

中国高中学生中的反应风格及其效应*

郭庆科^{1,2} 王昭¹ 韩丹¹ 时勤²

(¹辽宁师范大学心理系,大连 116029) (²中国科学院心理研究所,北京 100101)

摘要 通过测试正反向陈述的 NEO-FF 和 EPQ,探讨了中国高中生中默认、极端化、折中化、弹性反应风格的特点,及题目陈述方向的改变对其人格测验信效度的影响。结果发现反应风格在中国高中生中确实存在,折中化和弹性风格对测验的影响最为严重,其次为极端化风格,默认风格则可能不算一种偏差。NEO-FF 量表在使用反向陈述题目时信效度下降,说明由于教育水平低而使高中生理解反向题时存在困难。

关键词 反应风格,默认,极端化,折中化,弹性,题目正反向陈述。

分类号 B841

1 前言

反应偏差(response bias)是受试者回答问卷时不以题目内容为依据的一种系统行为倾向^[1-3]。反应偏差通常包括反应定势(response set)和反应风格(response style)两类。反应定势与题目的内容有关,指反应者总是有意无意地按某种系统化的方式或模式而不是根据自己的真实情况对题目内容做出反应的倾向,主要有掩饰、装坏、伪装等,其目的在于塑造特定的自我形象。反应风格则与题目内容相对无关,即被试不管题目的内容是什么,而总是偏向于特定的反应方式,主要有默认反应风格(Acquiescence response style,简称ARS,指总是回答“同意”或“是”的倾向)、极端化反应风格(Extremeness Response Style,ERS,指喜欢选择比自己的实际情况更极端的答案的倾向)、折中化反应风格(Indifference Response Style,IRS,指总是选择中性答案的倾向)、弹性反应风格(Flexibility Response Style,FRS,指选择带有倾向性的,但不极端的答案的倾向)等^[4-6]。反应风格的存在使测验结果不能体现所要测量的内容,降低了测验的信度和效度。

反应风格被认为代表了某种稳定的个人特点,与年龄、教育水平等有关,而且存在系统的团体差异^[5,6]。如Khalil发现在美国、埃及和沙特3国中,沙特被试的极端化反应,包括极端肯定和极端否定反应,是3国中最高的,其次是埃及,美国最少,弹性

反应则是美国最多沙特最少。在美国被试中没有发现反应风格与NEO-P的相关,而在沙特和埃及人中则有部分相关^[6]。Bachman和O'Mally发现黑人的极端反应多于英裔白人^[7]。

对折中化风格的研究主要是针对东亚和东南亚地区,如日本、中国的香港和台湾、印尼等被认为有选择中间答案的倾向,即有更强的折中化风格^[8,9],相应地,极端化反应就比西方被试少。Hurley在Michigan州立大学生中发现折中化、弹性和极端化反应风格与人格测验相关很低^[10]。但在1998年他又发现极端否定反应(“十分不同意”)与开放性存在显著正相关,而与神经质存在显著负相关,弹性肯定反应(“同意”)与神经质有显著正相关,而与开放性有显著负相关^[11]。Shiomi和Loo比较了加拿大和日本样本,发现折中化风格对测验信度和效度有一定影响,但不大^[12]。

默认反应风格对人格测验的影响学术界意见不一,Couch与Keniston认为默认的反应风格本身就是一种有用的人格变量(引自Weems等2001)^[13],也有人认为默认反应风格对测验的影响不可低估。检测默认风格的一种有效方法是计算同一量表中正向题与反向题的相关rPNS,由于正向题与反向题是同质的,因此应该有高的负相关。早期研究者Rorer发现使同一量表中的rPNS(甚至同一题目的正向陈述与反向陈述之间)没有相关,后来的研究则发现二者相关较高^[14,15]。Ray在日本、美国和新西兰人

收稿日期:2005-10-20

*国家自然科学基金资助项目(70471060)。

通讯作者:郭庆科, E-mail: guo_q_k@sina.com

中研究发现日本人中 rPNS较低而新西兰人 rPNS最高,但新西兰样本的效度并不高,而日本样本的效度并不低^[15]。

减少默认风格的一个方法是使量表中反向陈述题目数与正向陈述题目数平衡。但有人认为正向陈述题目和反向陈述题目测量的不是同一种特质,在同一量表中使用正向和反向陈述题目将导致因素结构复杂化,可能引入了更大的反应偏差^[16,17]。但也有研究支持反向陈述题目^[18,19]。Marsh认为教育水平是导致反向陈述题目质量不同的调节变量^[20]。

近年来随着对人格测量研究的增加,反应风格(偏差)问题重又受到国外学者的重视。中国心理学界早已注意到反应风格的存在,并认识到中国被试有不同于西方的特点,但相应的实证研究还较少。本研究选择 EPQ 和 NEO - FFI这两个有代表性的人格量表,探讨中国学生中反应风格的特点,以及反应风格与题目的正反向陈述对人格测验的影响。进而探讨如何采取措施提高人格测量的信度和效度。

