

情境判断测验的研究进展

陈 晶 车宏生

(北京师范大学 心理系 北京 100875)

摘 要:目前,情境判断测验的相关研究主要呈现两大趋势——效度研究和跨文化比较研究。其中,效度研究主要分为效标关联效度和构想效度研究,跨文化比较研究则主要探讨测验的文化公平性,以及对不同种族员工工作绩效的预测作用。文章还将介绍研究呈现出的新趋势——对情境判断测验本质的研究,即探讨测验形式特点和被试信息加工过程对测验结果的影响。

关键词:情境判断测验 效度研究 跨文化比较研究

中图分类号:B841.2

文献标识码:A

文章编号:1003-5184(2006)04-0078-05

1 引言

情境判断测验是一种有效的人事选拔工具,它通过向应聘者描绘出与工作相关的问题情境,让应聘者评价或选择与该问题相关的系列反应,从而判断应聘者是否具备该工作所要求的胜任特征。心理测量学领域对情境判断测验的应用可以追溯到 20 世纪 20 年代,近年来国内外研究者对它的关注日益增多,该测验之所以越来越被重视,主要是因为以下四点原因:1)具备更高的表面效度;2)开发成本较小,然而效标关联效度较高,尤其是针对管理岗位以及需要人际技能的岗位;3)与智力测验以及工作经验、工作知识有较高的相关;4)相比传统的认知能力测验,它对不同文化群体产生的负面效应更小。

作者总结了近年来国内外对情境判断测验的研究,发现目前该领域的研究主要集中于对效度以及跨文化比较的探讨。文章在简要介绍这些进展外,还要进一步探讨研究的一个新趋势——对测验本质的研究。

2 效度研究

2.1 效标关联效度

情境判断测验是一种应用于实务的工具,判断它是否有效的最直接的方法就是检验被试测验成绩与工作绩效的相关。McDaniel 等人在 39 个情境判断测验的研究基础上作了元分析,发现测验成绩和工作绩效的平均相关达到了 0.34^[1]。但是具体的研究得出的结论却不尽相同,Chan 和 Schmitt 发现被

试的测验成绩和核心技术掌握、工作奉献、人际协助、总体绩效表现的相关分别达到 0.30、0.38、0.27 和 0.30^[2]。Weekley 和 Jones 发现针对零售店小时工的预测效度达到了 0.22^[3],而对酒店管理员的预测效度仅达到 0.16^[4]。Motowidlo 等人针对七家电话公司的一般职员和管理人员进行了研究,考察了测验成绩和人际技能、问题解决、沟通和总体绩效的相关,最终得到分别为 0.35、0.28、0.37 和 0.30 的效度值^[5]。Phillips 用他所发展的销售技巧问卷来测查电话销售人员的绩效,发现测验成绩和工作职责、销售业绩、总体工作表现的相关分别为 0.24、0.18 和 0.16。后来他又针对电话服务人员进行了类似研究,发现相关分别达到了 0.45、0.41 和 0.43。Stevens 和 Campion 用这种测验来预测个体在团队中的绩效,结果表明测验成绩与团队协作的相关达到了 0.44,而与工作绩效的相关达到了 0.56^[6]。然而,Smiderle, Perry 和 Cronshaw 的研究却表明这种测验的信、效度都很差。

2.2 构想效度

除了效标关联效度,许多研究者还关心情境判断测验究竟测查了哪些特质或能力,以从根源上证实它对工作绩效的高预测作用。

2.2.1 智力

一些研究者认为情境判断测验之所以能够很好地预测工作绩效,就是因为它在很大程度上测查了智力,而智力是对工作绩效最好的预测变量。如

Weekley 和 Jones 的研究就发现 ,844 名零售店员工的情境判断测验成绩和智力测验成绩的相关达 $0.42^{[3]}$,而 1040 名酒店管理员的结果显示两者相关达 $0.48^{[4]}$ 。Stevens 和 Campion 则发现被试情境判断测验和能力倾向测验之间的相关达到 $0.81^{[6]}$ 。针对学生进行的研究也得到类似结果 ,如 Lievens 通过两项研究证明学生的测验成绩和智力显著相关^[7,8] 。对此 ,有研究者做出了解释 :McDaniel 认为 ,任何测量判断能力的测验都必然要和智力产生联系 ,Northrop 表明在情境判断中 ,智力是“判断”一个整合的部分 ,人们不可能分离这两者。所以情境判断测验要完全排除智力因素的干扰不现实。

