

- (中医栏目编辑:王松月)

# 护理医学

• 45 •

## 2 量表的形成

### 2.1 概念化

量表研制应该从对欲测内容进行合理的概念化开始,以便量表中的各个条目可以完整地捕捉到概念内容或其包含的维度。理论在对测量内容的概念化方面发挥着重要的指导作用<sup>[6]</sup>。概念化也可通过相关文献综述或对目标人群进行访谈以及专家咨询等建立概念框架<sup>[1,2]</sup>。对概念的解析还可以通过概念分析、概念综合和概念衍生等方法来实现<sup>[7]</sup>。

### 2.2 决定量表类型

量表一般分为传统汇总评分量表(或Likert型量表)和潜在特征量表<sup>[1]</sup>。量表研制所依据的测量理论主要有经典测试理论(classical test theory, CTT)和项目反应理论(item response theory, IRT)<sup>[4]</sup>。传统汇总评分量表基于CTT,其条目被认为是潜在测量概念的粗略可比性指标,条目的集合增强了量表获得接近真实分数的程度,量表测量所观察到的分数(X)是真实分数(T)加误差(e)的和<sup>[6]</sup>。潜在特征量表是基于IRT的,假设每个单独的条目对潜在变量或欲测概念都有自己的特征敏感性,强调量表中条目的反应水平,常用项目特征曲线表示,该曲线揭示了条目的难度和区分度。量表测量过程的设计和结果解释需要基于相应的测量框架,如常模参考测试或标准参照测试<sup>[8]</sup>。

### 2.3 条目生成

量表的各个条目应具有单维性,即每个条目仅属于一个分量表或维度,并且这些条目内在同质,它们共同构成对欲测概念操作性定义的测量<sup>[1]</sup>。一组好的条目应该从与感兴趣的概念相关的广泛的条目领域中随机选择<sup>[6]</sup>。参考现有的相关量表和文献、进行概念分析、实施质性研究及临床观察等都可作为条目生成的来源<sup>[1,9]</sup>。

### 2.4 决定条目特征

条目一般包括题干和反应选项两部分。通常,Likert评分量表易于被受访者理解,尤其适用于对观点、信念和态度等概念的测量,表现为一致性(非常不同意/同意)、频率(从不/总是)、重要性(非常重要/不重要)、或可能性(极有可能/不可能)等的连续反应选项<sup>[1,2]</sup>,一般为5~7级评分。另外,Guttman评分量表常涉及一系列对所测概念属性等级递增水平的条目,它对条目进行排序,使同意某个条目的被试者也会同意其之前(或之后)的条

目,例如儿童道德发展水平的测量等<sup>[6]</sup>。语义差异量表的选项主要由一系列相反的形容词词对组成(如好的与坏的、积极的与消极的),并被划分为5~9个(一般为7个)等值的评定等级,用于测量概念的语义内涵<sup>[8]</sup>。研究者应根据所测概念的性质选择不同的反应选项。

一般在量表开发初期建立一个足够大的条目池,美国学者DeVellis推荐条目池中的条目数量应该是最终量表中条目数量的3~4倍,或至少比最终量表的条目数大50%<sup>[5,6]</sup>。量表中条目最好随机排序,避免邻近效应(即对条目选项的反应会受到其前一个条目选项反应影响的趋势)导致人为地夸大估计量表的内部一致性<sup>[1]</sup>。条目编制时应选择与目标人群的受教育程度和阅读水平相符的词语,避免模棱两可的措辞、避免多重否定、避免过长的句子或短语、避免使用行业术语、避免双重条目(即一个条目中表达了两个或多个想法)等<sup>[1]</sup>。一般而言,针对大多数普通人群的量表在五至七年级的阅读水平较为合适<sup>[6]</sup>。量表中反向条目的纳入可以减少默认反应定势(即受访者倾向于同意所有的问题或者表示肯定的意思)的可能性<sup>[1]</sup>。然而,美国学者DeVellis认为,量表中设置反向条目的利大于弊,因为对条目陈述极性的颠倒可能会让受访者感到困惑,且有许多反向条目的例子在实际应用中表现并不好<sup>[6]</sup>。

