

新一代测验理论——认知诊断理论的源起与特征^{*}

刘声涛^{1,2} 戴海崎¹ 周 骏¹

(1. 江西师范大学 教育学院, 南昌 330027; 2. 福建师范大学 教育科学学院, 福州 350007)

摘 要: 认知诊断理论被视为新一代测验理论的核心, 是认知心理学与现代测量学相结合的产物。对认知诊断的研究已成为当前国外心理学研究的一个重要热点, 并引起国内学者的广泛关注。本文从认知诊断的源起、概念、特征, 及研究的基础、框架、意义和难点等七个方面对认知诊断的理论与技术作了一个简要述评, 以期推进我国心理学界认知诊断的研究工作。

关键词: 认知诊断 标准测验理论 新一代测验理论

中图分类号: B841.2

文献标识码: A

文章编号: 1003-5184(2006)04-0073-05

认知诊断理论被视为新一代的心理与教育测验理论的核心, 是认知心理学与现代测量学相结合的产物, 成为了当前国外心理测量学研究的一个重要热点。在考试研究发达的美国, 认知诊断引起了政府和著名考试公司 ETS 的关注, 吸引了一大批研究者的参与, 有大量的研究论文发表, 其中一批有代表性的论文已经结集出版成书, 曾召开过至少两次以认知诊断为主题的国际性大型会议。认知诊断研究也已引起我国学者的关注。早在几年前, 国内就有过一些与认知诊断有关的研究成果发表^[1-5], 但较为零散。近两年, 不少学者在不同场合多次呼吁, 要开展对认知诊断的系统研究。文章依据现有文献资料, 结合已有的研究经验, 从认知诊断的源起、概念、特征、及研究的基础、框架、意义和难点等七个方面对认知诊断的理论与技术作了一个简要述评, 以期推进我国心理学界认知诊断的研究工作。

1 认知诊断的源起

作为独立学科分支的心理和教育测量学已经经历了一个多世纪的发展, Mislevy R. J.^[6]指出, 到目前为止, 整个统计测验理论的发展可以分为两大阶段: 标准测验理论(Standard test theory)阶段和新一代测验理论(Test theory for a new generation of tests)阶段。

经典测验理论、概化理论、项目反应理论都属于标准测验理论。标准测验理论在能力水平的研究范式(Ability level paradigm)^[6]下产生, 将所测的心理特质视为一个心理学意义并不明晰的“统计结构”, 目

的在于从宏观的层次给个体一个整体的评估, 在单维的、线性的、连续的度量系统上, 指定一个标示位置的值得^[7]。有了这个值, 就可以进行安置、分配、评定等工作。以空间认知能力为例, 标准测验理论视野下的空间认知能力是要得到被试在整个测验上的一个整体的值, 作为被试空间认知能力的度量。

随着科学研究及社会生活的发展, 人们越来越不满足于只能得到个体的宏观层次的评价, 人们不仅要了解个体总的来说怎么样, 还需要了解个体具体的来说怎么样。而此时, 实质心理学尤其是认知心理学的发展, 也给人们进一步了解人的内在的心理机制提供了可能。比如, Embretson S. E 认为^[8], 影响人的空间认知加工过程难度的因素有目标物体旋转的角度和折叠时携带表面的数目, 人们想进一步地了解, 不同的被试解答具有不同难度来源的项目的潜能相同吗? 研究者对这一问题的回答就已不是给被试一个在项目上作答能力的整体评估值, 而是涉及了被试的作答过程和机制; 此时的度量系统已不是单维的, 而是多维的; 不是线性的, 连续的, 而是代表质的差异, 离散性的。总的来说, 人们已不满足于在总的能力水平层面上进行的测验实践, 而是想深入到作答的认知加工过程, 以揭示不同被试的认知加工特点。Mislevy R. J.^[6]将这种研究视野称之为认知水平的研究范式(Cognition level paradigm)。

1993 年, Frederiksen N 等^[9]编了一本书, 书名就叫新一代测验的测验理论(Test Theory for a New Gen-

^{*} 基金项目: 高等学校博士学科点专项科研基金项目(20050414001)。

eration of Tests),由此正式提出了新一代测验理论的概念。新一代测验理论强调测验应同时在两种水平的研究视野下进行。在新一代测验理论的框架中,强调宏观和微观、能力水平和认知水平的评估并举,将测验理解为一种整理和解释个体所知道的和所能做的事实的技术,而不是局限于能力的特质水平的概念。测验领域中,认知水平的研究是对能力水平的研究的深入和拓展,能力水平研究的目的在于能力的定位,而认知水平研究的目的在于认知能力结构和性状的诊断。

