

效度概念新认识

叶映华¹, 郑全全²

(1.杭州师范学院心理健康研究所, 浙江 杭州 310036; 2.浙江大学心理系, 浙江 杭州 310028)

【摘要】 本文主要讨论了效度概念的演变和存在的问题、效度概念新认识及如何寻找效度三个问题。效度概念的发展从内容效度、预测效度、构想效度和同时效度并存发展到现在由构想效度涵盖了其它所有效度。但一直以来,把效度概念作为一个法理学网络来理解,造成了我们对效度概念理解的很多误区。事实上效度概念是一种本体论的观点,重视现实参照并依据因果关系。我们应该通过建立一个理论假设、形成一个测验和确定测量的属性来寻找效度。

【关键词】 效度; 法理学网络; 本体论; 现实参照; 因果关系

中图分类号: R395.1

文献标识码: A

文章编号: 1005-3611(2006)06-0263-03

The New Understanding of Validity Concept

ZHENG Quan-quan, YE Ying-hua

Department of Psychology, Zhejiang University, Hangzhou 310028, China

【Abstract】 Three problems about the concept of validity were discussed here, which is the evolution of concept of validity and its existing problems, a new understanding of validity and how to construct validity. During the development of the concept of validity, content validity, predictive validity, construct validity as well as concurrent validity were prevailing until the emergence of construct validity in which all aspects of validity were involved. To date, the bias we put on the concept of validity were mainly originated from the way by which we have taken concept of validity as a nomological network. Actually, the concept of validity is a ontological claim which is seriously based on reference and causality. We should establish a theory hypothesis, form a test, and confirm the measure attribute, so we could obtain the validity.

【Key words】 Validity; Nomological network; Ontology; Reference; Causality

效度是心理和教育测验中主要的技术指标,是衡量一个测验是否客观、准确的重要标准之一。对于一个如此重要的概念,我们经历了一个从复杂到简单,从浅显到深刻,从模糊到清晰的理解过程。本文主要探讨了效度概念的演变和存在的问题、效度概念新认识 and 如何寻找效度概念三个问题。通过这些问题的探讨,希望能够对教育和心理测验研究方法中所存在的一些误区有更进一步的认识。

1 效度概念的演变和存在的问题

1.1 效度概念的演变

在上个世纪50年代以前,有关心理测验的效度还没有一个很精确的概念,这种情况一直延续到美国心理学会(American Psychological Association)心理测量委员会在50年代初意识到一个测验在发表前应该测量指定的一些属性为止。当时他们区分出了四种效度,包括内容效度、预测效度、构想效度和同时效度。其中一个最主要的革新是1955年Cronbach和Meehl所提出的构想效度的概念。Cronbach和Meehl认为构想效度指的是一个测验测量了理论构想的完美程度。而要想说明一个测验有构想效度,他们认为应该为测量构建一个法理学网络(nomological network),这个网络包括想测量的理论框架和如何测量的经验框架,以及两个框架之间关系的详细说明。

美国心理学会认为构想效度指个体在多大程度上具备测验成绩所反映出来的特点或构想。

Schwab提出构想效度指描述一个构想和同测量或操纵这个构想的操作程序间的相似性。他认为构想效度是研究过程中一个必要和主要的元素。

Messick和Kane认为,构想效度是效度的总和,并且三种效度(内容效度、效标关联效度和构想效度)实际上就是一种,因为它们都是验证测验有效这一过程中的必要部分,包括内容、实质、结构、概化、外化和推论六个部分。

Deborah Lee提到,效度指的是一个测验准确测量它所测量的东西的能力,包括内容效度、效标关联效度和结构效度。其中内容效度指的是与被测量的结构有关的数据信息收集的足够程度;效标关联效度指对测验工具(如问卷)的预测或估计能力的测量;结构效度指的是在经验证据不存在时,测量或推断抽象特征是否存在。