### 2.5 专家评价

专家评价是量表开发的必要阶段,它有助于最大限度地提高量表的内容效度<sup>[1,6]</sup>。专家小组成员需选择与所测概念和目标人群研究领域相关的资深专家,并至少纳入一名有量表研制经验的专家。专家组人数一般为3~5人<sup>[10]</sup>,国内文献较多采用德尔菲专家函询法纳入10~20名咨询专家对条目池内容进行筛选<sup>[11-13]</sup>。通常,每位专家应对每个条目的措辞是否清晰简洁,条目与每个维度的相关性,各维度与概念的相关性,以及条目是否适合目标人群等内容进行评价<sup>[1,6]</sup>。

数据收集后,需计算量表内容效度指数(content validity index, CVI),包括条目CVI(I-CVI)和量表CVI(S-CVI)。I-CVI的计算方法为对相关性的评分为1~4分的量表条目,评分为3或4分的专家人数除以专家总人数,即专家对相关性的比例。考虑机会一致性风险,当有5~8名专家时,I-CVI低于0.78的条目应进行修改或舍弃<sup>[1]</sup>。当专家人数 $\leq 4$ 时,I-CVI应达到1.00。S-CVI有

两种计算方法,首选平均S-CVI(S-CVI/Ave),即量表中所有条目I-CVI的平均值;另一种是专家普遍一致性S-CVI(universal agreement among experts,S-CVI/UA),它是所有专家评分为3或4分的条目数量除以总条目数量的比例<sup>[14]</sup>。一般S-CVI/Ave $\geq 0.9$ 及S-CVI/UA $\geq 0.8$ 即可<sup>[14,15]</sup>。另外,Cohen系数kappa(k)也是量表内容效度测量的有效方法,它是对机会一致性进行调整后的评价者间一致性共识指数,是对CVI的重要补充<sup>[14,16]</sup>。

## 2.6 预测试

预测试(或预调查)主要对量表在其实施过程中可能出现的问题进行早期识别和处理,其涉及的样本量较小,一般为20~150人<sup>[1,2]</sup>,也有学者纳入正式调查时目标人群样本量的10%作为预测试的样本量。量表内容的复杂性越大,样本量越大;目标人群的同质性越大,样本量也越大。数据分析时,研究者可进行条目初步分析,对无应答率高的条目、变异性差的条目、选中间选项多的条目以及极端选项比例高(地板效应或天花板效应)的条目应考虑删除或修改。

## 3 量表的评价

### 3.1 现场测试

现场测试旨在评估量表的心理测量学特征。其样本应充分代表目标人群,并具有一定的异质性(如年龄、教育背景、种族等差异),样本量需足够大<sup>[1,6]</sup>。Nunnally认为<sup>[2,6]</sup>,300例的样本量足以支持因子分析,而其他学者则建议受试者与条目的比率应在3:1至20:1之间,其中10:1被广泛采纳。在测量量表的重测信度时,其重测样本可以采用较小的样本量,如50~200例<sup>[17]</sup>。调查内容除量表本身外,还应受试者的一般人口学资料进行调查,或增加评估量表结构效度时可能需要测量的其他概念或变量的相关问卷或量表。一些自填式问卷可能受到社会赞许效应(即受访者倾向在调查中以不实意愿取代其真实意愿,以符合社会期望)的影响,产生不同程度的报告偏倚。通过声明量表并无正确或错误答案之分,可从一定程度上最小化受试者社会赞许效应。采用社会赞许效应量表,如马洛-克劳恩社会赞许效应量表(the marlowe crowne social desirability scale, MCSDS)可评估该效应对量表条目的影响程度<sup>[6]</sup>。

