心理学报 2010, Vol. 42, No.7, 791−801
Acta Psychologica Sinica DOI: 10.3724/SP.J.1041.2010.00791

应聘情境下作假识别量表的开发*

骆 方 刘红云 张 月

(北京师范大学心理学院; 应用实验心理北京市重点实验室, 北京 100875)

摘 要 在应聘情境中, 被试容易对人格测验作假。应对作假的常用方法是采用社会称许性量表对作假直接测量, 再去校正和识别作假效应。但是采用社会称许性量表测量作假存在很多问题, 因而基于作假的特殊性质开发了《作假识别量表》。采用探索性因素分析证实了量表的单维性, 解释率为 54.650%。概化理论检验表明测验信度较好, G 系数为 0.906, φ系数为 0.902。采用一个真实的应聘情境检验效度, 发现《作假识别量表》对作假更加敏感, 能够比较充分地测量作假。

关键词 作假; 社会称许性反应; 社会称许性量表; 工作称许性

分类号 B841

1 问题提出

近年来, 越来越多的企业使用人格测验进行人才选拔。与能力测验不同, 人格测验多采用自评方式, 由被试对自己的日常行为进行判断, 所以主试很难觉察被试回答的真实性。在工业与组织心理学(I/O)领域, 测验学家把被试在应聘过程中故意对人格测验进行夸大反应的趋势叫做作假(Faking)。由于作假的发生频率较高, 至今人们对人格测验在职业应聘中的有效性仍然持怀疑态度(Morgeson et al., 2007)。

研究者应对作假最常用的方法是在人格测验中嵌入社会称许性(Social Desirability, SD)量表, 直接对被试作假的程度进行测量, 然后再采取校正或者识别技术去除被试的人格测验得分中的作假效应。

SD 量表是社会称许性反应(Socially Desirable Responding, SDR)的测量工具, SDR 是指个体朝向社会期望的方向作答的反应趋势。1984 年 Paulhus 正式提出 SDR 的双成分模型: 自我欺骗(Self-Deception)和印象管理(Impression Management)。自我欺骗发生在无意识中, 它与心理适应、乐观、自尊及一般的胜任感等人格特质有较高的相关, 因而不能将其从人格测验的分数中排除。印象管理与作假性质相同, 都是被试的故意夸大行为, 区别在于: 印象管理发生在一般情景中, 被试是朝向社会期望的; 而作假发生在职业应聘情境中, 被试按照工作期望的标准来回答人格测验, 倾向于向主考方传递一种“优秀员工”的形象, 这是一种“工作称许性反应”, 而不是印象管理(Bradley & Hauenstein, 2006; Martin, Bowen, & Hunt, 2002; Mueller-Hanson, Heggestad, & Thornton III, 2006)。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, 如果被试正在应聘护士的职位, 他则会倾向于回答“同意”, 来说明自己具有护士的专业素质。

早期的 SD 量表(比如, Edwards SD 量表、Marlowe-Crowne SD 量表)是单一结构的, 直到 1991 年 Paulhus 开发了社会称许性均衡量表(Balance Inventory of Desirable Responding, BIDR), 它包含两个分量表分别测量自我欺骗和印象管理, 前者的内容都是有关性和侵犯方面的, 被试无意识中就会回避; 而后者的项目描述的都是极端夸大的行为(比如, 我做事情从来没有拖延过), 如果被试声称自己具有, 则一定是有意识的粉饰。BIDR 的测量框架得到了广泛的好评, 但是实际测量效果却并不理想。比如,

收稿日期: 2009−05−21

* 国家自然科学基金项目(30870784)资助。

通讯作者: 刘红云, E-mail: hyliu@bnu.edu.cn

Rosse, Stecher, Miller 和 Levin (1998) 发现印象管理与人格分数的相关达到 0.30 以上。Paulhus (2002) 承认“BIDR 的两个因素的内部相关超过 0.50, 多少遮住了两因素的光芒。” 很显然, BIDR 的两个分量表并不独立, 这违背了 Paulhus 编制量表的最初构想。

在职业应聘情境中印象管理不会发生, 它被作假取代, 但是无意识中产生的自我欺骗却会出现在任何测验情境中。由于印象管理和作假性质相似, 很多研究者主张在职业应聘情境中使用 BIDR 的印象管理分量表来测量作假, 本文认为这并不合适。首先, BIDR 的印象管理和自我欺骗量表的区分效度不高, 印象管理测量分数中含有自我欺骗的成分。其次, 印象管理量表的工作期望色彩不浓, 测量作假时敏感性会降低。因而, 基于作假的性质并吸取 SD 量表的编制经验来开发作假的测量工具十分必要。Goffin 和 Christiansen (2003) 的研究表明, 70%的人格测验的实际使用者希望使用作假识别量表。本研究的目的是编制在大多数工作应聘情境中都适用的《作假识别量表》, 并检验其有效性。

2 《作假识别量表》开发设计

SD 量表项目描述的情景有两类: 一类是受社会称许, 但是不经常有的行为, 如我从来不会迟到; 第二类是不受社会称许, 但是常有的行为, 如别人放在桌上的东西, 有时不经允许我就翻看了。这些行为是人类不可能具有的美德(或不可能去掉的陋习), 如果被试承认(或否定)自己具有这类行为, 则表明他(她)在夸大反应。这类项目不包含人格特质信息, 只能测量反应偏差。作假与 SDR 都是一种夸大的反应偏差, 因而可采用相似的项目来测量。

但是作假有自己独特的性质, 而且 SD 量表存在着结构效度不佳的问题, 因而为了提高作假识别的有效性, 本文将在编制过程中增加特殊的控制程序。首先, 与 SD 量表强调“受社会称许”不同, 这里强调“受工作称许”。虽然各行各业都有自己的独特性, 但是企业最看重的员工行为却存在着一般性, 这些行为对大多数工作来说具有较强的称许性。其次, 在职业应聘情境中, 被试无意识中会产生自我欺骗, 它与作假造成的结果相同——夸大了反应, 因而很容易与作假混淆测量。本文将通过实验操作挑选出对自我欺骗不敏感, 但是对作假有较大反应的项目。这些项目能够充分测量作假, 而且不受自我欺骗的污染。

具体的编制步骤如下:

