

• 综 述 •

医学模拟教育中复盘教学评估工具的范围综述

朱秀媛¹, 罗宇茵¹, 叶琴¹, 刘宁²

摘要:目的 总结并分析医学模拟教育中复盘教学评估工具的内容、开发过程、特点以及应用情况,为我国医学教学领域复盘评估工具的选用提供借鉴。方法 采用 JBI 范围综述指南为方法学框架,系统检索中国知网、万方数据知识服务平台、维普网、中国生物医学文献数据库、Embase、Cochrane Library、PubMed、Web of Science,检索时限为从建库至 2023 年 11 月 1 日。根据纳入、排除标准筛选文献。结果 最终纳入 21 篇全文文献,共涉及 8 种复盘教学评估工具,其中 10 篇关于评估工具的开发,11 篇关于评估工具的验证。常用工具主要为医学模拟复盘评价量表、客观结构化复盘评估工具、医学模拟教育追溯评级工具、实时复盘评估工具、同行复盘评估工具、复盘体验量表、引导者胜任力评价量表、有意义学习复盘评估量表。结论 模拟医学教育中针对复盘技能的评估工具种类多样,建议护理教育工作者全面考量和评估我国国情结合相关教学情景,合理选择评估工具,以促进有效和高质量的医学模拟教育复盘教学活动。

关键词:医学模拟教育; 复盘教学评估; 评估工具; 护理教学; 模拟教学; 范围综述

中图分类号:R47 **DOI:**10.3870/j.issn.1001-4152.2025.05.114

A scoping review of debriefing assessment tools in simulation-based medical education

Zhu Xiuyuan, Luo Yuyin, Ye Qin, Liu Ning. Department of Nursing, Zhuhai Campus, Zunyi Medical University, Zhuhai 519041, China

Abstract: **Objective** To summarize and analyze the content, development process, characteristics and application of debriefing assessment tools in medical simulation-based education, and provide reference for the selection of debriefing assessment tools in the field of medical education in China. **Methods** The JBI scoping review guideline was used as the methodological framework, and the CNKI, Wanfang Database, VIP.com, SinoMed, Embase, Cochrane Library, PubMed, and Web of Science were systematically searched from the establishment of the database to November 1, 2023. Literature was screened according to the inclusion and exclusion criteria. **Results** A total of 21 full-text articles were included, involving 8 debriefing assessment tools. In these 21 articles, 10 were about the development of tools and 11 were about the validation of tools. Commonly used tools were mainly Debriefing Assessment for Simulation in Healthcare (DASH), Objective Structured Assessment of Debriefing (OSAD), Simulation in Healthcare retriOaction Rating Tool (SHORT), Debriefing Assessment in Real Time (DART), Peer-Assessment Debriefing Instrument (PADI), Debriefing Experience Scale (DES), Facilitator Competency Rubric (FCR), Debriefing for Meaningful Learning Evaluation Scale (DMLES) and so on. **Conclusion** There are various types of assessment tools for review skills in simulated-based medical education. It is recommended that nursing educators comprehensively consider and evaluate my country's national conditions and combine relevant teaching scenarios to reasonably select assessment tools to promote effective and high-quality medical simulation education review teaching activities.

Keywords: simulation-based medical education; debriefing teaching assessment; assessment tools; nursing teaching; simulation teaching; scope review

医学模拟教育(Simulation-based Medical Education, SBME)是利用高仿真的模拟情景和模拟患者代替真实患者进行临床医学教学实践的医学教育模式^[1]。复盘(Debriefing)是经已经培训的复盘员采用基于证据的复盘模型主导的模拟实践后的反思过程,又称任务报告、事后情况说明、回馈讨论或引导性反

馈,是模拟教学或临床实践后导师与学员之间的学习性对话^[2-5]。复盘是医学模拟教学的基础及关键环节,是一种有意义的反思方法,可有效提高医学模拟教学效果。对复盘教学方法本身的评估极其重要,能够明确复盘者是否采用了正确的方法,并据此提高复盘者的技能掌握程度^[6-7]。选取有效、可靠的评估工具,可衡量复盘的运用效果,对复盘质量的提升具有重要意义。医学模拟教育中针对复盘过程的评估工具在我国开发与运用尚处于起步阶段,而现存工具多样,且其适用对象与范围、应用方法、应用效果存在较大差异。本研究采用 JBI 范围综述指南为方法学框架^[8],对医学模拟教育中复盘评估工具进一步梳理,分析其研究进展,旨在为评估工具的选择与应用提供参考,助力护理教育者开发或引入符合国情的评