2 研究方法

2.1 被试

FFI的被试是辽宁朝阳市一普通高中的 486 名高一学生,其中男 220 人,女 266 人,平均年龄 16.3 岁。EPQ 的被试是这所学校的高二学生 465 人,其中男 232 人,女 233 人,平均年龄 17.2 岁。

2.2 工具与施测

本研究的人格测量工具为五因素问卷 NEO - FFI和艾森克人格问卷 (EPQ)。NEO - FFI(NEO Five Factor Inventory,以下简称 FFI)是大五人格测验 NEO - PI的简化版,特点是简明而有效。也测量神经质 N、外倾性 E、经验开放性 O、宜人性 A 和认真性 C。FFI共 60 个项目,每个维度 12 题,采用 5 级评分^[22]。FFI已在中国得到应用^[21]。中国修订的成人版 EPQ 有 3 个人格量表(即内 - 外向量表 E、情绪性量表 N、精神质量量表 P),和一个说谎量表 L,共 85 个“是”“否”记分题目。其中 E 量表 21 题, P 量表 20 题, N 量表 24 题。

研究中将 EPQ 中的 L 量表删除,然后在把 FFI 和 EPQ 中题号是偶数的题进行语义反转(原来是肯定陈述句的改为否定陈述句,原来是否定陈述句的改为肯定陈述句),而奇数的题不变,形成的量表用于初测。然后把两问卷中题号是奇数的题语义反转而偶数题不变形成重测量表。

施测时告诉被试者要进行两次性格测试,以帮

助他们了解自己的性格。初测后半个月再进行重测。这样参加 EPQ 测试的每一被试都做了两次 EPQ,参加 FF 测试的人也做了两次 FF。两次测试的题目陈述正好相反。统计结果时将所有反向记分题目的得分反转过来。然后将初测中的正向陈述题与重测中的正向陈述题合并,产生正向陈述的 EPQ 和 FF,再将初测中的反向陈述题与重测中的反向陈述题合并,得全部为反向陈述的 EPQ 和 FF。这样我们共得到同一批 486 名被试做的正反两份 FF 和同一批 363 名被试做的正反两份 EPQ。正向陈述与反向陈述问卷题目完全一样,但陈述方向相反。

2.3 FFI中反应风格指标的确定

2.3.1 默认反应风格的统计 FFI和 EPQ 反应风格指标的统计都是在未根据记分键转换的数据中进行的。FFI中默认反应风格的统计有三种方法。第一种统计方法是将每题最初录入的数字相加。在 FFI中每题的 5 个选择项是“十分不同意”、“不同意”、“没意见”、“同意”、“十分同意”,最初录入的数字分别是 1、2、3、4、5。被试做了 60 对题目(一对题目指一个题分别以正向和反向陈述的形式呈现),如果他的回答非常客观,则他的任一对题目两次录入的数字之和都是 6。比如一个人在正向陈述时对一个题答“同意”(录入数字 4),他在反向陈述时就应答“不同意”(录入数字 2),其他情况亦然。如果一个人在一对题目上的得分和大于 6,就认为他更喜欢表达赞成的意见,即存在趋同风格,得分小于 6 则认为他更倾向表达反对意见,即存在否认风格。将每对题最初录入的数字相加,即可认为是默认风格的一个指标。记为“ARS1”。ARS1 的期望值为 360。

第二种方法是数出在每对正反向题上均回答“同意”或“十分同意”(即在正向题目录入的是 4 或 5,在反向题目录入的也是 4 或 5)的个数,作为趋同风格的指标,记为“ARS2”。再统计出在每对正反向题上均回答“不同意”或“十分不同意”(正反向题录入的都是 1 或 2)的个数,作为否认风格的指标,记为“ARS3”。

第三种方法是直接数出在所有题目中回答“同意”或“十分同意”的个数,作为趋同风格的指标,记为“ARS4”。再数出在所有题中回答“不同意”或“十分不同意”的个数,作为否认风格的指标,记为“ARS5”。

2.3.2 折中化反应风格的统计 统计时有两种方法。第一种方法是数出在一对正反向题上均回答

“没意见”(在数据中录入的都是 3)的个数。记为“RS1”。第二种方法是直接数出在所有题中回答“没意见”的总次数,记为“RS2”。

2.3.3 弹性反应风格的统计 第一种统计方法是数出在一对正反向题上均回答“同意”的次数,记为“FRS1”。再数出在一对正反向题上均回答“不同意”的次数,记为“FRS2”。第二种方法是求 FRS1 与 FRS2之和,记为“FRS3”,第三种方法是直接数出回答“同意”和“不同意”的所有次数,记为“FRS4”。第一、二种方法(前 3个指标)考虑了正向与反向题的对应,第三种方法是直接只数出总次数。