1) 广泛搜集行为描述, 满足 1)受工作称许, 但是不经常有的行为(正向), 或 2)不受工作称许, 但是常有的行为(负向), 并改编成量表项目。

2) 让评鉴者判断项目的称许性和普遍性。挑选具有较强的工作称许性, 并且有极端发生率的项目, 它们很少包含人格特质信息, 能够测量个体的夸大反应的偏差。

3) 采用被试内实验设计, 挑选出充分测量作假, 但又很少受自我欺骗污染的项目。

4) 采用因素分析考察量表的结构效度。

5) 采用多元概化理论确定构成量表的项目数量。

6) 采用 Logistic 回归检验量表的有效性, 并给量表划定分界线。

7) 检验作假识别量表的有效性。

3 《作假识别量表》的编制过程与结果

3.1 搜集行为描述

本研究于 2006 年 7 月通过滚雪球的抽样方式*, 在因特网上对 80 名不同职业领域的员工(工作年限不低于 2 年)进行了开放式调查, 平均工作年限为 4.8 年, 男女各半。问卷包括两个部分: a. 负向行为的部分: 请举出十件“工作中不该做但大多数人都难免会做的事”, 例如: 我干活的时候常想发懒。b.正向行为的部分: 请举出十件“工作中应该做但大多数人都做不到的事”。例如: 我上班总是很准时的。两部分内容的回答有些重叠, 因此把相似的描述都归到一起。最终形成 114 个项目的《作假识别量表》的最初版, 其中正向项目 61 个, 负向项目 53 个。

3.2 依据项目的称许性和普遍性筛选项目

挑选 60 名评鉴者(评鉴者来自一个在职研究生班, 平均工作 5.6 年。男性 12 人, 女性 18 人, 平均年龄为 32.3 岁), 分别对项目进行“称许度”和“普遍度”的评量, 并据此筛选项目。

1) 称许度: 评定项目的工作称许性, 即项目所描述的行为受工作称许的程度。

* 滚雪球抽样(Snowball sampling): 是指先随机选择一些被访者并对其实施访问, 再请他们提供另外一些属于所研究目标总体的调查对象, 根据所形成的线索选择此后的调查对象。

请被试假想自己是一名对大多数工作来说都很优秀的职员, 判断“下列句子所描述的行为, 你是否会做”? 在 Likert 5 点量尺上评定, “基本上不会”“可能不会”, “拿不准”, “可能会”, “基本上会”。正向项目的值越高, 表明越受称许; 负向项目的值越低, 表明越不受称许。统计各个项目的平均得分, 选择排名在前 2/3 的正向项目, 和排名在后 2/3 的负向项目, 它们的称许性很强。

2) 普遍度: 评定有多少职员会表现出该行为。

把项目改成第三人称, 再请被试根据自己的经验, 回答“您觉得对大多数工作的员工来说, 有多少人会做该句子描述的行为”。在 Likert 5 点量尺上评定, “很少人”, “比较少”, “一半”, “比较多”, “绝大部分”。正向项目的值越低, 表明该行为越少见; 负向项目的值越高, 表明该行为越常见。因而, 选择平均得分排名在后 2/3 的正向项目, 和排名在前 2/3 的负向项目, 前者是很少发生的正向行为, 后者是经常发生的负向行为。

最终有 50 个项目满足上面的筛选条件而入选, 构成《作假识别量表》初版, 其中正向项目 25 个, 负向项目 25 个。这些项目描述的都是工作称许性很明显, 却是很少见的正向或者常见的负向行为, 它们能够测量被试的反应偏差。

3.3 依据模拟实验研究结果挑选项目

被试: 某师范大学三年级 679 名学生, 分属四个公共课班级, 包含十多个文理专业。平均年龄 22.36 岁, 男生 201 人, 女生 478 人。

测试材料:《作假识别量表》初版的 50 个项目, 以及 20 个人格项目(用来平衡测验内容, 不参加随后的分析, 选自《大五人格测验》), 命名为《性格测验》。采用 Likert 5 点评定方式, 尺度为“很不符合、不太符合、中间状态、比较符合、很符合”。

施测: 在 2006 年 11 月开展了被试内设计的实验研究。每名被试两次接受人格测试, 前后间隔两周完成。一半学生第一次是在应聘条件下, 诱使被试作假反应; 第二次是在诚实条件下, 让被试尽可能诚实回答测验项目。另一半学生施测顺序相反, 先在诚实条件下完成, 后在应聘条件下完成。计算每名被试在《作假识别量表》上的得分(50 个项目的均分), 分别检验在诚实条件下和应聘条件下这两半学生的分数是否存在差异, 独立样本 t 检验的结果显示, 不存在显著的顺序效应(t=0.63, t=1.80, df=677)。

诚实条件下的指导语, 如下:

“《性格测验》是由若干描述性格特征的句子构成, 请你判断自己是否符合, 并在相应的数字上打勾。测试结束后, 你可获得测试结果, 据此你可以更加了解自己。我们对你的测试结果严格保密, 请真实回答!”

应聘条件下的指导语, 如下:

“假设这里是 2006 年校园招聘的现场, 现在将要给你施测《性格测验》。员工的性格特征直接影响工作绩效, 本测试将为一批企事业单位选拔出优秀的毕业生。现在将要对你进行性格测试来判断你是否优秀, 这是一个重要的选拔程序。该性格测试由若干描述性格特征的句子构成, 请你判断自己是否符合, 并在相应的数字上打勾。注意: 你对测试的回答直接决定着你能否被雇佣, 主考方很难判断真伪。请你把自己当作一名‘对这份工作特别期待’的求职者来回答, 揣摩求职者的心理, 而不是按照你的真实情况回答, 要想办法使主考方录用你。如果你的测验成绩足够优秀, 排名在前 5%, 你将在本次模拟招聘中胜出, 我们两周后会公布录用名单, 给你 20 元的奖励。如果你没有被录用, 我们也将对你的表现做出评定, 看看你能给主考方留下什么样的印象。测验中加入了一些探测说谎的题目, 请务必小心, 不要让主考方察觉出来, 否则你将被取消录用资格。”

