

人格测验中作假的控制方法

骆 方 张厚粲

(北京师范大学 心理学院 北京 100875)

摘 要 :被试很容易对人格测验作假 ,这严重影响了人格测验的有效性。目前测评专家已经提出了一些应对作假的方法 ,它们可被分为事前控制技术和事后识别技术两大类。前者包括迫选式量表 ,警告及假渠道技术等 ,后者包括作假识别量表 ,IRT 及反应时识别技术等。目前 ,在人格测验中嵌套使用作假识别量表 ,以及在测验指导语中加入警告是比较有效的两种方法 ,迫选式量表的发展也值得期待。由于研究者对作假的内部发生机制了解较少 ,这制约了 IRT 与反应时识别技术的发展。

关键词 :作假 ;社会称许性量表 ;迫选式量表 ;警告

中图分类号 :B841.2

文献标识码 :A

文章编号 :1003-5184(2007)04-0078-05

1 前言

心理测验主要有两大类 :认知测验和人格测验。认知测验测量个体具备一定认知能力的程度 ,因此不考虑猜测的因素 ,被试只有知道正确的答案才能够得分。而人格测验却不同 ,被试能够随意提升分数 ,例如人际关系测验的一个项目 :“当别人说话的时候 ,我会仔细的倾听。”如果被试回答“同意”则说明其具有高的人际技能。但是 ,人际技能低的个体 ,也可以回答“同意”。这是认知测验和人格测验的一个重要的区别 ,即在认知测验中被试不容易作假 ,在人格测验中却可以轻易地作假。

作假 (faking)是指个体为了表现出一个有利的自我表征 ,而故意对人格测验歪曲反应的趋势。社会称许性反应是心理学家常提到的另一种夸大的反应偏差 ,依据意识水平它可被划分为两种成分 :自我欺骗(无意识)和印象管理(有意识)。印象管理与作假都是指被试故意夸大反应的趋势 ,性质十分接近 ,区别在于 :被试作假是朝向测验情境的 ,而印象管理则是朝向社会期望的。因而 ,印象管理具有跨情景的一般性 ,而作假则发生在特定的情景下。比如职业选拔情境下 ,被试依据所应聘的工作性质作假 ,试图表现出一种“优秀员工”的样子 ,这是一种“工作称许性反应” ,而不是印象管理。有时被试作假和印象管理的方向可能是相反的 ,比如人格测验的一个项目“病人在痛苦呻吟时 ,我能够保持平静” ,如果被试

进行印象管理 ,则会选择“不同意”来说明自己具有同情心。但是 ,如果被试处在应聘护士工作的选拔情景下 ,他则会倾向于回答“同意” ,来说明自己具有护士的专业素质。目前 ,研究者对职业选拔情景下的作假已有了较多的探讨。

作假会严重影响人格测验的效标效度、结构效度 ,因而该如何控制作假一直是测评专家们关注的问题。目前 ,研究者已经提出来几种应对作假的方法 ,其中有一些已经很常用 ,而有一些方法尚在初始研究阶段。Nederhof(1985)^[1]把应对社会称许性反应的方法分为事前控制和事后识别两类 ,这里参照这个分类法 ,把应对作假的方法进行同样的分类来阐述。

2 应对作假的方法

2.1 事后识别技术

所谓事后识别是指在人格测验施测之前或过程中对被试作假不进行干涉和处理 ,等被试完成人格测验后 ,研究者再采用一定的技术手段 ,通过对测验反应的分析来测量和识别作假。这类技术主要有三种方法 :作假识别量表 ,IRT 和反应时识别技术。

2.1.1 作假识别量表

自 1917 年第一个人格测验开发起始 ,测评专家们就对人们能否诚实回答测验项目产生了怀疑 ,并从 1939 年开始逐渐对作假采取了一些行动 ,其中借助外部量表来测量作假的方法得到了更多人的认