在《心理和教育测量》(戴海崎、张峰等编)教材中,效度是指对于一个既定的目标,在做出决策和提供解释的过程中,测量的有用性程度,包括内容效度、效标关联效度(预测效度和同时效度)、构想效度。

效度概念的演变实际上就是构想效度概念演变的过程。从1955年Cronbach与Meehl提出构想效度概念以后,心理测量学家对这一概念就存在两种不同的看法。反对的人认为它会导致对测验效度的主观臆测,支持的人认为它涵盖了所有其它的效度证据。后一种观点已经占据了明显的上风,构想已经成为教育与心理测量中最重要、最核心的概念之一。这里,构想不再是效度证据三种来源的一种,而是被用来定义效度概念。这一改变表明,在主流教育与心理测量学界,今后已经不再存在“构想效度(construct validity)”这一概念。所谓效度,就是测验对构想进行测量的有效程度,或者指一个测验是否测量了它该测量的东西。

1.2 效度概念理解中存在的误区

Meehl和Messick等提出的用构想效度来涵盖其它所有

效度这一观点逐渐得到了越来越多的认同,如孙晓梅等认为效度发展经历了三个阶段,早期关注两个变量之间的相关,随后转向关注效度的多种类型,现阶段则注重考察测验的构想效度。但要判断一个测量是否有效,其实无须考虑太多,只需关注测验是不是测量了它应该测量的东西,如测量一个人的数学能力或语言能力。但是 Meehl 等人的文章中更关注的是测验得分的解释是否与一个既定的法理学网络相一致,而不关注更关键的问题,即测量的属性是否存在。

上述的研究历程说明了过去 50 年效度文献的一个显著特点。理论家们关注的效度概念与实践家们关注的效度概念有很大的不同。因为在过去的几个世纪里,效度的问题是从“是否测量”和“想测量什么”问题(Cattell, Kelley)发展到“测验得分之间的经验关系与一个法理学网络中的理论关系是否相匹配”问题(Cronbach & Meehl),最后发展到“解释和行为依据测验得分是否公正”的问题——不仅仅依据科学证据,同时依据社会和伦理使用测验的结论的问题(Messick)。因此,效度理论逐步把任何重要的与测验有关的问题看成同效度概念有关,并致力于把所有的问题统合在单个标题之下。然而,这样做的话,这个理论既不能服务理论导向的心理学者又不能服务于实践导向的测验者:理论导向的心理学者会在错综微妙的效度理论中迷失,反之实践导向的测验者不能从中获得包含实践含义的有效概念方案。

事实上大部分的效度文献都没有清楚地说明效度概念。效度不复杂、不具多面性,也不依据法理学网络和测量的社会后果。它是一个非常基本的概念,可以非常简单和恰当的表达,举例来说, Kelley 把效度定义为一个测验测量了它想要测量的东西。

当然对于用构想效度涵盖所有的效度是否合理、是否准确并不是本文讨论的重点。由于我们在效度认识上一直所存在的误区,本文只讨论效度究竟为何物,以及对其的一些新的认识。而且我们并不想过多的从概念上探讨效度,更多的是讨论测量或测验实践中对效度的理解。

2 效度概念新认识

如上所述,对效度概念的理解实际上是非常简单的。如果某种属性不存在,则不能测量,只有当某种属性存在时候,才能测量它;如果某种属性存在但不是产生测量程序结果变异的原因,那么同样不能测量任何东西或差异。因此,只有当属性存在且属性的变化是产生测量程序结果变异的原因时,一个测验对测量一种属性才是有效的。简单地说,效度指的是一个属性存在且是引起差异结果的原因。本文从三个方面阐述对效度概念的新理解。

2.1 本体论相对于认识论

效度概念究竟是一个本体论问题还是一个认识论问题?一直以来,对效度概念的认识都认为其是一个认识论问题,事实上效度概念主要的是关注一个属性的存在,而这个属性是影响测量结果的原因,也就是属性和结果之间有因果关系,因此效度概念主张的是本体论而不是认识论。

