

• 研究方法(Research Method) •

他评：一种有效的人格评价手段*

张登浩¹ 滕 飞² 潘 雪¹

(¹中国人民大学心理学系, 北京 100872) (²香港大学心理学系, 香港)

摘 要 他评法是一种有效的人格评价手段, 甚至在一些情境下比自评法更为有效。在现实准确性模型(the Realistic Accuracy Model, RAM)以及权重平均模型(the Weighted-average Model, WAM)的基础上, 对于影响他评准确性的因素进行了梳理, 涉及熟悉性、关系质量、特质可见度以及假设相似性等。针对目前研究中存在的问题, 提出要在中国文化背景下, 进一步拓展研究对象的取样范围, 在更多不同的情境下提高研究的生态效度, 同时采用更多的人格评价手段。

关键词 他评; 人格评价; 现实准确性模型; 假设相似性

分类号 R395.6

虽然在现代人格心理学中, 绝大多数研究都采用自我报告的方法来评估被试的人格特质。但实际上他评作为一种测量人格特质的方法比自评的方法历史更为悠久, 其历史甚至可以追溯到公元前 319 年, 当时的研究者 Theophrastus 在进行性格研究时就采用了他评的方法, 而在心理学中第一个使用他评方法的研究者是 Webb, 他曾在 1915 年使用该方法对一些实习教师的人格特质进行了研究(McCrae & Weiss, 2007)。获得他评数据的手段主要采取的是与目标对象自我报告的人格问卷相应的他评人格问卷来获得的, 一般来说都是将自评问卷中的第一人称“我”用“他/她”来代替从而形成相应的人格他评问卷(McCrae & Weiss, 2007)。当然如果已有的自我报告问卷通过这种改写以后并不完全适用的时候, 研究者也可以按照编制一般自我报告问卷的方法重新编写一个适合进行他评的专门量表。

虽然他评法历史颇为悠久, 但由于研究者一直认为自评法更加简便, 获取数据更为容易, 再

加上人们想当然的假设——我们自己是最了解我们自己的人, 因此自我报告法在人格研究领域受到更为广泛的应用, 根据 Vazire (2006)的统计, 在著名人格心理学期刊《人格研究杂志》(Journal of Research in personality) 2003 年的所有研究报告在评价人格特质时 98%都采用的是自评法, 而且其中 70%都是将自评法作为唯一的测量方法在使用, 而他评法应用的范围却非常有限。但大量的研究都不断表明, 观察者所作的评价, 不仅可以有效弥补自我报告的缺陷, 而且那些对于目标对象非常熟悉的观察者可以对目标对象的人格特质做出非常准确且令人信服的评价(例如 Connelly & Ones, 2010; Srivastava, Guglielmo, & Beer, 2010; Funder, 2012), 甚至在某些情境下他评法所做出的评价比自我报告法更加准确(Vazire, 2010; Oh, Wang, & Mount, 2011), 因此, 他评作为一种有效的人格评价手段应该受到更多的重视和应用。

1 人格自评的缺陷与不足

有效性是决定一种方法能否使用的重要标准之一。长期以来人格心理学家一直将自我报告法作为人格评价最主要的手段。这一方面是因为自我报告法有着非常明显的优势, 比如简便易行、成本较低等; 另外, 更重要的原因可能还是在于人们一直想当然地认为, 我们自己是最了解我们

收稿日期: 2013-04-18

* 国家自然科学基金(71272156)和全国教育科学规划教育部青年专项课题(EBA110328)成果。

通讯作者: 张登浩, E-mail: zdh@ruc.edu.cn

自己的(Paulhus & Vazire, 2007), 因此自评数据应该具有更高的可靠性。但实际上这一假设在很多时候面临着巨大的挑战(Vazire & Carlson, 2011; McDonald, 2008)。弗洛伊德认为, 潜意识是决定人类行为最主要的力量, 而潜意识在正常状态下根本无法进入人们的意识领域, 也根本不可能通过内省的方法进行自我报告。更多的心理学研究也同时发现, 个体的外显态度和内隐态度之间往往存在着巨大差异(例如 Rudman & Mescher, 2012), 这一方面为弗洛伊德的潜意识理论提供了一定的佐证, 同时也提醒研究者在预测个体行为时不能仅仅考虑其外显态度。

王登峰和崔红(2006)更是主张, 人格的内涵可以划分为4个不同的层次, 即个体的外显行为、内心情感和体验、稳定的动机和生理生化反应模式和神经系统活动的变化。其中, 除了外显行为无论当事人自己还是其他人都可以很容易地观察到之外, 内心情感和体验以及稳定的动机即便是当事人自己也很难非常准确的把握, 而神经系统活动的变化无论当事人自己还是其他人都无法觉察或观察。