结果分析: 每名被试都在两种条件下完成测验项目, 一种被要求“作假反应”, 另一种被要求“诚实反应”。《作假识别量表》的项目具有很强的称许性, 不含人格特质信息, 即使被要求“诚实回答”, 被试也容易在无意识中夸大反应, 所以被试在诚实条件下的项目得分代表自我欺骗。被试在应聘条件下的反应为被试故意作假行为, 但不可避免地自我欺骗也会发生。为了不使自我欺骗和作假混淆测量, 本研究拟挑选出对作假能够充分测量, 而且很少受自我欺骗污染的项目。为了满足该要求, 这里采用两个标准来挑选项目: 1)对每个项目求两种条件下测试分数的相关, 挑选相关系数低的项目, 这表明它们能够独立测量出作假和自我欺骗; 2)对每个项目求两种条件下平均数差异的效应值(effect size)*,

* 相关样本的科恩 d 值(Gravetter & Wallnau, 2006): $d = \frac{M_D}{s}$

其中, M_D 为两次测量分数的差值的平均数, s 为两次测量分数的差值的标准差, d 为效应大小指标。

挑选效应值大的项目，在这些项目上自我欺骗和作假的反应强度有较大的差别。

首先计算项目的各项统计指标(结果见表 1)，再逐步选择符合上述标准的项目。

表 1 《作假识别量表》初版的 50 个项目的各项统计值

项 目	选中	相关	诚实反应		作假反应		效应值
			平均数	标准差	平均数	标准差	
别人问我话时，我都能够对答如流。		0.25*	3.21	0.98	3.76	0.85	0.52
别人有困难时，我会毫不犹豫的尽力帮助他们。		0.22	3.58	0.97	4.15	0.79	0.56
出了差错时，我总是主动承担大部分的责任。		0.25**	3.64	1.02	4.21	0.76	0.55
发生争执时，我一定先反省自己。		0.53**	3.75	1.15	4.19	0.93	0.40
即使是公家的东西，我使用时也处处节俭。		0.29**	3.76	0.93	4.43	0.80	0.61
考虑事情时，我总是从集体利益出发。		0.17	3.46	0.72	4.21	0.85	0.60
情绪不好丝毫不会影响我的工作。	*	0.22	2.36	0.95	3.79	1.16	1.05
失败之后我总是立刻爬起来重新出发。		0.27**	3.55	0.97	4.15	0.96	0.54
为了集体目标，我会毫不犹豫的牺牲个人利益。		0.46**	3.02	0.87	3.77	0.77	0.75
我从不为自己的失误找借口。	*	0.23*	2.84	1.04	3.80	0.83	0.82
我从不议论别人的私事。		0.25*	2.97	1.13	3.54	1.10	0.46
我从未拿过不属于自己的东西。		0.45**	2.99	1.22	3.63	1.13	0.48
我对周边的人都是一视同仁的。		0.26**	3.33	0.97	3.85	0.88	0.50
我办事从来都是客观公正的。	*	0.39**	3.19	0.99	3.92	0.85	0.66
我几乎没有和家里人吵过嘴。	*	0.34**	2.59	1.11	3.47	0.96	0.70
我每天都记录自己的心得，反思提高。	*	0.28**	2.76	1.16	4.14	0.91	1.04
我一定不会把当天的事情拖到明天去做的。	*	0.29**	3.17	0.98	4.10	1.08	0.74
我与各式各样的人都能合作愉快。	*	0.24*	3.18	1.08	3.92	0.93	0.62
我在任何时候都能保持沉着、冷静。	*	0.07	3.14	0.93	4.21	0.80	0.97
我总是说到做到，从不食言。		0.25**	3.40	1.11	4.11	0.90	0.60
我做的每件事情，注意力都是非常集中的。	*	0.21	3.35	1.03	4.21	0.96	0.71
我做事没有半途而废的，即使遇到困难也不会放弃。	*	0.34**	3.35	1.02	4.26	0.83	0.79
一旦作出决定，我从不后悔。		0.42**	3.27	1.10	4.05	0.68	0.70
在弄清事实之前，我从不对任何人下结论。		0.29**	3.96	1.16	4.45	0.84	0.44
重要的事情没有做完，我是绝对不会做其他事情的。		0.18	3.61	1.10	4.21	0.94	0.51
即使对我的当面批评很正确，我也会觉得难堪。	*	0.38**	3.15	0.69	2.28	1.22	−0.72
别人放在桌上的东西，有时我不经允许就翻看了。		0.32**	1.95	1.07	1.79	1.06	−0.21
长时间地做同一件事情会令我烦躁。	*	0.37**	2.78	1.13	2.23	0.99	−0.50
存在利益冲突时，我也会有个别不利于对方的言行。	*	0.29**	2.98	0.97	2.36	1.01	−0.54
对某个人有意见，我往往不会和其开诚布公的谈。	*	0.28**	2.94	1.42	2.00	1.23	−0.60
对我而言，改掉一个坏习惯有些困难。		0.24*	2.45	1.08	1.90	1.00	−0.47
对一些不懂的事，我有时会装作很懂的样子。		0.5**	2.63	1.01	1.98	0.97	−0.56
跟生人打交道时，我起初会有些不自然。		0.41**	3.34	1.27	2.50	1.35	−0.55
见到显要人物，有时我会做出恭敬讨好的样子。		0.59**	3.00	1.11	2.44	1.23	−0.44
开会的时候，我会在下面搞些小动作。		0.36**	2.56	1.33	1.89	1.15	−0.48
看到别人在公共场合的不良行为，我一般不会上前制止。	*	0.25*	2.87	1.12	2.15	1.04	−0.57
朋友们的一些缺点，我没有诚恳的指出来。		0.32**	2.69	1.04	2.23	1.14	−0.40
如果好朋友超过了我，我会感到不舒服。		0.64**	2.95	1.09	2.35	1.05	−0.50
事情出了差错，我希望自己少承担一些责任。		0.47**	2.84	1.04	2.04	0.89	−0.68
适当的时候，我会占些小便宜。	*	0.28**	2.81	1.20	1.73	0.99	−0.79
为了达到自己的目的，我偶尔也会做几件损人利己的事情。		0.55**	2.16	1.36	1.50	1.09	−0.48
我同情有些人的遭遇，却没有给予真正的帮助。		0.36**	2.79	1.16	2.24	1.14	−0.44
我想要做一些事情，就是迟迟不肯动手去做。	*	0.39**	3.00	1.44	1.96	1.29	−0.64
我有时会找借口让自己少做些事。	*	0.19	2.65	1.17	1.64	0.98	−0.76
我有时生气并没有充分的理由。		0.41**	2.81	1.25	1.98	0.99	−0.62
我有时也会假公济私，为自己人做些事。	*	0.19	2.94	1.04	1.78	0.90	−0.94
我曾假装生病请假不去上班或上课。		0.68**	2.65	1.36	1.83	1.10	−0.57
我曾经在背后取笑过自己的朋友。		0.46**	2.29	1.08	1.70	0.75	−0.55
我做事情有时会虎头蛇尾。	*	0.23	2.75	1.21	1.72	0.90	−0.78
有时我会打肿脸充胖子。		0.38**	3.11	1.05	2.68	1.05	−0.39