假如我们把测量看成是一种认识论所认为的活动,则很容易犯错误,因为在测量之前我们似乎就已经知道测量的属性了,而事实上属性并不确定。只有当本体论确定了存在的属性后,测量程序才能够被用于寻找它所指的属性。举个简单的例子,如果智力的差异造成了 IQ 得分的差异,那么 IQ

得分的差异能够被用于发现智力的差异。从这个例子我们可以发现,本体论所主张的存在保证了认识论的正确认识活动。我们应该认识到本体论主张的真理必然地先于测量过程本身,因为这是测量成为可能的必要条件。

上面的这些问题确实使研究者感到非常迷惑,因为除了所包括的本体论的主张之外,测量没有普遍的特点。所有的测量程序所包含的唯一共同的东西是一个假设:存在一个属性,这一属性在长而复杂的一系列事件中导致了测量结果,在决定测量的价值时扮演了原因的作用。这不是复杂而模糊的概念,而是非常简单的思想。然而,如果一个人没有考虑这些,它将会得出非常表面化且具有认识论特点的复杂解释,这些与效度问题无关。

要确定一个测验是有效的,一个人应该持本体论的观点,即被测量的属性存在且影响测量程序的结果。很多研究者可能都觉得这比较难接受。但这并不是指效度概念本身是复杂的。

与测验有关的所有可能的活动的联合并不是效度(validity),而是确认(validation)。这两个术语在有些文献中可能会交换使用,但很明显它们不是同一个东西,因为效度是一种特性,而确认是一种行为。特别是,确认是研究者承担的寻找一个测验效度特性的一种行为。效度是一个概念,像真理;确认更像一个理论检验。效度是有关本体论的,确认是有关认识论的。两者不能被混淆起来。现在,许多的效度文献不是解决效度问题而是解决确认问题。虽然在描述、分类和评价确认策略上没有错误,但这种确认行为不能说明效度概念本身。事实上,如果一个人过于关注认识论问题,他会远离而不是更加接近效度概念。效度根本不是一种判断,它是被判断的特性。

2.2 参照相对于意义

在前面部分我们已经提到效度中的关键问题是一个属性的存在,这个属性是影响测量程序结果的原因,因此在理解效度理论时我们主张本体论不是认识论,获取足够的认识论经验只是为了借助一些基本事实。与此同时,这一问题还意味着,效度的主要关注点之一是参照:心理学家所指的属性必须存在于现实中,否则,测验不能有效的测量属性。这并不是指属性不能随着时间改变或心理属性是不能改变的存在物,而指的是分析我们所参照的理论术语时需要一些术语所指现象的现实位置,即测量应当包括被测量属性的现实存在。

但是很多哲学运动中所形成的广泛的效度理论并不这样认为,这些运动中最强大的是逻辑实证主义。他们认为理论术语的构建是无参照的,他们关注的只是理论术语的意义。但意义和参照是完全不同的。如晨星和晚星有不同的意义(即晚上最后一颗星星和白天最早的星星),但指的是同一回事情,只是参照点不同。因为实证主义者对形而上学有轻微的恐惧态度,他们希望解释理论术语的意义,而不让这些术语指向现实。

逻辑实证主义的努力遭到了失败,但正是这种失败把效度理论引入了今天的法理学网络。一个法理学网络就是某种体系的法律,把理论术语和其它术语及观察资料联系在一起。对实证主义者来说,这种网络用于为理论术语创造意义而不需要现实参照。他们认为理论术语的意义单独由术语在

