

文章编号: 1000-5862(2021)05-0441-11

# 近10年中国心理量表编制的现状、问题与建议

邓稳根<sup>1</sup> 张文丽<sup>1</sup> 郭磊<sup>2</sup>

(1. 海南师范大学心理学院 海南 海口 571158; 2. 西南大学心理学部 重庆 400715)

**摘要:** 量表编制对于构建心理与行为科学的知识体系至关重要。该文以2010年1月1日—2020年6月30日发表在5种心理学类期刊上的222篇量表编制文献为研究对象,系统地回顾近10年在这5种期刊上发表量表编制的文献状况,从量表编制的3个基本步骤(项目生成、内容分析和心理计量学分析)进行了分析,指出在当前文献中量表在编制过程中的主要局限。通过对这些局限的分析,期望能够有助于未来的研究,特别是在不同的概念和方法策略之间做出最合适的选择。

**关键词:** 量表编制; 项目生成; 内容有效性; 心理计量学

中图分类号: B 841 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2021.05.01

## 0 引言

心理测量是心理科学的一项基本活动,它使研究者能够获得关于人、对象、事件和过程的知识;心理量表是一种有效的测量工具,它由项目集组成,揭示了无法通过直接手段观察到的心理变量的水平<sup>[1]</sup>。许多研究者<sup>[1-4]</sup>一致认为,量表的编制过程涉及复杂和系统的程序,通常可以分为如下3个基本步骤:

第1步,项目生成。研究者对初始项目集的生成提供理论和方法支持<sup>[5]</sup>。初始项目生成的方法可以分为演绎法、归纳法和二者相结合的方法。演绎方法涉及基于广泛的文献回顾或已有量表生成项目<sup>[6]</sup>;归纳法基于目标人群(如焦点小组),采用访谈或其他定性探索性研究方法获得有关量表结构的定性信息后生成项目<sup>[7]</sup>。在项目生成中,研究者还关注各种参数(如量表的说明或指导语、项目数量、显示格式、项目用词、项目作答的变异性、项目对于不同群体的公平性等)对每个项目和整个量表设置的影响<sup>[1,4]</sup>。

第2步,内容分析。研究者需要评估新量表内容的有效性,以便保证初始项目集反映了期望的理论结构,为后续的结果推断提供有效的依据<sup>[8]</sup>。为了

保证内容的有效性,研究者通常会就操作性项目征求专家(量表编制的专家或目标结构的专家)或目标人群(量表的潜在使用者)的意见,使研究者能够确保研究中阐述的假设恰当地代表研究的结构<sup>[3]</sup>。

第3步,用心理计量学分析它涉及项目筛选的心理计量学分析、量表维度结构的探索和验证分析以及量表的信效度分析。项目筛选的心理计量学分析主要是对项目区分能力的考查,研究者通常用临界比率(CR)、题总相关、因子负荷、共同度大小和删除某项目后的信度改变量等指标来筛选项目<sup>[9]</sup>。量表的维度分析是开展量表信效度分析的前提和基础,通常采用探索性因素分析(EFA)和验证性因素分析(CFA)的方法对量表的维度结构进行探索和验证。效度直接关系到一个量表实际测到所要心理特质的程度,或者说它是指测验分数能够说明心理特质的程度<sup>[10]</sup>,这指的是从观察到的分数到这些观察应该包含有关心理特质信息的合理推论程度<sup>[11]</sup>。研究者通常报告量表的聚合效度和区分效度以及效标效度等。信度是指测验结果的一致性程度,通常使用重测信度、复本信度、分半信度、内部一致性系数和评定者信度等来衡量<sup>[12]</sup>。在心理计量学分析中,从目标人群中选取的样本数量和样本的异质性都会影响到分析结果的有效性和稳定性。因此,既要根据随机原则选取样本,又要在资源有限的情况下抽取合

收稿日期: 2021-01-17

基金项目: 国家自然科学基金青年课题(31900793)和海南师范大学教授(博士)科研启动基金(040924)资助项目。

作者简介: 邓稳根(1978—),男,江西遂川人,副教授,博士,主要从事心理统计与测量研究。E-Mail: dwengen@163.com

理的样本数量. 一个普遍经验法则是: 量表的每个项目至少应该有 10 名参与者, 理想的情况是 15:1 或 20:1<sup>[1-2,13]</sup>.

虽然有关量表编制的文献很多, 但是一些文献所涉及的量表编制过程存在一些局限. 这些局限将削弱量表的信度和效度, 限制了量表的适用性, 阻碍了它们在实践中的推广<sup>[14]</sup>. 从这个意义上说, 了解量表编制的现状及局限对该领域的研究至关重要. 鉴于一些研究者在编制量表过程中存在不同程度的局限, 中国一些心理学类期刊在审核量表类论文时非常谨慎, 如 2012 年 3 月 8 日《心理科学》在其投稿须知中明确指出“凡属测验类稿件, 必须提供经由中国心理学会、中国心理学会心理测量分会或本省一级心理学会组织专家(3 位或 3 位以上) 鉴定, 并加盖中国心理学会、中国心理学会心理测量分会或本省心理学会的公章(专家可不限本省单位) 的鉴定材料原件至编辑部, 稿件才能进入审稿程序.”

近年来, 研究者将心理量表广泛应用于教育、医疗、管理等领域中. 量表编制的质量会影响到研究和决策的准确性, 因此, 对已经编制好的量表进行梳理, 审视其编制过程, 才能够更好地了解这些量表的质量, 同时也为将来量表的编制提供一定的借鉴和指导. 本文拟系统地回顾中国近 10 年来发表在一些心理学类期刊上的关于量表编制的文献, 从量表编制的 3 个步骤对现有文献进行分析, 评估研究者在量表编制过程中存在的一些局限, 从而为将来量表编制的研究和实践提供一些建议.

## 1 方法

### 1.1 对象

本研究从 7 种心理学类期刊《心理学报》《心理科学》《心理发展与教育》《心理与行为研究》《心理学探新》《中国临床心理学杂志》《心理科学进展》中采用人工查阅的方法检索有关量表编制类文献, 文献检索时间限定为 2010 年 1 月 1 日—2020 年 6 月 30 日. 人工检索发现《心理学报》和《心理科学进展》近 10 年来没有发表与量表编制相关的论文, 故在结果分析中剔除了这 2 种期刊.

### 1.2 文献选取标准

(i) 选取测量非认知类心理现象的心理量表或

问卷的中文文献, 排除测量认知能力或教育成就等测验; (ii) 选取原创性量表编制类的文献, 去除只做量表信效度检验、适用性检验或量表修订/简化的文献.

### 1.3 统计分析工具

采用 Excel2019 和 Spss26 进行统计分析.

