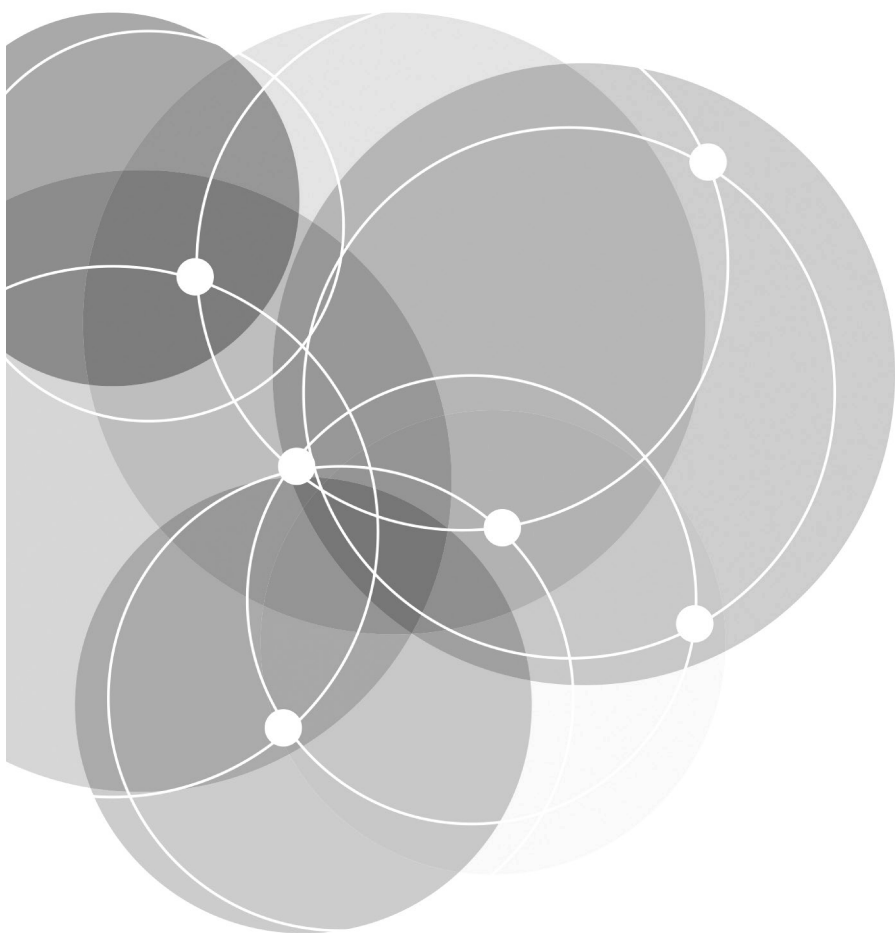


绪 论



1. 积极心理测评工具的特征、选择与评价

一、积极心理测评工具的特征

量表的标准化是量表编制中必需且重要的一环。也就是说,量表编制成稿之后经过标准化才算正式完成。编制后的量表能否投入使用,主要取决于其标准化特征和程度。

1. 取样(sampling)

按传统的测验学理论,对测验或评定的解释是,以被测或被评者的所属群体为标准,在编制测验或量表无法测量这个群体中所有成员时,便只能取样,即以样本代表全体(汪向东,王希林,马弘,1999)。样本具有代表性才能保障测评工具具有较好的信度和效度。否则可能导致测评工具的信度和效度不高,导致结果解释存在偏差。提高样本的代表性,除了要求样本的数量达到要求之外,还应该注意所测评内容(如积极心理测评中的生活满意度、生命意义感、情绪智力等)的所有有关影响因素(如年龄、性别、文化程度、民族、生活地区等)。