2.3.4 极端化反应风格的统计 统计方法与弹性风格相同。第一种方法是数出在一对正反向题上均回答“十分同意”的次数,记为“ERS1”。再数出均回答“十分不同意”的次数,记为“ERS2”。第二种方法是求 ERS1 与 ERS2之和,记为“ERS3”,第三种方法是直接数出在所有题目中回答“十分同意”或“十分不同意”的次数,记为“ERS4”。

2.4 EPQ中反应风格指标的确定

EPQ仅有“否”和“是”两种选择项,不存在折中化和极端化的反应,因此只能统计默认的反应风格。具体也有三种方法,也是在未根据记分键转换的数据中进行的。

第一种统计方法是将每题最初录入的数字相加。在 EPQ中每题有“是”和“否”两个选项,最初录入的数字分别是 1和 2。被试如果回答非常客观,则他的任一对题目两次录入的数字之和都是 3。如果一个人在一对题目上得分之和为 2,说明他两次都选择了“是”,如果他在在一对题目上得分之和为 4,说明他两次都选择了“否”。将 65对题目的得分相加,即得到默认风格的一个指标。记为“ARS1”。ARS1的期望值为 195。

第二种方法是数出在每个题目对的正反向题上均回答“是”的个数,作为趋同风格的指标,记为“ARS2”。再统计出在每对正反向题上均回答“否”的个数,作为否认风格的指标,记为“ARS3”。

第三种方法是直接数出在所有题目上回答“是”的个数,作为趋同风格的指标,记为“ARS4”。再数出在所有题中回答“否”的个数,作为否认风格的指标,记为“ARS5”。

从上述 FFI和 EPQ 中反应风格的统计方法可以看出,由于同时使用了正反向量表(即正向与反向陈述题各占一半),无论被试的人格特质水平是高还是低,其期望应答模式都是“肯定”和“否定”两

种相对立的回答数相同且各占一半。换句话说,被试的“肯定”和“否定”反应数量并不影响其测验得分,即各反应风格指标是独立于测验内容(即不受被试人格特质水平影响)的。

3 结果

3.1 反应风格指标间的关系

每一类 FFI反应风格指标都有多个,解释起来不方便,而且容易混淆。比如弹性风格中由于统计了 2(不同意)和 4(同意)的个数,所以很可能也体现了默认风格,极端化风格中由于统计了 1和 5,也可能和默认风格有关。因此我们对所有反应风格指标进行了因素分析。因素分析抽出特征根大于 1的因素 4个,经 OBL M N斜交旋转(我们不知道各反应风格间是否有相关)后各指标的因子载荷见表 1。表中的因子结构与预期基本一致。

但 ARS2、ARS3与弹性风格的 4个指标在同一因子上有高载荷,说明它们之间有共享方差或存在高相关,对数据进一步分析发现其高相关是由于 ARS2、ARS3中统计了“同意”和“不同意”两个选择项造成的,因此 ARS2、ARS3就不是纯粹的默认风格测量指标,二者从研究中删除。综合分析表明表 1中的 4个因子分别代表了弹性 FRS、极端化 ERS、折中 RS、默认 ARS四种反应风格。另外,FRS4在因子 2(ERS)上有反方向的载荷,这是因为 ERS多时 FRS就少,其他的反方向载荷也是由相同原因引起的。

表 1 各反应风格指标的因素分析

反应风格	因子 1	因子 2	因子 3	因子 4
ARS1			0.218	0.964
ARS2	0.881	-0.456	0.157	0.227
ARS3	0.840	-0.494		-0.447
ARS4	0.151		0.768	0.647
ARS5	0.165	-0.177	0.487	-0.825
RS1	0.382	-0.150	-0.857	
RS2	-0.237	0.176	-0.963	
FRS1	0.931			0.0195
FRS2	0.902			
FRS3	0.989	-0.123		-0.379
FRS4	0.367	0.716	0.410	
ERS1	0.440	-0.864		-0.293
ERS2	0.335	-0.903		
ERS3	0.413	-0.933		-0.145
ERS4	-0.190	-0.839	0.300	

注:绝对值低于 0.10的载荷没有显示

表 1 中 ERS1 代表的是极端否定反应, ERS2 则代表极端肯定反应, 二者却存在高相关 (0.787), 说明极端化反应风格是存在的, 与国外不同的是, 在中国被试中似乎不需要区分这两种极端化反应。我们还计算了弹性肯定风格 FRS1, 弹性否定风格 FRS2 的相关, 结果二者相关为 0.717, 这也说明弹性反应风格的存在, 而且没必要象国外一样区分弹性肯定和弹性否定两种风格。

