就是:并非所有的总体都测量到了相同的结构。测验编制者概念化一般能力的方式将影响怎样设计总体并显著地影响到怎样解释总体。因此,对总体分数不应该以相同的方式进行解释或运用。

用总体来解释测验分数涉及到一个人们怎样看待G因素的问题。许多人用不同认知任务中存在一个更高级的一般能力因素——G因素来解释总体分数,并认为这个因素表征了个体整体特征。例如,卡罗尔<sup>[10]</sup>认为:“G因素是一个整体是一个经济的观点”。Gustafson<sup>[11]</sup>对认知能力团体测验的层次分析后认为:高级的一般因素与流体智力的主要能力相同。

总体的建构方式也会对内容和解释产生巨大影响。智力测验的总量表不能完全地解释内部变化。但是,高内部相关的总体可以以相似的方式进行解释,也可以使用结构描述的方式进行解释。同时,相关数据分析表明,总体有重要分歧时,应该在对内容和理论基础进行仔细评价的基础上对测验进行解释。

### 2.2.4 注重从次级总体来理解测验分数

智力成套测验一般包括对次级能力进行测验的量表。对这些量表所测量的内容的清楚理解决定了解释的有效性。但,次级能力缺乏严格含义,所以对次级能力的思考较少。以心理测量学或神经心理学模式为基础的量表分数的解释不是建立在经验性基础上的。量表和模型结构之间的相关还不十分清楚。在单个测验上进行因素分析成为了量表结构效度的主要来源。因素分析的一个积极特征就是它保证了子测验之间的内聚性。当子测验分数加起来时,公共变量的影响将扩大。成套测验的因素结构与理论模式的符合就证明了量表的结构意义。因素

之间相关或者次级因素与高级一般因素的相关就是量表效度的指征。在结构效度方面,单套测验的因素分析与多套测验分析相比是有限的工具。任何成套测验的因素结构受到子测验的性质和种类的限制。单套测验的因素分析经常导致次级因素。因此,这个技术不能为量表解释提供充分的基础。对内容、任务要求和子测验成分的经验效度的分析也是理解次级总体的重要方法。有人提供了以经验为基础的对子测验和量表的分析。McGrew<sup>[12]</sup>等已经进一步把这种分析类型作为公共解释系统程序的一部分。使用这种方法,智力成套测验的量表的结构意义就通过怎样符合它们的成分子测验所测量的狭义能力来进行评价,也通过它们怎样适合Gf-Gc模型来进行评价,并且也通过判断量表在层次结构中属于哪个位置进行评价。量表的结构效度依赖子测验的每一成分与想测量的能力因素二者之间关系的强度。Carroll<sup>[5]</sup>, Woodcock<sup>[7]</sup>和 McGrew<sup>[12]</sup>都提供了这个问题的有关信息,但是,对这种分析的经验性基础工作还远没有完成。因此,从次级总体来理解测验分数时,因素和量表之间的区别很重要。

### 2.3 注重有用信息的抽取来解释测验分数的趋向

有用信息的抽取是解释智力测验分数的关键。有用信息的抽取与表征能力之间关系的模式,解释中的自上而下的方式,个人基准分,形成分析等方面有关。

#### 2.3.1 表征能力之间关系的模式

对解释问题,Gustafsson等人<sup>[9]</sup>在层次模型的两类型之间作了一个区别。一种类型,他们标记为“弱”,是自下而上建立的:因素首先从所观察到的变量中抽取出来的;然后,从这些因素的相关中抽取出来;一个或者多个二级因素也抽取出来;如此等等。在每一个层次水平,每一个因素都表征了能力。因而,不同水平的能力没有区别而是共享方差。尽管不同的水平提供了不同程度的细节,但每一个水平都是对能力的完整、独立的描述。另一种类型,标记为“强”,是自上而下的。首先,识别出一般因素,它的方差从所有的变量中除去。其次,次级因素从剩余方差之间的相关中抽取出来;这些因素表征了剩余能力。重复加工以建立更低水平的剩余因素。在

