心 理 学 报 2006, 38 (4): 626～632
Acta Psychologica Sinica

人格测验中题目正反向陈述的效应 *

郭庆科[1,2] 韩 丹[1] 王 昭[1] 时 勘[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 研究中将 EPQ和 NEO - FF中的一半题目进行语义反转 (肯定陈述的改为否定陈述,否定陈述的改为肯定陈述),而另一半的题目陈述方向不变,形成的量表用于初测。重测时则把初测问卷中的题目全部进行语义反转。测试后得到 363名大学生的正反 EPQ和 412名大学生的正反 NEO - FF数据。对比发现两套量表在使用正向题时信度略高,而使用反向陈述题目时效度略好;验证性因素分析发现正反向题能测量同一特质。正向题可能更易受反应偏差影响。

关键词 正向和反向陈述题目,反应偏差, EPQ, NEO - FFI。
分类号 B841

1 前言

反向陈述题目 (negatively - worded item)最初是为了减少默认反应偏差 (指不区分题目内容而随意做出同意反应的倾向)和不认真行为而在人格和态度测验中得到应用的。测量界对默认反应风格的关注由来已久, Cronbach将之称为 acquiescence, Couch与 Keniston则将喜欢正向陈述或反向陈述的倾向称为 yea-or nay-saying[1], Krosnick等将喜欢正向陈述的倾向称为 satisficing[2],并认为这类反应中投入了较少的认知努力。Nunnaly发现有些被试在 5点 Likert量表中也存在总做出肯定反应,即选择第 4和第 5等级的情况[3]。

测验中平衡综合使用正向、反向陈述题目被认为能减少反应偏差。同时也能平衡趋同偏差 (agreement bias)和否定偏差 (disagreement bias)对测验结果的影响。道理是如果只使用正向题,则存在趋同偏差的人得分会偏高,存在否定偏差的人则得分偏低。而只使用反向题时结果则相反。平衡使用正反向题则使那些存在趋同和否定偏差的人得分趋于测验分数的平均值,避免了极端高分和极端低分,而使测验结果更准确。同时,使用反向题目也有助于把那些不管题目内容而做出肯定反应的人筛选出来,以提高测量的精度。同时,由于对反向陈述题目反应时要先进行语义反转,这就使被试在对题目内容进行仔细理解后才做出反应。因此对题目内容的注意应该能减少反应偏差及默认反应风格。

但另一些测量学家对这种混合的题目形式提出了质疑。他们认为反向陈述题目和正向陈述题目不是测量同一种特质,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。如 Mook, Kleijn和 Ploeg调查乐观性这一单维双极特质时发现了双因子结构,积极词汇与乐观因素有关而消极词汇则与悲观性因素有关[4]。Roberts, Lewinsohn和 Seeley用 8个题目调查青少年的孤独感,发现 4个正向题测量的是一个因素, 4个反向题测量的是另一个因素[5]。

但同时也有研究对反向题有利,如 Bergstrom和 Lunz在中学生中的测试发现正向与反向题能测量同一特质[6], Gana等通过验证性因素分析证实了 Rosenberg自尊量表的单维结构,且没有发现反向题中有更大的方法效应[7]。

反向陈述题目的效应在不同人群中并不一致。有人发现教育水平可能是个调节变量,即反向题目在不同教育水平团体中的作用不同。如 Fried和 Ferris发现教育水平低的人对反向题目性质的认识没有受教育水平高的人好,在参加工作诊断调查 (JDS)时产生了无关因素[8]。Marsh在青少年中测试自尊量表 SES时发现了一个只与反向题目有关的独立因素,这一额外因素被称为人造因素 (arti-