注：**p<0.01, *p<0.05

依据表 1 中的数据结果, 剔除掉两种条件下分数的相关大于 0.40 的项目共 14 个, 这些项目测量的作假分数中含有较高的自我欺骗成分。然后在余下的 36 个项目中, 在正向和反向项目里各取效应值的绝对值最大的 10 个项目, 共 20 个项目(表 1 中带*的项目), 被试在这些项目上的作假反应比自我欺骗的强度高很多。据此可见, 挑选出来的这些项目对作假的发生比较敏感, 作假一旦发生, 项目分数就会有较大程度的提高, 而且分数中较少含有自我欺骗的成分。如果被试在这些项目上的得分较高, 表明被试有作假的嫌疑。

3.4　因素分析检验量表的结构效度

传统的 SD 量表包含“印象管理”与“自我欺骗”两种成分, 往往不能有效区分。本文的目的是编制“作假识别量表”, 但使用 SD 量表的编制程序, 只是增加了一些控制措施。一是选取受“工作称许性”的项目, 它们在职业应聘情境下对被试有更强的刺激, 可以增加测量作假的敏感性; 二是通过实验操作来选取对“自我欺骗”不敏感, 但对“作假”很敏感的项目, 这样编制出来量表可以独立测量作假, 而很少测量到“自我欺骗”, 这是与传统的 SD 量表最大的区别。因而, 本文预期《作假识别量表》具有良好的单维性。

对被试在应聘条件下的项目得分进行主成分分析。结果显示, KMO=0.90, 只抽取出一个因素, 特征值为 10.93, 解释率 54.65%。所有项目的因素载荷都比较高(>0.40), 这说明 20 个项目的区分度较好, 同时说明《作假识别量表》的结构效度较好。

3.5　《作假识别量表》的信度检验及项目个数的确定

Asendorpf 和 Ostendorf (1998)认为选择 20 个不涉及人格内容的项目就可以有效测量作假动机了。但是当人格测验的项目数量较少时, 20 个作假项目的加入会喧宾夺主。这里拟采用多元概化理论的 D 研究, 考察需要多少个项目能够充分测量作假动机。

概化理论是现代测量理论之一, 它是在经典测量理论的基础上, 引入实验设计和方差分析技术, 对测验情景中的各类误差进行分解和控制的一种测量理论。概化理论有两个研究阶段：G 研究和 D 研究(杨志明, 张雷, 2003)。G 研究是通过估计测量目标(即潜在特质)和测量侧面(即影响测量精度的因素)的方差协方差分量, 来确认测量目标和测量侧面的关系。D 研究则是通过对信度变化的评估来确定哪种决策更好, 从而为测验的改进提供依据。这里把正向和负向题视为两个因素来考察, 采用多元概化理论的 D 研究, 考察项目的数量不同时, 测验的信度系数(G 系数和 φ 系数)会发生怎样的变化。如果项目数量较小, 信度系数并没有明显的降低, 那么选择该数量的项目构成量表, 就是一个较好的测量决策。

本文采用 mGENOVA 2.1 程序, 首先进行 G 研究, 分别估计 10 个正向题和 10 个负向题以及整个测验的 G 系数和φ 系数。结果显示(见表 2), 正向题和负向题的信度较高, 整个测验的 G 系数为 0.91, φ 系数为 0.90, 是非常高的。然后, 变化正向题和负向题的数量(n1 和 n2)从 5 到 20, 考察信度系数的变化。当项目的数量为 10(正向和负向各 5 个)时, 正向和负向题的 G 系数和φ系数在 0.67~0.75 之间, 负向题的信度要低一些, 但是整个测验的 G 系数和φ系数都比较好, 为 0.82。而且随着项目数量的增加, 测验信度在逐渐增大, 但是当项目数量大于 20 时, 测验信度的增加并不明显。说明最佳的选择是 20 个项目(10 个正向, 10 个负向), 但是只用 10 个项目也可以达到较好的测量效果。所以, 从这 20 个项目中任意挑选 10 个及以上项目, 都可以构成一个信度较好的《作假识别量表》。

表 2　G 研究和 D 研究的结果

	正向项目的个数	负向项目的个数		正向	负向	总体
G 研究	10	10	G 系数	0.85	0.81	0.91
			Φ 系数	0.85	0.80	0.90
D 研究	5	5	G 系数	0.75	0.68	0.82
			Φ 系数	0.75	0.67	0.82
	7	7	G 系数	0.81	0.75	0.87
			Φ 系数	0.80	0.73	0.86
	15	15	G 系数	0.90	0.86	0.93
			Φ 系数	0.89	0.86	0.93
	20	20	G 系数	0.92	0.89	0.95
			Φ 系数	0.86	0.88	0.95

本文没有计算重测信度，因为重测信度考察的是同一测验在相同的条件下测量稳定特质的一致性，而作假是由于被试受应聘情景的刺激而产生的，属于暂时的被试变量，因而不适宜计算重测信度。

3.6 分界线的划定

从679名被试中随机抽一半被试取其诚实反应，再取另一半被试的作假反应合在一起形成一个独立样本，进行 Logistic 回归。自变量是个体在《作假识别量表》上的得分(20 个作假项目的均值)，因变量是值为 0 和 1 的识别变量(0 为诚实反应，1 为作假反应)，来考察个体在作假识别量表上的反应能否成功的识别作假者。结果显示，自变量的回归系数是 2.60，标准误为 0.34，在 0.01 的水平上显著。而且方程的解释率 Cox & Snell R Square 和 Nagelkerke R Square 分别为 0.39 和 0.47。Logistic 回归需要对自变量进行对数转换，方程的解释率比一般的线性回归要低一些，因此这里的结果比较理想，说明《作假识别量表》能够较好的识别作假反应。