国外文献中区分了赞成趋同和否定趋同两类默认风格, 从前文介绍的默认风格统计方法来看, 似乎 ARS1 低表示否定趋同, ARS1 高表示赞成趋同, 但分析 ARS4 和 ARS5, 却发现二者的相关只有 -0.128, 说明两种默认风格可能并不是同一维度的两极, 因此 ARS4 和 ARS5 的含义不能用 ARS1 来代表。因此我们保留 ARS4 和 ARS5, 分别作为赞成和否定趋同的指标。

以上结果表明对反应风格的分类和测量是合理的, 也为我们选择适当的风格指标进行分析提供了依据。如果同一类指标能互相代替, 则选择因子载荷最大的。根据这一原则, 我们最后选择的指标是 ARS4、ARS5、RS2、FRS3、ERS3。ERS3 与 ARS1、

ARS4、ARS5 的相关都低于 0.30, 说明这一极端化风格指标与默认风格指标是相互独立的。

各类指标中 ARS4、ARS5、RS2、FRS4、ERS4 是直接数出来的, 相关分析表明它们与同类的其他指标间相关很高, 可以相互代替。这一点有很重要的意义, 如果人格和态度调查表中正反向题平衡, 就可以采用本研究中的方法统计出各类反应风格。

EPQ 中有 5 个默认风格指标, 对这 5 个指标进行因素分析发现一个特征根大于 1 的因素, 这 5 个指标的载荷分别为 0.996, -0.805, 0.761, 1.00, 1.00, 我们选择 ARS4 作为分析指标, ARS4 与 ARS1 和 ARS5 的相关等于或接近 -1, 三者可以相互代替。

3.2 反应风格对测验信度的影响

3.2.1 反应风格高、低组测验信度的对比 研究中计算出 ARS4、ARS5、RS2、FRS3、ERS3 的平均数, 将反应风格得分大于平均数的定为高分组, 小于平均数的定为低分组 (高分组人数与低分组人数基本相同), 分别计算两组的 α 系数, 结果见表 2。表中列出了正反两份 FFI 的信度, 使信度比较更加全面, 下文也是如此。

表 2 反应风格高组与低组 FFI 各量表的信度

反应风格	正向陈述 FFI					反向陈述 FFI				
	N	E	O	A	C	N	E	O	A	C
ARS4 高	0.668	0.507	0.680	0.774	0.594	0.593	0.340	0.597	0.525	0.619
ARS4 低	0.619	0.419	0.690	0.791	0.627	0.530	0.140	0.524	0.432	0.545
ARS5 高	0.606	0.512	0.701	0.786	0.680	0.603	0.294	0.482	0.492	0.648
ARS5 低	0.618	0.498	0.674	0.774	0.578	0.545	0.226	0.590	0.508	0.450
RS2 高	0.613	0.477	0.688	0.785	0.597	0.499	0.273	0.580	0.462	0.459
RS2 低	0.669	0.526	0.698	0.776	0.654	0.616	0.301	0.567	0.486	0.675
FRS3 高	0.556	0.364	0.654	0.768	0.546	0.480	0.129	0.434	0.432	0.501
FRS3 低	0.696	0.541	0.708	0.747	0.696	0.636	0.348	0.655	0.520	0.670
ERS3 高	0.577	0.555	0.689	0.754	0.613	0.513	0.124	0.615	0.470	0.605
ERS3 低	0.713	0.487	0.681	0.735	0.662	0.638	0.421	0.524	0.467	0.617

注: 带“ ”标记的数字是反应风格高组与低组对比中较大的数字, 下同

从表 2 中可以明显看出 RS 和 FRS 高组测验的信度低, 原因是选择中间选项过多时测验分数的分布范围缩小, 导致误差方差占的比重增大。选择中间答案是亚洲人的特点, 但这样会导致题目的区分度下降, 使测验的信度和效度降低。本研究很好地说明了这一问题。因此中国被试者中可以考虑不设中性选项。

在正反向陈述 FFI 的 10 个分量表中, 默认风格 ARS4 和 ARS5 高组都有 7 个分量表信度高于低组,

有 3 个低于低组, 可见高组的信度更高。这说明默认风格在中国被试中可能并不是真正的“偏差”。而极端化风格 ERS3 高组有 5 个分量表信度高于低组 (平均为 0.552), 而低组也有 5 个分量表的信度高于 ERS3 高组 (平均为 0.595), 说明极端化风格对信度的影响并不大。

在 EPQ 中也以默认风格的平均数为界将学生分为高分组和低分组, 结果发现两组中默认风格高分组的信度反而更高 (见表 3), 这与 FFI 中的结论

一致,即默认风格不影响测验信度。

表 3 反应风格高组与低组 EPQ 各量表的信度

反应风格	正向陈述 EPQ			反向陈述 EPQ		
	E	P	N	E	P	N
默认风格高组	0.681	0.478	0.773	0.683	0.505	0.708
默认风格低组	0.776	0.510	0.761	0.639	0.394	0.515