## 2 结果

### 2.1 量表编制类论文年度发表量和引用次数

通过检索, 筛选出符合要求的文献共 222 篇. 图 1 描绘了各年度发表的量表编制类文献数量. 从图 1 可以看出: 文献数量呈逐年下降趋势, 这种趋势可以用一个指数函数进行拟合. 表 1 列出了各期刊在各年度内发表的量表编制类文献的数量及其在该年度内发表的总文献数量中所占的比例. 表 1 的结果显示《心理发展与教育》自 2012 年开始就没有刊发表量表编制的文献, 《心理科学》也大致如此. 《心理学探新》《心理与行为研究》《中国临床心理学杂志》则每年都会刊发不同数量的量表编制论文, 其中以《中国临床心理学杂志》刊登这类论文的数量最多. 上述期刊几乎都在 2012 年之后减少了量表编制类论文的数量.

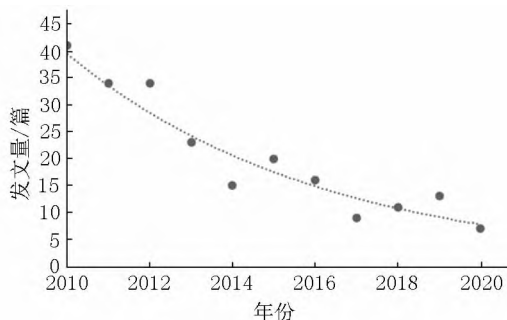


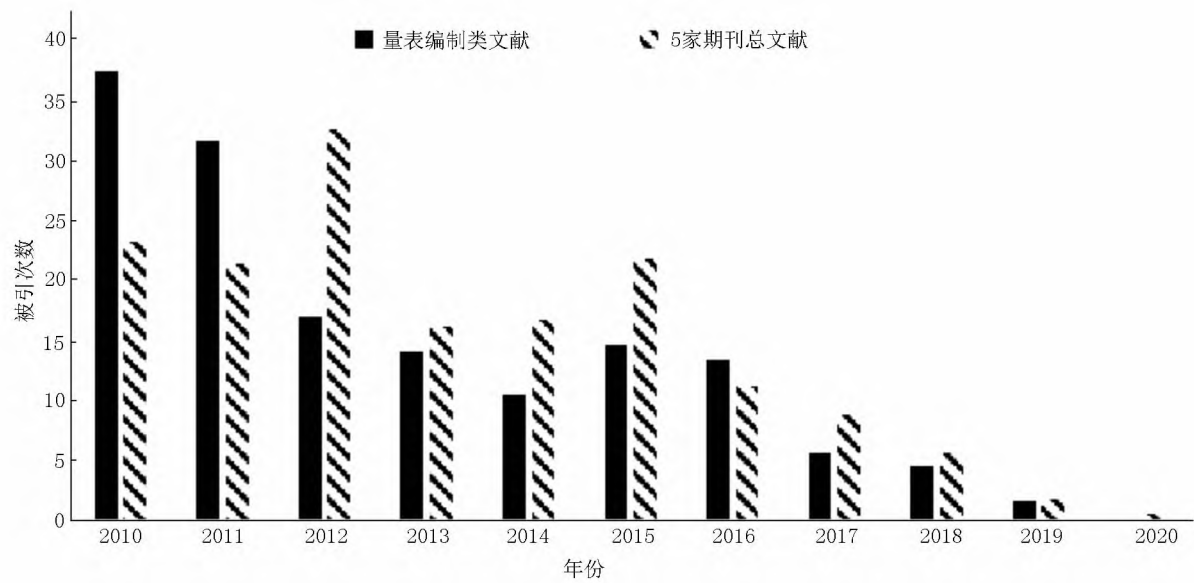
图 1 年度发文章量以及拟合趋势

本研究也对量表编制类文献的引用率进行了调查. 结果发现: 从 2010 年 1 月 1 日—2020 年 6 月 30 日 222 篇文献被引总次数为 4 290 次, 篇均被引次数为 19.32 次. 图 2 显示了各年度量表编制类文献的年篇均被引次数与 5 种期刊总文献的年加权篇均被引次数的对比情况. 从图 2 可以看出: 2010—2011 年的量表编制类文献年篇均被引次数远高于期刊的年加权篇均被引次数; 从 2012 年后(除 2016 年外), 量表编制类文献年篇均被引次数低于期刊的年加权篇均被引次数.

表 1 各年度各期刊问卷类文献中量表编制的文献数量及百分比

年份	《心理科学》	《心理发展与教育》	《心理学探新》	《心理与行为研究》	《中国临床心理学》
2010	5( 1.2)	10( 7.8)	7( 6.5)	3( 5.5)	16( 6.2)
2011	1( 0.4)	3( 3.3)	7( 6.4)	3( 5.8)	20( 7.6)
2012	1( 0.4)	0( 0)	5( 5.0)	13( 16.0)	15( 5.9)
2013	0( 0)	0( 0)	7( 7.4)	8( 5.6)	8( 3.0)
2014	0( 0)	0( 0)	2( 2.0)	7( 5.0)	6( 2.2)
2015	0( 0)	0( 0)	5( 5.0)	3( 2.3)	12( 4.5)
2016	4( 1.7)	0( 0)	2( 2.0)	2( 1.6)	8( 3.0)
2017	0( 0)	0( 0)	2( 2.1)	2( 1.6)	5( 1.9)
2018	0( 0)	0( 0)	2( 2.1)	3( 2.5)	6( 2.2)
2019	0( 0)	0( 0)	6( 6.7)	1( 0.9)	6( 2.2)
2020	0( 0)	0( 0)	2( 4.5)	1( 1.7)	3( 2.2)

注: 2020 的文献仅限于 2020 年 1 月 1 日—2020 年 6 月 30 日发表在上述期刊中的文献, 括号内的数字为百分比( % )。



注: 统计被引文量的时间截至 2020 年 6 月 30 日。

图 2 被引次数对比

## 2.2 项目生成

14.0% (  $n = 31$  ) 的研究单独使用了完全演绎的方法来生成项目, 11.3% (  $n = 25$  ) 的研究单独使用了归纳方法, 73.0% (  $n = 162$  ) 的研究采用了演绎和归纳相结合的策略, 1.7% (  $n = 4$  ) 的研究没有报告项目生成的方法。在使用演绎法生成项目的文献中, 通常采用文献回顾 (  $n = 141$ , 73.1% ) 和参照已有量表 (  $n = 102$ , 52.8% ) 的方法; 在使用归纳法生成项目的文献中, 常用的 4 种方法是: 专家小组的意见 (  $n = 136$ , 72.7% )、访谈 (  $n = 125$ , 66.8% )、开放性问卷调查 (  $n = 97$ , 51.9% )、焦点小组 (  $n = 7$ , 3.7% )。

