

语言测试的效度本质观

冯巨澜^a, 宋德云^b

(重庆大学 a. 外国语学院; b. 语言认知及语言应用研究基地, 重庆 400044)

摘要:对效度本质的不同认识,导致了不同的效度理论,产生了不同的效度验证实践。实证主义效度本质观强调相关关系和因果关系,注重确定性和普遍性。自然主义效度本质观强调研究结果和研究的其他部分之间的一致性,注重主体间不断互动的活动和过程。效度理论和效度验证实践研究中存在的争议主要有:效度与信度、效度的整体性与可分性、效度与效度验证等。在探讨实证主义和自然主义研究方法以及应用于语言教学研究的理论和实践时,应该尽力弄清潜在的研究假设,应该意识到确定研究证据的依据的不同。

关键词:语言测试;效度;本质观;研究范式

中图分类号:H310.4

文献标志码:A

文章编号:1008-5831(2009)05-0129-06

效度与效度验证是语言测试研究中最关注的问题之一。综观语言测试发展的历史可以发现,语言测试走向科学化、合理化是其必然趋势。这就必然对测试的信度、效度等概念和操作提出更加科学化、精确化的要求。应当指出的是,语言测试中测量方面的方法和理论几乎都是从心理测验中移植过来的,语言测试实际上是心理学、心理测量学、教学理论及某种语言理论的综合。语言测试本身并没有独立的效度理论,或者说,语言测试界对效度理论的发展并无自己的贡献。语言测试在信、效度理论方面全面继承了心理测量的衣钵。所以,我们在研究效度问题时必须把眼光扩展到整个的心理、教育教学及语言测验领域。

效度在其长期的发展过程中,自身也经历了十分复杂的演变历程。效度研究发展至今,其中的问题仍然很多,如效度概念的本质、含义、特点、作用等问题,都存在着争论甚至混乱。同时,人们似乎更专注于对效度一时一事的静态研究,而着眼于效度发展动态的相关研究并不是很多。对效度本质的认识和理解,是研究效度问题的基础。为此,笔者尝试从本体论和认识论角度,对效度本质的不同理解进行分析,以及对效度研究中的一些争议问题进行梳理和探讨,以期对语言测试的效度研究提供参考。

一、范式^①之争与对话

在教育学和心理学界,直到现在仍然存在着两大主要阵营:实证主义的、量化研究方法论(实证主义范式)的拥护者和自然主义的、质化研究方法论(自然主义范式)的提倡者^[1]。这一争论的核心是对研究的认识论基础的讨论,而它

收稿日期:2009-03-20

作者简介:冯巨澜(1968-),女,四川广安人,重庆大学外国语学院讲师,硕士,主要从事外语教学理论、语言测试研究。

①“范式”指的是常规科学所赖以运作的理论基础和实践规范,是从事某一科学的科学家群体所共同遵从的世界观和行为方式(Kuhn, J, 1968, Structure of Scientific Revolution, New York)。近年来西方社会科学界对这个概念的借用十分广泛,社会科学家们掀起了一场又一场“范式大战”。本文所简化的“两大范式”,相当于定性研究和定量研究。事实上“范式大战”并非如此简单。有关详细的范式讨论请参阅:(1)Lynch B K. Language program evaluation: theory and practice [M]. Cambridge: Cambridge University Press, 1996;(2)陈向明. 质的研究方法与社会科学研究[M]. 教育科学出版社, 2000;(3)李森. 现代教学论纲要[M]. 人民教育出版社, 2005。

对语言教学研究理论发展的关联在于,当我们强调某种证据优于另一证据时,我们所提出的假设是不同的。实证主义范式的思想基础是力图用自然科学的研究方法来研究社会科学问题,认为社会科学问题能够客观测量,它们独立于观察者的价值观和自我感觉。相反,自然主义范式认为,社会科学领域人们观察的结果永远不可能独立于观察者的价值观和经验。在语言教学研究界,无论采用何种观点,关键是要重视研究的本体论基础(要了解什么)和认识论基础(如何去了解),重视他们(本体论和认识论)是如何对研究方法的选择产生影响的^[2]。