3.3 反应风格高低组的正向与反向量表之相关

研究中我们有一个题目全为正向陈述的 FFI 和一个题目全为反向陈述的 FFI,两个 FFI 可视为两个特殊的副本,也可视为两个测验,或同一量表的不同形式的题目。这样两个 FFI 总分之间的相关可视为副本信度,在一定意义上可视为相容效度,也可以认为体现了题目的同质性,而同质性被认为是构念效度的一种指标。

因此我们可以推论,如果反应风格影响测验效度,则反应风格高组两个 FFI 之间的相关低,而反应风格低组两个 FFI 的相关高。反应风格高低分组 FFI 中各正反向分量表的相关系数见表 4。

表 4 反应风格高低组 FFI 各分量表的效度对比

反应风格	N	E	O	A	C
ACQ4 高组	0.579	0.345	0.370	0.047	0.496
ACQ4 低组	0.389	-0.009	0.026	-0.227	0.372
ACQ5 高组	0.510	0.244	0.216	-0.102	0.461
ACQ5 低组	0.582	0.224	0.256	-0.013	0.350
RS2 高组	0.466	-0.018	0.075	-0.190	0.341
RS2 低组	0.475	0.299	0.285	0.010	0.378
FRS3 高组	0.201	-0.180	-0.026	-0.291	0.056
FRS3 低组	0.646	0.394	0.301	0.147	0.579
ERS3 高组	0.353	0.051	0.100	-0.166	0.261
ERS3 低组	0.593	0.419	0.314	0.177	0.512

从表 4 中可见,FFI 中趋同默认风格 (ARS4) 高分组的相关反而大于低分组,而否定默认风格 (ARS5) 高分组的相关与低分组差不多,其他风格均是低分组的相关大于高分组。这表明折中化、弹性和极端化风格降低了正反向量表之相关,是真正的偏差。而默认风格则没有降低正反向量表之相关,这似乎也说明中国被试中“默认风格”不是偏差。但趋同默认和否定默认的作用并不一致,趋同默认还提高了正反向量表之相关,其原因尚需进一步探讨。

EPQ 中默认风格高分组 E、P、N 三分量表正反向陈述量表得分的相关为 0.722, 0.621, 0.634, 低分组的相关则为 0.731, 0.580, 0.694。两组差异不大。

虽然正反向量表之相关不是效度的直接证据,但我们可以推论,即折中化、弹性和极端化风格可能会降低测验的效度,而默认风格则不影响测验的效度。

3.4 反应风格与人格量表的相关

通常认为反应风格是一种较稳定的个体特征,但这种特征是否是一种人格特征?国外的研究尚未定论。为研究这一问题,我们计算各反应风格指标与 FFI 的相关 (见表 5)。从表中可见反应风格与 FFI 的相关有不少是显著的,但都是低相关,默认风格与 EPQ 中 E、P、N 三个量表的相关也不高 (分别是 -0.017, 0.140, 0.140)。说明我们还不能在反应风格和人格间建立联系。

表 5 反应风格与 FFI 各分量表的相关

反应风格	N	E	O	A	C
ACQ4	-0.055	0.130*	0.159*	0.107*	0.122*
ACQ5	-0.108*	0.013	-0.024	0.059	0.117*
RS2	0.119*	-0.115*	-0.114*	-0.128*	-0.180*
FRS3	0.124*	-0.187*	-0.310*	-0.229*	-0.135*
ERS3	0.070	-0.151*	-0.052	-0.177*	-0.083

注: * $p < 0.05$

3.5 正向与反向陈述条件下测验的信度和效度

克服默认风格的一个方法是使量表中正向与反向题平衡。如果被试在正向陈述与反向陈述题目上表现一致,则认为没有默认风格。但正向题与反向题上表现不一致也可能是其他因素造成的,有的研究者称这些因素为“正向陈述效应”和“反向陈述效应”,被试者受教育程度的不同也是导致在正反向题目上反应不一的原因。我们在大学生中发现正反向陈述题目的心理测量学性能非常一致,本研究采用相同的工具和方法将样本扩大到高中生,看是否能得出相同的结论。

表 6 正向与反向陈述 FFI 和 EPQ 的信度

陈述方向	NEO - FFI					EPQ		
	N	E	O	A	C	E	P	N
正向陈述	0.651	0.514	0.698	0.779	0.639	0.747	0.497	0.790
反向陈述	0.581	0.450	0.571	0.488	0.601	0.667	0.445	0.634

正向与反向陈述的 FFI 和 EPQ 的 alpha 系数见表 6。可以看出 FFI 和 EPQ 在正向陈述时的信度都高于反向陈述,而在大学生中正向与反向陈述则差异不大。这说明正向陈述情况下测验的信度更好。