在生成的项目中, 项目反应格式采用 2 种类型的占比为 0.9% (  $n = 2$  ), 3 种类型的占比为 0.5% (  $n = 1$  ), 4 种类型的占比为 11.3% (  $n = 25$  ), 5 种类型的占比为 48.2% (  $n = 107$  ), 6 种类型的占比为 14.0% (  $n = 36$  ), 7 种类型的占比为 9.5% (  $n = 21$  ), 9 种类型的占比为 0.5% (  $n = 1$  ), 10 种类型的占比为 1.0% (  $n = 2$  ), 14.1% 的研究 (  $n = 32$  ) 没有说明项目的反应格式。这些结果表明: 5 种反应类型格式是当前所编制量表的主要反应格式。

## 2.3 内容分析

29.7% (  $n = 66$  ) 的研究仅使用了专家意见来评估内容有效性, 11.7% (  $n = 26$  ) 的研究仅使用了目

标人群的意见来评估内容有效性,42.8% ( $n=95$ ) 的研究将这 2 种方法都使用了,15.8% ( $n=35$ ) 的研究对此无说明。

## 2.4 心理计量学分析

2.4.1 项目筛选的心理计量学分析 46.8% ( $n=104$ ) 的研究采用了结合临界比率(CR)小于 0.05 和题总相关系数大于 0.3 的策略筛选项目;67.6% ( $n=150$ ) 的研究采用了因子载荷大于 0.4,同时 25.2% ( $n=56$ ) 的研究结合共同度大于 0.2 的标准筛选项目。此外,还有少量研究者采用了鉴别指数(D 值) ( $n=14$ , 6.3%)、删除题目后信度的改变量 ( $n=2$ , 1.0%)、文本分类一致性( $n=1$ , 0.5%)、频率统计( $n=3$ , 1.5%)等方法筛选项目。

2.4.2 量表的维度结构和效度分析 10.8% ( $n=24$ ) 的研究仅采用了探索性因素分析(EFA)方法,9.5% ( $n=21$ ) 的研究仅使用验证性因素分析(CFA)方法,76.6% ( $n=170$ ) 的研究使用了结合探索性因素分析(EFA)和验证性因素分析(CFA)的方法。在统计的文献中 3.1% ( $n=7$ ) 的研究没有采用上述方法对量表维度结构进行分析。

在探索性因素分析中,在因素分析的适当性分析方面,73.7% ( $n=143$ ) 的研究运用了 KMO 检验和 Bartlett 球形检验相结合的方法,6.2% ( $n=12$ ) 的研究仅采用了 KMO 检验方法,20.1% ( $n=39$ ) 的研究没有进行因素分析适当性的分析。在因素的抽取方法方面,76.8% ( $n=149$ ) 的研究采用主成分法,1.6% ( $n=3$ ) 的研究采用主轴因子法,0.5% ( $n=1$ ) 的研究采用极大似然法,21.1% ( $n=41$ ) 的研究没有报告具体采用的方法。在因素的选取标准方面,40.2% ( $n=78$ ) 的研究仅采用了特征值大于 1 的标准,3.6% ( $n=7$ ) 的研究仅根据碎石图的变化作为因素选取的标准,32.5% ( $n=63$ ) 的研究则结

合使用特征值大于 1 和碎石图的变化来选取因素,0.5% ( $n=1$ ) 的研究则将碎石图和平行分析结合起来选取因素,同样有 0.5% ( $n=1$ ) 的研究将特征值和平行分析结合起来选取因素,0.9% ( $n=2$ ) 的研究则结合特征值、碎石图和平行分析来选取因素,25.3% ( $n=42$ ) 的研究没有说明采用哪种方法来选取因素。在因素旋转方面,56.7% ( $n=110$ ) 的研究采用正交旋转方法,18.6% ( $n=36$ ) 的研究采用斜交旋转方法,0.5% ( $n=1$ ) 的研究采用了斜交和正交旋转相结合的方法,24.2% ( $n=47$ ) 的研究没有报告因素旋转的方法。在选取的因素对方差的贡献率方面,92.8% ( $n=180$ ) 的研究报告了方差解释率,7.2% ( $n=14$ ) 的研究对此没有说明;在因素分析结果的呈现方面,95.9% ( $n=186$ ) 的研究报告了因子负荷的大小,4.1% ( $n=8$ ) 的研究没有报告该结果;37.0% ( $n=71$ ) 的研究报告了因子共同度,63.0% ( $n=123$ ) 的研究则没有报告该结果。

在验证性因素分析中,研究者主要采用 AMOS ( $n=79$ , 41.4%) 和 LISREL ( $n=72$ , 38.0%) 这 2 种软件进行分析,0.5% 的研究 ( $n=1$ ) 同时使用了 AMOS 和 LISREL 这 2 种软件,2.6% ( $n=5$ ) 的研究采用 Mplus 软件,1.0% ( $n=2$ ) 的研究采用 EQS 或 LATENTGOLD 软件,16.8% ( $n=32$ ) 的研究没有说明具体采用的软件名称。在模型-数据拟合评价方面,按数量排序前 7 的拟合指标分别为: RMSEA ( $n=189$ , 85.1%)、CFI ( $n=187$ , 84.2%)、 $\chi^2/d_f$  ( $n=173$ , 77.9%)、NNFI 或 TLI ( $n=139$ , 62.6%)、 $\chi^2$  ( $n=128$ , 57.7%)、 $d_f$  ( $n=121$ , 54.5%)、IFI ( $n=109$ , 49.1%)。具体如表 2 所示。在验证性因素分析结果的呈现方面,21.5% ( $n=41$ ) 的研究呈现了因素负荷结果,78.5% ( $n=150$ ) 的研究则未呈现该结果。

表 2 拟合指标不同区间值的文献数量及比例

$\chi^2/d_f$		RMSEA		值	文献数及比例		
值	文献数及比例	值	文献数及比例		CFI	NNFI	IFI
0 ~ <3	116(52.3)	0 ~ <0.05	49(22.1)	0 ~ <0.8	1(0.5)	5(2.3)	0(0)
3 ~ 5	52(23.4)	0.05 ~ 0.08	127(57.2)	0.8 ~ 0.9	23(10.4)	24(10.8)	12(5.4)
>5 ~	5(2.3)	>0.08 ~ 0.10	13(5.9)	>0.9 ~ 1.0	163(73.4)	110(49.5)	97(43.7)

注: 括号内数字为百分比(%)。

除了采用因素分析方法对量表的结构进行分析外,1.8% ( $n=4$ ) 的研究报告了聚合效度结果,3.6% ( $n=8$ ) 的研究报告了区分效度的结果,

44.6% ( $n=99$ ) 的研究报告了效标效度结果;1.4% ( $n=3$ ) 的研究同时报告了聚合效度和区分效度的结果,3.2% ( $n=7$ ) 的研究同时报告了聚合效度和