收稿日期: 2005 - 06 - 15
*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

facts)。他还发现这一因素在言语能力低的人中作用较明显,在言语能力高的人中则作用降低。Marsh也认为这是教育水平的调节作用[9]。

研究者关心的另一问题是正反向题对测验信效度的影响。Weems等发现采用正反向混合的题目形式将导致人为的极端反应,同时也将降低测验的信度[10]。Sandoval和Lambert发现在教师评定中将正向题混入反向题会提高信度和效度[11]。Schriesheim和Hill在大学生被试中发现使用反向题能提高测验效度但没有提高信度[12],Williams等也发现反向题与效标的相关比正向题更高[13]。

综上所述可知学术界对题目正反向陈述的效应尚存在争议,我们仍不知道使用正向题有利还是反向题有利。对正向和反向题的争议主要集中在以下几方面:一是正向题引起的反应偏差大还是反向题引起的反应偏差大,二是使用正向陈述题目时测验的信效度高还是反向题的信效度高,三是正反向题是否测量了同一种特质。本研究通过将题目语义反转产生同一量表的一式两份复本这一独特的方法对这一问题进行探讨。希望对实际工作有参考价值。

2　研究方法

2.1　被试

本研究以辽宁师范大学物理、化学、管理、教育、海华五个学院的1～3年级本科生为被试。其中初测时NEO-FFI问卷的被试为514人;EPQ问卷的被试为491人。重测时由于被试的遗失和无效问卷的排除,最后NEO-FFI问卷的被试为412人,其中男86人,女326人,一年级151人,二年级166人,三年级95人;EPQ问卷的被试为363人,其中男37人,女323人;一年级247人,二年级116人。

2.2　工具

本研究使用了测量人格的两个经典量表,即五因素人格问卷NEO-FFI和艾森克人格问卷(EPQ)。NEO-FFI(NEO Five Factor Inventory,以下简称FFI)是大五人格问卷NEO-PI的简化版,特点是简明而有效。也测量神经质、外倾性、经验开放性、宜人性和认真性五项特质[14]。FFI共60个项目,每个维度12题,采用5级评分。FFI已被我国学者证明有较好的信度和效度[15]。我国修订的成人版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。

2.3　研究程序

2.3.1　对FFI和EPQ的修改　简易大五人格问卷FFI和EPQ问卷都根据研究的需要做了修改。即先把L量表删除,然后在把FFI和EPQ中的一半题目进行语义反转(原来是肯定陈述的改为否定陈述,原来是否定陈述的改为肯定陈述),而另一半题目不变,形成的量表用于初测。然后把两问卷中在初测时没有进行语义反转题进行语义反转,而其他题目不变,形成重测量表。

2.3.2　施测与统计分析　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FFI测试的人也做了两次FFI。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,得全部为正向陈述的EPQ和FFI,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FFI。这样我们共得到同一批412名被试做的正反两份FFI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。

3　结果

3.1　采用正反向陈述时EPQ和FFI的信度对比

进行信度分析时发现EPQ和FFI个别题目的题-总相关不高,为不影响验证性因素分析结果,研究中将这些题-总相关低的题目删除。删改后EPQ中保留55题,FFI中保留51题,以下所有分析都是在删改后的量表中进行的。用SPSS 11.0软件计算出了EPQ和FFI各分量表在没改动时,及在采用正反向陈述时的α系数(见表1和表2)。

表1　正向陈述和反向陈述EPQ的α系数

量表类别	E量表	P量表	N量表
原EPQ	0.812	0.460	0.802
正向陈述EPQ	0.796	0.421	0.748
反向陈述EPQ	0.817	0.448	0.810

表2　正向陈述和反向陈述FFI的α系数

量表类别	N量表	E量表	O量表	A量表	C量表
原FFI	0.778	0.621	0.685	0.601	0.670
正向陈述FFI	0.817	0.653	0.693	0.642	0.747
反向陈述FFI	0.766	0.618	0.681	0.642	0.668

从表 1和表 2中可看出改动过的量表信度并没有明显降低,因此下文所进行的分析是有效的。正向陈述的 FFI量表信度都高于相应的反向陈述量表,而 EPQ中则反向陈述量表信度略高。但总体上看还是正向陈述量表的信度高一些。因此可以认为正向陈述题目改为反向题目后信度可能会降低,但幅度不大。