研究者通常使用误中率和误判率来表示分类是否准确。组织通常最重视的是误中率，即选拔出的人群中作假者所占的比例，他们并不关注有多少诚实者被错误的判定为作假者(误判)从而失去了被录用的资格。Logistic 回归可以估计个体的作假概率，如果大于临界概率，个体将被判为作假者，不再被组织录用。下面考察三种临界概率下，误中率和误判率的大小，以及相应的《作假识别量表》的得分(即分界线)。结果显示，当临界概率为 0.3 时，分界线较低(3.25)，个体容易被判定为作假者，因而被录用的人群中作假者数量降低，误中率很小(4.5%)，但是 21.8%的误判率较高，可能把一些高水平的诚实者挡在了门外。临界概率为 0.5 和 0.7 时，分界线升高，个体更容易被判定为诚实者，误中率增大，但是被误判的诚实者减少(见表 3)。可见，分界线低时，误中率低，但误判率高；分界线高时，误判率低，但误中率高。因而，误中率和误判率不能同时减小，该怎样划定分界线是一个两难的决策。

本研究在应聘条件下的测验结束后询问被试：你对大约多少的项目作了夸大反应？有五个反应项供选择(10%、20%、30%、40%、50%)。结果显示，无论在哪种临界概率下，被错误的判为诚实反应的被试基本上都是作假比例为 10%的轻微作假者。比如，0.5 的临界概率下被误中的 85 人中，有 69 人都是这种轻微作假者。在.7 的临界概率下误中的 151 人中，有 121 人是轻微作假的。这说明，当选择较高的临界概率(分界线为 4)时，虽然误中率较高，但在误中的人群中大多数是轻微作假者，严重作假者(约 20%)是能够较好识别的。而选择较低的临界概率(分界线为 3.65)时，更多的人(约 60%)被判为作假者而丧失了继续参与竞争的机会，这么大比例的被试如果被全部筛掉，使用人格测验意义就不大了，将起不到选拔优秀个体的作用。人格测验往往用作初筛工具，由于它本身并不精确，企业只用它筛选掉不合格的被试，然后再采用评价中心、面试等技术进行优中选优的选拔。因而，本研究建议采用较高的分界线 (4 分)，把严重作假者剔除掉，给轻微作假者进入下一轮选拔的机会。

表 3 不同临界概率下的作假识别表

临界概率	误中率	误判率	分界线
0.3	4.5%	21.8%	3.25
0.5	12.4%	13.6%	3.65
0.7	22.3%	4.1%	4.00

3.7 《作假识别量表》的效度研究

3.7.1 选拔测试

背景：某国际知名外企拟招聘“管理培训生”若干名，他们经过两年的集中培训和轮岗后，将在公司担任领导工作。该外企于 2006 年 11 月在北京、上海、杭州、武汉、南京五个城市的名牌大学里进行校园宣讲，吸引了近万名毕业生。

被试：该外企的人力资源部从近万份的简历中筛选出了 800 名学生进入第一轮的笔试。2006 年 12 月 18~19 日，该公司在五个城市举行了大规模的笔试，实际参加测试的人数为 568 人。被试的年龄在 19~27 岁之间，平均 23 岁。男生 205 人，女生 363 人。本科生 264 人，硕士生 304 人。

工具：测试工具之一为《管理者人格测验》，选自《中国青年人格问卷》*的四个维度(社会性、责任感、支配性和自我控制)，共 30 个项目。2006 年初对《管理者人格测验》的结构效度进行预研究，被试来自某测评网站共 897 名，每名被试完成测验后都会获得详细的测评报告，因而被试的态度积极认

* 《中国青年人格问卷》由北京师范大学心理学院，根据美国心理学家高夫所编制的“加州青年性格问卷”修订而成，共 330 道题。

真，回答真实可信。采用验证性因素分析检验结构效度，整体拟合指数 χ^2=588.32, *df*=399, χ^2/df=1.47, GFI=0.95, TLI=0.92, CFI=0.93, RMSEA=0.04。每个项目的因素载荷都在 0.4 以上，因素之间的相关在 0.12~0.61 之间。

从《作假识别量表》中挑选与管理行为有密切关系的 10 个项目，嵌套在《管理者人格测验》中一起施测。一共用时 15 分钟，施测完后请被试留下 e-mail。

3.7.2 后继研究

流程：“管理培训生”招聘工作在 2007 年 1 月全部结束。本研究在 2007 年 9 月给 568 人发 e-mail 以修订测验征求志愿者的名义邀请他们参加测试，目的是调查他们在诚实条件下的反应，测查个体的真实水平。间隔一周后发出第二次邀请，两周后给个体提供了详细的测评报告。

被试：在两周内共收回 234 份测验结果，虽然是匿名的调查，但是通过对 Email、高校信息、性别及年龄等的匹配，可以把被试的应聘分数和诚实分数对应起来。其中，男生 147 人，女生 187 人。本科生 154 人，硕士生 180 人。年龄在 20~27 岁之间，平均年龄是 23 岁。经过卡方检验，性别和学历与总体的构成没有显著性区别。经独立样本 *t* 检验，年龄也没有显著差异。

3.7.3 结果分析

1) 《作假识别量表》对作假识别的敏感性

被试在应聘与诚实条件下都完成了《管理者人格测验》和《作假识别量表》，下面分别对两种条件下的测验分数进行配对 *t* 检验，并求二者的相关。

结果显示(表 4)，《作假识别量表》的总分在两种条件下的相关比较低(*r*=0.36)，小于人格测验的相关(0.39~0.75)。相关系数的差异性检验显示，社交性和支配性维度的相关显著高于《作假识别量表》，*Z* 值分别为 2.98 和 6.47。这说明在应聘条件下《作假识别量表》更少测量到被试的真实反应。

《作假识别量表》总分的效应值为 1.16，比人格测验各维度的效应值(最大为 0.78)都大，说明作假发生时，个体在《作假识别量表》上有更大的反应。因而，《作假识别量表》对作假更加敏感，能够比较充分的测量作假。