一些语言测试研究者认为范式之间的对话是有可能的。例如,Reichardt 和 Cook^[3]就指出,早期的评估研究就是假定一个设计好的实验可以产生特定的、希望得到的结果。在这种假定下,评估的目的就是确认预期的效果会发生。当发现一些社会干预不能够完全达到,而且会出现一些难以预料的副作用时,评估者就会利用自然主义研究方法来帮助他们发现和描述这些现象。Reichardt 和 Cook^[3]认为,研究范式的选择不能够决定研究方法的选择。你可以认为社会现象是客观现实,不依赖于个人的主观意志(也就是采用实证主义观点),而同样可以选择自然主义研究的质性研究方法(如参与式观察)来调查那些现象。自然主义范式的研究者同样可以利用像频数以及其他实证主义量化研究的方法。而另一方面,选择量化的方法并不意味着一定是实证主义范式。量化方法有时也用“主观”的手段,如涉及到人们感情和信仰等方面的民意调查(通常被认为是主观的)。仅仅以量化来表示并不能保证其具有客观性。

如果一种研究范式与一些研究方法没有必然的联系,那么还有没有必要在两种范式之间进行选择呢?为何不同时选择两者,使之互补?这种观点类似于 Denzin^[4]提出的术语“三角测量法”,即为了避免单一渠道或单一方法的片面性,而采取多渠道、多方法收集信息。当然,同时采用两种范式,既使用实证主义的方法又使用自然主义的方法,可能会费时费力。大多数研究者很难接受同时使用两种不同的研究范式。正如 Trend^[5]指出,“大多数研究者都不能够自如地处理两种类型的数据,而且也没有关于糅合两种数据的过程技术描述”。

反对将实证主义和自然主义方法结合使用,不仅仅是因为费时费力和“缺乏过程技术描述”。对一些研究者来说,两种范式的哲学基础的不同才是反对将实证主义和自然主义方法结合使用的根本原因。Smith 和 Heshusius^[6]指出,两种方法的结合使用会混淆两种范式的重要区别,从而导致方法定义的混乱,因为两种范式不只是单方面地探讨有关研究的程序和技巧等方法,而且还要讨论关于验证逻辑(logic of justification)的方法。验证逻辑主要侧重于判断哪些因素应该作为研究的证据,而研究证据的确定是由较高水平的假设决定的,即本体论表明要

了解什么,而认识论则确定如何去了解。尽管这些术语似乎具有浓厚的哲学色彩,但却有利于我们进一步了解语言教学研究是如何受其影响的。

二、实证主义范式的效度本质观

实证主义范式的证据观来自于实证主义哲学。该范式的主要观点是,人们所要了解的事物特性是独立于人们的看法的,而且其真实性是可以被了解的。人们对自己的研究和评价,可以通过协调观察或测量等手段来与独立的现实世界之间建立有效的联系。“在这种情况下,对有效性的判断,限定了方法和技术的选择。实际上,为确保客观性等,从而强制使用了一定的程序。如不使用这样的程序,就会被批评,被认为非理性和主观性”^[6]。受此观点影响,语言教学研究领域很长一段时间都是采用传统的、定量数据实验方法来进行研究。这类方法有两种,真实的实验和半实验^[7]。在真实的实验测评方法中,学生被随机分为实验组和控制组。通过对实验组和控制组学业成就测试的比较,来判定课程与教学的有效性(也就是实验组胜过控制组)。半实验也可以比较实验组和控制组学业成就,但有些因素并非随机。也就是,学生与学校可能是自愿加入,也可能是由别人选定(如学校领导),而不是完全的随机取样。非随机取样的问题是,能否仅从体系的不同中获得实验的结果?如果实施一段教学过后,再去比较两组的成绩,我们无法确定两组的不同是实验的结果还是根本与实验无关。例如,成绩好的学生也许会自愿参加实验。他们比控制组的学生表现得更好可能是他们自身的原因而非实验的因素。好在,相关的统计和设计程序考虑到了学生的先前差异。