3.2 正反向陈述时的效度对比

3.2.1 正反向陈述时 EPQ和 FFI的因素结构 因素分析是检验测验构念效度的常用方法,验证性因素分析 (Confirmatory Factor Analysis, CFA)较探索性因素有更强的理论导向性,因此在因素数量和变量与因素的关系已经确定的情况下最好采用验证性因素分析。用 LISREL 8.53软件计算出删改后得到的 55题的 EPQ和 51题的 FFI的模型拟合度,见表 3。删改后的量表与数据拟合虽不太理想 (NNFI和 CFI小于 0.9),但基本能体现艾森克人格模型和大五人格模型的理论结构 (未删改的量表拟合度更差)。在此基础上再比较正向陈述与反向陈述量表的构念效度,其结果也列于表 3中。

表 3 删改后 EPQ和 FFI的模型拟合度

量表类别	χ^2	*df*	χ^2/df	RMSEA	NNFI	CFI	SRMR
原 EPQ	2305.89	1427	1.62	0.042	0.84	0.85	0.069
原 FFI	2393.81	1214	1.97	0.052	0.84	0.85	0.069
正向陈述 EPQ	2402.96	1427	1.68	0.045	0.85	0.86	0.072
反向陈述 EPQ	2456.22	1427	1.72	0.046	0.87	0.87	0.069
正向陈述 FFI	3952.15	1214	3.25	0.075	0.79	0.80	0.086
反向陈述 FFI	2868.20	1214	2.36	0.066	0.84	0.84	0.078

表 3中的结果表明 EPQ在采用正向与反向陈述时拟合度差不多,FFI量表在采用反向陈述题目时模型拟合度更好。因此从总体上看可认为反向陈述条件下的构念效度略好于正向陈述。但正反向陈述时的模型拟合度都低于原量表,这是因为删除题目时根据的是原量表中的题 - 总相关,而不是正向或反向陈述量表中的题 - 总相关。

这也说明对某些题目而言,其语义陈述方向的改变会对测验产生一定影响。

因子载荷的大小体现了题目测量相应特质时的有效性,因子载荷的平均值也可以看作是构念效度的一个指标。计算验证性因素分析中 EPQ和 FFI各分量表题目因子载荷的平均值。发现正向陈述 EPQ三个量表因子载荷的平均值分别为 0.41, 0.25, 0.33,反向陈述 EPQ三个量表的因子载荷平均值为 0.41, 0.28, 0.35。正向陈述 FFI各分量表题目因子载荷的平均值为 0.55, 0.41, 0.39, 0.35, 0.38,反向陈述各分量表题目的载荷则分别为 0.51, 0.42, 0.40, 0.34, 0.40。其中 EPQ量表中的 P、N分量表、NEO量表中的 E、O、C分量表在反向陈述时载荷稍大些。虽然整体上说这 8个分量表正、反向陈述时的因子载荷平均值差异并不大,但结合表 3中的结果,可以发现在反向陈述时 NEO和 EPQ的理论结构得到了更好的体现。

3.2.2 正反向陈述时的效标关联效度 为进一步检验正反向陈述情境下量表的效度,我们采用同伴提名法搜集到了 NEO和 FFI的效标。具体过程是向参加测试的大学生详细讲解 EPQ和 FFI各因素的含义,然后让他们从自己班里 (测试是以班级为单位的)选出与各因素高分特征最相近的 3个人。由于组织上的困难,最后 EPQ有 3个院系 290名学生参加了提名,FFI则只有 171名学生参加提名。提名后统计出被试者被提名的次数。被提名的次数越多,说明该生越被同学认为有某因素的高分特征。