效标效度的结果 5.0% ( $n = 11$ ) 的研究同时报告了效标效度和区分效度的结果 3.6% ( $n = 8$ ) 的研究同时报告了聚合效度、区分效度和效标效度的结果; 1.0% ( $n = 2$ ) 的研究同时报告了表面效度和交叉效度的结果 35.8% ( $n = 80$ ) 的研究没有报告上述任何一种结果. 在计算方法上 62.2% ( $n = 138$ ) 的研究采用了相关分析法(与成熟问卷进行相关系数检验)进行计算, 计算结果显示: 分量表相关系数大多在 0.30~0.65 之间, 总量表相关系数大多在 0.30 以上. 0.5% ( $n = 1$ ) 的研究采用了多质多法模型来考查量表的聚合效度和区分效度, 1.0% ( $n = 2$ ) 的研究采用计算的平均方差提取值来获取聚合效度和区分效度证据 0.5% ( $n = 1$ ) 的研究采用了 ECVI 值计算交叉效度.

2.4.3 信度分析 在量表信度的评估方面 26.1% ( $n = 57$ ) 的研究仅评估了量表的内部一致性(即克龙巴赫  $\alpha$  系数)大小 36.5% ( $n = 80$ ) 的研究同时评估了量表的内部一致性和重测信度的大小 0.5% ( $n = 1$ ) 的研究同时评估了量表的内部一致性、重测信度和组合信度的大小 0.5% ( $n = 1$ ) 的研究同时评估了量表的内部一致性和组合信度的大小, 11.7% ( $n = 26$ ) 的研究同时评估了量表的内部一致性和分半信度的大小 20.7% ( $n = 46$ ) 的研究同时评估了量表的内部一致性分析、分半信度和重测信度的大小 3.5% ( $n = 10$ ) 的研究则没有报告量表的信度 0.5% ( $n = 1$ ) 的研究是用概化理论计算信度. 表 3 列出了 3 种信度系数不同区间值的文献数量及比例. 在重测信度的结果报告中 95.3% ( $n = 121$ )

的研究报告重测信度的间隔时间大于 7 d 4.7% ( $n = 6$ ) 的研究未报告间隔时间. 此外 1.5% ( $n = 3$ ) 的研究报告了合成信度, 介于 0.75~0.85 之间.

表 3 3 种信度系数不同区间值的文献数量及比例

区间值	内部一致性	分半信度	重测信度
0.90~1.00	15(7.1%)	4(5.6%)	3(2.4%)
0.80~0.89	44(20.9%)	19(26.4%)	24(18.9%)
0.70~0.79	83(39.3%)	20(27.8%)	47(37.0%)
0.60~0.69	56(26.5%)	24(33.3%)	38(29.9%)
0~0.59	13(6.2%)	5(6.9%)	15(11.8%)

2.4.4 样本量和项目数 222 篇量表编制文献使用的样本量都大于经验法则所要求的最低样本数量, 即量表中每个项目至少有 10 名参与者. 在删除的项目数量方面 47.3% ( $n = 105$ ) 的研究删除了初始项目集 50% 及以上数量的项目 45.5% ( $n = 101$ ) 的研究删除的项目数少于初始项目数的 50%, 7.2% ( $n = 16$ ) 的研究没有明确说明删除的项目数量.

在量表各维度所包含的项目数量上, 在 3.2% ( $n = 7$ ) 的研究中的量表维度下量表题目存在数量小于 3 的情况, 在 34.2% ( $n = 76$ ) 的研究中的每个量表维度下量表题目数量大于等于 3, 在 29.3% ( $n = 65$ ) 的研究中的每个量表维度下量表题目数量大于等于 4, 在 25.2% ( $n = 56$ ) 的研究中每个维度下的量表题目数量大于等于 5 8.1% ( $n = 18$ ) 的研究未显示量表各维度所包含的项目数.

表 4 总结了量表编制过程的文献分析结果.

表 4 量表编制过程的文献分析结果总结

方法	数量/篇	比例/%	方法	数量/篇	比例/%
步骤 1: 项目生成			步骤 3: 心理计量学分析		
仅使用演绎	31	14.0	仅使用 EFA	24	10.8
仅使用归纳	25	11.3	仅使用 CFA	21	9.5
演绎 + 归纳	162	73.0	EFA + CFA	170	76.6
无说明	4	1.7	没有用 EFA 或 CFA	7	3.1
文献回顾	141	73.1	聚合效度	22	9.9
已有量表	102	52.8	区分效度	30	13.5
访谈	125	66.8	效标效度	128	57.7
焦点小组	7	3.7	内部一致性评估	211	95.0
开放性问卷调查	97	51.9	分半信度	72	32.4
专家小组	136	61.2	重测信度	127	57.2
步骤 2: 内容分析			关于步骤 3 的项目数		
专家评定	66	29.7	最终减少 50% 以下	101	45.5
目标人群评定	26	11.7	最终减少 50%	92	41.4
2 者都有	95	42.8	最终减少 50% 以上	13	5.9
无明确说明	35	15.8	未明确报告	16	7.2

## 2.5 量表编制过程中报告的主要局限

本研究发现在量表编制过程中主要存在 7 个方面的局限(见表 5)。

表 5 量表编制过程中报告的主要局限

局限性	<i>n</i>	%
样本特性局限:		
同质/方便样本	132	59.5
方法学局限:		
自我报告法	188	84.7
心理计量学局限:		
量表结构检验不科学	52	23.4
缺乏效标效度检验	97	43.7
缺乏量表结构的等价性	222	100.0
定性研究局限:		
仅使用演绎 + 仅使用归纳	56	25.2
未正式评估内容效度	35	15.8
范式局限:		
Likert 范式	178	80.2
项目局限:		
无反向计分项目	189	85.1
测量理论局限:		
经典测量理论	215	96.8

## 3 讨论

### 3.1 关于文献发表量和引用次数

本研究系统地回顾了近 10 年中国一些心理学类期刊上发表的量表编制类文献,发现自 2012 年以来,无论是文献的发表量还是引文量相对于其他类文献都有明显下降。这可能与在 2012 年《心理科学》在投稿要求中明确提出量表编制类文献要先进行鉴定之后各期刊相继提出同样的要求有关。各心理学类期刊之所以提出这种要求,根本上还是量表编制类文献在量表研制过程中存在诸多问题,导致编制的量表质量不高。同时,中国心理学会心理测量分会没有及时根据各期刊的要求,出台一些具体的可操作性标准主动开展鉴定工作,导致一些量表编制类研究不能得到发表。本研究认为,心理量表编制的鉴定工作可以借鉴中国心理学会临床心理学分会的注册系统做法。他们在取消了国家心理咨询师资