作者单位:遵义医科大学珠海校区 1. 护理学系 2. 基础部全科医学基础教研室(广东 珠海,519041)

朱秀媛:女,硕士在读,主管护师,767874722@qq.com

科研项目:贵州省研究生教育教学改革重点课题(黔教合 YJSJGKT[2021]034);遵义医科大学研究生教育教学改革与研究课题项目(ZYK95);遵义医科大学教育教学改革计划项目(XJJG2021-37)

收稿:2024-09-17;修回:2024-12-06

估工具,促进我国护理教育开展科学、有效的复盘教学活动,提升我国护理领域模拟教育的质量。

1 资料与方法

1.1 确定研究问题 初步确定研究问题如下:①医学模拟教育中复盘技能评估工具有哪些?②各工具的基本特征?③相关评估工具在医学教育领域的应用情况?

1.2 检索策略 根据研究目的,计算机检索中国知网、万方数据知识服务平台、维普网、中国生物医学文献数据库、Embase、Cochrane Library、PubMed、Web of Science,检索策略根据各个数据库自身特点进行相应调整,检索时限为从建库至 2023 年 11 月 1 日。检索数据库时,为提高查全率,采用主题词和自由词结合的方式进行检索。以 education, medical; simulation training; debriefing, simulation debriefer; assessment tool 为英文检索词。以医学教育;模拟训练,复盘教学法,复盘导师;评估工具为中文检索词。以 PubMed 为例,检索策略见图 1。

```
# 1 "Education, Medical" [MeSH Terms]
# 2 medical education [Title/Abstract] OR health care education
[Title/Abstract] OR medical
teaching [Title/Abstract] OR medical pedagogy [Title/Abstract]
# 3 # 1 OR # 2
# 4 "Simulation Training" [MeSH Terms]
# 5 simulation education [Title/Abstract] OR simulation pedagogy
[Title/Abstract] OR education by simulation [Title/Abstract] OR
simulation teaching [Title/Abstract] OR simulation teach [Title/
Abstract] OR simulation-based * [Title/Abstract]
# 6 # 4 OR # 5
# 7 debriefing [Title/Abstract] OR debrief [Title/Abstract]
# 8 simulation debriefer [Title/Abstract] OR simulation facilitator
[Title/Abstract] OR simulation instructor [Title/Abstract] OR
simulation educator [Title/Abstract] OR simulation teacher [Title/
Abstract]
# 9 # 7 OR # 8
# 10 # 6 OR # 9
# 11 assessment tool [Title/Abstract] OR assessment model [Tit-
le/Abstract] OR assessment scale [Title/Abstract] OR assess-
ment system [Title/Abstract] OR evaluation tool [Title/Abstract]
OR evaluation model [Title/Abstract] OR evaluation scale [Title/
Abstract] OR evaluation score [Title/Abstract] OR evaluation sys-
tem [Title/Abstract] OR self-assessment [Title/Abstract]
# 12 # 3 AND # 10 AND # 11
```

图 1 PubMed 检索策略

1.3 文献纳入与排除标准

1.3.1 纳入标准 ①研究工具为复盘方法的评估工具;②研究内容为工具的开发、更新与验证的原始文献;③评估工具应用于医学教育领域。

1.3.2 排除标准 ①无法下载全文;②综述或会议摘要;③重复文献;④关键信息缺乏;⑤未说明评估工具的内容或使用方法;⑥验证工具时未进行信效度检验;⑦仅将评估工具用于结局测量指标。

1.4 文献筛选与资料提取 由 2 名研究者严格按照纳入与排除标准对文献进行筛选,首先阅读题目和摘要进行初步筛选,对符合纳入标准的文献进一步查找并阅读全文,遇到有异议的文献,2 名研究者进行讨论或咨询第 3 名研究者,再决定是否将文献纳入。对最终纳入的文献进行分析,提取的信息包括:作者、国家、工具名称、量表语言、调查对象、样本量、工具名称、研究目的、评分方式和评估标准、维度/条目数、维度名称和量表信效度。