自我报告法面临的另外一个主要的问题就是社会赞许性问题。尽管研究者们采取了很多方式来试图减少个体在自我评价中的社会赞许倾向, 但很多时候这种方法收效并不明显。个体在进行自我报告时很容易出现对以往经验与行为的歪曲或者进行伪装, 而且有时候这种歪曲或伪装还是无意识的。他评法相对于自我报告法的一个很大的优势就在于可以比较有效地避免社会赞许性。一个人可能会受到社会赞许性的影响, 而故意歪曲或者通过有色眼镜来对自己过去的行为进行描述, 但是观察者没有理由也受到社会赞许性的影响, 而尽可能的夸大被评价者的优点同时尽可能的缩小其缺点(McCrae & Weiss, 2007; Paunonen & O'Neill, 2010)。

2 人格他评方法的有效性

他评法是否可以作为一种有效的人格评价手段来使用, 关键取决于其对被评价者人格评价的准确性。为了解决人格他评方法的准确性问题, 研究者(Funder, 1995, 2012)已经提出了三个得到广泛认同的准确性评价标准, 第一, 如果一个评价者所做的人格他评是准确的, 那么它应该能够

预测另一个评价者对同一个目标对象所做的人格评价, 即不同的他评之间应该具有一致性, 这实际上也是我们经常提到的评价者信度。研究发现, 在“大五”的五个维度上, 即外向性、愉悦性(Agreeableness)、公正严谨性(Conscientiousness)、神经质和开放性, 熟人对同一目标对象所做的评价之间的相关平均都在0.40以上, 甚至更高(Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008)。第二, 如果一个评价者所做的人格他评是准确的, 那么他应该能够预测目标对象自己所做的人格评价, 即人格的自评-他评之间应该具有一致性, 研究者经常使用自评-他评之间的一致性作为评价自评量表效度的重要指标。研究发现, 对大多数人格特质而言, 自评与非常熟悉的他人所做的评价之间基本上都有中等以上程度的相关(Connolly, Kavanagh, & Viswesvaran, 2007; Allik et al., 2010)。第三, 如果一个评价者所做的人格他评是准确的, 那么他应该能够预测目标对象相关的行为和结果。当然这三个标准并非完美, 自评-他评以及他评之间的高度一致性, 并非他评准确的充分条件, 只是必要条件, 比如他评之间的高度一致性很有可能是不同的评价者对目标对象拥有相似的刻板印象造成的。尽管如此, 如果他评法满足了这三个条件, 我们完全有理由认可其评价的准确性。因为研究者在评价人格自评的准确性时最多也只采用了这三个标准当中后两个标准进行评价而已。

Funder (1995, 2012)所提出的人格评价的现实准确性模型认为, 要实现人格评价的准确性, 必须经历一个复杂而特殊的过程(见图1)。在这一过程中首先涉及的是评价者观察目标对象人格特质的环境, 在这个环境中目标对象一定要能够清晰地将被评估的特质展现出来, 即所谓的相关性; 第二, 当目标对象表现出与被评估的特质相关的行为时, 评价者必须能够有效地意识到目标对象所表达的这种特质, 即所谓的可用性, 接下来, 评价者必须能够注意到与特质相关的行为线索, 因为人格评价的准确性不仅涉及到目标对象对于特质的表达, 也取决于评价者对于特质知觉的准确性, 即所谓的察觉性, 另外, 评价者在发现了相关的特质线索之后, 还必须能够将这些有效的线索恰当地结合起来, 从而对目标对象形成一个客观准确的形象, 即所谓的利用性。Funder (1995,

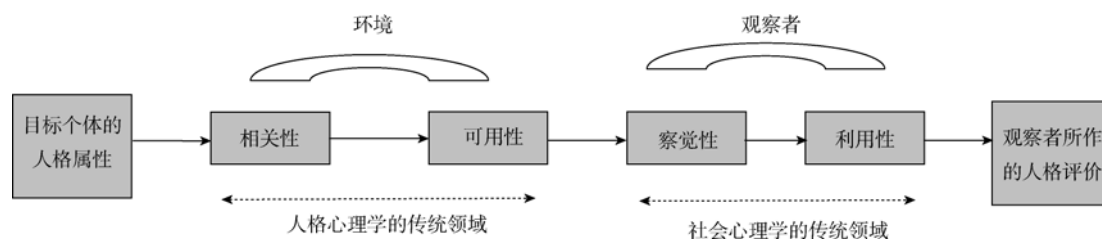


图1 人格评价的现实准确性模型(引自 Funder, 1995)

2012)认为只有这4个条件都得到满足的时候,评价者才可能对目标对象的人格特质做出准确地评价。在这个模型的指导下,研究者发现了许多影响人格评价准确性的重要变量,Funder (1995, 2012)将其归结为4类,即良好的特质、良好的评价者、良好的被评价者以及良好的信息。

Kenny (1991), Kenny 和 West(2010)在研究人际知觉的准确性时提出了权重平均模型,该模型认为,他评方法的准确性来源于不同评价者所观察到的行为的重叠程度以及他们所拥有的意义系统的相似性。如果不同评价者所观察到的目标对象的相同行为更多,他们对目标对象人际知觉的一致性程度就会越高,此外,如果不同评价者对于同一行为涵义的理解越相似,他们对目标对象人际知觉的一致性程度也就会越高。