2 对认知诊断概念的理解

目前,国外在认知诊断方面的应用研究很多,但并没有对认知诊断给出一个明确的定义。从文献资料来看,国外对认知诊断一词的应用也有广义和狭义之分。广义的认知诊断指建立起观察分数和被试的内部认知特征之间的关系。这种认知诊断既包括在心理学理论建构中的应用,也包括在教育教学领域中的应用。在心理学理论建构中,认知诊断是作为一种研究方法出现。如 Lohman D F 等^[10]举了一个例子。在心理旋转的任务中,有一个假设的加工成分模型,并且指出了哪些项目特征影响加工成分,图示如图 1:

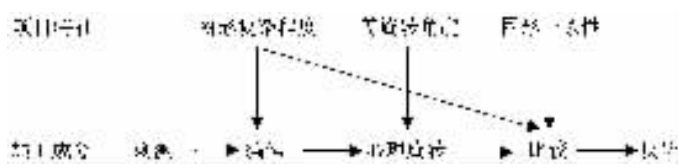


图 1 心理旋转任务项目特征及加工成分示意图

为了证实心理旋转这一加工成分是否存在,可以设计两组项目,这两组项目在其它项目特征上都一致,只是需旋转的角度不同。然后找两批被试进行实验。结果表明需旋转角度不同的项目组,被试完成的时间也不同。这种现象说明被试进行了与物理旋转类似的心理旋转,证实了心理旋转这一加工成分的存在。在教育教学中,认知诊断是按被试有没有掌握测验所测的技能或特质来对被试加以分类,比如在小学数学四则运算的测验中,认知诊断可以实现对学生加、减、乘、除各项技能的掌握情况的判断,给教师和学生提供进一步的教和学的信息。狭义的认识诊断即教育领域中用到的认识诊断的内涵。

3 认知诊断的特征

Cortier J E^[11]指出,传统测验和能实现诊断功能的测验有以下不同的特征。第一,传统测验测量广泛的能力,而诊断测验测量特定的特质。传统测验将测量对象视为“统计结构”,而诊断测验试图将这一“统计结构”进行分解,更深入地探测内部的心理特质。第二,传统测验假设不同水平的技能连续性变化,认知诊断测验测量的特质可能是“全或无”。在传统测验中,所测量的能力在一个连续性的量尺上变化,测验的目的是为了将所测能力在连续性量

尺上定位。在认知诊断测验中,所测量的特定的特质,如策略、技能等,可能会呈现有或无的状态。第三,传统测验中技能是单维的或多维的,认知诊断测验中技能的组成部分可以以任何模式联系起来。不同的被试或在不同的情境下可用不同的策略,技能的组成部分存在时间的或逻辑的相依。第四,传统测验应用项目一致性的几何模型,认知诊断测验中项目一致性的离散性特征模型是合适的。认知诊断所测量的特质之间不仅存在量的差异,更存在质的差异,因此,需用上离散性特征模型。

4 实现认知诊断的两大基础

认知心理学与心理测量学方法是实现认知诊断的两大基础。

首先,认知诊断的实施要以认知心理学模型为基础和核心。认知诊断的认知心理学研究即运用认知心理学的各种研究方法,对测量任务所涉及的知识、技能、策略、加工过程与成分等各认知变量做认知分析,获得测量任务的实质性的心理模型。应该说,测量任务的实质性的心理模型是任何一项认知诊断工作的起点。心理学理论建构中的认知诊断,需要有内在心理加工的假设模型,教育领域中的认知诊断,则需要已被证实的理论模型。从认知诊断和认知心理学的关系来看,认知诊断所应用到的