2.2.2 工作经验或工作知识

由于大量研究都发现情境判断测验和工作经验、知识存在一定相关 ,所以还有研究者认为情境判断测验可能测查了人们的工作经验或工作知识。这些研究包括 :Weekley 和 Jones 发现测验成绩和经验的平均相关达 $0.20^{[4]}$;Clevenger 对三个样本进行了研究 ,得出测验成绩和知识的相关分别为 0.13 、 0.19 和 $0.37^{[9]}$;McDaniel 的元分析结果表明测验成绩和经验的平均相关为 $0.07^{[1]}$ 。

2.2.3 其他新结构

已有研究发现 ,在排除智力和工作经验对绩效的预测作用后 ,情境判断测验还能解释一部分绩效变异。如 Weekley 和 Jones 采用回归分析来探讨录像版情境判断测验的递增效度时发现 ,情境判断测验解释的独特变异值比智力或者经验能够解释的都要大 ,事隔两年后再进行的研究也验证了该结果 ;Clevenger 发现在对于绩效变异的解释上 ,情境判断测验在智力、工作经验和工作知识之外还存在递增效度。这些证据都表明 ,情景判断测验测查的结构可能和智力、经验、工作知识都不一样。至于这种新结构究竟是什么 ,研究者存在不同意见。

一种观点认为情境判断测验测查的是内隐知识。这一概念由 Sternberg 提出 ,他认为这种知识通常是行动导向的、应用于实践的、由非正规途径获得的(如通过工作经验的积累获得) ,这种知识能够帮助人们更好地应对社会环境、更好地获得工作以及实际生活的成功。另一种观点认为情境判断测验可

能测查了情绪智力。情绪智力由 Goleman 提出 ,这也是一种非认知技能或能力 ,能够帮助人们有效处理真实社会情境。Grubb 的研究显示测验成绩和情绪智力之间的相关达到了 0.20 。还有人认为情境判断测验测查了社会智力 ,这一概念是由 Legree 和 Busciglio 提出的 ,他们认为社会智力像内隐知识一样 ,也能够帮助人们应对社会情境。尽管各领域研究者使用的术语不同 ,但一些 IO 心理学家认为针对内隐知识、情绪智力的测验等无论是从内容还是形式来看 ,都和情境判断测验十分相似 :首先 ,测查的概念独立于智力 ,且有助于人们应对真实的社会环境 ;其次 ,形式上都采用情境化的试题 ,在方法上存在共通性。所以 ,一些学者认为情境判断测验可能测查了类似新结构。

2.2.4 任何结构

近来有一种新观点认为情景判断测验没有具体测查某一特质或者能力 ,它只是一种测验方式 ,研究者可以根据不同的研究目的和设计 ,测查任何结构。持这种观点的学者包括 :Hanson、Horgen 和 McDaniel 等。

总的来看 ,目前学术界对情境判断测验的构想效度的意见分歧还比较大 ,但多数实证研究还是采用最后一种观点 ,即依据研究目的 ,设计不同的情境问题 ,从而考察不同的结构维度。此外 ,还很少有研究检验情境判断测验和人格特质、动机、价值观等方面的关系。尽管 McDaniel 在元分析中发现 ,情景判断测验成绩和宜人性、责任心、情绪稳定性之间的相关分别为 0.25 、 0.26 、 0.31 ,但是目前还没有研究者认为情境判断测验测查的是某些个性特征。