实证主义的另一个重要观念是因果关系。Cordray^[8]总结了这一关系的要求:“第一,原因 X 先于结果 Y;第二,X 的变化引起 Y 的变化;第三,除此以外的其他解释均不可信。”他们认为,可以通过证伪的方法接近现实,这样“至少可以让我们知道还有哪些错误没有被排除”。Campbell 首次提出了教育实验研究中的实验有效性问题,并与 Stanley 将有效性分为内部有效性和外部有效性^[9]。内部有效性是指实验中的自变量和因变量之间的因果关系的明确程度。外部有效性是指实验结果能够推广到样本的总体中和同类现象中的程度,即是实验结果的代表性和适用性。后来,Cook 和 Campbell 从内部有效性中分化出一部分命名为统计结论有效性,即统计方法适切性所引起的统计结论的有效程度,主要反应统计量与总体参数之间的关系;又从外部有效性中分化出一部分称为构想有效性,指出变量之间关系构想的准确性以及实验变量操作定义与推论时的定义一致性程度。他们还探讨了影响有效性的因素,指出影响教育实验研究内部有效性的因素有 12 种:(1)历史;(2)成熟;(3)测试;(4)测试工具;(5)统计回归;(6)差异性选择;(7)实验者流失;(8)选择一成熟的相互影响;(9)实验处理扩散;(10)控制组的

补偿性竞争;(11)处理的补偿性均等;(12)控制组的不满而士气低落^[10]。

实证主义范式及其本质观主要注重两大方面:确定性和普遍性^[1]。内部效度主要关注确定性。在语言教学中,内部有效性主要研究教学的结果与教学任务之间的相关程度。而外部有效性主要关注普遍性,即主要研究教学结果的代表性和适用性。实际上,确定性与普遍性是一对相悖的概念。因此,一些语言教学研究家们对外部有效性提出了质疑。Cronbach^[11]对因果关系和有效性概念的描述已经超出了传统的逻辑实证主义的范畴了:“有效性不仅仅依赖于对数据的收集和分析,同时它还需要有一套对结果的陈述和交流的方法。有效性与其说是客观的,不如说是主观的。”

三、自然主义范式的效度本质观

Cronbach^[11]对有效性的描述,代表了一些自然主义范式研究者的证据观。自然主义范式起源于19世纪后半期的现象学及对社会探究的解释主义方法。通常,自然主义者信奉的现实世界不是客观的,事实与价值是紧密相连的,以及“现象只能放到周围环境中才能被真正地理解”^[12]。自然主义范式对实证主义研究的传统权威提出了挑战。自然主义研究追求一种自然发生的、变数的设计,与实证主义研究事先安排的或建构的设计相反。也就是说,研究设计在研究者的调查中会随时出现,允许在调查实验过程中改变信息收集方式和想法。同时,自然主义研究不希望控制条件和变量。重点放在观察、描述、解释和理解在真实世界而不是控制条件下事件是如何发生的。这种方法视被实验的项目为一个不断变化的动态过程,而不是静止的、无变化的处理。自然主义实验者收集数据主要是通过诸如深度访谈、参与式观察以及日志等形式进行的。

一些更加激进的建构主义者和批判理论者则认为,不论是“效度”还是其他类似的概念都不适合自然主义范式的研究。这类概念以事物是独立的、自足的客观存在为前提,认为研究者可以识别并验证其客观真实性^[13-14]。而自然主义范式研究的一个重要理论前提是:“客体”不是一个固定不变的实体,它是一个与主体相互配合、适应、转换和变化的另外一个“主体”。研究者对事物的理解不是简单的主体对客体的认知,而是主体与主体在一定社会文化环境中的重新相互建构。这是一个复杂的运动过程,主体间的理解受制于各自所处的研究情境。