### 3.2 心理测量学评估

3.2.1 条目分析 基于CTT,条目之间的相关性可以反映条目与所测概念真实分数之间的相关性。通过条目间相关系数矩阵可分析条目间相关程度,条目-量表相关系数常采用校正后系数。通常,相关系数为0.30~0.70的条目性能较好,予以保留<sup>[1]</sup>。此外,计算每个条目的均值和方差,也是一种有用的条目性能检查方法。若条目得分均值偏于一个或另一个极端值,说明该条目的区分度较差<sup>[1]</sup>。若条目具有相对较大的方差,说明其样本的异质性较好<sup>[6]</sup>。也可采用临界比值法从条目的区分度方面筛选条目<sup>[18]</sup>。

3.2.2 信度分析 信度是指量表测量结果的可靠性、稳定性和一致性。内部一致性信度反映量表条目之间的一致程度,与条目的同质性有关,主要采用克隆巴赫 $\alpha$ 系数进行评价,条目间相关性越大, $\alpha$ 系数越高<sup>[1,6]</sup>。对于二分类式选项的量表,应采用Kuder-Richardson公式20(KR-20)计算该信度,它是 $\alpha$ 系数的特殊版本<sup>[3,6]</sup>。通常,量表信度系数在0.70左右即可接受(尤其对于分量表而言),但系数为0.80及以上则更好<sup>[1]</sup>;若大于0.90,应考虑精简量表的条目数量<sup>[6]</sup>。但对用于医疗诊断、学业成绩评定等领域的量表,其信度系数应为0.90及以上<sup>[1,6]</sup>。增加量表的条目可提高其内部一致性信度<sup>[6]</sup>。也有学者推荐采用系数 $\omega$ (coefficient omega, $\omega$ )评价内部一致性信度,因为它使用条目共享的公共因子上的条目载荷矩阵来计算相关方差,比 $\alpha$ 系数的计算方法更准确<sup>[2,6]</sup>,但其计算过程较为复杂,目前仍未被广泛使用。

重测信度反映量表在一段时间内重复测量时得分的稳定性<sup>[1,3]</sup>。其假说为,对属性不易改变的概念进行测量,受试者在两次测试中的得分差异都来自于测量误差,分数差异越小,信度越高。护理研究者常采用皮尔逊相关系数(r)估计重测信度,但心理测量学家更推荐组内相关系数(the intraclass correlation coefficient, ICC)作为重测信度的首选指标<sup>[1]</sup>。在实践中,重测信度并不适用于所有情况,许多概念或特征(如态度、知识、技能等)的确会随着时间的推移而改变;如果被测因素确实发生了变化,重复测量的结果其实并不是对信度的测量。另外,记忆干扰也会导致可能虚高的重测信度值。因此,应基于被测概念属性的可变性、测量过程的复杂性和参与者的特征来确定两次测量之间的最佳时间间隔,一般为1~2周<sup>[3,17]</sup>。

平行测试信度涉及在两个不同场合,对相同的受试者,使用测量同一概念的两个平行(或替代)量表进行测试,然后评价两次测试结果的一致程度,即组内相关系数ICC<sup>[1]</sup>。当研究人员希望在短时间内完成测量并希望避免受试者对条目的记忆干扰时,可采用此方法。在护理领域,由于量表的平行版本较难获得,很少有量表研制者进行平行测试信度分析。分半信度(或折半信度)采用与平行测试相同的逻辑,评价单个量表中一半条目与另一半条目之间的关系,它是确定量表内部一致性的原始方法<sup>[3]</sup>。但分半信度可能存在因分半方法不同产生不同信度系数的问题。

**3.2.3 效度分析** 效度反映量表的准确度、正确性和有效性,它是指某量表是否有效地测量到了它想要测量的内容,即实际测量结果与预想结果的符合程度<sup>[1,6]</sup>。表面效度是指量表看起来是否是在测量目标概念。尽管它常常不被认为是量表的关键测量特性,但它可以是内容效度的先驱或一个方面<sup>[1,3]</sup>。内容效度是指量表条目充分捕捉被测概念结构的程度,涉及条目抽样的充分性,比表面效度具有更结构化和更严格的程序<sup>[1,6]</sup>。内容效度的评价方法已在专家评价部分阐述。