格考试后,迅速出台了一系列注册标准,开展了相关的培训,使得心理咨询工作不但没有萎缩,反而得到了更好的发展。建议心理测量分会也能主动开展培训和鉴定工作,提高研究者量表编制水平,促进量表编制的发展。

### 3.2 关于项目生成

本研究发现,大多数量表编制研究(如方晓义等<sup>[15]</sup>的研究)采用了演绎和归纳相结合的方法生成项目,这与 R. S. Devellis 的建议相一致<sup>[1]</sup>。在演绎方法中,文献回顾和参照已有量表是最广泛使用的方法(如王馥芸等<sup>[16]</sup>的研究),这与一些研究者<sup>[17-19]</sup>的观点基本一致,它们不同之处在于中国量表编制者参照已有量表来生成项目的比例很高。专家小组的意见、访谈和开放问卷调查(如姚本先等<sup>[20]</sup>的研究)则是最广泛使用的归纳方法,与一些研究者<sup>[7,18-19]</sup>的综述结果不同之处在于,后者在最广泛使用的归纳方法中并没有专家小组的意见这一项。

作为量表编制第 1 步的项目生成十分重要,而生成项目的质量取决于测量结构的定义是否恰当,若未能充分定义测量结构的概念域,则会导致与测量结构的概念定义不佳的问题。如导致混淆该量表结构的概念定义和其他结构概念的定义,可能生成缺陷或污染的项目,从而影响该测量结构与其他测量结构的关系<sup>[14]</sup>。文献回顾有助于厘清目标测量结构的性质和多样性,而通过与目标人群的访谈和开放问卷调查则可以获得与所测对象相符合的原始结构和真实信息。

参照已有量表为项目生成带来便利,但带来的潜在问题是知识产权的争议,同时与中国研究者缺乏足够的量表编制训练有关。在中国市场上专门介绍量表编制的书籍很少,一些研究者获得量表编制的知识来源于侧重讲述心理计量学分析和量表使用的心理测量学书籍,不仅里面涉及量表编制的内容很少,而且相应的内容更多与认知和教育测验的编制相关联。对非认知心理构念的测量量表编制的特异性考虑不足。借鉴已有量表的项目还可能带来的一个问题是:原量表项目测量的结构概念与现有量表的测量结构概念并不一致,从而导致前面所述的概念混淆问题。依赖专家小组意见生成项目可以弥补研究者量表编制经验的不足,但过度依赖专家意见可能带来主观性。在量表结构概念界定不清晰的

情况下,专家对量表结构概念的理解与真实要测量的量表结构概念可能并不一致,从而影响项目内容的有效性。

综合以上分析,建议未来应该以文献回顾为核心,结合对目标人群的访谈和开放问卷调查来生成量表项目。在生成项目时尽量少借鉴已有量表的项目,以保证项目的原创性。同时,专家小组的意见可以作为项目生成的一个辅助方法,而不能主导项目的生成。

在项目反应格式方面,本研究发现:大多数量表采用5~7点Likert反应格式,尤其以5点反应格式居多。相较于反应选项数较少的情况,它的一个潜在好处是:能潜在地揭示被调查者之间更微妙的心理差异,从而对他们做更精细的区分;同时,与数量较多的选项相比,它又能减少因难以区分各选项之间过于细微的差异而导致随机作答产生的随机误差。因此,大多数研究者在项目反应格式方面,推荐5~7个反应选项数<sup>[21]</sup>。然而,对于认知能力水平偏低、认知判断处于“非黑即白”阶段的年龄较小被试,则更适宜使用2个反应选项。对于反应选项数设置为偶数还是奇数为宜,研究者也存在一定的争议,赞成设置为偶数的研究者认为可以减少社会赞许效应,而赞成设置为奇数的研究者认为中间选项更有利于心理特质水平符合项目中点的被调查者作答。本研究认为:受中国传统中庸文化的影响,中国人在对项目选项进行选择时更可能依据项目中点,而不是依据项目低点或高点为参照作出反应,若缺乏中间选项则会造成被试选择的困难。因此,本研究建议反应选项数应设置为奇数为宜。

### 3.3 关于内容效度

本研究的结果表明:将专家评定和目标人群评定2种方法结合已经成为分析内容效度的最广泛使用的方法,其中专家评定是删除不适当项目的最常用定性方法,这与一些研究者<sup>[7,48]</sup>的结果一致。D. M. Hardesty等<sup>[22]</sup>强调专家仔细分析初始项目集的重要性,他们建议:任何使用新的、改变的或以前未检查过的量表项目的研究,至少应该由1个专家小组来评判,并且他们制定了一系列决策规则来决定项目是否保留,如“总分决定规则”,即所有专家对一个项目评定的总分被认为是否保留项目的一个有效规则。内容有效性判断不仅要考虑专家的意见,

还要考虑目标人群的意见。目标人群的意见是内容有效性的基础<sup>[17]</sup>;还有一些研究者强调<sup>[2,23]</sup>,新的量表必须经过目标人群成员的事先审查。但目标人群意见的使用需要进行预试,预试可以确定受测者对量表每个项目的看法和反应,从而使研究人员能够在推广使用该量表之前发现并消除量表中存在的潜在问题。

### 3.4 关于心理计量学分析

本研究还发现:研究者通常根据因子负荷高于0.4的标准筛选项目,其次是根据临界比率小于0.05和题总相关系数高于0.3的标准筛选项目。这是Likert类型量表在编制时筛选项目的通常做法。但特别需要注意的一点是:一些量表编制者在计算项目总分时认为项目总分就是初始项目集中所有项目的总分,本研究认为这种做法不一定合理,合理的做法是只有属于同一个理论假设维度的项目得分才可以相加。如果根据理论假设,量表是由多个相互独立的特质维度构成,那么不同维度的项目得分不适合相加。如EPQ中不同量表维度的项目得分是不适合加总的。在量表的结构假设不清晰的情况下,更适合采用因子负荷并结合共同度来筛选项目。