可。其基本原理是 :在人格量表中嵌套进作假识别量表 ,被试在该量表上的得分代表其作假程度。作假识别量表主要有两类 :

1)社会称许性量表(Social Desirability Scale ,SD 量表) ,它是目前使用最广泛的作假识别量表。它的项目是 (1)受称许 ,但是不经常有的行为(比如 ,我从来不会掩饰自己的错误)(2)不受称许 ,但是常有的行为(比如 ,我有时候想偷看别人的日记)。个体对不可能的美德肯定 ,或者否定人们共有的陋习 ,都计一分。因此 ,每一个人的作假分数都可以获得。

采用 SD 量表能够获得被试的作假分数 ,这不仅可以帮助人们识别出作假者 ,而且可以通过偏控制法或者减法去除人格测验分数中的作假成分 ,获得被试真实的人格分数。偏控制法是进行作假变量对人格分数的回归 ,残差可被视为被试的真实分数 ;减法则直接把作假分数从人格分数中减去。然而 ,有研究发现基于这两种方法得到的校正分数 ,并不能够代表被试人格的真实分数^[2]。Goffin 和 Christiansen(2003)^[3]认为关于校正分数的作用还有待进一步的研究。

尽管不少研究者对 SD 量表的效度提出了质疑 ,但是因为它能够直接测量作假 ,使用方便 ,所以受到了很多研究者和应用者们的青睐。

2)其他的作假识别指标。尽管 SD 量表最为常见 ,还是有一些其它形式的识别工具逐渐被开发出来。

第一种是一个关于图形的知觉任务 ,被试被要求在较大的图画中识别一个目标。该目标往往是图画中不存在的图形 ,被试如果有作假的动机 ,则会宣称自己能够看出这些不存在的图形 ,分数升高。

第二种是伪造项目量表(bogus item scales)。比如 ,Anderson 等(1984)^[4]让被试自我回答是否参与过一些工作任务时 ,加入一些不存在的工作任务(比如 ,念大学时参与校办改革等)。被试声称完成过这些不存在任务的多少 ,代表其作假分数。结果发现有高的作假分数的被试 ,会在 13 种职业中的 8 个选拔性的人格测验上有同样高的分数。

这两种识别指标的有效性得到了一些研究的支持 ,但是它们同人格测验的形式差别较大 ,表面效度较低 ,被试比较容易觉察出来 ,这将使它们的识别作用丧失。

2.1.2 IRT 识别技术

最近有一些研究使用 IRT 技术来识别人格测验中的作假 ,比如个人拟合测量(I_z)和混合 Rasch 模型(MRM)分析。这种技术一般会先建构被试作假的反应模型。

1)个人拟合或者适合度测量(I_z)

在 IRT 分析中 , I_z 技术可以识别有独特反应模式的被试 ,这些被试被认为具有作假的嫌疑。 I_z 指标是标准化的对数似然比 ,在作假研究中用来反映作假者与大部分诚实者的反应差别。如果测验足够长 , I_z 类似 Z 分数 ,大的负的 I_z 值能够说明被试进行的是作假反应。

这种技术来自认知测验领域 ,目前仅有两个研究使用 I_z 来识别人格测验中的作假。Zickar 和 Drasgow(1996)^[5] ,Ferrendo 和 Chico(2001)^[6]的研究表明 , I_z 指标能够用来识别作假者 ,但是其复杂的识别程序并不利于广泛应用。

2)混合 Rasch 模型分析

MRM 分析是一种相对较新的 IRT 技术 ,它假设所有被试并非来自同一总体 ,而是存在不同的子总体 ,他们以不同反应方式来完成测验。MRM 分析会依据被试的反应模式是否类似 ,把被试归到不同的类别中 ,每个类别有不同的 Rasch 模型的项目参数。MRM 分析用来识别被试类别的技术 ,不同于其他的分类技术(比如聚类分析 ,鉴别分析) ,它不仅仅基于均值的差异来分类 ,还整合了被试的项目反应模式。换句话说 ,两个人总分相同 ,如果反应模式不同 ,也会被分在不同的组内。目前仅有一个研究 Zickar 等^[7]使用 MRM 分析来识别人格测验中的作假和诚实人群 ,误中率和命中率都是可以接受的。