实际上,不管是 Funder (1995)的现实准确性模型还是 Kenny (1991), Kenny 和 West (2010)提出的权重平均模型,还是更多的研究者对于影响人格评价准确性的相关变量的研究都至少说明了,人格他评方法的准确性是可能的。

当然,这里还有一个很重要的问题,尽管人格他评具有准确性,但如果他评方法所获得的数据与自评的数据一致,即上面提到的第二个标准,那么我们依然没有特别充分的理由舍弃自我报告法而使用他评法,因为自我报告法简便易行的优势非常明显。虽然并非所有的相关研究都证实了他评法与自我报告法相比有更加准确,但的确已有相当多的研究证实了这一点。Watson 和 Humrichouse (2006)考查了新婚夫妻人格特质在两年内自评和配偶评价发展与变化情况,虽然这两者的评价前后两次相关的稳定性都比较高,但是在第二次评价中配偶评价更加的消极,而且配偶所做的人格评价与个体自己报告的婚姻满意度呈现显著的相关,而自我报告的人格特质则与婚

姻满意度没有显著的相关。Oh 等(2011)也对五因素人格特质与工作绩效的关系进行了元分析,结果发现在观察者评定的基础上,所有的五因素特质都是整体绩效非常显著的预测指标,而且五因素人格特质的观察者评定在预测整体绩效方面其作用要远远高于自我报告。Connolly 和 Ones (2010)整合了对来自 263 个独立样本的 44178 名目标对象个体评价的元分析发现,观察者评定是行为更有效的预测指标,当对个体的学业成就和工作绩效进行评价时,他评比自评表现出了更好的预测效度。Redzanowski 和 Glück (2013)对来自奥地利克拉根福大学 17 个不同院系的 179 名教师的聪颖性(wisdom)采用自评和他评的方法进行了考察,结果发现,两种不同量表所获得的自评结果以及他评的结果之间相关并不显著。对于这一结果研究者认为,虽然我们想当然地认为自评具有更强的可靠性,但是两种自评方法所获得的结果之间虽然相关,但相关程度却很低的事实至少说明研究者对于聪颖性概念的理解以及操作化方面存在巨大的分歧,而聪颖性作为一种更容易被他人而不是个体本人所了解和评价的人格特质可能更适合于采用他评的方法进行评价(Staudinger & Glück, 2011)。据此,越来越多的研究者认为虽然不能说他评一定要优于自评,但是至少在一些情况下他评的确比自评更加准确(Connolly et al., 2007)。这对于我们重新审视已有的主要以自评方法所获得的相关研究结果无疑具有极为重要的意义。

3 使用人格他评数据的情境

由于人格他评数据具有一定的有效性,因此,有研究者认为在任何评定人格的情境下都可以使用他评数据,而且在一些不能使用自我评定的情况下,依然可以使用他人评定的方法来进行人格

评价(McCrae & Weiss, 2007)。首先, 当目标对象不能进行自我评定的时候, 比如儿童(Wang, Chen, Petrill, & Deater-Deckard, 2013)、丧失认知能力或判断能力的个体、历史人物、类人猿或宠物等(Meagher, 2009), 此时他评成为惟一可以使用的手段(苏彦捷, 2007), 这也是目前使用他评数据较多的情境之一; 其次, 他评数据可以作为自评数据的一种重要的补充, 例如在招聘情境中, 由于个体的求职意愿非常强烈, 掩饰自己的真实人格特点而努力迎合招聘者要求的倾向比较突出, 此时个体自评数据的可靠性会受到较大的影响, 如果同时参考观察者评价则可以有效地提高人格评定的有效性; 第三, 当研究者试图通过对目标对象的人格特质进行多方位的评价从而提高评价的准确性时, 研究者通常会将他评作为一种非常重要的评价手段来使用(刘电芝等, 2011; Schneider & Schimmack, 2009)。Colbert, Judge, Choi 和 Wang (2012)针对前人在研究人格与领导力的关系时只使用自评作为人格测量的手段提出, 应该将自评和他评的方法结合起来使用。他们通过对 155 名被试的研究发现, 将这两种评定方法结合起来的时候, 与单一使用人格自评相比, 此时人格特质可以解释更多的领导力的变异, 而且在所有可以解释领导力变异的人格特质中, 他评的外向性所解释的比例最高, 其次是自评的开放性和他评的开放性。他评数据所具有的良好效度, 成为目前越来越多的研究者开始使用他评数据的重要原因。

此外, 虽然很多研究者都认识到了自评法存在的一些突出问题, 但一直坚持认为自评法具有更加简便、节省费用等无可比拟的优势。对此, Vazire (2006)提出了不同的看法, 他认为, 随着网络和电子邮件的日益普及和广泛使用, 观察者可以非常简单地点击邮件中的链接地址在网络上完成问卷调查, 而不需要完成传统的纸笔测验, 这使得他评法的效率也大大提高。此外, 为了降低他评法给付给被试的费用而带来的花费增加的问题, Vazire (2006)认为可以通过不给付报酬的方式来解决这一问题, 因为不给付报酬同样也可以得到较高的问卷回收率, 给付报酬并不是提高问卷回收率的重要保证, 而且给付报酬会鼓励被试作假, 甚至一些被试为了获得更多的报酬而自己包办观察者评价, 即便是真正的观察者也可能会因