心理测评常用的取样方法分简单随机抽样、整群抽样、分层比例抽样和机械抽样等。

(1) 简单随机抽样: 包括直接抽选法、抽签法、随机数字表法。

(2) 整群抽样: 先将总体分为多个群,然后随机抽取其中若干个群,对这些群内所有个体或单元均进行调查。

(3) 分层比例抽样: 先将总体中所有的单位按照某个标志分成若干类(组),然后在各个类中分别随机抽取样本。

(4) 机械抽样: 先将抽样总体单位按照一定顺序排队,根据总体单位数和样本单位数计算出抽选间隔(抽选距离),然后按照一定的间隔抽选样本单位。由于抽选间隔相等,所以也叫等距抽样。

不管用何方式,都要使样本有足够代表性。因此,在选择测量工具时必须注意编制此工具的样本代表性和所要评估的对象情况是否在样本内。只有受测者

的情况在这些方面与标准样本相符,所测结果与标准样本才有可比性。

2. 常模(norm)

常模是一种供比较用的标准量数,由标准化样本测试结果计算而来,其形式大致有如下几种。

(1) 均数。某一受评者所得分数与标准化样本的平均数相比较时,便能确定其水平的高低。

(2) 标准分。用标准分作常模,可提供更多的信息。标准分能说明受测者的评定成绩在标准化样本的成绩分布图上所在位置。许多评定量表采用这种常模可衍化出其他形式,如 T 分。

(3) 百分位。将得分低的排列在下,得分高的在上,计算出样本分数的各百分位范围。如 P25,说明样本中 25% 的得分在该受评者的得分之下(或至多和他的得分一样),另有 75% 人数的得分比他要好。以此类推。

(4) 划界分。在积极心理测评工具中,特别是某些用于特殊人群的测评工具,由于其样本常常不是正态分布,不具备制定标准分常模的条件,而多采用划界分常模。划界分是指用一具体量数,对评定结果进行划界。

除上述常模形式外,还有其他性质的常模,例如年龄常模(按年龄分组建立、在儿童和老年人量表中常用),区域常模和各种疾病诊断常模,等等。从可比性来看,常模越特异越有效;从适应性来讲,则通用常模使用更方便。

3. 信度(reliability)

信度是指测验或量表的一致性、稳定性及可靠性,一般多以内部一致性来加以表示该测验信度的高低。信度分析的常用具体方法有重测信度、复本信度、分半信度、信度系数法四种(后两种可归为内部一致性信度)。

(1) 重测信度(test-retest reliability): 同一组受测者在两次不同时间作同一套量表评定,对两次结果作相关性检验,以评估量表结果的稳定性。重测信度所考查的误差来源是时间的变化所带来的随机影响。在评估重测信度时,必须注意重测间隔的时间。重测信度的时间间隔的选择存在两难问题: 缩短两次测试的时间间隔,受测者较容易回忆出测试的题目;而延长两次测试的时间间隔,则受测

者较容易受外部影响而变化(风笑天, 2006)。一般情况而言, 重测间隔在两周到6个月之间比较合适。

(2) 分半信度(split-half reliability): 将一套量表的各项目按奇、偶数号分成两半, 对其评定结果进行相关性检验。

(3) 同质性信度(inter-item consistency): 所谓同质性信度是指评定量表内部所有项目间的一致性, 这里讲的是分数的一致, 而不是项目内容或形式的一致。量表内各项目分数相关越高, 则量表项目就越同质。另外, 分半相关法也是估计同质性信度的一种方法。

(4) 评定者信度(inter-rater reliability): 数名不同评定者采用同一套量表对相同受评者进行评定, 对所得结果进行一致性检验, 以评估评定量表评分的客观性。对心理测验而言, 一般都采用客观性项目, 且有一套相当标准化的评分程序, 因此由评分引起的误差变异是可以忽略的。而评定量表相当多的是主观项目, 且评定者在评分时或多或少掺杂主观判断成分, 故评定者之间的评分误差变异难以避免。一般要求在成对的受过训练的评定者之间平均相关系数达到0.9以上, 才认为评分是客观的。

除以上指标外, 还有正复本相关、因素信度、测量标准误等信度检验方法。

4. 效度(validity)

效度即有效性, 指一个测量工具能够准确测查出所要测量的变量的程度。积极心理测评工具的效度检验方法与心理测验一样, 分为三类, 即内容效度(content-related validation)、效标效度(criterion-related validation)和结构效度(construct-related validation)。