在认知测验领域使用 IRT 技术来识别作假已经得到广泛认可 ,但是人格测验中的作假有独特的性质 ,IRT 技术是否适用值得商榷。首先 ,认知测验的作假人数和被作假的项目数量较少 ,但是人格测验的作假比例非常大 , I_z 指标异常的被试很可能是诚实者 ,而不是数量可观的作假者。其次 ,MRM 分析的内隐假设——作假者的反应模式是类似的——也可能是不成立的 ,因为每个人的作假动机不同 ,他们的反应存在相当大的变异。

2.1.3 反应时识别

有一些研究者从信息加工的角度,探讨作假者的反应机制。他们认为作假会改变反应的潜伏期,依据被试反应时的差别,可以对作假者和诚实者进行区分。

然而,过去几十年的研究结果却是相互矛盾的。有的研究发现,作假者的反应时更长,而另一些人的研究结果则刚好相反——作假使得反应时更短。研究者都提出了相应的认知模型来解释自己的研究结果。比如,语义练习模型,假设作假者根据项目内容与自己要伪装的形象进行语义上的分析做出反应;诚实者则要与自我图式进行符合性判断,后者的认知加工更复杂,因此作假者的反应时要短。自我图式模型,认为作假者首先根据自我图式对项目内容作出反应,当项目内容与图式中积极的一面一致时,给出诚实的反应,没有必要作假;不一致时,个体才需要进行作假,因此反应时延长。改编图式模型,认为作假者和诚实作假者的认知加工过程是一样的。两种反应都基于对项目内容与所参照的图式之间关系如何做出评估之后进行反应,诚实作答参照的图式是自我图式,而作假者参照的图式是经过加工改编的理想回答者图式。当项目内容与参照的图式一致时,反应时较短,不一致时反应时增长。因此,对同一个被试而言,改编的图式和自我图式对反应时有同样的影响,即不管是作假还是诚实反应,其反应时都有可能长,有可能短。

反应时作为作假研究的测量指标本身也有很大的问题¹⁾难以消除被试本身和项目本身所带来的污染变量的影响,包括阅读速度、肌肉运动速度、对指导语的回忆速度等;项目的长度、顺序和词汇的难度水平等。²⁾人们作认知判断的速度是非常快的,需要有很高精度的计算机记录,才能保证反应时的测量误差较小,而在人格测验的应用情境中往往很难做到这一点。

2.2 事前控制技术

一些研究者试图在编制测量工具时或者在施测过程中增加一些控制,来阻止被试发生歪曲反应,而不是作假发生后再进行识别。这类技术主要有三种:迫选式(forced-choice format)量表,警告以及假渠道技术。

2.2.1 迫选式(forced-choice format)量表

这种人格量表迫使被试在两个称许程度相等的句子或答案间选择,来抑制作假反应发生。它提供了一种使得测验作假降到最小化的可能。这种方法始自40年代,Edwards(1954)开发了第一个标准化的人格迫选量表《爱德华个人偏好量表》(EPPS)。

迫选量表早期受到很多批判,主要是因为它的自模式数据(ipsative,量表的总分是个常数)不能进行个体内分数的比较,因为一个维度的分数升高,另一个维度的分数必然降低。此外自模式数据不能进行方差分析,因素分析等处理⁸⁾。模拟研究证明至少用于配对的人格量表超过30个时,迫选量表的分数才接近规范的数据形式(normalized,比如likert式)⁹⁾。

2.2.2 警告(warning)