为报酬的原因而不严肃认真地对待问卷, 因此除非是非常困难且很浪费时间的任务, 一般任务完全可以不给付被试费用。对他评法的另外一个质疑就是观察者有可能不愿意合作, 心理学研究中一般都使用大学生作为被试, 而且很多作为被试的大学生参加实验是为了获得相应的学分, 而作为大学生的观察者, 特别是他们的父母, 作为成年人不大可能因为学分而参加实验。但事实证明, 观察者也是相当配合的, Vazire (2006)在他们的多次研究中就发现, 观察者的问卷回收率达到了 76%~95%, 研究者认为, 回收率之所以这么高, 一个很重要的原因就是, 观察者很多时候认为自己是非常了解被观察者的, 而且这样一个评价可以让观察者有机会重新审视被观察者, 帮助他们更深刻地理解对方, 而且 Vazire (2006)在研究中所使用的问卷只有 100 个题目, 仅仅只需花费 10 分钟左右的时间, 对观察者而言并不是一个难以承受的负担。除此之外, 为了提高他评法的回收率, Vazire (2006)还使用了一些简单的策略, 比如为了防止观察者将研究者所发的邮件当作垃圾邮件删除, 他们均使用学校所属的电子信箱来发送邮件, 而且邮件的主题中都附加上被评价者的名字来作为提示。最后, 传统的观点认为他评法的有效性比较低, Vazire (2006)认为可以通过向观察者强调其评定的机密性, 减少他们对评价结果的忧虑所带来的干扰。当然, 和自评方法一样, 他评法也无法完全消除被试不诚实作答所带来的问题。

4 影响人格他评准确性的主要因素

按照 Funder (1995, 2012)的现实准确性模型以及 Kenny (1991), Kenny 和 West (2010)的权重平均模型, 人格特质他评的准确性会受到多个因素的影响, 目前研究者关注较多的因素主要有以下几个。

4.1 熟悉性对人格他评准确性的调节作用

对于熟悉性效应的证明既包括横向的跨情境研究和追踪研究, 也包括实验室情境中的研究和自然关系中的研究(张登浩, 王登峰, 2010)。横向跨情境研究基本上都证实了人格的准确性评价中的熟悉性效应(Biesanz, West, & Millevoi, 2007; Schneider & Schimmack, 2009)。但在横向跨情境研究中除了熟悉性之外, 很难排除其它与人格评

价准确性相关的因素的影响。鉴于横向跨情境研究的不足,许多研究者建议采用追踪研究来探讨人格评价准确性问题(Biesanz et al., 2007; 张登浩, 王登峰, 2010)。许多追踪研究都发现人格评价的准确性随着熟悉程度的增加而增加。例如, Paulhus 和 Reynolds (1995)将以前不熟悉的3个班级的196名同学组成每组四到六名学生的组,在7周当中,这些小组的成员每周见面1次,每次20分钟。在第一次见面之前,所有的被试都进行了FFM (the five-factor model personality)特质的自我评价,并且在第1次和第7次见面之后对他们的团体成员进行评定。研究结果确认了熟悉性效应,在FFM的大多数维度上自评-他评的一致性在第7周的时候都要高于第1周。

然而,追踪研究的结果却并不像跨情境研究那样一致,还有相当多的追踪研究并不支持熟悉性效应。例如,上面提到的 Paulhus 和 Reynolds (1995)的研究就发现,在外向性维度上第一周时一致性最高,然后会有一些轻微的下降。Watson 和 Humrichouse (2006)考察了新婚的年轻人在两年中自我评价和配偶评价的一致性,结果发现配偶评定出现了与自我评定完全不同的模式,在公正严谨性、愉悦性、外向性和开放性方面都出现了显著的下降。

对此, Biesanz 等(2007)认为以前的追踪研究大多数都是在实验控制的情境当中进行的。虽然实验室情境可以对熟悉性进行实验操控,但是这些发现可能会很难概化到自然关系中去,自然关系当中的研究虽然提供了更加有意义的情境,但它很难对熟悉性的程度进行控制,而且很难对刚开始熟悉的过程进行探讨(Kurtz & Sherker, 2003)。许多研究者都一致认为大学生室友是解决这一两难选择的一个很好的办法,大学生室友一般都是随机分配且原来互相都不认识,研究者可以从最初的陌生到后来的熟悉阶段的任意一点采集数据来进行人格的自评-他评一致性的追踪研究(Kurtz & Sherker, 2003; 张登浩, 王登峰, 2010)。

另外,关于熟悉性效应的研究结果不一致的原因可能还在于不同研究者对于熟悉性的定义不同。Starzyk, Holder, Fabrigar 和 MacDonald (2006)提出了构成熟悉性水平的6个维度:持续时间、交往的频率、对目标的了解、身体上的亲密程度、自我表露以及社会网络的熟悉程度。而在很多研