(1) 内容效度, 又称表面效度, 指量表项目反映所测量内容的程度。内容效度评估方法有专家判断法、统计分析法(评分者信度、复本信度、折半信度、再测法)和经验推测法(实验检验)。

(2) 效标效度, 是指一个量表对处于特定情境中的个体的行为进行预测的有效性, 也就是对于我们所感兴趣的行为能够预测的程度。检验实证效度一般是将量表评定结果与某一标准行为(即效标)进行相关检查。如将个体的抑郁量表与抑郁症病人的临床典型症状、诊断性量表与临床诊断进行相关分析。如果二者相

关程度(或一致性程度)高,则说明效标效度理想;反之认为效标效度不理想。

(3) 结构效度,又称构想效度,反映编制的量表所依据理论的程度。如编制一个心理弹性量表,必定依据有关的心理弹性理论。同时,所编制的量表是否符合原来预计的理论框架,也可用结构效度来检验。

量表的效度研究,不管是在量表编制过程中,还是编制之后都要持续地进行,并且是一项长期的、永无止境的工作。量表的效度材料越丰富,对量表的功能认识就越全面。

5. 方法标准化

积极心理测评工具必须按一定的规则实施,其记分方法和方式、分级的标准、结果的换算也要按一定的规则进行,只有这样,评定结果才具有较高的真实性。

积极心理测评工具标准化特征,并不是每个测评工具都要像标准化测验那样应全部具备,有的量表可能多一些,标准化程度要高一些;有的可能少些,标准化程度就不那么好。至于每个具体量表做了哪些标准化工作,如常模样本、施测方法、信度和效度资料,都应尽可能详细地在使用手册上说明。

二、积极心理测评工具的选择与评价标准

借鉴 Riezen 和 Segal (1988)提出的一整套评价量表的原则,我们认为积极心理测评工具的选择与评价标准主要包括以下几个方面。

1. 量表的功效

量表的功效是指所使用的量表能够全面、清晰地反映所要评定的内容特征和真实性的程度。有的量表可评定多个方面的特质,而另一些量表则只限于评定 1~2 种特质。有的量表适用于所有年龄和各种类型的人群,而另一些量表可能只限于某一年龄阶段或某一特殊人群。质量好的量表应该对项目描述具体、清晰,等级划分合理,定义明确,以反映出心理行为的细微变化。出现的频度或严重程度分极最好采用 3~7 级划分。量表应尽可能简短,又不损失必要的细节。

2. 敏感性

敏感性是指选择的量表应该对所评定的内容敏感,即能测出受评者某种特质、行为或程度上的有意义的变化。量表的敏感性既与量表的项目数量和结果表达形式(如因子分)有关,又受量表的标准化程度和信度高低影响。此外,评定者的经验和使用量表的动机也影响量表的敏感性。

3. 简便性

简便性是指所选择的量表应尽量简明、省时和方便实施。然而实际上,量表简短、省时就难以做到全面。使用者不加训练和采用非标准化方法就会降低量表的信度,影响结果的可靠性。几个量表同时配合使用,能弥补单一量表这方面的缺陷。

4. 可分析性

使用量表的目的就是要对测评对象的特质、行为或现象作质与量的评价。一般而言,量表应具有其比较标准,或者是常模,或者是描述性标准。量表中单项分、因子分及总分都是常用分析指标,总分常反映受评者总的情况和变化;而单项分、因子分则是分析这些方面变化的主要指标。把单项分或因子分画成曲线或构成廓图,受评者的某方面心理特质、行为特征或社会背景情况等特点会更直观、清晰。

参考文献

1. 汪向东,王希林,马弘,1999. 心理卫生评定量表手册(增订版)[J].中国心理卫生杂志社.
2. 风笑天,2006. 社会研究方法(第一版).北京:高等教育出版社.
3. Riezen H. V. & Segal M, 1988. *Comparative Evaluation of Rating Scales for Clinical Psychopharmacology*[M]. Elsevier.