认知心理学研究成果中可操作化的那部分成果。其次,认知诊断的实现要以心理测量学方法为工具。要最终实现对被试个体差异的分析和解释,仅有认知分析获得的理论依据是不够的,还必须强调测验设计,即根据测量对象的实质性心理模型来选择、编制项目和测验,达到详细描述被试间差异的目的^[12]。测验设计的核心在于通过项目的设计,来引发被试内在认知特点的外在表现,从而实现对内认知特征的判断。测验设计好后,需要借助合适的测量模型,使得理性思辨的成果获得实证性的测量数据。具体的做法是将各种认知变量直接融合入测量模型,借助各种现代统计方法揭示个体认知变量的特征。认知诊断所需测量模型与传统测验有很大不同,在传统测验中,有一套测量模型就能解决具有统计特征的个体能力定位问题,如经典测验理论中的真分数模型及由此而引申的信度、效度、难度、区分度的计算就构成一套解决能力定位和测验质量分析的模型。而在认知诊断中,出现的是一族模型,每个模型对应于各自特殊的应用对象,模型间从表现形式到诊断功能都有较大的差异。比如,Fisher^[13]提出的线性逻辑斯谛克特质模型(linear logistic trait model)与Tatsuoka等^[14]提出的规则空间模型(rule space methodology),两模型所呈现的表现形式和能解决的问题就有很大不同。模型的丰富性根源于认知分析成果的内在结构的丰富性。

5 现有认知诊断研究的框架

为了弄清楚认知诊断应该或可能从哪些方面进行,有必要对已有的认知诊断的成果作一番整理,以确定目前认知诊断的研究框架。

认知诊断研究是一个系统的研究工作,它的整体实施流程包括对测量对象的认知分析、编制项目、确定评分规则、统计分析结果的反馈、程序编制、预测效度研究等几个过程,在这个过程中需要有学科专家、心理学家、计量学家、计算机专家的合作,是一项庞大而复杂的研究工作。因此,目前已有的研究往往将研究重点放在认知诊断的某方面的研究上,包括学生式思维模拟(student modeling)^[15]、概念网络(concept networks)^[15]、模型开发与测验分析。

学生式思维模拟一般将被试内部认知特征用数字系统或计算机技术进行表征。以Haertel, E. H.等^[16]的一个例子来说明。有一个二级记分的项目

A,它所要求的能力是a,那么对于任务A被试的能力状态可以用集合表达。另有一个二级记分的项目B,它所要求的能力是b,那么对于任务B被试的能力状态可以用集合表达。现在有任务C,是任务A和任务B的组合,任务C所要求的能力c就是能力a和能力b的并集,c的状态也可以用集合表达。仅当a和b呈现(1,1)的组合时,c才取1。当c取0时,a和b的状态可能呈现(0,0)(0,1)(1,0)三种组合方式之一。当任务C误答时,研究者还有必要去讨论到底是缺乏a还是缺乏b或者两者都缺乏。象这样用数字系统来表征内部特征的做法,其目的有二:一是便于对能力状态及能力状态之间关系的描述。当任务所要求的分任务多的时候,这个优点是明显的;二是便于后续的数量化的分析。

概念网络多用于陈述性知识的认知诊断,陈述性知识的本质是命题,命题网络或图式表征的建构。概念网络反映了概念关系的功能作用和它们之间的联系,可用于考察学习者对有关知识综合贯通的表征能力。用概念网络做认知诊断首先要建立专家的概念网络。建立的方法有间接的方法和直接的方法。间接的方法是请学科专家参与建立概念网络的整个过程。Johnson, R. J.等^[17]介绍了一种间接的方法,首先,找出内容领域的全部概念,请专家就概念和内容领域的相关在5点量表上评定,选择若干个评定平均数与专家评定相关高的项目作为目标概念,从入选的目标概念中确定一个目标概念,选择和目标概念高度联系的5%~10%的其它概念,确定有关和无关的概念对;上述两步循环直到每个概念都被确定为目标概念;最后,请专家对所有项目对在5点量表上评定概念对相关性。通过上述过程就可以建构起领域知识的概念网络。直接的方法就是请学科专家直接给出内容领域的概念网络。有了专家的概念网络,可将被试的概念网络和专家的概念网络相对照,来诊断被试知识结构中的缺陷。

为了能够对各认知变量进行量的分析和刻画,计量学家还致力于建构能融合不同的认知变量的模型,并且运用各种现代统计方法使得模型能识别,即模型中的参数都能被估计出来。这就是模型开发与测验分析。国外有研究者统计,到2002年为止,至少有14种认知诊断的模型^[18]被开发并被用于认知诊断。就已开发应用的这些模型看,认知诊断的测