本研究的结果显示:大多数量表编制文献(如刘影等<sup>[24]</sup>的研究)结合使用了探索性因素分析(EFA)和验证性因素分析(CFA)方法对量表的结构进行分析,这与A. Clark等<sup>[2]</sup>强调研究人员需理解和应用因素分析的重要性相一致,也与R. F. DeVellis<sup>[1]</sup>指出结合使用EFA和CFA对量表维度结构进行分析的建议相一致。EFA在通过探索观察变量之间的关系来有效地识别衡量潜在变量或因素方面具有非常适用的价值,然而,与其他方法相比,它具有更强的主观性<sup>[25]</sup>。L. Koen等<sup>[26]</sup>认为:在根据EFA结果进行项目筛选时,带有一定程度的主观性是必要的。通常的做法是根据对其中因子负荷最大项目的解释来标记每个因素,它的筛选标准为0.30~0.50<sup>[27-28]</sup>。为了在不同的目标人群样本中得到更一致的结果,需要选取新的样本进一步作CFA,对EFA中探索得出的因素结构(包括因素之间的相关性)进行验证,应结合使用多个拟合指标(重点关注CFI、RMSEA、 $\chi^2/d_f$ 等指标)的结果考查模型数据的拟合情况。如果模型拟合数据(如 $N_{NFI}$ 或 $T_{LI} \geq 0.90$ 、 $C_{FA} \geq 0.90$ 、 $R_{MSEA} \leq 0.08$ <sup>[29]</sup>),那么研究人员就可以

得出因素结构一致的结论;如果模型不拟合数据,那么需要使用修正指数对其进行修改<sup>[30]</sup>。在可能的情况下应选取新的样本对新的模型进行交叉验证,以提供更一致的心理计量结果。除采用因素分析获得测验结构的证据外,本研究还发现:大多数量表编制文献还会考查量表的效标效度(如赵欢欢等<sup>[31]</sup>的研究)。考虑到效度对保证收集数据的质量和量表的适用性十分重要,未来的研究应该考虑以不同的方式来评估新量表的有效性,从而提高量表编制的心理计量学严谨性。在 CFA 的操作软件方面,本研究发现:大多数研究使用的是 LISREL 和 AMOS 软件。由于 LISREL 软件基本停止了更新,使用又较复杂,大多数国外研究者较少使用该软件,同时尽管 AMOS 操作简单,但在模型修正等方面不太便利,因此,它更适合初学者使用。相对而言,国外越来越多的研究者使用 Mplus 软件进行分析,它的好处是:经常更新,使用不复杂,模型的修正便利,与其他数据分析软件的兼容性也较好。因此,本研究推荐使用 Mplus 软件作为量表编制者进行 CFA 分析的工具。

在信度方面,大多数量表编制文献报告了分量表或全量表最终版本的内部一致性,即 Cronbach's  $\alpha$  系数(如傅绪荣等<sup>[32]</sup>的研究)。这与一些研究者的分析结果一致<sup>[7]</sup>。R. F. Devellis<sup>[1]</sup>认为:内部一致性是最广泛使用的可靠性度量,它与量表内各项目的同质性有关。鉴于其重要性,未来的研究应该将内部一致性评估作为测量可靠性的重点,同时由于心理特质相对稳定,因此,还应考查量表的重测信度(如张瑞雪等<sup>[33]</sup>的研究),在项目个数为偶数且同质的情况下也应考查分半信度(如张少华等<sup>[34]</sup>的研究)。由于验证因素分析方法的盛行,求取量表的组合信度也越来越便利,因此,有必要报告量表的组合信度以便得到量表信度更全面的证据。在信度系数的计算方面,本研究发现:少数编制量表的信度值低于 0.60,这低于通常量表所要求的信度标准<sup>[21]</sup>,量表的信度过低会影响到量表测量结果的可靠性,从而影响其使用价值。

本研究的结果表明:在样本数量方面,大多数研究中的样本数量大于经验法则所要求的数量(即样本人数与项目数量最低所需比例 10:1)。在项目数量方面,大多数研究者编制的最终量表删除了初始项目集中的大量项目,在 47.3% 的研究中最终量表的项目数较初始项目集减少了 50% 及以上。如刘视

湘等<sup>[35]</sup>编制了一个由 231 个项目组成的初始项目集,最终形成的量表包含的项目数仅有 12 个。项目既可能因为内容的有效性不高而被删除,也可能因心理计量学分析结果未达到相应的要求而被删除。因此,量表编制者应充分预计在量表编制过程中可能会有项目数量损失,尽可能编制一个较大的初始项目集。R. F. Devellis<sup>[1]</sup>提出,初始项目集应该是所需项目数量的 3~4 倍。

### 3.5 关于量表编制的局限

与 F. F. R. Morgado 等<sup>[36]</sup>研究类似,本研究发现有文献中量表编制过程也存在类似的 7 个方面的局限。这些局限会影响到量表的质量,阻碍它们的推广使用。

(i) 样本特征局限。这里所指的样本特征局限是采用方便取样的原则收集数据而导致量表推广的局限。J. C. Nunnally<sup>[3]</sup>建议:为了扩展新量表的可推广性,应在数据收集方面考虑样本的多样化。未来的量表编制研究应该考虑这一建议,招募异质的、真正随机的样本来评估新量表的结构效度和信度。

(ii) 方法学局限。这里所指的方法学局限是大多数量表项目采用自我报告方法进行作答。N. D. M. Mahudin 等<sup>[37]</sup>认为:这可能增加受试者的偏见、社会赞许性、需求特征和反应集的可能性,从而影响测量结果的有效性。因此,未来的研究还可以加入其他客观或独立的措施,以补充在编制新量表时对所研究变量的主观评估,并改进对结果的解释。

(iii) 心理计量学局限。心理计量学方面的第 1 个局限涉及量表的结构分析方面。一些研究仅采用 EFA 或 CFA 对量表的结构进行分析,未选取 2 个不同的样本分别采用这 2 种方法对量表的结构进行验证,这可能导致所得到量表结构存在样本依赖性和方法依赖性,无法进一步推广使用。在采用 EFA 方法对量表的结构进行探索时,一些研究并没有根据 KMO 值和 Bartlett 球形检验等结果来考查因素分析的适当性。在 CFA 中,一些研究中所获得的模型的数据拟合指数值并没有达到相应的标准,这些局限都影响结果的有效性。此外,大多数 EFA 和 CFA 分析结果没有列出相应的因素负荷值,这也影响了所编制量表结构的可重复性。S. Sveinbjornsdottir 等<sup>[19]</sup>指出:只有通过量表结构和信度验证,才能证明新量表的有效性和可靠性。对量表作相关分析来



考查效度大小所选择的工具或变量不合理是第2个心理计量学局限。M. F. F. D. Morgado 等<sup>[38]</sup>的研究提供了一个有效的替代方案。研究者使用在社会人口调查问卷上收集的信息(如教育水平和体力活动强度)与量表的结构相关联。心理计量学的第3个局限是所有的量表编制者都没有对量表的等价性或项目是否存在偏差进行检验(可能量表编制者另写其他论文进行这方面的研究)。不同的被试群体不仅存在特质水平的差异,而且也可能具有不同的特质结构。不加证明地认为所有被调查的被试都有相同的特质结构,这可能导致量表的不合理使用。