究中对于熟悉性的考量或者主要考查的是评价者与目标对象交往的持续时间、或者考查对目标对象的了解程度,考查的指标差异较大,因此出现一些不一致的结果似乎也就难以避免了。

4.2 关系质量对人格他评准确性的调节作用

“情人眼里出西施”这非常典型的反映了关系质量对于人格他评准确性的影响。当评价者与目标对象之间的关系比较好的时候,目标对象会认为评价者更有可能了解他们的思想、感情和行为,而且评价者也会对目标对象更加偏爱,对其人格特质的评价可能会更加积极(Boyes & Fletcher, 2007)。Watson 和 Humrichouse (2006)在对新婚夫妻的研究中发现,配偶评定的变化与婚姻满意度的变化存在着显著的正相关,出现了所谓的“蜜月效应”,即在蜜月期间,配偶的评价与自我评价相比更加积极。McCrae 和 Weiss (2007)也指出,父母在对自己的子女进行人格评价的时候,会使用更多社会赞许的特质。张登浩和王登峰(2010)对中国大学新生室友的研究也发现,在第15周时,关系质量对于外向性和处世态度维度的自评-他评一致性具有显著的调节作用,而且与被试的自评相比,良好的关系质量会导致室友将被试评价为更加善良、人际关系更好。实际上,个体在对伴侣进行评价的时候经常会出现所谓的“积极幻觉(positive illusions)”,即个体在对伴侣的特征进行评价时经常会出现比客观状况更加积极的现象(Murray, Holmes, Dolderman, & Griffin, 2000),例如个体经常会将他们的伴侣评价的过分“善良”和“智慧”,而且这种积极幻想与个体和伴侣之间的关系质量存在着显著的相关,积极幻想可以有效地预测伴侣之间更高的关系满意度、信任水平、相爱水平以及更低的冲突水平。

4.3 特质的可见度(visibility)

一些人格特质与其它人格特质相比更容易被评价,比如外向性、健谈的(talkative)等,这些人格特质所描述的行为相对很容易被观察者直接观察到,而像开放性、神经质等人格特质所描述的行为不大容易被观察到(Connolly et al., 2007; Hall et al., 2008; Permack, 2011),因此,观察者在评价这些不同的人格特质时难易程度会不同,准确性也会受到不同的影响,可见度越高的特质越容易被评价。研究者认为高可见度的特质往往描述的是那些比较外在的倾向,比如行为,而低可见度

的特质往往描述的是那些比较内在的倾向,比如思想和情感,对于观察者而言很难观察到也就很难应用到人格评价的过程之中。因此,对于一般观察者而言,对于每一种可见度不同的特质了解的程度是不同的,很多研究发现,在“大五”人格模型中,外向性的可见度比较高(例如 Permack, 2011)。按照 Vazire 的自我-他人的知识不对称模型(the Self-other Knowledge Asymmetry Model, SOKA),可见度低的特质自我评价更为准确,而那些具有评价性启示的特质应该被他人评价的更好。因此,担忧或感觉焦虑的倾向应该被自我评价的更准确,而智力被熟人评价的更为准确(Vazire, 2010)。

4.4 假设相似性(Assumed similarity)

假设相似性指的是人们倾向于认为他人与自己很相似的倾向(Beer & Watson, 2008),或者与某个典型的他人很相似的倾向(Cronbach, 1955)。假设相似性、相似性以及自评-他评的一致性如图 2 所示。很多研究都发现,观察者所做的自我人格评价与他们对他人所做的人格评价之间存在显著的相似性(Watson & Humrichouse, 2006)。Ready, Clark, Watson 和 Westerhouse (2000)发现,假设相似性水平高的特质其自评-他评的一致性更低,研究者认为这一现象表明,将自己的特质投射到他人身上可能是一个比较普遍的倾向,特别是当评价者对被评价者缺乏相关信息的时候,假设相似性对其所做的评价会产生较大的影响,有研究发现在零熟悉情境下,对神经质而言假设相似性($r=0.32$)要显著的高于外向性($r=-0.07$) (Beer & Watson, 2008),而对于已婚夫妻、恋人以及朋友关系的研究发现,“大五”的五个人格特质的假设相似性大多都在 0.28 以下,但对于积极情感和消极

情感特质的评价,假设相似性大多都在 0.30 以上(Watson, Hubbard, & Wiese, 2000)。因此有研究者认为一般来说对于可见度高的特质,因为获取相关信息比较便利,因此评价者更有可能依据所获得的真实信息进行评价,而对于那些可见度低的特质而言,评价者更有可能依据一个以自我为基础的启动效应(the self-based heuristic)进行评价(Beer & Watson, 2008)。但也有研究存在不同的结果, Lee 等人(2009)以人格的 HEXACO 六因素模型基础的研究就发现,在六个大维度层面上,假设相似性最高的是诚实-谦虚(Honesty-Humility) ($r=0.44$),其次为经验的开放性(Openness to Experience) ($r=0.35$),其余分别为外向性($r=0.11$)、愉悦性(Agreeableness) ($r=0.06$)、公正严谨性(Conscientiousness) ($r=0.05$)以及情绪性(Emotionality) ($r=-0.07$),而很多研究发现,在大五人格模型中,外向性特质可见度一般来说最高(例如 Permack, 2011)。因此,完全从可见度的角度来解释假设相似性似乎并不完全准确。