尽管自然主义范式对是否应该使用和如何使用“效度”这一概念有不同的意见,但是绝大部分研究者(尤其是后实证主义观的研究者)仍旧沿用“效度”这一词语来讨论研究结果的真实性问题。然而,大家都同意,自然主义范式所使用的“效度”这一词语不论是在概念定义、分类方法还是使用范畴上都和实证主义范式的研究很不一样。前者使用的“效度”所指的“关系”,是研究结果和研究的其他部分(包括研究者、研究的问题、目的、对象、方法和情境)

之间的一种“一致性”。当我们说某一研究结果是“真实可靠的”时候,我们不是将这一结果与某一个可以辨认的、外在的客观存在相比较(事实上这一“存在”并不存在),而是指对这个结果的“表述”是否“真实”地反映了在某一特定条件下某一研究人员为了达到某一特定目的而使用某一研究问题以及与其相适应的方法对某一事物进行研究这一活动(或过程)^[15]。

自然主义范式中对话言测试影响较大的是Maxwell的效度观^[1]。作为一个后实证主义者,Maxwell认为,自然主义范式的研究可以继续使用“效度”这个词语,但是其定义和分类必须采取与实证主义范式不同的研究思路。自然主义范式的研究者应该从自己从事研究的经验出发,介绍自己在研究的过程中是如何思考、甄别和处理效度问题的。他强调考察在具体研究中通常会出现什么类型的效度问题(他称之为“效度威胁”,即对研究的真实性可能形成威胁的因素),然后反省自己是如何处理这些问题的。从研究者自己的角度来探讨效度问题比站在这之外或之上评头论足更具有可信度和说服力,因为这样更加贴近研究的具体实践。Maxwell^[16]将研究中的效度问题分成五种类型:(1)描述型;(2)解释型;(3)理论型;(4)推论型;(5)评估型。对语言测试效度理论产生类似影响的还有Messick的效度观和Henning对效度的定义。Messick^[17]认为,效度就是“对在测试分数基础上所做出的推断或所采取的行动是否充分地、适合地支持经验证据和理论基础这一问题的综合评价性判断”。此定义涉及到了效度理论层面的关系,即认为效度是一个整体的概念。Henning^[18]也提出,“效度一般指某个测试或其任何一个组成部分是否恰当地测量了预期测量的内容”。此定义实际上提出了效度与效度验证的关系问题。

综上所述,自然主义范式中的“效度”这一概念是用来评价研究报告与实际研究的相符程度,而不是像实证主义范式那样对研究方法本身的评估。实证主义范式假设研究对象是一个客观的实体,只要研究者遵循一定的方法和操作程序就可以保证获得可靠的数据和研究结果。而自然主义范式认为,客观的、固定不变的实体是不存在的,研究是一个主体间不断互动的过程。因此,“效度不是一个商品,可以用方法买到……效度就好像是品质、性格和质量,只能在与一定的目的和环境的关系之中加以测量”^[19]。“效度”不可能按照某种严格的、预定的程序被生产出来,只能依赖研究中存在的各种关系因素。当我们说某一结果的效度比较“高”时,我们不仅仅指该研究使用的方法有效,而是指对该结果的表述再现了研究过程中的所有部分、方面、层次和环节之间的协调性、一致性和契合性^[15]。

四、效度研究中的一些争议问题

对效度本质的不同认识和理解,必然会带来实践研究中的不同观点和方法。即使是相同的效度本质观,在具体的研究中也会有一些争议问题。下面

对一些主要的争议问题进行梳理。

(一) 效度与信度

信度与效度是语言测试的两大基本属性,效度体现的是测试的真实性,而信度则体现了测试的一致性^[20]。效度研究侧重探寻某测量工具在多大程度上反映所要测量对象的真正特征,而信度研究基本以相关分析为技术,以组间相关系数(interclass correlation coefficient)为基础,研究测量结果的信度系数估计问题^[21]。现代语言测试着重研究它们之间相互补充的关系。正如 Bachman 所指出,“很多有关信度与效度的论述强调两者之间的区别,而不是他们的相同之处。但我认为,如果把两者看成是同一个问题的两个方面将能使人们更好地理解它们”^[22]。Davies 认为,语言研究中的信度与效度就像形式(指信度)与内容(指效度)的关系^[15]。内容即意义所在;然而没有形式,内容就会消失。同样,效度给测试带来了生命力,但作为一个整体,测试也需要信度。