我们将正向与反向陈述的 FFI 和 EPQ 分别进行验证性因素分析 (CFA),模型的拟合度见表 7。

表 7 正向与反向陈述 FFI和 EPQ的模型拟合度

模 型	χ^2	df	χ^2/df	RMSEA	NNFI	CFI	SRMR
原 FFI	4151	1700	2.44	0.054	0.82	0.82	0.072
正向陈述 FFI	4338.65	1700	2.551	0.065	0.86	0.87	0.086
反向陈述 FFI	4247.08	1700	2.498	0.062	0.65	0.66	0.080
正反向合并 FFI	20008.96	7010	2.854	0.072	0.73	0.73	0.094
正向陈述 EPQ	2563.69	2012	1.274	0.029	0.90	0.90	0.055
反向陈述 EPQ	2067.23	2012	1.027	0.012	0.97	0.97	0.050
正反向合并 EPQ	9625.67	8252	1.166	0.026	0.91	0.91	0.055

从表 7 中可见反向陈述 FFI在 χ^2 、RMSEA、SRMR 三个指标上略优于正向陈述 FFI,但 NNFI和 CFI太差,因此可以认为正向陈述的 FFI有更好的模型拟合度(删除几个载荷低的题目后正向 FFI的两个拟合指数 NNFI和 CFI都能达到 0.90 以上,而反向陈述量表则仍然较低)。这说明 FFI改为反向陈述后构念效度变差了,潜在的因素结构发生了变化。为对此进一步检验,我们将正向 FFI与反向 FFI合并在一起进行 CFA(其中 60 个题为正向陈述,60 个题为反向陈述,在模型中仍设 5 个因素,每一个因素由相应的 12 个正向题和 12 个反向题测量),结果模型拟合度更差,有一半以上的题目在正向陈述时载荷较高,而反向陈述时载荷变得接近 0 甚至变成负值,这说明题目改为反向陈述后所测量的特质发生了变化。

但正向反陈述 EPQ 量表的拟合度都可以接受,相对而言反向陈述拟合度反而更好。将正向与反向陈述 EPQ 放在同一个验证性因素分析模型中,模型也是拟合的,而且正反向题的载荷方向也多数一致。这说明 EPQ 的正反向题目可以认为是测量相同特质的。

造成 FFI和 EPQ 结果不一致的原因可能有以下几个:一是被试在测试 EPQ 时只需对每题做出“是”或“否”的判断,且选项“是”在前“否”在后,而 FFI则需要 5 个等级间做出区分,后者更为复杂,且选项“十分不同意”在前而“十分同意”在后;二是 EPQ 的被试是高二学生而 FFI的被试是高一学生,高一学生在理解反向陈述题目时可能存在认知困难。

与此相一致的另一结果是,EPQ 中正反陈述的 E、P、N 三量表的相关为 0.696, 0.599, 0.589。FFI 中正反陈述的 N、E、O、A、C 五个分量表的相关为 0.475, 0.177, 0.204, -0.063, 0.372。这说明 EPQ 量表中正反题目可以认为测量了相同的特质,而 FFI中则不是。

4 讨论

4.1 中国高中生反应风格的特点

从上述结果中可以看出中国高中生中也存在折中化、弹性、和极端化反应风格,折中化和弹性反应风格降低了测验的信度,很可能也降低了测验的效度,极端化反应风格虽没有降低测验的信度,但很可能降低了测验的效度。就对测验的影响程度而言,折中化和弹性风格要比极端化风格更严重。这与中国被试的特点有关,受传统文化的影响,中国被试倾向于用不极端的方式表达自己的观点。这种行为在测验中会导致其所选答案与真实情况不符,导致题目的区分度和方差下降,从而降低了测验的信度和效度。因此为提高测验的信效度可以考虑不设中性选项,鼓励被试做出倾向性更强的选答。

国外研究表明极端化风格不影响测验信度而影响测验效度,本研究也支持这一结论。由于极端化风格是一种稳定的行为倾向,这种倾向又会使测验分数变异增大,因此不会使测验信度降低。但这种个体倾向却干扰了对被试人格特质的测量,因而降低了测验效度。

本研究中使用的默认风格指标没有影响测验的信度和效度。说明中国被试中可能并不存在“默认风格”。对默认风格的第一个指标 ASR1 所做的描述统计发现,FFI中 ARS1 的实际均值为 355(标准差为 21),低于期望均值 360,EPQ 中 ARS1 的实际均值为 196.6,略高于期望均值 195,说明中国被试从总体上讲并不存在趋同默认,否定默认也不明显。之所以会出现不少人在题目的正反陈述上都选了“同意”或“十分同意”的情况,是因为测验中的不认真行为造成的,不代表一种系统的反应倾向,即“默认风格”。