2 结果

2.1 文献筛选结果 按检索策略初检得到文献 8 993 篇,导入 EndNote X9 去除重复文献后,获得文献 7 109 篇。阅读标题及摘要,对文献进行排除后剩 642 篇。查找和阅读全文,对文献进行排除后,纳入文献 19 篇。阅读纳入文献的参考文献后,补充文献 2 篇。最终共纳入文献 21 篇^[9-29]。

2.2 纳入研究基本特征 见表 1。

2.3 针对复盘技能评估工具的名称及内容要点

2.3.1 DASH 由 Brett-Fleegler 等^[9]于 2012 年开发,是一种评价复盘行为质量的锚定评价量表(Behaviorally Anchored Rating Scale, BARS)。该工具有助于复盘后行为有效性的质量评分。量表包含建立有吸引力的学习环境(4 个条目)、保持良好的学习环境(3 个条目)、以结构化的方式组织复盘(3 个条目)、鼓励学生参与讨论(5 个条目)、识别和探讨学生表现的差距(2 个条目)、帮助学生达到或维持良好的未来表现(3 个条目)6 个维度 20 个条目,并要求评分者将观察到的表现与定义的要素进行比较。各条目采用 Likert 7 级评分法,从 1 分(非常无效/有害)到 7 分(非常有效/出色)评价,评分越高说明复盘行为的完成质量越高。维度 1 用于评估模拟课程的介绍部分,而维度 2~6 则专注于评估模拟教育后的复盘过程。量表 Cronbach's α 系数为 0.89,组内相关系数(Intra-class Correlation Coefficient, ICC)为 0.74。

Olvera-Cortés 等^[10]将其翻译成西班牙语,并应用于心力衰竭远程模拟场景后的虚拟复盘教学活动中。通过对比 30 名导师、338 名本科生和 7 名模拟评分员在虚拟复盘中的表现,探讨该工具在教学评价中的实际应用效果。根据评价人的不同,将 DASH 分为 DASH-RV、DASH-SV 和 DASH-IV 3 个版本,分别由受过专业培训的评估员对导师进行评分、学生对导师进行评分以及导师自我评分。各版本 Cronbach's α 系数均为 0.97, DASH-RV 的评分者信度为 0.78。DASH 量表及其各个版本显示出较高的内部一致性和良好的评分者间一致性,从而确保了其测量结果的可靠性。但在工具研制过程中,仅运用了调查研究方法,可能会影响其开发过程的严谨性。

2.3.2 OSAD 由 Arora 等^[11]于 2012 年开发的结

构化工具,OSAD提供了一种循证的、手术参与者知情的方法来进行手术复盘,以评估外科手术复盘的质量。量表分方法、环境、参与度、反应、反思、分析、诊断、应用8个条目,各条目采用Likert 5级评分法,从1分(做得很差)到5分(做得很好)评价,总分8~40分,得分越高表明复盘质量越高。量表ICC=0.881。

Abegglen等^[12]研究者将该工具翻译成德语版本(GOSAD),并在57个录制的复盘视频中应用此工具,以检验其测量学特性,当有3名评分员评分时,ICC为0.848;当有2名评分员评分时,ICC为0.790;单个评分员评分时,ICC为0.650。Cronbach's α 系数为0.865,复合信度 $\omega=0.930$ 。