有研究者发现,在人格测验的指导语中加入警告“该测验中含有可以识别作假的指标/该测验有成熟的技术可以识别作假”,会大大降低被试的夸大反应,也会降低作假人群的比例。Dwight和Donovan(2003)¹⁰⁾最近的元分析揭示,警告的呈现会降低被试作假的程度,测验分数降低0.23个标准差。也有研究发现,如果不指出“作假将会受到严厉的惩罚”,警告将不会有实质性的作用。

Ellingsor(2003)¹¹⁾认为识别出作假者后,警告并给予重测的机会,是一种很有效的控制作假的方法。但是这种警告可能会使被试在测验过程中的焦虑程度提升。要求被试重做测验,可能会使被试认为测验是不公平的,进而影响到他们完成测验的动机。因此,要特别注意评估过程中被试的态度变化,了解他们对评估的满意度,知觉到的公平性,焦虑程度,完成测验的动机等等。

2.2.3 假渠道技术(bogus pipeline)

其实就是一种假的测谎仪。研究者给被试接上电子仪器,并声称通过此设备的生理指标可直接测量他们的人格。这个仪器被形容为“直达心灵的通道”,具有测量一个人的自我洞察力的功能。然后研究者假称想研究被试的自我洞察力,于是让被试猜测仪器对他完成每一个项目反应所显示的读数。这些猜测即可作为被试对每一个项目的真实反应。用这种方法的原理是:被试因为不想被仪器查出自己没有自我洞察能力而感到难堪,所以他们所给的猜测必然比一般的自陈式测验更要真实,更少受作假

动机的影响。

许多其它的方法都是假渠道技术的翻版。例如,直截了当的告诉被试会让他们填一个测谎量表,并会在计分后通知他们得分是否显示他们曾经作假,如没有作假,才会让他们回答真正的测试题。

3 讨论

对三种事后识别技术的比较研究发现,作假识别量表比 IRT 与反应时识别技术都更加有效^[6],它是目前最常用的方法。IRT 与反应时技术都需要研究者对被试的作假反应模式有清晰的假设,然而目前研究者对作假的内部发生机制的认识还比较浅,往往直接套用认知测验中作假的研究模式,这将造成研究结果的较大偏差。因为,被试在认知测验中的作假形式比较单一,主要是指抄袭,而且作假人数较少,被作假的题目数量也较少,而被试在人格测验中的作假则要复杂的多,被试的作假动机与测验情境、个人特征等相关,作假的人数和作假的项目数量可能会很大。因而,研究者对作假的内部机制的探讨如果不能深入的话,将会制约 IRT 与反应时识别技术的发展。

作假识别量表包括两种:社会称许性(SD)量表和其他的作假识别指标。SD 量表和人格测验的项目形式是相同的,具有较高的表面效度,不容易被被试觉察出来,是比较好的测量作假的指标。但是,SD 量表是社会称许性反应的测量工具,在前面已提到社会称许性反应与作假并不相同,在一些特定的测验情境下,SD 量表识别作假的效度较差,比如,Ones 等(1996)^[12]发现从被试的人格分数上控制掉印象管理的得分后,并不能够提高人格测验对工作绩效的预测效度。因而,特定测验情境下被试作假的识别工具不能完全用 SD 量表来代替,尤其是职业选拔情景下的作假识别工具亟待重新开发。

Donovan 等(2003)^[13]调查了职业选拔情景下被试作假的基本发生率,结果显示将近一半的被试承认夸大了一部分人格特征,这个比例是相当高的,将使人格测验无效。可见,在作假发生之前尽量多的使用一些事前控制策略是很有必要的。在三种事前控制技术中,迫选量表的相关研究较为丰富,研究者针对它的自模式的数据形式进行过很多探讨,最近 Stark 等(2004)^[14]试图使用 IRT 的 GGUM 模型解决