网络中的位置决定,而不是由指向实际的参照决定。根据这一观点,理论术语可以不需要参照而获得意义。这个观点却被 Cronbach 和 Meehl 用来改进他们的结构效度理论,他们的理论主要依赖于一个法理学网络。这个网络为心理学构想提供了一个明确的定义,同时它也产生了结构效度观点本身,指的是法理学网络和经验数据的一致性。也就是说,如果测验得分间的经验关系与结构间的理论关系相匹配,也就是实践与理论的结果相一致,那么一个测验在构想上被认为是有效的。举例来说,假设法理学网络认为构想“焦虑”可能与构想“压力”有关,而与构想“犯罪行为”无关。进一步说,研究得到了一个焦虑测验和一个压力测验间的相关是 0.5,而与监狱里度过的时间之间相关是 -0.4,这就是经验关系和理论关系间的一个很好的匹配,而他们认为这种匹配组成并界定了效度概念。

然而,这个观点还是有很多问题,它使科学程序出现倒退。我们先画一个关系网络图,中间有一些空点,然后就认为空点必定是暗含的构想,构想的特点可以从已经画出的网络图中读出。当然,考虑一种属性与另一种属性的关系非常重要,这是理论构建过程的关键一步,但若认为这样一个网络关系图明确界定了问题中的属性,或者认为测量程序的效度来自被测量属性和其它属性的关系则有点牵强。长度的界定不依据与宽度的关系,更不用说米尺的效度。第二个问题是,限定理论术语的意义需要一个严格限制的法理学网络,只有少数心理学的法理学网络足够详细来完成限定理论术语的工作。因为一个人若要逃避现实主义的形而上学,他不能说“智力是对测量有因果影响的一个真实的属性”,而只能说“智力是一个理论术语,根据它与法理学网络中的其它构想的关系界定”。

总之,一个实证主义者对构想效度的理解需要依据一个描述性理论,必须援引法理学网络。然而,结构效度的法理学网络解释是不充分的,像我们上面所提到的。停止对法理学网络的讨论仅仅指个体应当回到心理学属性现实解释中来。然而,在一个现实解释中,关键问题是参照而不是意义。因此,此类问题如“IQ 测验对智力是否有效?”只能够在先验假设之下提出来,这个先验假设就是在使用智力术语时事实上已存在所指定的属性。

2.3 因果关系相对于相关关系

在上面的讨论中我们得出若不考虑法理学网络,用参照代替意义,重新介绍本体论观点,所有的这些使得构想效度理论更合理。同时,人们用构想效度理论来消除非理论的、经验主义者的效标关联效度思想(包括预测效度和同时效度),这是很重要的一项工作,因为效标关联效度是测量理论和效度理论中最主要的错误之一。效度存在于一个测验和一个效标相关关系中的思想已经阻碍了很多对效度的理解,这种阻碍一直在延续。

效标关联概念中最错误的观点是认为效度指的是相关。事实上,效度关注的是具有方向性的因果关系。当然,因果关系是一个沉重的概念,许多研究者都怕使用它。但测量是一个因果概念,不是一个相关概念,效度同样也是。有三点理由:首先,若以相关系数衡量效度的话,世界上的任何事情,尤其是在社会科学中,往往都有相关,但是在现实生活中,任何事情遭遇零相关的可能性本身就非常小。因此,若把相关