4.2 反应风格与人格特质的相关

反应风格被认为是稳定的个人特点,对该特点的分析很有意义。国外有的研究发现反应风格与人

格特质间存在相关,有的则没有。本研究则没有发现二者间存在实质性的联系。但这不是说反应风格与人格间的关系不存在,如借助其他的方法,如行为观察,他人评定等,可能会发现更有意义的结论。

4.3 高中生中题目正反向陈述的效应

我们在大学和高中生中研究了题目肯定陈述和否定陈述的效应(肯定陈述题大多数是正向记分的,否定陈述题则多数是反向记分)。在大学生中发现题目改为反向陈述后所测量的特质不变,而且有相同的因子结构。而在高中生中则发现了不一致的结果。在FFI所测试的高一学生样本中,FFI量表的题目如果陈述方向改变,则所测特质也改变。正向陈述的FFI量表基本能体现原先的理论结构,而反向FFI则不能。即在高一学生中反向陈述题目不如正向陈述。但EPQ的结果则与大学生相当一致,即题目的正反向陈述并不影响其心理测量学性能。

这一结果说明教育水平确实是个调节变量,高中尤其是高一学生由于理解能力的局限,对反向陈述题目可能不能正确地理解,或不易完成从五个等级中选出符合自己的答案这种认知操作。而高二学生在测试EPQ时就不存在这一问题,可能是高二学生的理解能力提高了,也可能是EPQ只要求做出“是”和“否”判断这两种简单的认知操作。因此对高中生群体似乎应少用反向陈述的题目,即使用也要使题目易于理解。

4.4 反应风格与测验设计

反应风格是影响人格测验信度和效度的重要因素,测验应如何设计才能减少反应风格的发生?针对本研究可以做以下几方面的思考:是否采用反向陈述题目?反向陈述题目一方面可以减少默认风格的影响,另一方面又增加了题目内容的饱和度,但对教育程度较低的人可能不适用。而默认风格又不影响测验的信度和效度,似乎说明不必采用反向陈述题目。但是,尽管默认不影响测量精度,但默认的人还是有的,如果只使用正向陈述题目,这些人的测验分数就不能代表其人格特点。建议还是平衡使用正反向题目,但反向题一定要易于理解。就象本研究所发现的,正反向题目数平衡的另一大好处是可以鉴别出存在反应风格的被试。

本研究只涉及将肯定句改为否定句后的效应问题,没有涉及将特质词汇由积极词改为消极词后的效应(尽管肯定句多为正向记分,而否定句多为反向记分,与使用积极词汇和消极词汇时的记分方向

一致),后者还需要进一步研究。

题目选择项的数量是奇数还是偶数,以多少选项为宜?本研究似乎表明设中性选可能会引入折中化风格,从而降低测验的信度和效度,因此中国被试似乎应将题目选择项的数量设为偶数。从逻辑上讲,题目的选择项越多则对被试的区分就越精细,但选择项太多就会增加测验任务的复杂性,对教育水平低和年龄小的被试可能不适宜,因此选择项又不宜过多,至于多少为宜尚需要进一步研究。

还有一点需要测验编制者注意的是,国外研究表明题目内容越含糊,反应风格的影响就越大。

5 结论

本研究的证据表明中国高中生在确实存在反应风格。比如在FFI中选择第二个选项的次数与选择第四个选项的次数(即FRS1与FRS2)有高相关,选择第一个选项的次数与选择第五个选项的次数(即ERS1与ERS2)也有高相关。而且反应风格确实影响测验的信度和效度。因此如果能采取有效的措施减少反应风格,将能提高人格测量的准确性。

控制反应风格的方法有多种,比如使正向记分题与反向记分题数平衡以减少默认风格,不设中性选项以避免折中化风格,不使用模糊的语义陈述以减少极端化风格,使选择项的社会期望性相匹配以消除社会期望性风格等。针对本研究的结果,我们认为还是应尽量平衡使用正反向题目,但反向题必须易于理解,题目选择项不宜太多,可能考虑不设中性选项。如果正向题与反向题平衡,就可以采用本研究中的方法直接统计出极端化、弹性、折中化等反应风格。

中国高中生的反应风格有不同于西方的特点。比如弹性和折中化风格对测验的影响最为严重,默认风格可能并不存在,极端化风格可能只有一种(不存在极端肯定风格和极端否定风格的区别),弹性风格也可能只有一种等。

反应风格与人格特征间存在相关,但相关不高。探讨人格特征与反应风格需要借助其他方法。

参 考 文 献

- 1 Cronbach, L. J. Response sets and test validity. *Educational and Psychological Measurement*, 1946, 6: 475 ~ 494
- 2 Rorer L. G. The great response style myth. *Psychological Bulletin*, 1965, 63: 129 ~ 156
- 3 Paulhus D. L., Reid D. B. Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*,