3 跨文化比较

情境判断测验的另一个研究趋势就是跨文化比较。一些西方国家通常面临着多文化融合的问题 ,由于语言文化不同 ,不同的亚文化群体对相同测验工具的理解和反应通常不一致 ,而传统的认知能力测验很大程度上依赖于语言理解能力 ,因而常被指责带有文化歧视性 ,所以组织通常苦于寻找既不失公允 ,又能有效测查来自不同文化群体的应聘者的真实水平的测验。此外 ,随着全球经济一体化 ,不少跨国公司希望能够在当地招募雇员 ,这也涉及到测

验工具在不同文化群体中的适用性问题。

研究者希望情境判断测验能够解决上述问题,所以一直以来都有人在探讨它是否具有更低的文化歧视性以及在各亚文化群体中的效度。

在 Weekley 和 Jones^[3 4]、Motowidlo 和 Tippins^[10]的研究中,不同种族被试的情境判断测验成绩差异要小于认知测验的成绩差异。Chan 和 Schmitt 对比了录像版和纸笔版的情境判断测验,发现纸笔版测验对阅读理解能力的依赖更大,而且表面效度更低,所以黑人和白人之间的成绩差异在纸笔版测验中更大^[11]。Phillips 用情境判断测验来预测电话服务人员的绩效,发现测验对黑人的预测作用是显著的,而对西班牙人的预测作用则不显著。Motowidlo 等人考察了测验成绩和人际技能、问题解决、沟通和总体绩效的相关,发现除了对人际技能的预测效度在黑人样本中没有得到验证外,其它维度在所有样本中都得到了验证^[5]。

4 研究新趋势——对测验本质的探讨

目前对情境判断测验的研究还不够深入。实证结果对于测验的效标关联效度没有得出统一结论,研究者对情境判断测验的构想效度也存在较大分歧,归根结底还是人们对情境判断测验这一工具本身知之甚少。目前情境判断测验的形式很丰富,在项目开发、答题和计分方式、问题呈现等方面都各不相同,而在效度研究和跨文化比较研究中,多数研究都仅仅停留在分析测验成绩和其他结构得分的表面相关关系上,没有从深层挖掘测验的信息。Campbell 曾经提出,这种笼统的分析不能帮助人们区分测量结构和理论结构,因为在没有弄清楚测量工具时,面对一个不好的研究结果,人们不知道究竟是所选择的测量工具没有反映出理论构念的特点,还是所选择的理论构念不足以真实地说明问题。但是如果对测量工具足够了解,人们就能够利用测量工具更好地去探究真理。

4.1 测验形式对测验的影响

McDaniel 提出,影响测验结果的可能因素有:测验开发是否使用了工作分析、问题情境的详尽程度、项目的认知因素负荷程度,以及选择同时效标还是预测效标来验证效度。另外一些研究者也注意到了

这方面的问题,并在一定程度上对测验形式的影响作用进行了探讨。

1) 情境呈现方式的影响。情境判断测验通常有两种情境呈现方式——文字呈现和录像呈现。在传统的纸笔测验中,情境通过语言描述呈现在题本上,而在录像测验中,问题情境通过演员表演出来。有限的使用录像测验的研究结果表明,录像测验的预测效果更好。如 Jones 将录像版测验用于选拔酒店接待员、订票代理、航空服务员和客户服务代理,发现效度分别达 0.38、0.28、0.28、0.36^[12]。McHenry 和 Schmitt 在两个研究中用录像测验调查客户服务人员,得到的效度为 0.34 和 0.40^[12]。Chan 和 Smith 的研究同时应用了录像测验和纸笔测验,发现前者对被试的阅读理解能力要求更低,因而认为录像测验可以部分地排除认知能力的影响^[11]。