现代语言测试研究中的另一个观点是,把信度看作是效度的一部分^[15]。即信度是效度的必要条件,但不是充分条件;一个有效度的考试一定有信度,而一个信度很高的考试不一定有效度^[23]。信度可以独立存在,而效度与信度共存,即有效度就有信度^[21-24]。对统计学家来说,效度与信度有一种固定的关系,即一项测试的最大效度是其测试信度的平方根^[23]。同时,在有关解释学方法对测试效度的讨论中,有学者宣称信度并不像传统上所认为的那么重要了^[25]。

(二) 效度的整体性与可分性

多年来,效度研究的焦点逐渐演变。最初,其焦点放在具体参照标准的预测上,正如 Guilford^[26]的描述:“概括地说,一个测量总是需要相对于它所关联的那个东西有效。”之后,效度研究的焦点渐渐转移到有限几种效度类型上,即内容效度、效标关联效度和构想效度。然而在使用这些不同标准和方式对语言测试的效度进行多方面的探索时,人们逐渐发现,尽管不同种类的证据的相对重要性因测试而异,但根据其中任何一种标准所得到的信息本身都不足以表明对测试的某种特定解释或使用的效度,效度只有通过收集一切有关的信息并给以解释才能证明。Cronbach^[27]强调,“三种效度证明的类型(即内容、效标关联和构想效度)并不是可以任意选择的,对于几乎所有的测验都应该结合这三种方式共同建立一种解释,分开来任何一种都无法代表”。于是,专家们开始倾向于把不同类型的效度看成是包含在这个整体性效度概念中的不同的方面^[22]。

Messick^[17]提出的构想效度概念,是整体性效度理论中影响最大的一种,语言测试学界也对此倍加关注。Messick 用渐进矩阵(progressive matrix)的概念表达了自己的效度思想,认为效度的概念还应该涵盖测验结果的诠释与使用两方面,应扩展到影响后果之层面,因此这个 2X2 的矩阵包含两个维度:合

释及使用为一个维度(interpretation and use),证据基础及后果基础(evidential basis and consequential basis)为另一个维度。二者交互后便形成四种情况,这四种情况就是效度应该涵盖的层面。

与此同时,坚持效度可分性的理论也在不断涌现。Maxwell^[16]将研究中的效度问题分成五种类型(见自然主义范式的效度本质观)。Hammersley^[13]认为效度应该是多元的,而不是只有一种“效度”。他提出了“复杂的现实主义”的观点,认为“知识”是一些具有合理效度的信念;知识是人的建构,具有丰富多样的形态。所谓的“知识宣称”必须建立在可信和可能的基础之上,与研究者关于世界的假设具有一定的相容性。而研究者的假设是多元的,对于同一个现象往往有相互不矛盾的多种说明,因此其效度也应该是多样的。因此,我们应该抛弃对唯一“效度”的坚持,承认效度的可分性、多样性、丰富性和变化性。

(三) 效度与效度验证

效度与效度验证的关系问题实际上就是理论观念与实际操作的问题^[20]。有研究者将效度验证归纳为对“效度威胁”的排除^[1,16]。在实证主义范式的研究中,研究者可以在研究设计中通过随机抽样和控制组等手段事先将所有的“威胁”全部排除。在语言测试中,与此相对应的是“基于论证的效度验证法”(an argument-based approach)^[28-29]。这一方法基于效度验证即论证(假设检验)的思想,认为效度验证包括两个步骤:首先,提出效度验证观点,包括两方面的问题,试图对测试分数做哪些解释和使用?还有哪些与预期相反的解释和使用?其次,收集有关证据支持所提议的解释与使用,反驳相反的观点^[30]。