由于有效的提名少,导致被提名次数呈严重偏态,研究中将被提名的次数低于 1次者编码为 0,被提名次数在 1次或 1次以上者编码为 1 (还可以将被提名次数为 2或 3的编码为 1,但这样会使编码为 1的人太少,故研究中没有采纳)。再计算与 EPQ和 FFI各相应分量表得分的点二列相关。结果见表 4和表 5。

表 4 提名次数与 EPQ的点二列相关 (N = 290)

量表类别	E量表	P量表	N量表
正向陈述 EPQ	0.296**	0.146*	0.069
反向陈述 EPQ	0.293**	0.223**	0.069
编码为 1的人数	119	59	85

注 表中 *表示 $p < 0.05$, **表示 $p < 0.01$,下同

表 5 提名次数与 FFI的点二列相关 (N = 171)

量表类别	N量表	E量表	O量表	A量表	C量表
正向陈述 FFI	0.087	0.323**	0.166*	0.096	0.114
反向陈述 FFI	0.159*	0.365**	0.179*	0.071	0.089
编码为 1的人数	56	57	88	120	99

根据表 4、表 5,发现正向陈述时有 4个分量表、反向陈述时有 5个分量表的效度系数达到 0.05的显著性水平。特别是 EPQ中的 P分量表,FFI中的 N、E和 O分量表在反向陈述时的效度大于正向陈述。尽管在单一样本的相关系数差异的显著性检验

中,正、反向陈述量表的效度系数差异都不显著,但由于反向陈述时达到显著性水平的相关系数多于正向陈述,可以认为反向陈述的效标关联效度略好于正向陈述。

3.3 正反向陈述题目的同质性检查

关于反向陈述题目的另一个争论是反向题与正向题可能测量的并不是同一种特质。为检验这一论点,我们把 EPQ的各正向陈述分量表与反向陈述分量表合并,产生正反向 E量表、正反向 P量表、正反向 N量表;对 FFI也做同样处理。正向与反向题合并后再做 CFA。所设定的模型与 3.2.1节中的模型相同,但题量增加了一倍。即设定正向题与反向题从属于同一因素。模型的拟合度见表 6。

表 6 正反向题合并后 EPQ和 FFI的模型拟合度

量表类别	χ^2	df	χ^2/df	RMSEA	NNFI	CFI	SRMR
合并 EPQ	10261.55	5882	1.74	0.047	0.84	0.85	0.072
合并 FFI	10449.55	5039	2.07	0.057	0.84	0.84	0.074

对比表 3中的结果可知正向与反向题合并后模型拟合度并没有明显变差,这说明正向与反向题测量同一特质的假设成立。为便于进行正向与反向题的对比,表 7中列出了 FFI合并模型中正向与反向题的因子载荷。为节省篇幅,EPQ 的结果没有表出。

表 7 CFA中 FFI正向与反向题合并后各题目的载荷

各分量表题目序号	N量表		E量表		O量表		A量表		C量表	
	正向	反向	正向	反向	正向	反向	正向	反向	正向	反向
1	0.75	0.49	0.25	0.24	删除	删除	0.44	0.28	0.36	0.50
2	0.44	0.41	0.27	0.41	0.39	0.15	0.49	0.42	0.50	0.46
3	0.58	0.41	0.39	0.40	0.37	0.44	0.41	0.47	0.60	0.48
4	0.75	0.71	0.37	0.38	0.07	0.13	0.33	0.29	0.32	0.26
5	0.55	0.57	0.22	0.22	0.33	0.37	0.42	0.52	0.29	0.49
6	0.43	0.45	0.16	0.25	0.16	0.26	0.23	0.23	0.27	0.33
7	0.67	0.57	0.55	0.54	删除	删除	0.38	0.31	0.48	0.32
8	删除	删除	0.61	0.55	删除	删除	0.31	0.34	0.20	0.30
9	0.21	0.38	0.66	0.63	0.22	0.37	删除	删除	0.16	0.27
10	0.74	0.62	删除	删除	0.53	0.53	0.21	0.26	0.50	0.49
11	0.42	0.45	0.59	0.56	0.76	0.70	0.28	0.18	0.47	0.47
12	删除	删除	删除	删除	0.67	0.69	0.36	0.45	删除	删除