表1 纳入研究的基本特征

作者	国家	语言	调查对象	样本量	工具名称	研究目的
Brett-Fleegler等 ^[9]	美国	英语	模拟导师	114	DASH	工具开发
Olvera-Cortés等 ^[10]	墨西哥	西班牙语	模拟导师/学生/模拟评分员	30/338/7	DASH-IV DASH-SV DASH-RV	工具验证
Arora等 ^[11]	英国	英语	外科和麻醉医生以及手术室护士	33	OSAD	工具开发
Abegglen等 ^[12]	瑞士	德语	模拟导师	14	G-OSAD	工具验证
Runnacles等 ^[13]	英国	英语	儿科医生	16	儿科OSAD	工具开发
Runnacles等 ^[14]	英国	英语	儿科模拟导师	10	儿科OSAD	工具验证
Zamjahn等 ^[15]	美国	英语	不同学科模拟导师/复盘评分员	4/2	eOSAD	工具开发
Rivière等 ^[16]	法国	英语	复盘员/复盘评分员	11/2	SHORT	工具开发
Baliga等 ^[17]	澳大利亚	英语	复盘评分员	8	DART	工具开发
Baliga等 ^[18]	澳大利亚	英语	复盘评分员	12	DART	工具验证
Saylor等 ^[19]	美国	英语	模拟教育及复盘专家(阶段1/2)	15/5	PADI	工具开发
Saylor等 ^[20]	美国	英语	模拟教育及复盘专家	13	PADI	工具验证
Reed等 ^[21]	美国	英语	护理学生	130	DES	工具开发
Tosterud等 ^[22]	挪威	挪威语	护理学生	146	DES	工具验证
Almeida等 ^[23]	葡萄牙	葡萄牙语	护士	103	DES	工具验证
Farrés-Tarafa等 ^[24]	西班牙	西班牙语	护理学生	290	DES-sp	工具验证
Xie等 ^[25]	中国	汉语	护理学生	34	DES	工具验证
Leighton等 ^[26]	美国	英语	评分员/护理模拟导师	7/18	FCR	工具开发
Forbrig等 ^[27]	德国	德语	护理模拟导师	100	FCRG	工具验证
Bradley等 ^[28]	美国	英语	护理模拟导师	3	DMLES	工具开发
Bradley ^[29]	美国	英语	护理模拟导师	234	DMLI	工具更新与验证

注:DASH(Debriefing Assessment for Simulation in Healthcare)为医学模拟复盘评价量表,DASH-IV(DASH Instructor Version)为医学模拟复盘评价量表模拟导师版本,DASH-SV(DASH Student Version)为医学模拟复盘评价量表学生版本,DASH-RV(DASH Rater Version)为医学模拟复盘评价量表评分员版本;OSAD(Objective Structured Assessment of Debriefing)为客观结构化复盘评估工具,G-OSAD为德语版OSAD,eOSAD为电子版OSAD;SHORT(Simulation in Healthcare retrOaction Rating Tool)为医学模拟教育追溯评级工具;DART(Debriefing Assessment in Real Time)为实时复盘评估工具;PADI(Peer-Assessment Debriefing Instrument)为同行复盘评估工具;DES(Debriefing Experience Scale)为复盘体验量表,DES-sp为西班牙语版DES;FCR(Facilitator Competency Rubric)为引导者胜任力评价量表,FCRG为德语版FCR;DMLES(Debriefing for Meaningful Learning Evaluation Scale)为有意义学习复盘评估量表;DMLI(Debriefing for Meaningful Learning Inventory)为有意义学习复盘清单。

英国学者Runnacles等^[13]于2014年修订了OSAD,形成了专门针对儿科领域的OSAD(儿科OSAD),其8个条目分别命名为导师的方法、建立学习环境、学生的参与度、反应、描述性反思、分析、诊断和应用。该工具不仅能指导和评估儿科急症模拟后的复盘情况,还能作为复盘过程中的辅助备忘录,促进参与者的反思。Runnacles等^[14]于2016年通过10名儿科模拟专家验证了该工具的可靠性和有效性,测得总量表Cronbach's α 系数为0.78,ICC为0.643。Zamjahn等^[15]于2018年开发出OSAD的电子版本(eOSAD),其条目名称及评分方法与OSAD相同,预复盘阶段和正式复盘阶段的ICC分别为0.955和0.853。该工具能够提供共享链接,使所有参与人员能够随时随地在线访问,及时导出和分析数据,从而显著提高了使用的便利性。

OSAD在信度和效度方面表现良好,具有较高的内部一致性、重测信度、内容效度、复合信度和结构效度,以及良好的收敛效度。儿科OSAD在开发阶段未进行信效度检验。然而,在验证阶段,检验结果显示工具的信度良好,尤其是内部一致性方面表现突出。在收敛效度方面,量表与外部标准显示出较强的相关性。不过,评分者间的ICC仍需提升。eOSAD在预复盘和复盘阶段的ICC值均显示出评分具有高度的一致性和可靠性。