量学模型有两个基础性的模型^[19],一个是 Fisher^[13]提出的线性逻辑斯谛克特质模型(linear logistic trait model),另一个是 Tatsuoka 等^[14]提出的规则空间模型(rule space methodology)。在这两个模型基础上又发展了许多其他模型。目前,认知诊断研究中较新的模型有融合模型(fusion model)^[19]、DINA 模型(deterministic input "noisy" and "gate model")^[20]、NIDA 模型(noisy inputs, deterministic, "and" "gate model")^[20]。不同的模型是针对于不同的测验分析实践开发的,有了模型之后,就可以在相应的应用对象上进行测验分析了。如, Tatsuoka 等将规则空间模型用于小学数学测验的认知诊断分析中。

以上谈到的是对现有的认知诊断研究工作的整理,从整理结果中可以看出致力于认知诊断的研究者重点想解决的问题以及解决问题的方式,对人们理解该领域纷繁复杂的文献是有帮助的。

6 认知诊断的意义

认知诊断是当代测验发展的新追求,有重大的意义。首先,认知诊断能实现测验的最重要的功能:促进发展。现代认知心理学的测量观的基本观点是:运用认知分析的方法描述心理活动的内在机制,据此设计各种形式的测验以探测被试心理活动的机制与相应机制之间是否一致或存在缺陷,以便提出补救措施,促进发展。认知诊断是实现测量与发展之间的循环促进作用的关键环节。

其次,认知诊断有利于提高测验的内容效度。目前,人们通常运用经典测验理论或项目反应理论编制教育和心理测验,这二者是依据项目的统计特性来指导测验编制,对内在的知识结构重视不够,难以对测验的内容效度进行分析。认知诊断依据认知心理学的研究成果编制测验,测验的内容效度能得到保证。

7 认知诊断研究的难点

认知诊断测验的提出由来已久,然而,到目前为止,认知诊断更多地是停留在研究领域和小规模的应用领域,在社会上大规模的考试中的应用并不多见。在考试研究和应用发达的美国,目前大规模的认知诊断测验也仅有一种:PSAT。在我国,有关认知诊断测验的研究刚刚起步,仅有几篇论文见于学术刊物。总观国际国内认知诊断研究的发展现状,笔者认为:认知诊断提出了一个很好的理念:目标明

确、路径清晰、结果令人向往,但要得到理想的研究结果却有很大的难度,以致至今未见令人兴奋的实际应用。细细分析,难度主要还是来自于两大基础研究。首先,从认知心理学研究来看,认知心理学家用计算机来类比、模拟人的大脑的内部加工机制,就目前水平来看,其研究结果很多是解释性、描述性的,在测验领域实际应用时的可操作性不强。其次,从测量学研究来看,开发出合适的可实用的统计模型的工作也很难。复杂的模型,能更好地容纳认知心理的研究结果,所得参数能更精确地解释心理特质,但是参数估计等技术难题很难解决;简单的模型,参数估计技术好解决,但参数不能很好地解释心理特质。最后,认知心理学和测量学两学科的沟通与结合不是很好。复杂的数学模型令认知心理学家望而却步,认知分析工作需要丰富的认知心理学的知识、理论和实践经验才能胜任,对于测量学家来讲,认知心理学的那套话语还很陌生,不仅要接受而且要娴熟运用更需较长时期的学习。在这种情况下,两条研究路线沿各自的方向前进,要真正实现“认知和测量相结合”,还需真正有志于此方向的研究者融两家之长,付出艰苦卓绝而富有成效的创造性的劳动。笔者以为,有志于认知诊断的研究者要踏踏实实地钻研认知诊断的已有研究,补充必要的心理学和统计测量学的知识,找到合适的切入点,以此为开端,探索认知诊断各环节的工作,认知诊断研究应该吸引更多的、各有关研究领域研究者参加,形成分工协作的研究团队,这样才能有效推进我国认知诊断的研究工作。

参考文献

- 1 余嘉元. 运用规则空间模型识别解题中的认知错误. 心理学报, 1995, 5.
- 2 戴海崎, 张青华. 规则空间模型在描述统计学习模式识别中的应用研究. 心理科学, 2004, 4.
- 3 康春花, 戴海崎. 采用 LLTM 作测量与认知结合研究的初步探讨. 心理科学, 2001, 5.
- 4 戴海崎, 刘声涛. 瑞文测验项目认知难度因素分析及 LLTM 拟合验证. 心理与行为研究, 2004, 2.
- 5 邓太平, 戴海崎等. IRT 计时模型在瑞文标准推理测验分析中的应用研究. 2004 年北京世界心理学大会论文.
- 6 Mislevy R J. Foundations of a new test theory. In: N Frederiksen et al. Eds. Test theory for a new generation of tests. Hills-

dale NJ : LEA ,1993. 19 – 39.