这一问题,虽然结果并不完善,但是其进一步的研究成果值得期待。另外,在指导语中加入警告也是一种较好的方法,尽管它的有效性还处在争论中,但是使用起来非常方便,负作用较小,因而在高利害测验情景下,最好在人格测验中加入严厉的警告。

近年来,人格测验已经逐渐在我国临床、教育、工业与组织中得到了应用,但是由于被试作假的存在,人们对它的有效性仍然持怀疑态度。作假研究是个很老的研究课题,几乎是伴随着第一个人格测验的降生而出现的。很多优秀的测量专家都探讨过作假的问题,然而由于作假的异常复杂性,近二十多年来,并没有长足的进步。除了测量专家,如果能有更多的认知心理学家和人格心理学家加入作假的研究行列,对作假的内部机制的研究将会更加深入,作假的识别技术将会有全新的发展。

参考文献

- 1 Nederhof A J. Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 1985, 15: 263 - 280.
- 2 Ellingson J E, Sackett P R, Hough L M. Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 1999, 84: 155 - 166.
- 3 Goffin R D, Christiansen N D. Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment* 2003, 11(1): 340 - 344.
- 4 Anderson C D, Warner J L, Spencer C C. Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology*, 1984, 69: 574 - 580.
- 5 Zickar M J, Drasgow F. Detecting faking using appropriateness measurement. *Applied Psychological Measurement*, 1996, 20: 71 - 87.
- 6 Ferrendo P J, Chico E. Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales. *Educational and Psychological Measurement*, 2001, 61: 997 - 1012.
- 7 Zickar M J, Gibby R E, Robie C. Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model Item Response Theory. *Organizational Research Methods* 2004, 7(2): 168 - 190.

- 8 Bowen C ,Martin B A ,Hunt S T. A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *International Journal of Organizational Analysis* , 2002 ,10(3) :240 – 260.
- 9 Baron H. Strengths and limitations of ipsative measurement. *Journal of Occupational & Organizational Psychology* ,1996 ,69(1) : 49 – 57.
- 10 Dwight S A ,Donovan J J. Do warnings not to fake reduce faking ? *Human Performance* 2003 ,16(1) :1 – 23.
- 11 Ellingson J E. The viability of retesting and warnings for managing faking tendencies. In :E. D. Heggstad(Chair) ,Ed. *Practical considerations for implementing personality testing in organizations*. Symposium conducted at the annual meeting of the Society for Industrial and Organizational Psychology. Orlando , FL 2003.
- 12 Ones D S ,Viswesvaran C ,Reiss A. Role of social desirability in personality testing for personnel selection : The red herding. *Journal of Applied Psychology* ,1996 ,81 :660 – 679.
- 13 Donovan J J ,Dwight S A ,Hurtz G M. An assessment of the prevalence ,severity ,and verifiability of entry – level applicant faking using the randomized response technique. *Human Performance* 2003 ,16(1) :81 – 106.
- 14 Stark S ,Chernyshenko O S ,Drasgow F. An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions :The multi – unidimensional pairwise – preference model. *Applied Psychological Measurement* 2005 ,29(3) :184 – 203.

Methods of Coping with Faking of Personality Tests

Luo Fang Zhang Houcan

(Psychology College ,Beijing Normal University ,Beijing 100875)

Abstract :Personality tests are easy to be faked and faking have adversely affected the validity of personality tests. Psychometrical specialists have brought forward some methods to coping with faking. Two main categories are distinguished ,which are the pro – preventing and post – detecting techniques. The former include forced – choice format scale ,warning ,bogus pipeline. The latter include detecting faking scale ,IRT and reaction – time detecting techniques. Now detecting faking scale nested in personality test and warning added in the test direction are the more effective methods and the developing of forced – choice format scale will be good. Because researchers less know the inner generating mechanism of faking ,it restricts the developing of IRT and reaction – time detecting techniques.

Key words :faking ;Social Desirability Scale ;forced – choice format scale ;warning