表 4　不同施测条件下测验分数比较

	施测条件	平均数	标准差	效应值	*t*	*df*	*p*	相关
社交性	应聘	4.33	0.66	0.62	8.64**	233	0	0.58**
	诚实	3.88	0.80					
责任感	应聘	4.04	0.53	0.71	8.19**	233	0	0.40**
	诚实	3.65	0.58					
支配性	应聘	3.37	0.85	–0.06	–0.98	233	0.33	0.75**
	诚实	3.42	0.85					
自我控制	应聘	4.37	0.47	0.77	8.35**	233	0	0.39**
	诚实	3.93	0.70					
《作假识别量表》	应聘	4.19	0.51	1.16	11.89**	233	0	0.36**
	诚实	3.57	0.55					

2)《作假识别量表》对作假测量的有效性

个体作假的目的是要挤进测验成绩排名的高端，而在低端作假者的比例会明显减少，作假强度也会明显减弱(Marcus, 2006)。因而，高分端人群中作假者数量多、强度大，低分端人群中作假者数量少、强度低，如果《作假识别量表》有效，则两组人群在该量表上的得分应该有显著差异。

按照被试在人格测验上的应聘分数的高低把个体分为高低组(各 117 人)，然后对两组在《作假识别量表》上的得分进行独立样本 *t* 检验。结果显示，高分组(4.45)比低分组(3.92)的得分高(*t*=7.01, *df*=232, *p*<0.01)，说明《作假识别量表》能够较好的测量作假。

3)《作假识别量表》有助于录用决策

由于人格测验的效标效度并不理想，企业往往只把人格测验用作初筛工具，用它筛选掉不合格的被试，然后再对入选者采用评价中心、面试等技术进行优中选优的选拔。这里将重点讨论使用《作假识别量表》是否有助于录用决策，被试为 234 名在应聘条件和诚实条件下都完成了测验的应聘者。

首先，参照人格测验在人事选拔中的使用过程，设定录用步骤如下：

① 合成个体在应聘条件下《管理者人格测验》上的总分，然后依据测验的常模转换成标准分(均值为 0，标准差为 1)，可以反映个体的相对优秀程度。

② 把在《作假识别量表》上的得分超过分界线的个体从总体中剔除。由于被试都是名校的优秀毕业生，在诚实和应聘条件下的得分都比一般被试的得分要高，所以分界线取高一些定为 4.5 和 4。

分界线划为 4.5 时识别出作假者 66 人，剔除后剩余 168 人。分界线为 4 时识别出作假者 147 人，剔除后剩余人数 87 人。如果不使用《作假识别量表》，则在该步骤不剔除被试。

③ 筛掉人格测验得分不合格的被试，即把人格测验得分小于零(即低于一般水平)的个体剔除。在分界线为 4 时，剔除 38 人；在分界线为 4.5 时剔除 59 人；不使用作假识别量表时，剔除 61 人。最终选入的人数分别为 49 人、109 人和 173 人。

然后，把 234 名被试在诚实条件下的人格测验得分等分为高中低三个水平(各 78 人)，考察入选者中高和低水平的被试所占的比例。

表 5 不同条件下入选人员的特征对照表

是否使用《作假识别量表》	入选人数				配合度检验		入选者的真实水平	
	低分者	中间	高分者	合计	χ^2	*df*	*M*	*SD*
使用，分界线为 4	4(8.1%)	21(42.9%)	24(49.0%)	49	14.25**	2	0.48	0.66
使用，分界线为 4.5	9(8.3%)	44(40.4%)	56(51.4%)	109	32.83**	2	0.42	0.77
不使用	43(24.9%)	64(37.0%)	66(38.4%)	173	5.63	2	0.33	0.80

注：**p<0.01, *p<0.05

结果显示(表 5)，不使用《作假量表》进行作假识别时，在入选的 173 人中有 42 名是低水平个体，占 24.9%，66 名是高水平的个体，占 38.4%。配合度检验不显著，表明入选人员的真实水平符合自然比例，即高中低三个水平的人数比仍然为 1:1:1，说明由于被试作假，《管理者人格测验》没有起到选拔的作用。

使用《作假量表》且分界线分别为 4 和 4.5 时，在入选的 49 和 109 人中低水平个体所占比例为 8.1%和 8.3%，高水平个体为 49.0%和 51.4%。配合度检验均显著，说明入选的高水平被试更多，使用《作假量表》比不使用更有助于有效选拔。

而且，当分界线为 4.5 时入选的人员更多，给应聘者接受下一阶段精确选拔的机会更多，而且平均分为 0.42，仍然较大程度上高于平均水平(等于 0)，起到了一定的选拔筛选作用，因而建议采用较高的分界线。

4 讨论

4.1 作假的实验研究

作假的研究范式通常是：通过对指导语的操作，创建一个诚实组和一个作假组，通过两个组的反应对比，来了解作假的性质。指导语有两种类型：一类是提供“装好”与“诚实”的指导语，随机分配两组被试，对一组要求“夸大反应，尽可能表现出最好的形象”，对另一组要求“尽可能诚实回答，结果是匿名的”。另一类是“假装应聘”和“诚实”的指导语，一组被试仍然被要求“尽可能诚实回答”，一组被试则被告知“假装正在应聘某份工作，尽量作假回答而被录用”。目前有很多研究者对前者提出了批评，认为在该指导语下被试会依据社会标准而非工作标准对反应作假，而后者则能够诱导出与职业应聘情景中的作假更接近的歪曲反应(Zickar, Gibby, & Robie, 2004)。

本文的 3.3 部分就是采用后一种设计，给被试设置了“应聘”和“匿名”两种测试情境，从而诱发作假和自我欺骗的发生，然后再来挑选项目。本研究对应聘条件下的指导语作了精心安排，期望能够最大可能的模拟应聘者的心境。一是强调“个体要装扮成对这份工作特别期待的求职者”来反应，这使个体明确自己要扮演怎样的角色；二是给排名在前 5%的个体提供一定的奖金，这是一种物质奖励；三是“强调如果你没有被录用，我们也将对你的表现做出评定，看看你能给主考方留下什么样的印象”，个体可以据此了解自己的作假能力，这是一种精神奖励。Mueller-Hanson, Heggestad 和 Thornton III (2003) 认为奖励作假的方式会诱发出更加真实的作假动机。