结构效度,是指量表测量结果与设计该量表时所假定的理论(概念和关系)之间的符合程度。当一个概念存在多个子主题时,结构效度可评价量表是否有对应的各维度和条目来准确对应各子主题<sup>[1,3]</sup>。该方法适用于反映较为抽象概念的量表,也被广泛认为是量表研制过程不可或缺的重要环节<sup>[2]</sup>。结构效度的检验方法通常有两种:因子分析和假设检验效度。因子分析包括探索性因子分析(exploratory factor analysis, EFA)和验证性因子分析(confirmatory factor analysis, CFA)。通常, EFA有助于提出一个概念的基本结构框架,而CFA可对该框架的有效性或模型的拟合度进行评价<sup>[6,8]</sup>。假设检验效度关注在多大程度上可以证实关于量表测量的分数与其他概念(或同一概念)的测量分数相关的假设,分为聚合效度、区分效度和已知群体效度<sup>[1]</sup>。

准则效度(或效标效度)是指量表得分与某种外部准则(或效标)间的关联程度<sup>[1]</sup>。量表与准则之间须在实践经验上有一定联系<sup>[6]</sup>。准则效度有两种类型,即同时效度和预测效度<sup>[1]</sup>。效度分析依数据类型不同而采取不同的

统计分析方法,如相关分析、回归分析、敏感性、特异性、t检验、受试者工作特征曲线(receiver operating characteristic curve, ROC)、ROC产生的曲线下面积指数等<sup>[19]</sup>。然而,当概念较为抽象时,往往很难找到一个合适的准则进行比较,该方法在护理领域的应用相对缺乏。

### 3.3 量表的优化与验证

对量表进行心理测量学评估后,还应根据其分析结果修订与完善量表条目。若内部一致性信度系数 $\alpha$ 较低,应考虑修改或添加条目,但决定条目的去留还应综合考虑其内容本身或CVI值。量表研制的最后环节是最终量表的验证,其方法有CFA、假设检验效度和准则效度分析。若因条件有限无法开展单独的验证研究时,量表开发者应尽量利用现场测试研究的样本数据开展其他恰当的效度测试研究<sup>[1]</sup>。但当量表进行EFA分析时纳入的样本量较少(<150例)时,选择目标人群的其他单独样本进行CFA将有助于解释量表的可推广性<sup>[6]</sup>。

### 参考文献

- [1] Polit DF, Beck CT. Nursing research: Generating and assessing evidence for nursing practice[M]. Lippincott Williams & Wilkins, 2017.
- [2] Kalkbrenner MT. A practical guide to instrument development and score validation in the social sciences: The MEASURE Approach[J]. Practical Assessment, Research, and Evaluation, 2021, 26(1): 1-20.
- [3] Gray JR, Grove SK, Sutherland S. Burns and grove's the practice of nursing research-E-book: Appraisal, synthesis, and generation of evidence. 8th edition ed[M]. Elsevier Health Sciences, 2017.
- [4] Boateng GO, Neilands TB, Frongillo EA, et al. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer[J]. Front Public Health, 2018, 6:149.
- [5] 罗伯特·F·德威利斯. 量表的研制:理论与应用(第3版)[M]. 席仲恩,杜珏,译. 重庆:重庆大学出版社,2016.
- [6] Robert F. DeVellis. Scale Development: Theory and Applications[M]. Los Angeles: Sage publications. 2017.
- [7] Walker L, Avant K. Strategies for theory construction in nursing[M]. Boston: Person. Prentice Hall, 2019.
- [8] Waltz CF, Strickland OL, Lenz ER. Measurement in nursing and health research. Fifth Edition ed[M]. Springer Publishing Company, LLC, 2017.