而在自然主义范式的研究中,“效度威胁”不可能事先被识别并通过统一的技术手段而加以排除。这是因为自然主义范式的研究者认为,自己所研究的事物不是一个脱离主体而单独存在的客观实体,不能单方面地被认知或证实,只能被主体在与其互动的关系中重新构建。因此,其效度也只能在这一过程中得到此时此地的、逐步的检验。由于自然主义范式的研究将“效度”看作是某一特定条件下的产物,其“效度威胁”也是具体、个别和动态的,因不同情况而有所不同。在语言测试中,相对应的效度验证方法是基于证据(evidence-based approach)的效度验证^[31]。这一方法进一步发展(而不是否定)了“基于论证的效度验证法”,认为效度验证即收集效度证据的过程。收集效度证据的方法主要有^[17,32]:(1)内容分析;(2)项目或任务的实验分析;(3)维度分析;(4)测试分数与其他测试和行为的关系;(5)测试行为的差异;(6)测试后果。这种基于论证的效度验证比较适用于各种测试的解释和使用。同时,尽管解释性论证不能对效度做出任何绝对性决定,但可以在整体上提高该分数解释的合理性^[33]。

五、结语

应用语言学曾经一度钟情于实证主义范式的最有可能的原因是,实证主义范式对要了解什么和如何去了解,给予了明确的相关方法和意义标准。使用确定的、规范的、客观的、较少需要解释的研究程序确实会引人注目。如果在研究中总带有讲故事似的、主观的以及猜测的词语或概念,会使读者觉得较为随意或不真实。然而,即使是最客观的研究程序,如量化方法的构想效度研究,也可能需要主观的解释。而且,一些专用的统计模式,像 Pearson 积矩相关等,也会显示出研究者的社会偏见^[34]。在承认了这些副作用的情况下,实证主义的研究方法确实有他的权威性和方便之处。

当然,自然主义研究也可以成为应用语言学的主要方法之一。如果现实世界不是客观的,如果没有确定的标准,那么我们的研究和评估就不能拒绝有意义的自然主义方法的可能性,就不能放弃探究理性的其他方法。Feyerabend^[35]曾提出,任何研究必须允许智慧的自由,避免任何预先的、固定的方法的观念和唯理性。

从研究程序和技巧角度将实证主义研究和自然主义研究结合起来是有可能的,但却不可能融合他们各自的验证逻辑。这两种范式,对要了解什么和如何去了解的看法是完全不同的。Smith 和 Heshusius^[6]认为,自然主义(质性)研究者和实证主义(量化)研究者之间已经“停止了对话”,部分原因是由于一些研究者如 LeCompte 和 Goetz^[36-37]等为质性研究“创造了具有同样的客观性的方法,而且对量化研究也会适用”。从这一点来看,并不是两种范式的相互适应,而是自然主义的技术和程序被实证主义范式所采纳^[38],但对所要了解的事物特性以及什么是证据的观点却各自为政。

范式对话,使得应用语言学家在进行效度研究时,不得不考虑一些重要的问题。研究者们如 Guba^[2]、Lincoln^[12]、Smith 和 Heshusius^[6]等的争论,使我们清楚地了解到,在效度验证方法中哪些应该成为证据,我们必须清楚研究的哲学基础。为了尝试实证主义和自然主义两种范式的方法,我们需要清楚地知道哪些可以成为证据。从实证主义的角度,用自然主义方法来进行效度验证就会有问题,因为自然主义分析的是没有建立在预定的程序中的发现。同样,从自然主义的角度,传统的依赖量化数据和实验设计也是不合适的,因为从随机的试验数据来汇报意义推论数据,本身就不能确定发现的效度。最后,如果我们打算混合应用实证主义和自然主义的方法,我们的结果肯定会被与其他的效度验证(可能是完全的实证主义或完全的自然主义)进行比较,在应用语言学领域内建立自己的可接受性。