2. 积极心理团体干预的效果评估方法与数据分析

王梦南¹ 倪士光²

(1. 清华大学心理学系, 北京, 100084;

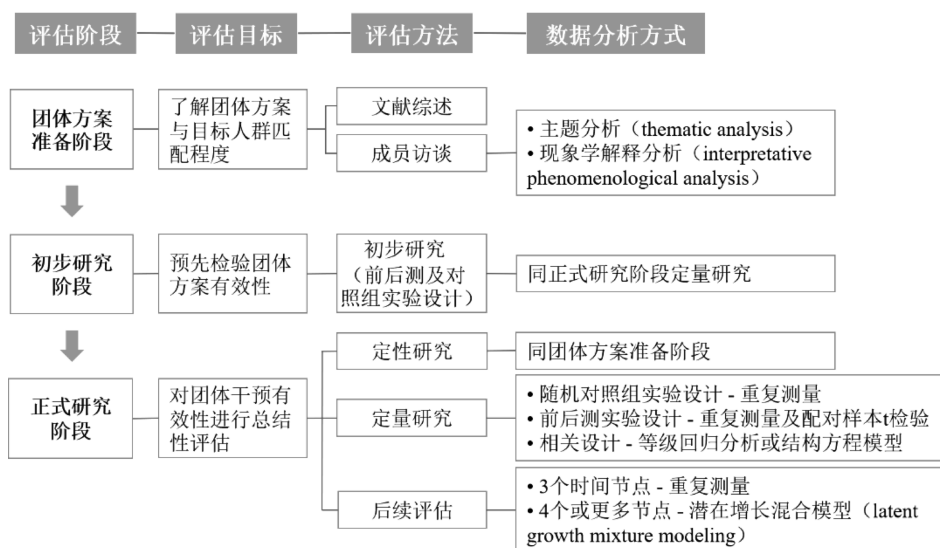
2. 清华大学深圳国际研究生院, 深圳, 518055)

【摘要】 近年来团体干预在中国发展迅速,但由于团体干预效果的评估方法的规范性匮乏,团体干预的效度亟待更为充分的研究和验证。本文以积极心理团体干预(Group-Based Psychological Interventions, G-PPIs)为例,梳理了团体效果评估的团体方案准备、初步研究和正式研究评估等三个阶段,总结了效果评估的研究设计及数据分析方法,为研究者们进行团体干预的效果评估提供规范性、科学性的方法参考。

积极心理干预是以积极心理学为理论指导的心理干预方案,主要针对提高积极情感、积极认知或积极行为(Sin & Lyubomirsky, 2009)。团体心理干预是在团体情境中提供心理辅导、心理咨询或心理治疗的一种干预形式。在帮助那些有着共同发展主题和相似心理困扰的人时,团体辅导相比较个别咨询而言更加经济有效(樊富珉, 2005)。随着积极心理学在团体情境中的适用性为更多人所了解,越来越多的团体心理干预采用了积极心理学取向。积极心理团体干预是在团体情境下,运用积极心理干预的理念和技术提供心理指导的干预方式。

虽然积极心理团体干预运用逐渐广泛,但是能否客观地评价团体心理干预的效果仍是影响团体心理干预研究效度的重要议题。王茹婧等(2017)对国内 89 篇团体咨询研究文章进行分析后指出,目前国内团体咨询效果评估呈多样化特点,一部分研究存在误用、滥用效果评估方法从而得出团体咨询效果显著的结论,一部分在研究设计上需要进一步改进。为有效提升积极心理团体干预效果评估的科学性和客观性,我们有必要进行量化的研究设计及数据分析。它不仅为研究者提供了团体干预研究设计的风险与收益的证据,也提供了团体干预研究设计的潜在问题,帮助研究者们做出更加严密的设计,完善团体干预方案。因此,本文旨在

为积极心理干预,主要是积极心理团体干预的效果评估,提供研究设计及数据分析方法的参考和指导。团体咨询的效果评估通常包含团体方案准备、初步研究和正式研究三个阶段(Sreevani & Reddemma, 2013),本文将总结各阶段的评估目标和方式,并提出针对不同评估阶段的研究设计及数据分析方法。具体如图所示。