- 1991, 60: 307 ~ 317
- 4 Susan R. Acquiescence and conservatism. *Child Development*, 1973, 44: 811 ~ 814
- 5 Greenleaf E A. Measuring extreme response style. *Public Opinion Quarterly*, 1992, 56: 328 ~ 350
- 6 Khailil E A. Extremeness, flexibility, and indifference response sets: a cross-cultural study. the Annual Conference of the American psychological Association, 2001
- 7 Bachman J G, OMalley P M. Yea-saying, nay-saying, and going to extreme: black-white difference in response style. *Public Opinion Quarterly*, 1984, 48: 491 ~ 509
- 8 Stening B W, Evertt J E. Response styles in a cross-cultural managerial study. *The Journal of Social Psychology*, 1984, 122: 151 ~ 156
- 9 Chen C, Lee S Y, Stevenson H W. Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 1995, 6: 170 ~ 175
- 10 Hurley J R. Constructive thinking and firm disagreement versus neuroticism and mild agreement: Asymmetric correlations of content with response style. *The Journal of Psychology*, 1996, 132: 461 ~ 475
- 11 Hurley J R. Timidity as a response style to psychological questionnaires. *The Journal of Psychology*, 1998, 132(2): 201 ~ 210
- 12 Kunió Shiomí, Robert Ioo. Cross-cultural response styles on the Kirton adaptation innovation inventory. *Social Behavior and personality*, 1999, 27: 413 ~ 419
- 13 Weems G H, Onwuegbuzie A J. The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development* 2001, 34 (3): 166 ~ 176
- 14 Ray J J. Reviving the problem of acquiescent response bias. *Journal of Social Psychology* 1983, 121: 81 ~ 96
- 15 Ray J J. Acquiescent response bias as a recurrent psychometric disease: Conservatism in Japan, the U. S. A. and New Zealand. *Psychologische Beitrage*, 1985, 27: 113 ~ 119
- 16 Mook J, Kleijn W C, Ploeg H M. Positively and negatively worded items in a self-report measure of dispositional optimism. *Psychological Reports*, 1993, 71: 275 ~ 278
- 17 Roberts R E, Lewinsohn P M, Seeley J R. A brief measure of loneliness suitable for use with adolescents. *Psychological Reports*, 1993, 72: 1379 ~ 1391
- 18 Bergstrom B A, Lunz M E. Rating scale analysis: Gauging the impact of positively and negatively worded items. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA (ERIC Document Reproduction Service No. ED 423 289), 1998. 1 ~ 20
- 19 Gana K, Alaphilippe D, Bailly N. Factorial structure of the french version of the rosenberg self-esteem scale among the elderly. *International Journal of Testing*, 2005, 5(2): 169 ~ 172
- 20 Marsh H W. Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, 1996, 70(4): 810 ~ 819
- 21 Costa P T, McCrae R R. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual. Odessa, FL: Psychological Assessment Resources, 1992
- 22 XU S L, Wu Z P, Wu Z Y et al. Age Differences of psychological well-being of Chinese adults. *Chinese Journal of Psychological Health*, 2003, 17(3): 147 ~ 151
(许淑莲,吴志平,吴振云等. 成年人心理幸福感的年龄差异研究. *中国心理卫生杂志*, 2003, 17(3): 147 ~ 151)

Response Styles of Chinese Middle School Students and Their Effects on Personality Tests

Guo Qingke^{1,2}, Wang Zhao¹, Han Dan¹, Shi Kan²

(¹Department of Psychology, Liaoning Normal University, Dalian 116029, China) (²Institute of Psychology, Chinese Academy of Science, Beijing 100101, China)

Abstract

Response style is an important form of bias and error variance in psychological testing. But there is no study explored the characteristics of response styles in Chinese people, and to which extent they affect reliability and validity of psychological testing. In this study, one form of NEO-FFI with 60 positive narrated items was administered at one time to 486 middle school students. A negative form of the measure was administered to the same student group two weeks later. In addition, a positive and negative narrative form of EPQ was also administered to 465 middle school students in two weeks interval. Four kinds of response styles are calculated: acquiescence, extremeness, indifference and flexibility.

The several results were found. First, response styles do exist in Chinese students. Second, Response styles that affect reliability and validity of personality tests most seriously are indifference and flexibility response styles. Extremeness response style reduces test reliability and validity, but acquiescence style does not. Third, The negative form of NEO-FFI has lower in reliability and validity than the positive form, suggesting that middle school students may have difficulties understanding negative narrated items. These results indicate that response styles of Chinese students are different from those of western countries. Whereas indifference and flexibility are most important in Chinese students, extremeness is less important. Acquiescence as a response style does not exist.

Key words response style, acquiescence, extremeness, indifference, flexibility.