4.2 《作假识别量表》项目的性质

一般来说, 受工作称许的行为同样受社会的称许, 反之亦然。但是, 二者的细微差别却会影响到SD 量表和《作假识别量表》的测量敏感性。

传统的 SD 项目描述的多是社会禁忌和人品道德等方面的行为, 比如发生争执时, 我一定先反省自己; 为了达到自己的目的, 我偶尔也会做几件损人利己的事情; 别人放在桌上的东西, 有时我不经允许就翻看了。当个体违背了这些行为时, 会受到道德规范的压力, 而产生焦虑感。因而个体即使诚实反应, 在这些项目上也会有较高的得分。

而《作假识别量表》的项目描述的多是组织最看重的品行, 比如: 我有时也会假公济私, 为自己人做些事; 存在利益冲突时, 我也会有个别不利于对方的言行; 适当的时候我会占些小便宜等。这些项目描述的多是在生活中已经司空见惯, 但在工作中又被反复强调的行为, 个体只有在应聘情景中遇到这些项目才会感到压力。此外,《作假识别量表》有较多的项目是与工作效率和质量相关的, 比如, 情绪不好丝毫不会影响我的工作; 我每天都记录自己的心得, 反思提高; 我一定不会把当天的事情拖到明天去做的; 我做的每件事情, 注意力都是非常集中的; 我在任何时候都能保持沉着、冷静; 我做事情有时会虎头蛇尾; 我想要做一些事情, 就是迟迟不肯动手去做等等。这些项目在传统的 SD 量表中很少见, 生活中由于个人能力参差不齐, 社会舆论对个体具有良好的工作行为的要求会有所降低。《作假识别量表》因为包括了这些与工作密切相关的行为, 才能更加有效地测量作假。

此外, 作假和自我欺骗都是夸大反应的趋势, 又采用相似的程序编写项目, 因而二者很容易被混淆测量, 为了避免该问题的发生, 本研究特别进行了被试内设计的实验研究来挑选项目。设置的挑选标准是: 诚实和应聘条件下的反应相关低, 及效应值大的项目。最后挑选的项目描述的多是在生活中已经司空见惯, 但是在工作中又被反复强调的行为, 它们不容易激起被试的自我欺骗, 却能够唤起被试强烈的作假动机, 因而能够充分测量作假, 而且不受自我欺骗的污染。

同时, 3.3 部分的实验操作的指导语中还强调“本测试将为一批企事业单位选拔出优秀的毕业生”, 其目的是突出“工作的一般性”, 而不是针对某一特殊职业的, 据此选择的项目适用于大多数工作的职业应聘情景中。

4.3 《作假识别量表》的有效性检验

Dwight 和 Donovan (2003) 担心测验中加入警告会使得个体识别出测谎题而故意诚实回答。在 3.3 部分的实验研究的指导语中有严厉的警告“测验中加入了一些探测说谎的题目, 请务必小心, 不要让主考方察觉出来, 否则你将被取消录用资格”, 个体却仍然有较大的反应率, 说明被试并不能识别出《作假识别量表》的极端项目。因为人们心中的优秀员工形象是比较完美的, 他们对行为的发生率并不敏感。

在 3.4 部分, 对作假反应的得分进行了因素分析, 20 个项目聚在一个因子上, 解释率 54.65%, 而且所有项目的因素载荷都比较高(>0.40), 这说明《作假识别量表》的结构效度较好。同时, 3.5 部分采用多元概化理论对构成量表的项目数量进行了探讨。结果表明, 采用 10 个项目就可以有较好的信度指标(G 系数和 φ 系数)。在 3.6 部分采用 Logistic 回归, 探讨了《作假识别量表》得分对作假反应的识别效果。结果表明, 方程的解释率比较理想, 对作假者的识别率也比较高。

此外, 本研究在一个真实的职业应聘情景下检验了《作假识别量表》的有效性。结果表明, 《作假识别量表》比人格测验更加敏感, 在应聘条件下, 分数有更大程度的提升。应聘情境下作假者多数集中在测验成绩分布的高端, 经检验在应聘分数高低端的个体的《作假识别量表》得分有显著的差异, 能够把这种作假现象反映出来。这些表明《作假识别量表》比较充分的测量了作假, 但是职业应聘情景存在多样性, 仍然需要大量的实证研究来论证其有效性。

4.4 《作假识别量表》的使用

在临床领域, 人们使用 SD 量表的得分进行分数校正得到真实特质的分数; 或者划定分界线, 识别作假者。在职业应聘情景中, Ellingson, Sackett 和 Hough (1999) 发现基于偏控制法得到的校正分数, 不能够代表个体在作假情境下真分数。Barrick 和 Mount (1996) 的研究结果发现, 偏控制的方法导致了预测效度的轻微缩减, 说明校正移走了特质成分。Goffin 和 Christiansen (2003) 认为关于校正分数的作用, 以及对录用决定的影响还有待进一步的研究。

在 I/O 应用领域, 测量到作假动机后除使用校正技术外, 还可以采用识别技术。本研究在编制《作假识别量表》时, 特别挑选了个体在作假和诚实条

件下反应差别比较大的项目，因而作假者在这些项目上的得分较高，划定分界线是可以把作假者识别出来的。

研究表明，有 60%~80%的被试在回答人格测验时会作假，这么大比例的作假者如果被全部筛掉，使用人格测验就毫无意义了，根本起不到选拔优秀个体的作用。但是，如果制定较高的分界线，把近20%~30%的严重作假者识别出剔除掉是可取的。

个体作假的目的是要挤进测验分数分布的高端，在低端作假者的比例会明显减少。Mueller-Hanson 等(2006)认为使用人格测验作选拔适合采用筛选出模式。筛选进模式是指按照从高到低的排名，录用少量的高分者。筛选出模式则是按照排名从下往上排除掉不符合最低标准的人员，该模式录用的人数是比较多的。使用筛选进模式录用的个体几乎都是作假者，而筛选出模式可以筛掉不符合最低标准的人员，虽然作假者不会遭淘汰，但该模式可以筛选掉那些虽然诚实反应了但真实水平低的个体。筛选出模式可以在选拔的早期使用，从成千上万的应聘者中排除掉不合格的人选，然后再使用面试、评价中心等手段进行细筛。