分析正向与反向题的载荷,发现二者间有相当高的一致性,从绝对数量上看有时载荷会不一致,但基本上有一个规律,即正向载荷大时反向载荷也大(但不是每个题都如此)。因此从表 6的结果看可以认为正向与反向题测量了同一特质,这是个总体结论,而从表 7的结果看,并不是每个题都能得出这一结论。

另外还计算了各分量表的正向题总分与反向题总分的相关。EPQ中 E、P、N三量表正向与反向题总分相关分别为 0.996、0.872、0.938, FFI五个量表的相关系数分别为 0.681, 0.672, 0.569, 0.545, 0.624。这也佐证了题目在正向和反向陈述时测量相同特质的结论。

4 讨论

4.1 题目正反向陈述条件下测验的信度

本研究通过语义反转和重测得到了采用正反向陈述的两套内容相同的问卷。EPQ在反向陈述时信度更高,FFI则在正向陈述时信度更高。总体上看是正向陈述时测验信度略高。但正向与反向陈述条件下测验信度的差别不太大。

4.2 题目正反向陈述条件下测验的效度

研究中先对原 EPQ和 FFI量表中的题目根据题 - 总相关进行删改,然后进行验证性因素分析,发现删改后的量表基本能体现理论上假定的因素结构。然后再对正向陈述量表和反向陈述量表进行验证性因素分析,结果发现反向陈述条件下模型的拟合度优于正向陈述条件。说明采用反向陈述题目时有更好的构念效度。但对比验证性因素分析中因素载荷的平均值,发现反向陈述题目与正向陈述题目差别不大。

不少研究得出了正向题目会导致更大的反应偏差的结论。如 Samuelstuen将学习策略量表中的正向题量表和反向题量表分开,并设置一个正向陈述

因子和反向陈述因子,结果发现正向题量表在正向陈述因子上载荷大,而反向题量表在反向题因子上的载荷小,说明题目存在正向陈述效应(偏差),而反向陈述效应则小或没有[16]。又如 Sung - Woo Bae发现亚洲人的自尊得分在正向陈述量表上低于其他人种,而在反向题量表上则与其他人种没有差别,这也说明亚洲人的默认偏差更强[17]。反应偏差被认为是一种系统误差,因此影响测验的效度而不影响信度,本文发现正向陈述题目信度略高而效度略低,说明正向陈述题目更易引起反应偏差,这与国外的研究是一致的。但在本研究中反应偏差的影响似乎并不严重。但关于这种反应偏差的心理机制及如何控制等还需要深入研究。

从提名次数与 EPQ和 FFI的相关分析可知,反向陈述时量表的效标关联效度并不低于正向陈述量表。但研究中的相关系数都较低,有的相关并没有达到显著水平。原因可能是有的特质含义抽象,并不能很好地被别人识别,再加上对同学的熟悉度不够,和个人好恶的影响,导致提名不够准确。

4.3 反向与正向陈述题目是否测量同一特质

关于反向陈述题目的另一争论是是否与正向题目测量相同特质。本研究将每一分量表中相应的正向和反向题合并,再进行验证性因素分析,发现正反合并 EPQ和 FFI的模型拟合度与原模型没有太大差别,说明题目语义反转后仍然测量同一特质的假设成立。但有些题目的载荷在陈述方向改变后会有较明显的改变,说明语义的反转可能会改变题目测量相应特质时的有效性,因此若研究者在研究中需要改变题目的陈述方向时,要先进行试测并做项目分析。

国外的研究发现文化程度是影响反向陈述效应的调节变量,性别因素则基本没有影响。本文没有将性别作为调节变量研究,在取样时也没有要求男女人数相同,只是在文理科和年级上进行了匹配。因此虽然本研究的样本男女比例失衡,或许并不影响结论的普遍性,今后的研究中我们会对这一问题做进一步探讨。本研究的结论适用于具有大学文化程度的被试,至于是否适用于其他群体则需另做研究。另外本研究采用 EPQ和 FFI作为实验材料,涉及的都是一般性人格特质,研究结论能否推广到情感与态度类特质,尚需检验。