有研究者认为,假设相似性对于人格评价准确性的影响不仅在陌生人之间存在,而且在熟悉的个体之间也普遍存在。对于这一现象的理解可以通过理解亲密关系中的自我中心主义(egocentrism)来解释(Simmonds, 2010)。自我中心主义指的是从自我的角度来定义他人的品质、价值观和经验。自我中心主义者会从自己的态度和观点出发去对他人做出假设,有研究发现,在令人满意的长期关系中,自我中心主义使得个体将伴侣看作是自我的一部分,从而想当然的认为伴侣与自己在某些实际上并不相似的领域具有相似的特点(Simmonds, 2010)。

个体在对他人的人格进行评价的时候除了会

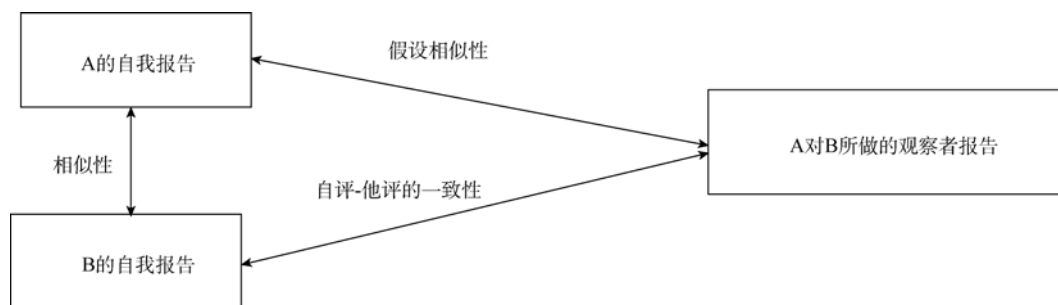


图 2 二元关系中人格自评与他评之间的三种关系(引自 Lee et al., 2009)

依据自己的人格特质进行评价之外,还会依据头脑中所储存的典型他人的形象进行评价,比如当我们了解到一位陌生人是大学生的时候,由于我们对其了解的信息非常有限,此时为了做出一个更为准确和全面的评价,我们会根据自己已有的关于一般或典型大学生的特点对该大学生进行评价。实际上此时影响人格他评准确性的是我们对某一群体的刻板印象。有研究发现,自评-他评一致性水平在那些比较中性的特质上要比那些社会赞许性比较高的特质上更高一些(John & Robins, 1993)。

5 总结及未来研究的展望

虽然大量的研究已经证实了他评不仅可能是准确的而且其准确性或预测能力在某些情况下还超过了人格的自评,但已有的研究也存在着一些难以忽视的问题。

首先,已有研究中所采用的被试基本上都来自于大学生,特别是零熟悉情境的实验室研究中,对于其它群体的他评准确性的研究很少看到。虽然我们不能据此去广泛地质疑已有的相关研究结果,但不可否认的是,没有考虑到被试样本来源的多样性不能不说是一个重要的问题。大学生基本都处在成年早期的阶段,从被评价者的角度来说,他们的人格相对而言还具有相当的可变性,评价难度较大,而从大学生作为评价者的角度来说,他们的社会经验和阅历还相对缺乏,对他人所表现出的特质相关行为的理解和认识可能还有待提高,这两方面的因素叠加可能会低估他评的准确性。

其次,已有的关于人格他评准确性的研究基本上都是在西方文化进行的,以中国人为被试的研究还非常少。王登峰等认为,中西方对于人性的不同看法使得中国人行为的防御性或掩饰性更强,中国人的言行在经过了反复的包装和修饰之后,往往只是反映了个体对环境要求的回应,而很少能够体现个体内心真实的体验和动机,只有在与自己最熟悉、最亲近的人在一起时,个人的行为才主要由个体的内在需要所决定(王登峰,崔红, 2006)。因此,对于中国人而言准确评价他人的人格会更加困难,实际上我们的研究也证实了这一点,我们对大学新生室友的研究就发现,与开学第二周相比,第十五周的自评-他评的一致性水

平虽然大多都有所提高,但绝大多数却没有达到统计上的显著性水平(张登浩,王登峰, 2010)。因此亟需在中国文化背景下对这一问题进行研究。

再次,很多关于零熟悉情境的研究都采用了严格的实验室研究来考察哪些因素会影响他评的准确性。在这些研究中很多采用被评价者的照片、简短的视频材料或评价者和被评价者简短的交谈等作为评价的线索来考察他评的有效性。虽然严格的实验控制条件使得研究者对于所发现的因果关系更具有信心,但同时我们不得不注意另外一个问题,一些非常常见的人际交往情境在实验室环境中是很难重现的,例如与关系非常好的朋友的非常亲密的交流(Vazire & Carlson, 2011)。此外,个体在人际交往的过程中往往会为了各种各样不同的目的而去展示或塑造各种不同的形象,因此如何透过个体在不同情境下所展示出来的不同形象或表现出的各种不同行为去把握其背后的稳定的人格特质,需要我们在更多不同的情境下去进行研究,并将这些研究结果进行整合。