团体干预效果评估的各阶段研究设计及数据分析方法

一、团体效果评估阶段

1. 团体方案准备阶段

主要目的是了解团体方案设计与目标人群行为需求的匹配程度。团体方案设计需要根据目标人群的亚行为特征进行调整,不能直接采用未进行本土化的外国方案(王茹婧等, 2017)。本阶段效果评估建议采取质性评估方式。以文献综述及成员访谈的形式,了解目标人群需求的影响因素以及相关领域内已有的有效干预方式,帮助制定与目标人群需求更为契合的团体干预方案。

2. 初步研究阶段

初步研究是在正式干预实施之前,对目标人群中的少量成员进行试点,目的

是预先检验团体干预方案的有效性,并根据初步研究结果修改和完善方案,以获得最佳效能,减少方案的潜在风险。王茹婧等人(2017)的研究发现,仅有 3.4% 的国内团体研究使用了初步研究。鉴于初步研究在保障团体方案有效性上的重要作用,建议在进行正式团体干预前设计初步研究,并以量化方式进行评估,如设置对照组或前后测等方式,通过数据分析保障正式研究的可预测性、可行性和有效性。

3. 正式研究阶段

在正式研究阶段,效果评估的重点是对正式团体干预方案进行总结性评估。通过测量大样本群体,收集反映团体成员变化的证据,并依此判断团体方案是否适用于样本对应群体。值得注意的是,在进行总结性评估时应注重评估角度的多元化。评估对象应该包括团体及个人层面,评估执行者也应当加入团体领导者、督导及观察员等。在收集证据时,既有定性的数据,也有定量的数据。在进行效果评估时也应适当结合过程因素,如团体氛围或成员投入情况。若条件允许,建议进行后续评估,对研究效果的持续性进行检验。

二、团体效果评估方法及数据分析

1. 定性方法(quantitative study)

定性数据包括成员主观报告、绘画作品等,其中成员访谈是收集定性数据的主要方式。成员访谈包括对成员个体的半结构式访谈和焦点小组访谈,围绕明确的访谈主题通过开放式问题了解成员的需求或体验。

访谈或其他质性数据可以通过主题分析(thematic analysis, TA)或者现象学解释分析(interpretative phenomenological analysis, IPA)的方法进行分析,并依此对团体方案进行修正完善。

2. 定量方法(qualitative study)

对团体干预效果进行量化评估的方式包括标准化测验、自编问卷和行为量化等。在标准化测验中,全人健康量表(holisticwell-being scale, HWBS)是测量积极心理团体干预有效性的方式之一(Chan, 2014)。该量表具有较好的内部一致

性和跨时间的稳定性,内部一致性 α 系数范围为 0.63~0.87;探索性和验证性因素分析的结果均表明 HWBS 的两维度及七因子的结构效度良好。行为量化包括某特定行为发生情况的量化数据,以及生理指标如皮质醇水平、血压、皮肤电等,相关数据可直观反映个体的情绪或其他心理变化。当已有量表不能准确测量相关问题或不适用于目标人群时,许多研究者亦采用自编量表。但在使用自编量表时,需注意信度及效度的检验,并在研究报告相应指标以确保量表的有效性。

3. 实验设计

(1) 随机对照实验设计(randomized controlled trial, RCT)

随机对照实验运用了严谨的实验设计,将团体成员随机分配到实验组和对照组,最大限度排除干扰因素,以验证团体干预的有效性是否高于其他干预方法。值得注意的是,随机对照实验可能产生伦理问题,即被分配到对照组的人们与实验组虽然有同样的问题或需求却无法得到有效干预。此时可以采用等待序列随机对照实验(waiting-list randomized controlled trial, WL-RCT),将目标人群随机分为 2 组,一组作为实验组在较早时间点 T_0 接受干预,另一组在正常时间点 T_1 接受干预,在中间等待(T_0-T_1)期间后者作为前者的对照组。