看成有效性的决定特性,在某种程度上说对任何事情的测量都变得有效了,这是不合理的。但这种不合理性在因果理论中不会出现,因为任何事情都具有因果关系是不大可能出现的。其次,效标关联效度把效度的程度与相关系数的水平等同起来,相关越高,效度越高。若两个变量的相关非常完美,则效度也非常完美。这种推论实际上也存在问题。举例来说,假设某个人测量打雷的出现。阅读文献可能会查到其与闪电的出现有一个完美的相关,因为它们都是云层光电发射的结果。然而虽然它们有很好的相关,但却不是同一回事。当一个测验有效的测量了打雷的出现时,它没有有效的测量闪电的出现,原因很简单,因为它根本没有测量闪电的出现。所以完美的相关不一定有完善的效度。这种不合理性在因果理论中不会出现,因为闪电出现的变化不是打雷出现的变化原因。第三,相关是依据总体的统计,它对总体被测量的属性的差异量非常敏感,但有时候这一点也会有很多不合理性。假设测竿子的长度,测量工具使用的是米尺,同时假设这一测量没有误差,测量结果和真实长度之间的相关对大部分总体是一致的。但在一个总体中它们的相关却是零,这个总体就是相同长度的竿子。因此,根据数据结论个体会认为在这样的总体中,米尺对测量长度是无效的。这是一个奇怪的结论。在因果关系中,这种荒谬性不会出现。因为因果关系是有方向性和条件性的:因果叙述认为如果属性间有差异,这些差异就会造成测量结果的差异;同样的,如果属性间没有差异,那测量结果也不会有差异。

3 如何寻找效度

现有的一些教材和文献认为构想效度的获得可以通过相关分析、专家评判、因素分析和结构方程模型获得,并且认为构想效度的确定通常需要评价多种不同的资料,一般采用两种途径:一是考察它与度量同类构想的测验是否相关;二是考察它与不应有关的东西无关。大多数心理学研究在考察效度时都是以相关系数作为衡量标准的,Westen 等曾经提出两个计算效度的简便的公式 $r_{\text{detering-CV}}$ 和 $r_{\text{contrast-CV}}$ 。

通过对效度概念的重新认识,我们知道效度并不是特别复杂的事物,它仅仅向我们说明了一个属性,由一个理论术语所指派的属性。由于这个属性存在,对这个属性的测量能够用一个给定的测验完成,属性差异是测验得分差异的原因。通过对这个概念的理解,我们可以简单和有效的获得效度,具体的说,效度的获得应该考虑以下几点:首先要有一个理论假设。许多效度研究关注创建一个相关系数表格,然后检查它们是否有一个正确的方向。虽然这种宏观水平的相关是重要的,但效度研究的首要目标不是创建一个正确方向的相关,而是提供测量结果形成过程的一个理论解释。也就是说,效度研究的第一步是形成一个有关属性差异和测量得分差异间因果关系的理论解释。形成一个测验。测验结构线性推论的结果非常清晰。纯粹的经验主义的方法,依据外部相关的最优化,是不大可能产生被认为是效度测量的测验。因为关注预测性指标对效度没有好处,其原因是测量同一属性的条目间会有相关,但这些相关的条目对预测没有作用,它们产生了多重线性。最优的预测性条目是条目间无相关,而与效标有高相关。这也说明最佳的测量效度与最佳的

(下转第 269 页)