5 结论

综合上述分析可知,反向陈述题目在人格心理测量中的性能要略好于正向题目,在编制相关的心理测验时可以积极采用。但在使用反向题目时要使题目的语义简明易懂,与被试者的受教育水平相适合,尽量不采用双重否定的形式,同时要保证被试答题时态度认真。

虽然反向陈述题目的测验信度略低于正向陈述题目,但效度却较之略高。由于效度反映的是系统误差的大小,因此我们可以认为正向题目更易受反应偏差等系统误差的影响,而反向题目则不然。就是说,由于采用了否定陈述,反向题目需要被试进行更深入的加工,内容饱和度更大,因此被试的回答也就更为客观。关于这一点尚需要从测验心理机制上进行进一步探讨。

反向陈述题目是我们在编制人格或态度问卷时经常采用的形式。但不少人在做因素分析时发现反向题目的载荷低,或导致模型拟合度差。出现这一情况时可能说明量表中正向题目的社会称许性(social desirability)太高,或被试的回答不认真,对题内容没有进行充分理解,出现这种情况应尽量采取措施克服。那种为提高测验信度而删除反向题目的做法可能并不恰当。

本研究的结论是鼓励使用反向题,但并不是说人格和态度测量中所有反向题都优于正向题,实际应用中要考虑测验的目的,测验对象的年龄和教育程度等因素。至于在何种条件下使用正向题和反向题,正向题与反向题如何搭配,如何更好地避免默认偏差等问题尚需要深入探讨。综合本研究和以往研究的结论,测验中平衡使用正向和反向陈述题目仍是较为理想的选择。

另外,本研究中正向陈述和反向陈述量表的构念效度都不如删改后的原量表,因为删除题目时根据的是原量表中的题 - 总相关。这也说明对某些题目而言,其语义陈述方向的改变会对测验产生一定影响。因此如果研究中需要改变题目的陈述方向,应先试测并作项目分析。

参 考 文 献

1 Weems G H, Onwuegbuzie A J, Eggers S J. Characteristics of respondents who respond inconsistently to positively-and negatively-worded items on rating scales. Assessment and Evaluation in Education, 2003, 17(1): 45～60

2 Krosnick J A, Narayan S, Smith W R. Satisficing in surveys: Initial evidence. In: Braverman M T, Slater J K (eds). Advances in Survey Research. San Francisco: Sage, 1996: 29～44

3 Toner B. The impact of agreement bias on the ranking of question-

naire response. Journal of Social Psychology, 1987, 127(2): 221～222
4 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
5 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
6 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
7 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
8 Fried Y, Ferris G R. The dimensionality of job characteristics: Some neglected issues. Journal of Applied Psychology, 1986, 71: 419～426
9 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
10 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
11 Sandoval J, Lambert N M. Reliability and validity of teacher rating procedures in the assessment of hyperactivity as a function of rating scale format. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada. (ERIC Document Reproduction Service No. ED 160 614), 1978. 1～20
12 Schriesheim C A, Hill K D. Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. Educational and Psychological Measurement, 1981, 41: 1101～1114
13 Williams R L, Bush V J, Park S H et al. Work Ethic Scale for middle school students. (ERIC Document Reproduction Service No. ED 452 257), 2001: 1～15
14 Costa P T Jr, 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. 1～15
15 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)
16 Samuelstuen M S. Psychometric properties and item-keying direction effects for the Learning and Study Strategies Inventory-High School Version with Norwegian students. Educational and Psychological Measurement, 2003, 63(3): 430～446
17 Bae Sung-Woo, Brekke J S. The Measurement of Self-Esteem Among Korean Americans: A Cross-Ethnic Study. Cultural Diversity & Ethnic Minority Psychology, 2003, 9(1): 16～30