这种讨论不是要解决实证主义研究者和自然主义研究者之间的争论,也不是提倡结束这场范式的对话。在此,笔者建议要了解这样的争论,而不是盲目地将量化的和质性的数据和方法揉和起来,似乎

两种范式的哲学假设是相同的。尽管多数研究者也许会接受现实世界是依赖于感觉的以及“任何事物都是相对的”主张,但这一领域的文献资料表明,直到最近,大多数效度理论研究包括效度验证所追求的仍然是实证主义观念中认为是有效的和可靠的效度理论。因此,我们在探讨实证主义和自然主义研究方法以及应用于效度理论和效度验证时,应该尽力弄清潜在的研究假设。也就是说,如果采用实证主义范式,则需要强调相关关系和因果关系,注重确定性和普遍性;如果采用自然主义范式,则需要强调研究结果和研究的其他部分之间的一致性,注重主体间不断互动的活动和过程;或者混合实证主义和自然主义的方法,那么则需要在应用语言学领域内建立自己的可接受性,需要确立验证证据有效性的标准和依据。

参考文献:

- [1] LYNCH B K. Language program evaluation: theory and practice [M]. Cambridge: Cambridge University Press, 1996.
- [2] GUBA E G, LINCOLN Y S. Fourth generation evaluation [M]. Newbury Park, Calif: Sage, 1989.
- [3] REICHARDT C S, COOK T D. Beyond qualitative versus quantitative methods [C]// In COOK T D, REICHARDT C S. Qualitative and quantitative methods in evaluation research. Beverly Hills, CA: Sage, 1979: 7-32.
- [4] DENZIN N K. The Research Act [M]. Chicago: Aldine, 1970.
- [5] TREND MG. On the reconciliation of quantitative and qualitative analyses: a case study [C]// COOK T D, REICHARDT C S. Qualitative and Quantitative Methods in Evaluation Research. Beverly Hills, Calif.: Sage, 1979: 68-86.
- [6] SMITH J K, HESHUSIUS L. Closing down the conversation: the end the quantitative - qualitative debate among educational inquirers [J]. Educational Researcher, 1986, 15 (1): 4-12.
- [7] COOK T D, CAMPELL D T. Quasi - experimentation: design and analysis issues for field settings [M]. Boston: Houghton Mifflin, 1979.
- [8] CORDRAY D S. Quasi - experimental analysis: a mixture of methods and judgment [C]// TROCHIM W M K. Advances in quasi - experimental design and analysis. New Directions for Program Evaluation, 1986, 31: 9-27.
- [9] CAMPBELL D T, STANLEY J C. Experimental and quasi-experimental designs for research [M]. Chicago: Rand McNally, 1966.
- [10] 梅雷迪斯·D·高尔. 教育研究方法导论[M]. 许庆豫,译. 南京:江苏教育出版社,2002.
- [11] CRONBACH L J. Designing evaluations of educational and social programs [M]. San Francisco: Jossey - Bass, 1982.
- [12] LINCOLN Y S, GUBA E G. Naturalistic inquiry [M]. Beverly Hills: Sage, 1985.
- [13] HAMMERSLEY M. What 's Wrong with ethnography? [M]. London & New York: Routledge, 1992: 52-54.
- [14] WOLCOTT H F. On Seeking - and - Rejecting - Validity