参 考 文 献

- 1 Hammill DD. On defining learning disabilities: An emerging consensus. *Journal of Learning Disabilities*, 1990, 23(2):74-84
- 2 李晓东,等. 初中二年级学生学习困难、人际关系、自我接纳对心理健康的影响. *心理发展与教育*, 2002, 2: 68-73
- 3 凌辉. 学绩不良儿童行为问题及自我概念的对照研究. *中国临床心理学杂志*, 2001, 9(4):271-272
- 4 凌辉. 父母养育方式与学习不良儿童行为问题及自我意识的相关研究. *中国临床心理学杂志*, 2004, 12(1):50-52
- 5 Alloway TP, Susan E, Gathercole SE, Willis C, et al. A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 2004, 87:85-106
- 6 周世杰. 学习困难儿童的工作记忆研究. *中国临床心理学杂志*, 2004, 12(3):313-317
- 7 Kail R. Developmental changes in speed of processing during childhood and adolescence. *Psychological Bulletin*, 1991, 109:490-501
- 8 Kail R. Speed of information processing: Developmental change and links to intelligence. *Journal of School Psychology*, 2000, 38:51-61
- 9 Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychological Review*, 1996, 103:403-428
- 10 Swanson HL, Sachse-Lee C. Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology*, 2001, 79:294-321
- 11 Geary DC, Hoard MK, Byrd-Craven J, et al. Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, 2004, 88:121-151
- 12 龚耀先,蔡太生. 中国修订韦氏儿童智力量表手册. 长沙:湖南地图出版社,1993
- 13 范晓玲,龚耀先. 4-6年级多重成就测验的编制. *中国临床心理学杂志*, 2005, 13(3):253-257
- 14 周世杰,张拉艳,杨娟. 工作记忆成套测验的编制及在小学生中的初步效度分析. *中国临床心理学杂志*, 2005, 13(3):261-264
- 15 龚耀先. 非文字智力测验手册. 湖南医科大学, 1997
- 16 Siegel LS. Phonological processing deficits as the basis of a reading disability. *Developmental Review*, 1993, 13:246-257
- 17 Swanson HL. Reading comprehension and working memory in learning-disabled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Child Psychology*, 1999, 72:1-31
- 18 Swanson HL, Ashbaker MH. Working memory, short-term memory, speech rate, word recognition and reading comprehension in learning disabled readers: Does the executive system have a role? *Intelligence*, 2000, 28(1):1-30
- 19 Keeler ML, Swanson HL. Does strategy knowledge influence working memory in children with mathematical disabilities? *Journal of Learning Disabilities*, 2001, 34(5):418-434
- 20 Passolunghi MC, Siegel LS. Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology*, 2001, 80:44-57

(收稿日期:2007-01-31)

(上接第265页)

预测性并不是同一回事。确定测量的属性。测验效度的大部分工作应该关注测验构想环节,而不应该关注测验分析环节。目前效度研究的问题不是难以发现测量了什么,而是难以确定所要测量的属性。我们甚至可以认为效度不是也不可能单纯是方法论问题,当然这并不是说方法论的和心理测量的技术与效度研究无关。但对“测验是如何工作?”这一问题的理解首先是属性的真实存在性而不是方法论。所以我们绝对不会认为一个效度测验就是一个矩阵结构。效度的构想是一个属性的存在对得分的因果影响,任何矩阵表都不会是这种因果影响的替代物。因此,属性反映的因果过程来自于真实存在的心理学过程,而不是方法论原理。

参 考 文 献

- 1 Denny B, Gideon JM, Jaap van H. The Concept of Validity. *Journal of Psychology Review*, 2004, 111(4):1061-1071
- 2 Scott WO, Robert JV. The empirical assessment of construct validity. *Journal of Operations Management*, 1998, 16:387-405
- 3 Edward FM. Integration of construct and external validity by means of proximal similarity: Implications for laboratory experiments in marketing. *Journal of Business Research*, 2004, 57:142-153
- 4 Gray B. On assessing the construct validity of two multidimensional constructs: Occupational commitment and occupational entrenchment. *Human Resource Management Review*, 2001, 11:279-298
- 5 Drew W, Robert R. Quantifying Construct Validity: Two Simple Measures. *Journal of Personality and Social Psychology*, 2003, 84(3):608-618
- 6 Deborah L. Survey Research: Reliability and Validity. *Journal of Library Administration and Management*, 2004, 18(4):211-212
- 7 孙晓敏,张厚燊. 效度概念演进及新进展. *心理科学*, 2004, 27(1):234-235
- 8 张宝均. 传统效度理论与统一构想效度的比较研究. *北京林业大学学报(社科版)*, 2003, 2(4):69-72
- 9 郭靖,龚耀先. 小学生(四-六年级)学习能力倾向测验的初步编制:效度检验. *中国临床心理学杂志*, 2005, 13(3):271-273
- 10 张凯. 测量是理论的组成部分-再谈构想效度. *云南师范大学学报*, 2004, 2(5):7-14
- 11 谢小庆. 1999年新版《教育与心理测试标准》介绍. 沈阳出版社

(收稿日期:2006-05-29)