2.3.3 SHORT SHORT是一种评估模拟教育后简短复盘的工具,由Rivière等^[16]于2019年开发,并已应用于急诊和家庭医疗医师的复盘实践中。作为综合评分量表,SHORT包括5个条目(环境、复盘结构、协助、分析、知识迁移)和1个专家整体评分。采用Likert 5级评分法,1分为“有害”,2分为“无功无

过”,3分为“必须改进”,4分为“可以改进”,5分为“专家”,分数越高说明复盘的效率和熟练度越高。如果任何条目或整体评估的得分 ≤ 3 分,则应考虑对复盘员在相关方面的表现提供额外培训。

SHORT是特殊类型的复盘工具,讲师通常重点放在一个特定的问题上。通过利用连续的模拟课程和复盘来建立整体反馈,SHORT能够通过连续的短期模拟课程一个接一个地解决问题。该量表的概化系数(Generalizability Coefficient, G系数)较高,为0.78~0.82,表明该工具具有较好的信度和可靠性。然而,尚未对该工具进行效度检验和跨文化验证。

2.3.4 DART DART工具是由高级儿科和围生期教育中心的教师根据观察到有效的、非医疗行业的模拟和复盘实践开发的,是复盘表现的实时客观衡量标准^[30]。DART由Baliga等^[17]于2022年开发用于评估跨学科医疗模拟复盘活动。DART详细记录了基本信息(包括导师姓名、模拟场景和日期)、时间记录(涵盖模拟场景和复盘的开始与结束时间、持续时间、场景结束与复盘开始的间隔、复盘过程中首次播放教学视频的时间、复盘时长在模拟场景中的占比、学习目标时长在复盘中的占比、复盘时长与模拟场景时长的比率、视频暂停次数以及播放视频的时间)和互动分析等关键信息。互动分析的评估可以根据实际情况对最多15个关键节点进行评分。在每个节点,都需要对复盘过程中的3个部分,即根据复盘过程中的导师提问(Instructor Questions, IQ),导师陈述(Instructor Statements, IS),学员答复(Trainee Responses, TR)的表现进行打分。打分方式是通过勾选1~10个方框分数,分数越高表示表现越好,并使用累积计分法统计结果。最后,计算导师提问与导师陈述的比例(IQ: IS)以及学生与导师输出的比例(TR: [IQ+IS]),可以更有效地分析和评估教学过程中导师与学生之间的互动模式。2023年,Baliga等^[18]将DART应用于8个复盘视频,对45名跨学科学习者的复盘质量进行了评估,测得DART的Cronbach's α 系数为0.852~0.978。作为一种定量评估工具,DART能够提供有关医学模拟复盘的可靠数据。它既可以作为实时工具单独使用,也可以与定性工具(如DASH或SHORT)结合使用。DART工具在开发阶段仅采用了专家评审这一方法,这可能限制了其开发过程的严谨性。

2.3.5 PADI 是由Saylor等^[19]开发的自我和同行评估工具,旨在专门评估复盘人员有效开展复盘会议的能力和复盘效果,适用于各医疗学科的复盘活动。PADI包括复盘的结构和组织、语言和非语言沟通、为复盘环节设定环境和基本规则、谈论化解、回顾模拟体验、反思行动、促进学生将模拟体验与临床实践相联系、总结8个条目,各条目采用4级评分,根据每个项目的完成百分比进行自我和同行评估,1分为表

现很差,4分为表现很好,总分范围为8~32分,分值越高代表模拟导师的复盘能力越强。该工具ICC为0.973,单项测量ICC为0.818。Saylor等^[19-20]已将PADI应用于模拟导师中,通过整合同行评审意见,促进了参与模拟教育和复盘的导师之间的互动。在开发过程中,该量表采用了文献检索、专家评审和调查研究3种不同的开发方法,确保了工具的全面性。此外,整体和单项测量的ICC值均较高,这显示了测量过程具有高一致性和高可靠性。