- in qualitative research [C]// EISNER W, PESHKIN A. Qualitative inquiry in education: the continuing debate. New York: Columbia University, 1990.
- [15] 陈向明. 质的研究方法与社会科学研究[M]. 北京: 教育科学出版社, 2000.
- [16] MAXWELL J. Understanding and validity in qualitative research [J]. Harvard Educational Review, 1992, 62(3): 279-300.
- [17] MESSICK S. Validity [C]// LINN R L. Educational measurement (3RD ed.). New York: American Council on Education/Macmillan, 1989: 13-103.
- [18] HENNING G. A guide to language testing [M]. Rowley, Mass: Newbury House Publisher, 1987: 89.
- [19] BRINBERG D, JOSEPH E M. Validity and the research Process [M]. Newbury Park: Sage, 1985: 13.
- [20] DAVIES A, ELDER C. Validity and validation in language testing [C]// HINKEL E. Handbook of research in second language teaching and learning. London: Lawrence Erlbaum Associates, Inc., 2005: 796-811.
- [21] 汪顺玉, 席仲恩. 复杂结构语言试卷信度的多元概化理论分析[J]. 重庆大学学报(社会科学版), 2008(6): 130-135.
- [22] BACHMAN L F. Fundamental considerations in language testing [M]. Cambridge: Cambridge University Press, 1990: 160.
- [23] BELLAK A S, HERSEN M. Research methods in clinical psychology [M]. New York: Pergamon, 1984: 48.
- [24] CHAPELLE C. Validity in language assessment [J]. Annual Review of applied Linguistics, 1999, 19: 254-272.
- [25] LYNCH B K. Language assessment and program evaluation [M]. Edinburgh, Scotland: Edinburgh University Press, 2003.
- [26] GUIFORD J P. New standards for test evaluation [J]. Educational and Psychological Measurement, 1946, 6: 427-439.
- [27] CRONBACH L J. Essentials of psychological testing (4th ed.) [M]. New York: Harper & Row, 1984.
- [28] KANE M T. Validating high-stakes testing programs [J]. Educational Measurement: Issues and Practices, 2002, 21(1): 31-41.
- [29] KANE M T. Current concerns in validity theory [J]. Journal of Educational Measurement, 2001, 38(4): 319-342.
- [30] BACHMAN L F. Statistical analyses for language assessment [M]. Cambridge: Cambridge University Press, 2004.
- [31] WEIR C J. Language testing and validation [M]. Palgrave: Macmillan, 2005.
- [32] MESSICK S. Validity of psychological assessment: Validation of inferences from person's responses and performance as scientific inquiry into scoring meaning [J]. American Psychologist, 1995, 9: 741-749.
- [33] SHEPARD L A. Evaluating test validity [J]. Review of Research in education, 1993, 19: 405-450.
- [34] SMITH J K. The evaluator / researcher as person vs. the person as evaluator / researcher [J]. Educational Researcher, 1988, 17(2): 18-23.
- [35] FEYERABEND P. Against method [M]. London: Verso, 1988: 4-5.
- [36] LECOMPTE M D, GOETZ J P. Problems of reliability and validity in ethnographic research [J]. Review of Educational Research, 1982a, 52(1): 31-60.
- [37] LECOMPTE M D, GOETZ J P. Ethnographic data collection in evaluation research [J]. Education Evaluation and Policy Analysis, 1982b, 4: 317-400.
- [38] BERATTA A. Toward a methodology of ESL program evaluation [J]. TESOL Quarterly, 1986, 20(1): 144-145.

Perspectives of the Nature of Validity in Language Testing

FENG Ju-lan, SONG De-yun

(a. College of Foreign Language; b. Research Center of Language, Cognition and Language Application, Chongqing University, Chongqing 400044, China)

Abstract: Different understanding of the nature of validity leads to diverse validity theory and validation practice. The positivistic paradigm focuses on certainty and generalization. The naturalistic paradigm emphasizes the degree to which the evaluator and evaluation audience can agree on, and feel confident about, an interpretation of the evaluation findings. The challenges of validity research are: reliability and validity, the unitary and the divisible, and validity and validation, etc. The key notion is that although it is clear that important differences remain between positivistic and naturalistic perspectives on validity, validity is establishing what counts as evidence. We have seen that the criteria for doing this can differ depending on the research perspective being used.

Key words: language testing; validity; perspective of the nature of validity; paradigm

(责任编辑 胡志平)