这个方式中,次级量表的方差一部分归因于一般能力,一部分归因于剩余主要能力因素。从高级到低级因素之间存在相关。解释不仅考虑在弱模式中的单一水平而且还应该考虑所有水平。对测验解释者,强类型模式在原理上具有一定的优势。然而,在实际中,根据强层次模型来解释会产生一些问题。例如,在概念水平,剩余能力分数毕竟是人为的,它们与能力的实际情况的解释不一致,它们受到层次结构中几个水平因素的影响。剩余能力分数也会产生严肃的测量问题,剩余分数是基本差异分数并且比原始量表分数具有更低的信度。对每一种能力都需要大量的可信和有效的变量。因此,解释的剩余能力方式更适宜于量表而非子测验。

### 2.3.2 自上而下的方式

这种类型的解释策略首先通过从总量表分数推论一般能力开始。然后,对次级水平上的分数进行检验。这通过量表分数的配对比较或者量表分数与总量表分数的比较进行。如果在量表分数之间,有统计上的差异或者在量表分数之间变异程度很大,那么变异就可以解释为在次级能力上反映了临床的重要性。现代测验在它们手册和其他记录形式中,为这种类型的轮廓分析提供了大量的指南。这种方式与强层次模型的相似性是清楚的。决定主要能力的非g成分的高低的方法是分析量表分数的变异性。如果量表分数发生了显著变化,那么自上而下方式的标准特征就反对解释总量表分数。在原则上,处理变异的方式,依赖一个人对总体的意义和性质的态度。

### 2.3.3 个人基准分

与剩余能力分数更接近的是个人基准分,也就是在量表或者子测验中和高级能力的测量中有差异的分数。计算这个分数的合适的方式就是把获得的量表或者子测验分数和从高级能力获得的预测分数进行比较。这个预测将随着量表或子测验的高因素负荷的功能而变化。在个人基准分和没有调整的分数之间,量表和子测验几乎没有差异。反之,具有高一般能力负荷的子测验和量表预测分数几乎远远高于或者低于总体平均数。Glutting<sup>[13]</sup>等人已经报告了对个人基准记分法研究的令人失望的结果。因

为,首先,在子测验上的分析,缺乏对可靠分数进行区别的信度和特殊性。其次,在子测验与相同标准而不是与预期的子测验的一般能力负荷的分数比较中采用了简单化的方法。最后,一些研究已经评价了个人基准的子测验或量表分数的效度,而不考虑在高级能力因素上的分数。因为,个人基准分数已经剥去了能力变异的主要来源,所以,它们需要对表征了其它来源的分数的一致性进行分析。当好结构的剩余能力分数与表征高水平的能力层次的分数同时进行分析时,它们就与外部标准具有了稳定的和信息性的关系。

### 2.3.4 形成分析

形成分析方式很适合弱层次模型如,Gf-Gc模型,因为它把一般能力与另外的一般能力集中进行比较而不是与一般因素进行比较。用与量表有联系并在标准分之上的直接方式对主要能力进行推论,忽略了决定量表分数是否显著地不同于总体分数的步骤。以配对比较为基础来对强和弱进行推论。因为形成方式缺乏统计的显著性并且总体频率指标通过自上而下的方法产生且在临床人群上试图建立起诊断性的有用轮廓的结果感到失望,所以这种方式不受欢迎。为了在标准化的样本中运用事件群组分析的方式,Glutting<sup>[14]</sup>等人已经提出了“核心”子测验轮廓。在正常人群中这些轮廓发生频率相对较高。提供核心轮廓的合理性是因为正常个体经常具有大量的可变性,也因为这是另外一种帮助实践者判断一个特殊的轮廓是普通还是不寻常的方式。Konold<sup>[15]</sup>等人已经提供了用于这个目的算法。来自任何测验的核心轮廓,最引人注目的特征是它们中的大多数是接近标准的,也就是说,核心轮廓主要表征了一般能力的不同水平。但,要注意核心轮廓的合理性把稀有性与意义性混淆了。