2.3.6 DES 由Reed等^[21]于2012年研制,可供护生描述他们在复盘期间的主观体验,以帮助护生确定最佳复盘实践。DES分为体验和重要性2个分量表,各分量表包括分析想法和感受(4条目)、学习和建立联系(8条目)、导师引导复盘的能力(5条目)、导师的适当指导(3条目)4个维度20个条目。各条目采用Likert 5级评分;体验量表从“非常不同意”(1分)到“非常同意”(5分)评分;重要性量表从“非常不重要”(1分)到“非常重要”(5分)评分。体验分量表Cronbach's α 系数为0.93;重要性分量表Cronbach's α 系数为0.91。2015年,Tosterud等^[22]将DES翻译成挪威语并进行了文化调适和验证,量表维度不变,将条目调整为18个。通过对146名护理本科生的调查,结果表明该量表在评估学生复盘情况方面具有良好的潜力,总量表Cronbach's α 系数为0.91。2016年,Almeida等^[23]将DES翻译成葡萄牙语,量表维度和条目维持原量表不变,并最终纳入103名护士作为受试对象,进行了信效度检验,测得体验量表Cronbach's α 系数为0.94;重要性量表Cronbach's α 系数为0.96。2022年,Farrés-Tarafa等^[24]将DES翻译为西班牙语版本(DES-sp),测得体验量表Cronbach's α 系数为0.926;重要性量表Cronbach's α 系数为0.933。同年,Xie等^[25]汉化和调适了DES形成简体中文版,研究团队在不改变原有量表格式的基础上,加入了形容词及修饰语,使之更易于理解。对34名大学二年级护生进行测试后,测得体验量表Cronbach's α 系数为0.90,重要性量表Cronbach's α 系数为0.92;量表ICC为0.86。DES经过多国跨文化检验,无论是量表整体还是其分量表,都显示出良好的信度。

2.3.7 FCR 由Leighton等^[26]于2018年基于从新手到专家理论构建的评估工具,旨在区分护理模拟导师的能力水平,并评估引导者在复盘阶段的胜任力。FCR涵盖了模拟教育中的5个阶段,即准备(7个条目)、预复盘(4个条目)、引导(6个条目)、复盘(8个条目)和评估(4个条目)共29个条目。该工具采用Likert 5级评分,从初学者(1分)、高级初学者(2分)、熟练(3分)、精通(4分)至专家(5分)。各阶段的总分及评分层次划分如下:准备阶段总分0~35分,其中0~14分为初学者到高级初学者层次,15~

27分为熟练层次,28~35分为精通到专家层次;预复盘阶段总分0~20分,其中0~8分为初学者到高级初学者层次,9~15分为熟练层次,16~20分为精通到专家层次;引导阶段总分0~30分,其中0~12分为初学者到高级初学者层次,13~23分为熟练层次,24~30分为精通到专家层次;复盘阶段总分0~40分,其中0~16分为初学者到高级初学者层次,17~31分为熟练层次,32~40分为精通到专家层次;评估阶段总分0~20分,其中0~8分为初学者到高级初学者层次,9~15分为熟练层次,16~20分为精通到专家层次。达到精通到专家层次的引导者可以为初学者到高级初学者层次的引导者提供指导。该量表测得重测信度为0.43~0.92,ICC为0.62~0.90,G系数为0.80~0.99。Forbrig等^[27]于2023年将FCR翻译成德语版(FCRG),未对量表结构和评分方法进行修订,测得总量表Cronbach's α 系数为0.980,ICC为0.986。最初,FCR主要适用于重症监护室的护士,但随着其不断发展,现已被广泛应用于护理教育和培训领域。FCR在信度方面表现非常出色,具有高内部一致性、高重测信度和高分者间信度。结构效度也是良好的,但收敛效度较低。

2.3.8 DMLES 由Bradley等^[28]于2016年设计,共31个条目,每个条目通过二元选项进行评分,即“存在”或“不存在”,旨在观察和评估有意义学习复盘(Debriefing for Meaningful Learning, DML)行为。DML作为一种复盘方法,不仅在模拟环境中,也在其他临床环境中得到应用,其目的在于回顾患者护理、培养反思性思维,并促进有意义的学习^[31]。DMLES特别评价了与DML方法一致的可观察行为。Bradley等^[28]将DMLES应用于通过DML培训的护理复盘员中,测得总量表Cronbach's α 系数为0.88,评分者间信度为0.86。为了全面探讨复盘者如何理解和制订DML,Bradley^[29]于2018年对DMLES进行修订,并将其命名为DML清单(DMLI)。在保留原DMLES条目内容的前提下,将其扩充到57个条目,条目1~52使用“总是”“有时”“从不”3级评分法进行评分,条目53~57以“是”或“否”二元选项评分。DMLI用于评估复盘者是否应用了与DML一致的行为,量表Cronbach's α 系数为0.964。