另外,按照 Funder (1995, 2012)所提出的人格评价的现实准确性模型,影响人格评价准确性的变量包括四类,即良好的特质、良好的评价者、良好的被评价者以及良好的信息。但从已有的研究来看,前两类变量即良好的特质和评价者受到的关注相对更多一些,而后两者特别是良好的被评价者受到的关注很少。研究发现,行为的一致性水平较高或行为的掩饰性水平较低的被评价者,他评的准确性可能会更高。但对于那些行为一致性水平较低或行为掩饰性水平较高的被评价者,我们应该如何去把握他们的人格特点呢?非言语行为、微表情可能在这种情况下会对我们有所帮助。

最后,目前在研究人格评价准确性时基本上都采用相关的人格问卷来进行的。随着人格评价手段的越来越丰富,比如许多研究者通过分析个体的面部结构、音乐品味、个体的 Facebook 主页的内容甚至是其卧室的整洁程度来评价个体的人格特点。特别是对于电子社会媒体越来越多的使用打开了一个令人激动且富有挑战性的机会,让我们能够获得人格鲜活的社会交往关系,就像它正在发生一样(Back et al., 2010; Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011)。因此在未来的研究中,应该采用更多的评价手段来探讨人格评

价的准确性问题。

随着社会的发展,不同文化之间的交流日益频繁,人际交往越来越成为日常生活必不可少的一部分,虽然交往的媒体越来越多样化,但不管怎样,建立、维持和发展一段良好的人际关系都需要我们对交往的对象做出评价,而这一评价是否能够更加准确必然影响着一段人际关系的开始和健康发展,因此,如何让他评这样一种人格评价手段更准确?如何通过各种干预方式促使个体对他人人格特质的评价更加准确?都是相关领域的研究者所必需面临的问题。

在对人格进行评价的时候,到底是自评更准确还是他评更准确?这的确还是一个很难做出抉择的问题。但通过已有的研究我们至少可以得出这样的结论,即在某些情况下,他评的确比自评更为准确。但同时我们又不得不面临另外一个问题,自评和他评的数据到底应该如何使用呢?这实际上是一个更加难以做出明确回答的问题。之所以难以回答是因为这实际上牵扯到了准确性评价的标准问题。人格评价准确性研究在上世纪50年代戛然而止,在80年又开始重新兴起,都源于准确性的标准问题。Funder(1995, 2012)所提出的现实准确性模型就是在综合已有的实用主义和建构主义角度的基础上提出来的,不仅仅考虑评价者,同时也关注被评价者。现实准确性模型遵循的是批判现实主义哲学,他认为缺乏完美的、一贯正确的标准来判定真理,所以不能强求人们认为所有有关现实的解释都是同样正确的。因此,虽然现实准确性模型所提供的三个准确性标准也不完美,也就是说即便我们同时满足了这三个标准,也无法确定我们的评价就一定是准确的。但是除非我们放弃关于准确性的研究,否则作为研究者来说,必须依据手中的信息或能够收集到的信息做出尽量合理的选择。因此,对于自评数据和他评数据,我们很难说到底应该使用哪一个,因为每一个都有其优势,但同时也是不完美的(Funder & Dobroth, 1987)。因此,到目前为止,大多数的研究者比较认可的做法是将他评数据作为自评数据的一个有益的补充,他们都为我们理解人类行为提供了不同的视角,将两者结合起来,使得我们所做的人格评价尽可能满足更多的准确性标准,尽可能提高人格评价的准确性,同时也更有可能对个体的行为做出更加准确的评价

(Connelly & Ones, 2010; Vazire & Carlson, 2011)。

参考文献

- 刘电芝, 黄会欣, 贾凤芹, 龚茜, 黄硕, 李霞. (2011). 新编大学生性别角色量表揭示性别角色变迁. *心理学报*, 43(6), 639-649.
- 苏彦捷. (2007). 动物个体差异研究对人格心理学的贡献. *心理科学进展*, 15(2), 260-266.
- 王登峰, 崔红. (2006). 行为的跨情境一致性及人格与行为的关系: 对人格内涵及其中西方差异的理论与实证分析. *心理学报*, 38(4), 543-552.
- 张登浩, 王登峰. (2010). 人格自评—他评一致性 & 关系质量的调节作用. *中国临床心理学杂志*, 18(3), 349-352.
- Allik, J., Realo, A., Möttus, R., Esko, T., Pullat, J., & Metspalu, A. (2010). Variance determines self-observer agreement on the Big Five personality traits. *Journal of Research in Personality*, 44(4), 421-426.
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21(3), 372-374.
- Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment*, 90(3), 250-260.
- Biesanz, J. C., West, S. G., & Millevoy, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, 92(1), 119-135.
- Boyes, A. D., & Fletcher, G. J. O. (2007). Metaperceptions of bias in intimate relationships. *Journal of Personality and Social Psychology*, 92(2), 286-306.
- Colbert, A. E., Judge, T. A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly*, 23(4), 670-685.
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136(6), 1092-1122.
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15(1), 110-117.
- Cronbach, L. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin*, 52(3), 177-193.
- Funder, D. C. (1995). On the accuracy of personality