Effects of Positively and Negatively Worded Items in Personality Measurement

Guo QingKe[1,2], Han Dan[1], Wang Zhao[1], Shi Kan[2]
([1] *Department of Psychology, Liaoning Normal University, Dalian* 116029, *China*)
([2] *Institute of Psychology, Chinese Academy of Science, Beijing* 100101, *China*)

Abstract

In personality measurement, some examinees show the tendency of simply agreeing with the scale items regardless of the item content. This tendency is called acquiescence' by Cronbach (1946, 1950), and yea-or nay-saying' by Couch and Keniston (1960). To diminish this kind of response bias, negatively worded items are adopted. Negatively worded items require more attention to item content, and also can balance the impact of yea- or nay-saying on overall test scores.

But studies on the change of item narrative direction have provided inconsistent findings. Whereas some studies show that the use of negatively worded items can increase reliability and validity (Sandoval and Lambert, 1978; Schriesheim and Hill, 1981; Williams et al., 2001), others show that the use of negatively worded items can confound factor structures (Weems et al., 2001) and may result in a separate factor irrelevant to the trait being measured' (Ibrahim, 2001). When both positively and negatively worded items are used to measure a uni-dimensional trait, factor analysis will exact two factors, one is related to positively worded items and the other is related to negatively worded items.

In this study, Chinese EPQ and NEO-FFI were used as personality measures, but items narration directions were changed in both measures. In one form (form Ⅰ) of EPQ (or NEO-FFI), half of items were positively worded, and half of the items negatively worded. In another form (form Ⅱ) of EPQ (or NEO-FFI), positive items in form Ⅰ were converted into

negative items, and negative items into positive ones. These two test forms of EPQ (or NEO-FFI) were administered to auniversity student group with the time interval of two weeks. Thus we obtained the data from 363 university students who took positively worded EPQ and negatively worded EPQ, and the data from 412 students who took positively worded NEO-FFI and negatively worded NEO-FFI.

The result showed that positively worded sub-scales of EPQ and NEO-FFI were a little more reliable than their negative counterparts, α coefficients for three positive EPQ sub-scales (E, P, N) were 0.80, 0.42, 0.75, for three negative EPQ sub-scales were 0.82, 0.45, 0.81; α coefficients for five positive NEO-FFI sub-scales (N, E, O, A, C) are 0.82, 0.65, 0.69, 0.64, 0.75, and for five negative NEO-FFI sub-scales were 0.77, 0.62, 0.68, 0.64, 0.67.

Confirmatory Factor Analyses showed that negative EPQ and NEO-FFI exhibited better fit than the positive ones: for negative EPQ, model $\chi^2 = 2456.22$, RMSEA = 0.046, NNFI = 0.87, SRMR = 0.069, for positive EPQ $\chi^2 = 2402.96$, RMSEA = 0.045, NNFI = 0.85, SRMR = 0.072; for negative NEO-FFI $\chi^2 = 2868.20$, RMSEA = 0.066, NNFI = 0.84, SRMR = 0.078, for positive NEO-FFI $\chi^2 = 3952.15$, RMSEA = 0.075, NNFI = 0.79, SRMR = 0.086.

Based on classmate nominations as criterions, sub-scales of negative EPQ and NEO-FFI showed higher correlations with criterions than positive ones. When all positive and negative items of EPQ (or NEO-FFI) factor analyzed in one CFA model, the fit measures did not become significantly lower than the positive or negative item model, and loadings of the positive and negative items were highly consistent. These results suggest that positive and negative forms of an item measure the same construct.

In conclusion, positive worded items may produce higher reliability, but lower construct validity and criterion-related validity than negative worded ones do. Negatively and positively worded items measure the same construct, but positively worded items may be more affected by response set bias.

Key words positively and negatively worded items, response bias, EPQ, NEO-FFI.