(iv) 定性研究的局限性。定性研究的第1个局限是仅使用演绎方法或仅使用归纳法生成项目。演绎方法是基于已有文献中有关信息演绎出量表结构的概念,它可能并不完全包含当前被试者心理结构,而归纳法是根据有限的目标群体样本归纳出心理结构信息,它可能未能包含心理结构的全部概念域甚至产生有偏的概念域。因此,只有将这2种方法结合才能产生可靠和有效的心理结构的概念域。M. J. Sewitch 等<sup>[39]</sup>认为,没有定量评估内容有效性是量表编制的第2个局限。本研究注意到,一些研究没有清晰地报告是如何对这些项目进行内容分析的,这种模糊性可能增大研究者在根据内容作出保留项目决定时的主观性,从而降低了量表编制的科学性。内容有效性是确保使用最终量表形式做出任何推论有效性和可信度的重要一步,它是所有量表在编制过程中的必然要求,精确地给出一个具体内容有效性的值能够让研究者清晰地看出量表的有效性高低。

(v) 范式局限。大多数文献采用 Likert 范式编制量表。一些个体在作答 Likert 式量表时会作假,尤其是在高利害的测评情景中。不仅作假会影响量表的预测效度和结构效度,而且作假会对个体成绩排名产生影响,从而严重影响到有关组织录用决策<sup>[40-43]</sup>。此外, Likert 量表的编制程序是建立在项目反应的优势过程假设的基础上,即假设随着个体特质水平的升高,赞成项目的概率水平也升高。有些研究者<sup>[44]</sup>认为:该假设与在被试对一些项目进行反应时的实际心理过程是不一致的。项目反应过程的错误限定可能影响测验的效度。

(vi) 项目局限。主要局限是大部分研究中编制的量表没有反向评分的项目。虽然一些方法学者认为,为了避免参与者之间的默许,反向评分是必要

的,但应谨慎对待这一建议。有些研究者<sup>[45]</sup>认为:反向评分的项目可能会让参与者感到困惑,并且反向评分的项目结构可能与量表结构不同,反向评分的项目往往是因子分析中拟合程度最差的项目,或者会产生一个与项目措辞有关的方法学因素,从而影响量表的结构效度。未来的量表编制者应该注意这方面的问题,在避免参与者之间的默许和防止其他一些与使用反向分数相关的问题之间权衡取舍。

(vii) 编制理论局限。大多数研究者基于经典测验理论(CTT)编制项目。虽然 CTT 有一定的优势(如计算方便和简单、所需要的样本量小等),但基于 CTT 的心理计量学分析结果严重依赖于所选取的样本,并且 CTT 没有对测量误差进行分解。未来可以结合概化理论(GT),精心设计量表,对测量误差进行分解。根据 G 研究和 D 研究的结果,确定相对优化的量表信度系数。同时也可以根据项目反应理论考查量表的信息函数值大小,根据不同的测量目标要求将测量误差控制在期望的水平内。

## 4 结论

本研究回顾了近10年来在一些中国心理学类期刊上发表的量表编制类文献,根据量表的编制步骤对量表编制的现状进行了分析,并指出其局限和提出一些改进建议,从而得出:(i)自2012年以来,量表编制的文献及其被引用量都呈相对下降趋势;(ii)所编制的量表在项目筛选、量表结构分析、信效度检验等方面存在一些不足,这些不足影响了量表的质量。

## 5 参考文献

- [1] Devellis R F. Scale development: theory and applications [M]. London: Sage Publications Ltd 2003.
- [2] Clark L A, Watson D. Constructing validity: basic issues in objective scale development [J]. Psychological Assessment 1995 7(3): 309-319.
- [3] Nunnally J C. Psychometric theory [J]. American Educational Research Journal 1978 5(3): 83.
- [4] Pasquali L. Instrumentação psicológica: fundamentos e práticas [M]. Porto Alegre: Artmed 2010.
- [5] Hutz C S, Bandeira D R, Trentini C M. O que é avaliação psicológica-métodos, técnicas e testes [EB/OL]. [2020-12-