- judgment: A realistic approach. *Psychological Review*, 102(4), 652–670.
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177–182.
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52(2), 409–418.
- Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, 14(9), 483–488.
- Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42(6), 1476–1489.
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61(4), 521–551.
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98(2), 155–163.
- Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self-and other perception: A meta-analysis. *Personality and Social Psychology Review*, 14(2), 196–213.
- Kurtz, J. E., & Sherker, J. L. (2003). Relationship quality, trait similarity, and self-other agreement on personality ratings in college roommates. *Journal of Personality*, 71(1), 21–48.
- Lee, K., Ashton, M. C., Pozzebon, J. A., Visser, B. A., Bourdage, J. S., & Ogunfowora, B. (2009). Similarity and assumed similarity in personality reports of well-acquainted persons. *Journal of Personality and Social Psychology*, 96(2), 460–472.
- McCrae, R. R., & Weiss, A. (2007). Observer ratings of personality. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 259–272). New York: The Guilford Press.
- McDonald, J. D. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1(1), 1–18.
- Meagher, R. K. (2009). Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science*, 119(1–2), 1–14.
- Murray, S. L., Holmes, J. G., Dolderman, D., & Griffin, D. W. (2000). What the motivated mind sees: Comparing friends' perspectives to married partners' views of each other. *Journal of Experimental Social Psychology*, 36(6), 600–620.
- Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96(4), 762–773.
- Paulhus, D. L., & Reynolds, S. (1995). Enhancing target variance in personality impressions: Highlighting the person in person perception. *Journal of Personality and Social Psychology*, 69(6), 1233–1242.
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). London: The Guilford Press.
- Paunonen, S. V., & O'Neill, T. A. (2010). Self-reports, peer ratings and construct validity. *European Journal of Personality*, 24(3), 189–206.
- Permack, D. R. (2011). *The role of emotional intelligence in the judgmental accuracy of neuroticism*. Unpublished doctoral dissertation, The University of Guelph.
- Ready, R. E., Clark, L. A., Watson, D., & Westerhouse, K. (2000). Self- and peer-reported personality: Agreement, trait ratableity, and the "self-based heuristic". *Journal of Research in Personality*, 34(2), 208–224.
- Redzanowski, U., & Glück, J. (2013). Who knows who is wise? Self and peer ratings of wisdom. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(3), 391–394.
- Rudman, L. A., & Mescher, K. (2012). Of animals and objects: Men's implicit dehumanization of women and likelihood of sexual aggression. *Personality and Social Psychology Bulletin*, 38(6), 734–746.
- Schneider, L., & Schimmack, U. (2009). Self-informant agreement in well-being ratings: A meta-analysis. *Social Indicators Research*, 94(3), 363–376.
- Simmonds, A. M. (2010). *Assumed similarity in alcohol consumption and relationship satisfaction in dating and married couples*. Unpublished doctoral dissertation, The University of North Carolina.
- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, 98(3), 520–534.
- Starzyk, K. B., Holder, R. R., Fabrigar, L. R., & MacDonald, T. K. (2006). The personal acquaintance measure: A tool for appraising one's acquaintance with any person. *Journal of Personality and Social Psychology*, 90(5),

- 833–847.
- Staudinger U. M., & Glück J. (2011). Psychological wisdom research: Commonalities and differences in a growing field. *Annual Review of Psychology*, 62, 215–241.
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, 40(5), 472–481.
- Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281–300.
- Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, 20(2), 104–108.
- Wang, Z., Chen, N., Petrill, S. A., & Deater-Deckard, K. (2013). Observed personality in childhood: Psychometric and behavioural genetic evidence of two broad personality factors. *European Journal of Personality*, 27(1), 96–105.
- Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, 78(3), 546–558.
- Watson, D., & Humrichouse, J. (2006). Personality development in emerging adulthood: Integrating evidence from self-ratings and spouse ratings. *Journal of Personality and Social Psychology*, 91(5), 959–974.

Observer-rating: An Efficient Method in Assessing People's Personality

ZHANG Denghao¹; TENG Fei²; PAN Xue¹

(¹ Department of Psychology, Renmin University of China, Beijing 100872, China)

(² Department of Psychology, Hongkong University, Hongkong, China)

Abstract: Observer-rating is efficient in assessing people's personality, and even more efficient than self-rating in certain circumstances. Based on realistic accuracy model (RAM) and weighted-average model (WAM), the focal observation summarized factors that might influence the accuracy of observer's rating such as familiarity, quality of relationship with the target, the visibility of a trait, and assumed similarity. Moreover, we suggested several ways that might help in solving the problems of the current investigations. For instance, sampling in a wide range, enhancing the ecological validity in different circumstances and adopting multiple strategies to assess personality should be considered in future research.

Key words: observer-rating; personality assessment; realistic accuracy model; assumed similarity