7 漆书青 戴海崎 ,丁树良 .现代教育与心理测量学原理 .江西 江西教育出版社 ,1998.307.

8 Embretson S E. Psychometric models for learning and cognitive processes. In : N Frederiksen et al. Eds. Test theory for a new generation of tests. Hillsdale NJ :LEA ,1993. 126 – 150.

9 Frederiksen N , et al. .Test theory for a new generation of tests. Hillsdale NJ :LEA ,1993.

10 Lohman D F ,Jppel M J. Cognitive Diagnosis :From Statistically Based Assessment Toward Theory Based Assessment. In :N Frederiksen et al. Eds. Test theory for a new generation of tests. Hillsdale NJ :LEA ,1993. 41 – 71.

11 James E. Corter. Using clustering methods to explore the structure of diagnostic tests. In :Paul D. Nichols ,Susan F. Chipman , Robrt L. Brennan Eds. Cognitively diagnostic assessment ,Lawrence Erlbaum Associates ,Inc ,1995.307.

12 Embretson S E. Introduction to the Problem of Test Design. In : Embretson S E. Eds. Test design :developments in psychology and psychometrics. Harcourt Brace Jovanovich , Publishers , 1985. 3 – 17.

13 Hans Spada , Barry McGaw. The Assessment of Learning Effects with Linear Logistic Test Models. In Susan E. Embretson Eds. Test design : developments in psychology and psychometrics. Harcourt Brace Jovanovich ,Publishers ,1985. 169 – 194.

14 Tatsuoaka K K. Architecture of Knowledge Structures and Cognitive Diagnosis :A Statistical Pattern Recognition and Classification Approach. In P D Nichols ,S F Chipman ,R L Brennan Eds. Cognitively Diagnostic Assessment. Hillsdale ,NJ :Erlbaum ,1995. 327 – 360.

15 Chipman S F , Nichols P D , Brennan R L. Introduction. In : Paul D. Nichols ,Susan F. Chipman ,Robrt L. Brennan Eds. Cognitively diagnostic assessment ,Lawrence Erlbaum Associates ,Inc ,1995. 1 – 18.

16 Haertel E H , Wiley D E. Representations of ability structures : implications for testing. In :N. Frederiksen et al. Eds. Test theory for a new generation of tests. Hillsdale NJ :LEA ,1993. 359 – 384.

17 Johnson R J , Goldsmith T E , Teague K W. Similarity ,structure ,and knowledge :a representational approach to assessment. In :Paul D. Nichols ,Susan F. Chipman ,Robrt L. Brennan Eds. Cognitively diagnostic assessment ,Lawrence Erlbaum Associates ,Inc ,1995. 237 – 243.

18 Hartz S M. A Bayesian Framework for the Unified Model for Assessing Cognitive Abilities :Blending Theory with Practicality. Unpublished doctoral dissertation ,University of Illinois at Urbana – Champaign 2002.

19 Hartz S , Roussos L , Stout W. Skill Diagnosis :Theory and Practice. User Manual for Arpeggio software. Princeton NJ :Educational Testing Service 2002.

20 Robert Henson , Jeff Douglas. Test Construction for Cognitive Diagnosis. Applied Psychological Measurement ,2005 ,29(4) : 262 – 277.

Test Theory for a New Generation of Tests
——the Origin and Features of Cognitive Diagnostic Theory

Liu Shengtao^{1 2} ,Dai Haiqi¹ ,Zhou Jun¹

(1. Education College Jiangxi Normal University ,Nanchang 330027 2. Education and Science College ,Fujian Normal University ,Fuzhou 350007)

Abstract Cognitive diagnostic theory , the core of test theory for a new generation is the product of the combination of cognitive psychology and modern measurement theory. Studies in cognitive diagnosis have got its momentum in psychological research abroad , and scholars at home have paid much attention on it. In order to advance cognitive diagnosis study in China , the article deals with seven important issues on cognitive diagnostic theory and technology such as the origin , concept , characteristics ,study bases ,framework , significance and study difficulties.

Key words cognitive diagnostic theory , standard test theory , test theory for a new generation of tests