- [9] Morgado FFR, Meireles JFF, Neves CM, et al. Scale development: ten main limitations and recommendations to improve future research practices[J]. *Psicol Reflex Crit*, 2017, 30(1): 3.
- [10] Ikart EM. Survey questionnaire survey pretesting method: An evaluation of survey questionnaire via expert reviews technique[J]. *Asian Journal of Social Science Studies*, 2019, 4(2): 1.
- [11] 张毅,汪健健,邓艳红,等. 护士人文素养自评量表的研制及信效度检验[J]. *中华护理杂志*, 2022, 57(9): 1120-1128.
- [12] 柏丁兮,高静,杨直,等. 老年慢病共存患者治疗负担量表的研制[J]. *中国全科医学*, 2023, 26(13): 1613-1620.
- [13] 蔡真真,陈媛,闫丽娟,等. 护士信息胜任力评估量表的研制及信效度检验[J]. *护理学杂志*, 2022, 37(24): 43-46, 63.
- [14] Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations[J]. *Res Nurs Health*, 2007, 30(4): 459-67.
- [15] 史静琚,莫显昆,孙振球. 量表编制中内容效度指数的应用[J]. *中南大学学报(医学版)*, 2012, 37(2): 152-155.
- [16] Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication[J]. *J Caring Sci*, 2015, 4(2): 165-78.
- [17] Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations[J]. *Qual Life Res*, 2014, 23(6): 1713-20.
- [18] 吴明隆. 问卷统计分析实务:SPSS操作与应用[M]. 重庆:重庆大学出版社, 2010.
- [19] Waltz CF, Strickland OL, Lenz ER. Measurement in nursing and health research. 4th ed[M]. Springer publishing company, 2010.

(收稿日期:2023-05-19)

## 手术室多元化护理在白内障超声乳化中的应用

罗丽花

(福建省三明市中西医结合医院手术室,福建三明 365001)

**摘要:** **目的** 探讨手术室多元化护理在白内障超声乳化手术患者康复治疗中的应用效果及影响。**方法** 选取在福建省三明市中西医结合医院诊治并行超声乳化手术的282例白内障患者为研究对象,选取年限为2021年1月~2022年12月。采用简单化随机法将282例患者分为对照组和观察组,每组各141例。所有患者均行超声乳化手术进行治疗,对照组患者在术前、术中、术后采取手术室常规护理措施进行干预,观察组实施手术室多元化护理措施进行干预。观察并比较2组患者护理前后的视功能生存质量、心理焦虑和抑郁评分、视力功能指标(眼压和黄斑变形区域面积)和并发症发生情况。**结果** 与对照组比较,观察组患者的视功能生存质量评分(CLVQOL)明显更高( $P<0.05$ ),焦虑(SAS)评分和抑郁(SDS)评分明显降低( $P<0.05$ ),视力功能指标(眼压和黄斑变性区域面积)均有所改善( $P<0.05$ ),术后并发症发生率无明显差异( $P>0.05$ )。**结论** 经手术室多元化护理措施干预后,白内障超声乳化手术患者的康复效果明显提高,患者的负面焦虑和抑郁情绪明显得到缓解,患者的视功能生存质量显著改善,视力功能明显好转,且患者术后的不良反应发生情况显著降低,安全性良好。

**关键词:** 手术室多元化护理;白内障;超声乳化手术;康复治疗

**中图分类号:** R472 **文献标志码:** A **文章编号:** 1004-6879(2024)01-0049-04

白内障是指由多种因素导致的人眼晶状体透明度减弱或颜色呈乳白色,进而引起的光学质量改变,影响因素包括代谢异常、遗传、局部营养障碍或晶状体老化等<sup>[1-3]</sup>。白内障早期发病症状并不明显,其临床表现主要为视力降低、光敏感度减弱、屈光改变等临床不良症状,

若不及时治疗,中后期会出现晶状体浑浊加重或完全失明,严重影响患者的生活质量<sup>[4]</sup>。目前,临床上针对白内障的治疗以手术治疗为主,药物治疗为辅,包括超声乳化手术等,虽然手术治疗能够有效改善患者的视力功能及临床症状<sup>[5,6]</sup>。但手术治疗及药物治疗对患者产生的