- [19]. [https://www.researchgate.net/publication/301540536\\_O\\_que\\_e\\_avaliao\\_psicologica\\_-\\_metodos\\_tecnicas\\_e\\_testes](https://www.researchgate.net/publication/301540536_O_que_e_avaliao_psicologica_-_metodos_tecnicas_e_testes).
- [6] Hinkin T R. A review of scale development practices in the study of organizations [J]. *Journal of Management*, 1995, 21(5): 967-988.
- [7] Kapuscinski A N, Masters K S. The current status of measures of spirituality: a critical review of scale development [J]. *Psychology of Religion and Spirituality*, 2010, 2(4): 191-205.
- [8] Arias M R M, Lloreda M J H, Lloreda M V H. *Psicometría* [M]. SA: Alianza Editorial, 2014: 25-66.
- [9] 周露, 郝加虎, 唐玲, 等. 妊娠期生活事件问卷编制及项目筛选 [J]. *中国公共卫生*, 2015, 31(8): 1026-1029.
- [10] Churchill G A. A paradigm for developing better measures of marketing constructs [J]. *Journal of Marketing Research*, 1979, 16(1): 64-73.
- [11] Podsakoff N P, Podsakoff P M, Mackenzie S B, et al. Are we really measuring what we say we're measuring? Using video techniques to supplement traditional construct validation procedures [J]. *Journal of Applied Psychology*, 2013, 98(1): 99-113.
- [12] 戴海崎, 张锋, 陈雪枫. *心理与教育测量* [M]. 广州: 暨南大学出版社, 1999.
- [13] Hair Junior J F, Black W C, Babin N J, et al. *Análise multivariada de dados* [M]. São Paulo: Bookman, 2009.
- [14] Mackenzie S B, Podsakoff P M, Podsakoff N. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques [J]. *Social Science Electronic Publishing*, 2011, 35(2): 293-334.
- [15] 方晓义, 袁晓娇, 胡伟, 等. 中国大学生心理健康筛查量表的编制 [J]. *心理与行为研究*, 2018, 16(1): 111-118.
- [16] 王馥芸, 李志强, 秦启文. 中国成年人自省问卷编制 [J]. *中国临床心理学杂志*, 2020, 28(3): 465-470.
- [17] Bastos J L, Celeste R K, Faerstein E, et al. Racial discrimination and health: a systematic review of scales with a focus on their psychometric properties [J]. *Social Science and Medicine*, 2010, 70(7): 1091-1099.
- [18] Ladhari R. Developing e-service quality scales: a literature review [J]. *Journal of Retailing and Consumer Services*, 2010, 17(6): 464-477.
- [19] Sveinbjornsdottir S, Thorsteinsson E B. Adolescent coping scales: a critical psychometric review [J]. *Scandinavian Journal of Psychology*, 2010, 49(6): 533-548.
- [20] 姚本先, 王莲, 全莉娟, 等. 当代小学生理想问卷的编制及其应用 [J]. *心理与行为研究*, 2015, 13(6): 763-770.
- [21] Michael F R. *Scale construction and psychometrics for social and personality psychology* [M]. Thousand Oaks, CA: Sage Publications Ltd, 2011.
- [22] Hardesty D M, Bearden W O. The use of expert judges in scale development: implications for improving face validity of measures of unobservable constructs [J]. *Journal of Business Research*, 2004, 57(2): 98-107.
- [23] Malhotra N K. *Pesquisa de marketing. Uma orientao aplicada* [EB/OL]. [2020-12-20]. [http://disciplinas.stoa.usp.br/pluginfile.php/2022219/mod\\_folder/content/0/Resumo%20do%20livro%20Pesquisa%20de%20Marketing.%20Uma%20orienta%C3%A7%C3%A3o%20Aplicada\\_.pdf?forcedownload=1](http://disciplinas.stoa.usp.br/pluginfile.php/2022219/mod_folder/content/0/Resumo%20do%20livro%20Pesquisa%20de%20Marketing.%20Uma%20orienta%C3%A7%C3%A3o%20Aplicada_.pdf?forcedownload=1).
- [24] 刘影, 桑标, 朱仲敏, 等. 中学生学业情绪调节问卷的编制 [J]. *中国临床心理学杂志*, 2018, 26(1): 1-5.
- [25] Roberson R B, Elliott T R, Chang J E, et al. Exploratory factor analysis in Rehabilitation Psychology: a content analysis [J]. *Rehabilitation Psychology*, 2014, 59(4): 429-38.
- [26] Koen L, Free D B, Nadine E, et al. Development of the self-regulated learning teacher belief scale [J]. *European Journal of Psychology of Education*, 2009, 24(1): 79-96.
- [27] Gottlieb U, Brown M, Ferrier L. Consumer perceptions of trade show effectiveness [J]. *European Journal of Marketing*, 2013, 48(1/2): 89-107.
- [28] Khine M S. Knowing, knowledge and beliefs [EB/OL]. [2020-12-20]. <http://dx.doi.org/10.1007/978-1-4020-6596-5>.
- [29] 温忠麟, 侯杰泰, 马什赫伯特. 结构方程模型检验: 拟合指数与卡方准则 [J]. *心理学报*, 2004, 36(2): 186-194.
- [30] Reise S P, Waller N G, Comrey A L. Factor analysis and scale revision [J]. *Psychological Assessment*, 2000, 12(3): 287-297.
- [31] 赵欢欢, 许燕, 张和云. 中国人敬畏特质的心理结构研究 [J]. *心理学探新*, 2019, 39(4): 345-351.
- [32] 傅绪荣, 汪凤炎. 整合智慧量表的编制及信效度检验 [J]. *心理学探新*, 2020, 40(1): 50-57.
- [33] 张瑞雪, 肖蓉. 物质主义倾向量表的编制及其信效度 [J]. *中国临床心理学杂志*, 2019, 27(4): 685-689.
- [34] 张少华, 桑标, 刘影. 青少年情绪表达灵活性问卷的编制 [J]. *中国临床心理学杂志*, 2018, 26(2): 239-243.
- [35] 刘视湘, 董洪杰. 农转居居民社区感问卷的编制 [J]. *中国临床心理学杂志*, 2019, 27(1): 44-48.

- [36] Morgado F F R ,Meireles J F F ,Neves C M ,et al. Scale development: ten main limitations and recommendations to improve future research practices [J]. *Psicologia Reflexo E Crítica* 2018 ,30( 1) : 3.
- [37] Mahudin N D M ,Cox T ,Griffiths A. Measuring rail passenger crowding: scale development and psychometric properties [J]. *Transportation Research Part F Psychology and Behaviour* 2012 ,15( 1) : 38-51.
- [38] Rocha M F F D ,Neves B C A N ,Fernandes Tavares ,et al. Development and validation of the self-acceptance scale for persons with early blindness: the SAS-EB [J]. *Plos One* 2014 ,9( 9) : e106848.
- [39] Sewitch M J ,Abrahamowicz M ,Dobkin P L ,et al. Measuring differences between Patients' and Physicians' health perceptions: the Patient-Physician discordance scale [J]. *Journal of Behavioral Medicine* ,2003 ,26( 3) : 245-264.
- [40] Jackson D N ,Wroblewski V R ,Ashton M C. The impact of faking on employment tests: does forced choice offer a solution? [J]. *Human Performance* 2000 ,13( 4) : 371-388.
- [41] Marcus B. Relationships between faking validity and decision criteria in personnel selection [J]. *Psychological Test and Assessment Modeling* 2006 ,48( 3) : 226-246.
- [42] Komar S ,Brown D J ,Komar J A ,et al. Faking and the validity of conscientiousness: a Monte Carlo investigation [J]. *Journal of Applied Psychology* 2008 ,93( 1) : 140.
- [43] Honkaniemi L ,Tolvanen A ,Feldt T. Applicant reactions and faking in real-life personnel selection [J]. *Scandinavian Journal of Psychology* 2011 ,52( 4) : 376-381.
- [44] 邓稳根 ,戴海琦 ,黄海 ,等. 展开模型对 EPQ 反应数据的拟合调查 [J]. *心理与行为研究* ,2009 ,7( 4) : 300-303.
- [45] Cicero David C ,Kerns John G ,McCarthy Denis M. The aberrant salience inventory: a new measure of psychosis proneness [J]. *Psychological Assessment* ,2010 ,22( 3) : 688-701.

## The Current Situation ,Problems and Suggestions of Psychological Scale Development in China over the Last Ten Years

DENG Wengen<sup>1</sup> ZHANG Wenli<sup>1</sup> GUO Lei<sup>2</sup>

( 1. School of Psychology ,Hainan Normal University ,Haikou Hainan 571158 ,China;

2. Faculty of Psychology ,Southwest University ,Chongqing 400715 ,China)

**Abstract:** The scale development process is critical to building knowledge in psychology and behavior science. Based on 222 articles published in five psychological journals from January 1 2010 to June 30 2020 ,the literature status of the scale published in these five journals in the past 10 years are systematically reviewed. Then these documents are analyzed from the three basic steps of the scale compilation that are project generation ,content analysis and psychometric analysis ,and the main limitations such as sample characteristics limitations ,method limitations ,psychometric limitations ,and qualitative research limitations and limitations of project preparation are pointed out in the current literature. Through the analysis of these limitations ,it is helpful for future research ,especially in making the most appropriate choice between different concepts and methods.

**Key words:** scale development; item writing; content validity; psychometrics

( 责任编辑: 冉小晓)