## 3 展望

总之,智力测验的解释从来就不容易,智力测验的解释实际上仍然要回答几个基本问题:智力测验解释的理论基础是什么,测量到了什么,测量的程度如何,信息如何抽取。在新技术和新理论的推动下,这几个方面的手段增多了,程度加深了,解释的水平提高了。但是在解释方面,未来重要的进步,可能主

要来自心理学上新手段的发展,如,在层次结构的几个水平上进行解释,仔细地设计子测验,在低水平上对能力进行更精确测验,对高水平能力的影响进行控制等等。这些方面仍有待于深入的研究。

### 参考文献

- [1] 郑日昌. 心理测量[M]. 湖南: 湖南教育出版社, 1987. 240- 241.
- [2] Sternberg R J. Handbook of Intelligence [M]. New York: The Press Syndicate of the University of Cambridge, 2000. 477- 490.
- [3] 张阔, 胡竹菁. 略论智力测验发展的现状与趋势[J]. 心理学探新, 2002, 22(2): 36- 40.
- [4] Das J P, Naglieri J A, Kirby J R. Assessment of cognitive processes: The PASS of intelligence[M]. Needham Heights, MA: Allyn & Bacon, 1994. 22- 30.
- [5] Carroll J B. Human cognitive abilities: A survey of factor-analytic studies[M]. New York: Cambridge University Press, 1993. 10- 30.
- [6] Keith T Z. Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence test[A]. In: Flanagan D P, Genshaft J L, Harrison P L. Contemporary intellectual assessment: Theories, tests, and issues [C]. New York: Guilford, 1997. 373- 402.
- [7] Woodcock R W. Theoretical foundations of the WJ- R measures of cognitive ability [J]. Journal of Psychoeducational Assessment, 1990, (8): 231- 258.
- [8] Gorsuch R L. Factor analysis (2nd ed.) [M]. Hillsdale, NJ: Erlbaum, 1983. 272- 275.
- [9] Gustafsson J E, Undheim J O. Individual difference in cognitive functions[A]. In: Berliner C, Calfee R C. Handbook of educational psychology [C], New York: Macmillan, 1996. 186- 242.
- [10] Carroll J B. No demonstration that g is not unitary, but there's more to the story: Comment on Kranzler and Jensen [J]. Intelligence, 1991, 15: 423- 436.
- [11] Undheim J O, Gustafsson J E. The hierarchical organization of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL) [J]. Multivariate Behavioral Research, 1987, 22: 149- 171.
- [12] McGrew K S. Analysis of the major intelligence batteries according to a proposed comprehensive Gf- Gc framework [A]. In: Flanagan D P, Genshaft J L, Harrison P L (Eds.). Contemporary intellectual assessment: Theories, tests, and issues [C]. New York: Guilford, 1997. 151- 179.
- [13] Glutting J J, McDermott P A, Konold T R. Ontology, structure and diagnostic benefits of a normative subtest taxonomy from the WISC- standardization sample [A]. In: Flanagan, D P, Genshaft J L, Harrison P L (Eds.). Contemporary intellectual assessment: Theories, tests, and issues [C]. New York: Guilford, 1997. 349- 372.
- [14] Glutting J J, McDermott P A, Prifitera A, McGrath E A. Core profile types for the WISC- and WAI: Development and their application in identifying multivariate IQ- achievement discrepancies [J]. School Psychology Review, 1994, (23): 619- 639.
- [15] Konold T R, Glutting J J, McDermott P A. The development and applied utility of a normative aptitude- achievement taxonomy for the Woodcock- Johnson Psycho- Educational Battery- Revised [J]. Journal of Special Education, 1997, (31): 212- 232.

## Developmental Trend of Interpretation of Intelligence Test Score

Zhang Zhongming Li Hong

(Psychology School of Southwest China Normal University, Chongqing 400715)

**Abstract:** Interpretation of intelligence test score is one of important research domains of intelligence psychology. Understanding is the fundamental purpose of interpretation of intelligence test score. With the development theories and psychometric theories, interpretation of intelligence test score show some characteristics and tendencies: pay more attention to the intelligent testing theoretical foundation, pay attention to plenty of validity and how can reliable information about those abilities be extracted from the test results.

**Key words:** intelligence; test score; interpretation; scales; tendency