2) 问题提问方式也可能对测验结构产生影响。情境判断测验中,问题提出一般有两种方式:即面对系列反应,让被试进行评价或者选择。Nguyen 认为,如果让被试进行评价,那么考察的可能就是认知能力或者知识^[13]。为了考察被试能力,可让被试选择他们“会做”和“应该做”的反应。“会做”通常表明了被试真实的行为,而“应该做”表明了被试经过理性分析后得出的行为。Robins 同时使用了“会做”和“应该做”的问法,发现两种问题形式下的测验成绩相关达到 0.77,这表明两种提问方式可能导致不同的反应。Hanson 认为,理想状态下测验应该使用“会做”的问法,因为这代表了被试真实的反应,然而在实际应用中,这种问法就有可能导致结果受到污染,因为有的被试为取得好成绩或好印象,会更多暴露自己经过理性分析后的反应。Ployhart 和 Ehrhart 细致地探讨过相关问题,发现“会做”和“应该做”的成绩相关为 0.20,而且只有“会做”对学习成绩存在显著预测效度^[14]。目前该方向的研究不多,而弄清楚该问题有助于对测验结果真实性进行一定程度的控制。

3) 正确答案确定方式的影响。McDaniel 在元分析中总结到,正确答案的产生途径有三个:专家确定、通过预测验的被试反应来确定以及实证确定。Weekley 和 Jones 对比了第一种和第三种方式,发现

两种记分方式下结果之间的相关为 0.53 ,前者对绩效的预测为 0.33 ,后者则达到了 0.38^[4]。Krokos 发现若采用实证方式来确定答案 ,测验在效标关联效度上有大幅提高。

4.2 情境判断——信息加工过程

情境判断是一个复杂的过程 ,需要被试准确地理解情境和问题 ,利用情境中的信息 ,结合自己过往的经验 ,有效地处理问题。Robins 曾经提到 ,完成情境判断测验是一个认知信息加工的过程 ,“判断”仅仅是其中一步 ,所以研究不能仅仅关注最后经由“判断”得出的成绩 ,还要探讨前面的数个步骤及其影响因素。通过过程研究才能够了解被试面临一个假设的情境时是如何做出判断和选择的 ,这种反应和面临真实情境时的反应有何不同 ,从而进一步了解情境判断测验的本质。

Robins 认为 ,情景判断测验的信息加工过程包含了三个步骤 :对情境的知觉和理解、对过去类似经验的记忆提取、选择一个有效的反应 ,这三个步骤并不按序发生 ,而且互相会有影响。当时 Robins 并没有验证他的假设 ,但这一观点后来被 Krokos 采纳 ,并且进行了验证。Krokos 的信息加工过程包括 :情景知觉、记忆提取、新想法的产生、作出决策。她设计了三种场景 :针对学生的学习生活困境、针对领导者的管理困境、针对客户服务人员的工作困境。数据模型对三种情境都进行了拟合 ,但最终只有情境知觉对测验得分有较大的影响 ,而其他过程对测验得分的影响都很微弱。Krokos 的研究当然也存在缺陷 ,如测验题目太少 ,导致成绩总分变异较小 ,此外 ,在决策制定的测量指标上 ,她选择了一般在进行风险决策或经济决策时应用的策略 ,而这些策略通常不适合一般的工作场景。未来研究中 ,一定要审慎选择对过程的测量指标。

4.3 情境因素对测验的影响

有研究者从情境多样化的角度探究情境因素的影响 ,发现它对测验质量起着很大的作用。从传统的信息加工的观点来看 ,测量工具反映的情境特点属于误差或者方法偏差 ,应该从测验得分中排除。在情境判断测验中 ,如果测查同一结构的各个题目反映的情境异质性太强 ,则有可能干扰测验测查的