因而本文建议采用人格测验作选拔时，使用《作假识别量表》识别严重作假者，并且采用筛选出模式排除掉不合格的人员，然后再使用其它评价方法作精确的选择。

4.5 有待进一步研究的问题

首先，被试取样范围和样本量都比较小，而且很多被试都是没有工作经验的大学生。如果能有更加多样化的被试参加研究，研究结果会更加丰富。

其次，可能在一些特殊职业上有更加复杂的作假产生，识别作假也更加困难，《作假识别量表》需要在更多的职业应聘情景下检验其有效性，并判断分界线划定为多少更为合适。

最后，在人格测验中嵌套使用《作假识别量表》，是一种应对作假的事后识别技术。今后还应该尝试编制一些能够事前控制作假发生的人格测验，比如迫选量表等，如果它们与作假识别量表共同使用，将能够更加有效的应对作假。

5 研究结论

通过对作假与 SDR 的关系进行了辨析，基于作假的特殊性质，开发了《作假识别量表》。该量表共有 20 个项目，它们 1)受工作称许的程度高，取样范围广泛；2)只测量反应偏差，很少测量到人格特质信息；3)描述的多是在生活中已经司空见惯，但是在工作中又被反复强调的行为。在模拟研究和一个真实的职业应聘情境中检验量表的有效性，结果表明该量表对作假的发生敏感，作假一旦发生，量表分数会升高很多。

参 考 文 献

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*(4), 955–967.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*(3), 261–272.

Bradley, K. M., & Hauenstein, N. M. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlation and factor structure. *Psychology Science, 48*(3), 313–335.

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*(1), 1–23.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155–166.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and assessment, 11*(1), 340–344.

Gravetter, F. J., & Wallnau, L. B. (2006). *Statistics for the behavioral sciences* (7th edition). Wadsworth Publishing.

Marcus, B. (2006). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science, 48*(3), 226–246.

Martin, B. A., Bowen, C.-C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*, 247–256.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683–729.

Mueller-Hanson R. A, Heggestad E. D., & Thornton III G. C. (2006). Individual differences in impression management: an exploration of the psychological processes underlying faking. *Psychology Science, 48*(3), 288–312.

Mueller-Hanson, R., Heggestad, E. D., & Thornton III, G. C. (2003). Faking and selection; considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*(2), 348–368.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes (pp.17–59). San Diego: Academic Press.

Paulhus, D. L. (2002). Socially Desirable Responding: The Evolution of a Construct. In Brown H.I., Jackson D.N. & Wiley D.E., The role of constructs in psychological and educational measurement (pp.49–69). Mahwah NJ. Erlbaum.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634–644.

Yang Z. M., & Chang L. (2003). *Generalizability Theory and Its Applications.* Beijing, China: Educational Science Publishing House.

[杨志明, 张雷. (2003). *测评的概化理论及其应用.* 北京:教育科学出版社.]

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model Item Response Theory. *Organizational Research Methods, 7*(2), 168–190.

The Developing of Faking Detection Scale in Application Situation

LUO Fang; LIU Hong-Yun; ZHANG Yue

(*School of psychology, Beijing Normal University; Beijing Key Lab of Applied Experimental Psychology, Beijing* 100875, *China*)

Abstract

Using personality tests in personnel selection becomes increasingly popular in industries in China. However, the validity of personality tests is challenged due to the possibility of getting faking answers (Morgeson et al, 2007). Methods dealing with faking in personality tests are available. The most commonly used one is to detect faking by using social desirability (SD) scale, and then calibrate the results to remove the faking effect.

However, authors of this study argue that faking is conceptually different from socially desirable responding (SDR), and using SD Scale to measure faking effect is inappropriate. The reasons include: (1) although both faking and SDR are the tendency to exaggerate, the former is influenced by job desirability, while the later is influenced by social desirability. (2) SD Scale can not measure Impression Management and Self-Deception independently and has poor construct validity (Paulhus, 2002).

This research developed a “Faking Detection Scale for General Positions” in seven steps based on the procedures used to develop the SD scale and the special tributes of faking.

First, behavior descriptions which are strong job-desirable and extremely infrequent or high frequency were collected and rewritten into 114 items. Second, trained raters rated the desirability and occur-probability of these items and selected 50 items for the draft scale. Third, an experimental study was conducted with 679 subjects to further select items. Subjects were paid for 20 RMB if they successfully faked in simulated selection situations. As a result, 20 items which sufficiently measure faking and are not contaminated by Self-Deception remained in the scale. Forth, principal component analysis was used to examine the construct validity of this scale. Results showed the first factor explained 55% of the total variance. Fifth, multivariate generalization theory was used to study the optimal number of items included in the scale. The results showed that when 10 items remained in the Faking Detection Scale, the scale has good reliability, with $G=0.823$ and $\varphi=0.818$. Sixth, the scale’s validity was further examined using logistic regression. The Cox&Snell R Square and Nagelkerke R Square were .386 and .468, repectively. These results showed that the scale was valid and can effectively detect faking candidates. However, there is a dilemma when setting the cut-off score, because it is impossible to decrease the false judgment and false selection rate at the same time. Results from this study showed that setting up higher cut-off score to detect serious faking candidates and exclude unqualified candidates using select-out mode can help organization make valid recruitment decisions and accomplish personnel selection goals. Finally, this scale’s validity was verified in real occupational selection situations with 234 subjects. Faking Detection Scale (effect size was 1.163) is far more sensitive than personality scale (the largest effect size was .767) and can more effectively measure faking effect. This scale has been validated and can effectively detect faking candidates in real occupational selection situations.

By exploring the quality of items, it is found that most faking detection items described common behaviors in daily life and repeatedly emphasized in work-settings. Subjects showed a high consciously inflation on these items when they were faking; when faking did not happen, they would unconsciously exaggerate the reaction to a less extent because of Self-Deception. So the scale could sufficiently measure faking effect with less contaminative by Self-Deception.

As a conclusion, this study differentiated faking from SDR and developed a scale to detect faking. This scale has 20 items which is very sensitive to faking. Its validity was verified in simulative and one real occupational selection situation. More studies are needed to further explore whether this scale can be applied to varied situations.

Key words faking; socially desirable responding; social desirability scale; job desirability