重复测量(repeated measures)可用于 RCT 及 WL-RCT 的数据分析。

(2) 前后测实验设计(pre-post test)

前后测实验设计通过比较成员接受团体干预前与接受团体干预后的数据验证团体干预方案的有效性。根据是否设置对照组,前后测实验设计可分为实验组前后测实验设计和实验组、对照组前后测实验设计。实验组前后测实验设计由于操作简便同时保证有效性,常被用于仅面向少数成员的初步研究的效果评估中。实验组、对照组前后测实验设计则更为严谨,能较为准确地反映团体干预方案的有效性。

重复测量和配对样本 t 检验可用于前后测实验设计的数据分析。

4. 相关设计(correlational design)

由于条件或伦理限制,可能出现不可以设置对照组或只接受一种干预的实验组的情况,此时可采用相关设计来检验干预的有效性。成员在干预前接受前测

(T_0),在干预后接受后测(T_1)。通过比较成员前后测获取的心理健康数据,我们可以了解成员的心理健康是否发生了变化。同时在 T_1 对成员的顺从性和满意度进行检验。通过分析成员的健康变化与顺从性、满意度的相关,我们可以对团体干预的有效性进行评估。

对于小样本团体,可以采用逐步回归分析(hierarchical regression analysis)验证成员心理健康变化和一致性、满意度的相关性,成员一致性和满意度在心理健康水平改变中所占的方差标志着干预的有效性。对于较大样本,则需采取更高级的统计方式,如结构方程模型(structural equation modeling),除前述方差值外,还需参考拟合优度指标(goodness of fit indices)以验证有效性。

5. 后续评估(follow-up assessment)

由于团体心理干预常需要一段时间起效,只用前后测方式检验两个时间节点的变化可能不足以证明团体干预的有效性。因此如预算、时间及人力允许,建议进行后续评估。后续评估可以解释由团体心理干预引发的心理发展轨迹。对于 3 个时间节点的评估,可以用重复测量检验干预效果;对于 4 个或更多的时间节点,则可以采用潜在增长混合模型(latent growth mixture modeling)进行检验。

虽然近年来团体咨询在我国已进入专业化发展时期,但相关研究的质量及多样性均有待提高。国外的团体研究经历了从效果研究到过程—效果研究,从单一到丰富的主题,从简单到严谨的研究方法的过程(邵瑾,2015)。国外团体研究的历史及发展趋势可以为国内的团体研究提供指引。由于团体干预是一个动态发展的过程,因此有必要在效果评估中增加过程因素的研究,如团体过程、团体氛围、成员互动等。效果评估也应贯穿团体干预的全程,不同阶段采用不同的侧重点和研究方法。

参考文献

1. Chan C. H., Chan T. H., Leung P. P., Brenner M. J., Wong V. P., Leung E. K., Chan C. L., 2014. Rethinking well-being in terms of affliction and equanimity: development of a holistic well-being scale[J]. *Journal of Ethnic and Cultural Diversity in Social Work*, 23(3-

- 4), 289—308.
2. DeLucia-Waack JL, 2014. 团体咨询与团体治疗指南[M]. 李松蔚, 鲁小华, 贾烜, 等, 译. 北京: 机械工业出版社.
3. Sin N. L. & Lyubomirsky S, 2009. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: A practice-friendly meta-analysis [J]. *Journal of clinical psychology*, 65(5), 467—487.
4. Sreevani R, Reddemma K, Chan CL, et al, 2013. Effectiveness of integrated body-mind-spirit group intervention on the well-being of Indian patients with depression: A pilot study [J]. *Journal of Nursing Research*, 21(3), 179—186.
5. 邵瑾, 樊富珉, 2015. 1996—2013 年国内团体咨询研究的现状与发展趋势[J]. *中国心理卫生杂志*, 29(4), 258—263.
6. 王茹婧, 樊富珉, 李虹, 倪士光, 2017. 2001—2016 年国内团体心理咨询效果评估: 现状、问题与提升策略[J]. *中国临床心理学杂志*, 25(3), 577—583.