结构。Soonmook 等人认为 ,测验题目情境的多样化可以看作以测量工具为基础的方法的多样化 ,当用这种工具测查多种特质时 ,就可以沿用多质多法模型的思想 ,分离特质和工具的影响。他们利用结构方程模型 ,通过设定测验项目特质相关而方法不相关(CTUM 模型)来探讨情境多样性对测量结构的影响。最后发现情境的变异确实会对整个测验项目的一致性产生影响。Carl 利用录像版测验对 5426 名警官应聘者进行了研究 ,发现情境特点的影响作用是显著的。当同样测查控制力时 ,如果情境存在激发过度控制的情节 ,如攻击、违反法规、强暴妇女等 ,那么被试就会表现出更强的控制行为 ,而如果情境的激发因素不足 ,如情境体现了合作性、同理心等特点的话 ,被试就会表现出较低甚至不足的控制行为。Chan 和 Smith 认为情境判断测验的结构是多维性的 ,这导致测验的 α 系数通常较低 ,所以情境判断测验的信度不适合用内部一致性指标来评价。作者认为 ,除了测验结构的多维性对 α 系数会产生影响 ,情境多样性对 α 系数也会产生较大的影响 ,因为表面分散的问题情境可能干扰了潜在的测量结构。鉴于人格存在跨情境的一致性 ,而能力则不然 ,未来的研究可以探讨在情景判断测验中 ,情境的异质性的影响作用是否会不同。总的看来 ,目前对情境判断测验本质的研究还很少 ,但这是一个新的趋势。由于已有的研究存在不同的缺陷 ,所以相关问题的前景还不明朗 ,研究空间还很大。

参考文献

1 McDaniel M A ,Finnegan E B ,Morgeson F P et al. . Use of situational judgment tests to predict job performance : A clarification of the literature. *Journal of Applied Psychology* ,2001 ,86 (4) :730 - 740.

2 David Chan ,Neal Schmitt. Situational judgment and job performance. *Human Performance* ,2002 ,15 (3) :233 - 254.

3 Weekley J A ,Jones C. Video - based situational testing. *Personnel Psychology* ,1997 ,50 :25 - 49.

4 Weekley J A ,Jones C. Further studies of situational tests. *Personnel Psychology* ,1999 ,52 :679 - 700.

5 Motowidlo S J ,Dunnette M D ,Carter G W. An alternative selec-

tion procedure : The low – fidelity simulation. Journal of Applied Psychology ,1990 ,75(6) :640 – 647.

6 Stevens M J ,Campion M A. Staffing work teams : Development and validation of a selection test for teamwork settings. Journal of Management ,1999 ,25(2) :207 – 228.

7 Filip Lievens ,Tine Buyse ,Paul R. Sackett. The operational validity of a video – based situational judgment test for medical college admissions : Illustrating the importance of matching predictor and criterion construct domains. Journal of Applied Psychology ,2005 ,90(3) :442 – 452.

8 Filip Lievens ,Pol Coetsier. Situational tests in student selection : An examination of predictive validity ,adverse impact ,and construct validity. International Journal of Selection and Assessment , 2000 ,10(4) :245 – 257.

9 Clevenger J ,Pereira G M ,Wiechmann D et al. . Incremental validity of situational judgment tests. Journal of Applied Psychology ,2001 ,86(3) :410 – 417.

10 Motowidlo S J ,Tippins N. Further studies of the low – fidelity simulation in the form of a situational inventory. Journal of Occupational and Organizational Psychology ,1993 ,66 :337 – 344.

11 Chan D ,Schmitt N. Video – based versus paper – and – pencil method of assessment in Situational Judgment Tests : Subgroup Differences in Test Performance and Face Validity Perceptions. Journal of Applied Psychology ,1997 ,82 :143 – 159.

12 张艳秋 ,凌文铨. 情境判断测验的应用和进展. 引进与咨询 ,2003 ,4 :18 – 20.

13 McDaniel M A ,Nguyen N T. Situational judgment tests : A review of practice and constructs assessed. International Journal of Selection and Assessment ,2001 ,9 :103 – 113.

14 Robert E. Ployhart ,Mark G. Ehrhart. Be careful what you ask for : Effects of response instructions on the construct validity and reliability of situational judgment tests. International Journal of Selection and Assessment ,2000 ,11(1) :1 – 16.

The Research Trends of Situational Judgment Test

Chen Jing Che Hongsheng

(Psychology Department of Beijing Normal University ,Beijing 100875)

Abstract :With the development of situational judgment test , researchers pay more and more attention to it. This paper introduces the research approaches in situational judgment test in recent years , which can be summarized as validity research , cross – cultural comparison and the impact of the nature of the test including test characteristics , information – processing and diverse situations ’ impact. These trends are described in detail in this paper.

Key words Situational Judgment Test ; validity research ; cross – cultural comparison