2.4 评估工具优势、局限性及适用范围 见表2。在测量特性方面,以下评估工具尚未进行效度检验:DASH、eOSAD、SHORT、DART和PADI。儿科OSAD在开发过程中未进行信效度检验,而其跨文化验证结果显示评分者间的信度为中等。DES、FCR和DMLES已经完成了信度和效度的检验,特别是DES,已在多个国家进行了信效度检验,且结果显示其信效度均佳。此外,OSAD、SHORT、DART和PADI在开发和验证过程中样本量有限,因此需要进一步增加研究对象以增强评估量表的可靠性和有效

性。在评估时间方面,DASH、OSAD、SHORT和DES所需的时间较短,而PADI所需的时间较长;对于其他量表,未报告具体的评估时间。在应用范围上,DASH和PADI因其广泛的适用性,被用于不同学科的复盘活动。OSAD主要应用于外科和儿科的复盘。SHORT工具适用于快速评估。DES、FCR和DMLES已被证实在护理专业的复盘教学中具有有效的应用价值。评估量表的优势、局限性及适用范围各不相同,在选择复盘评估工具时应综合考虑。

3 讨论

3.1 评估工具的研发对复盘教学活动的意义 复盘技能的评估工具是一种基于证据,评估特定背景中的复盘质量,以便优化复盘的有效性^[9,11]。本研究通过范围综述,系统梳理了8种复盘技能评估工具的开发、验证和应用情况。本研究发现每个工具都有其独特的用途、适用情境、特色之处、积极效应及潜在不足。具体来说,DASH工具关注于复盘行为的完成质量;OSAD则侧重于衡量复盘教学的实际效果;SHORT工具旨在评估复盘者的效率和熟练度;DART则专注于测量导师和学员在复盘过程中的参与度;PADI关注复盘行为的完成度;DES则着眼于评估学员的复盘体验;FCR工具用于衡量引导者在复盘阶段的能力表现;DMLES则进一步评估复盘行为的完成情况。这些多样化的评估工具共同为提升复盘的质量和效果提供了有力支持。

上述工具通过构建清晰、逻辑严谨的流程或框架,促进复盘活动有序进行,这种结构确保了活动的连贯性和实效性。这些工具激励学生在讨论、反思与分析中积极参与,进而促进知识的内化与应用。反思与分析是这些工具的共同特征,他们协助学习者审视自身表现,识别优劣,并探讨改进措施。工具同样强调学习环境的重要性,无论是构建引人入胜的环境还是为复盘活动设定场所,均旨在营造有利于学习与反思的氛围,同时这些工具致力于推动学习者将所学知识与应用于实际,实现知识的转化与应用。部分工具还涉及评价与反馈环节,有助于学习者认识自身表现,并根据反馈进行相应调整。

3.2 评价工具在开发和验证中存在的问题 本研究发现,现有的评估工具仍存在一定的局限性。首先,在开发过程中,纳入的工具多采用文献检索和专家评审的方法,但较少考虑学生的意见。这些工具在信度检验中多采用内部一致性和组内相关系数等方式,而在效度检验中则多采用因子分析法和内容效度指数等方法。然而,目前有5种工具(DASH、eOSAD、SHORT、DART、PADI)尚未经过严格的效度检验,因此,这些工具需要接受更深入的信效度检验,以确保其准确性和可靠性。其次,当前的评估工具大多由国外研发,除了DES外,其他工具尚未进行本土化验

证。由于各国国情和医疗环境的差异,这些工具在我国的实际应用中可能不完全适用。最后,部分评估工具对使用者的要求较高,评估人员在使用这些工具之

前需要接受专业培训,掌握正确的使用方法,否则可能会导致评估结果的错误解读。

表 2 评估工具优势、局限性及适用范围

工具名称	优势	局限性	适用范围
DASH	①应用广泛,信度佳;②具有导师、学生及评分员 3 个版本,可多角度对复盘教学进行评估	①评分人员在使用 DASH-RV 前必须接受专业培训,操作便捷度欠佳;②工具的研制过程欠严谨性	①通用性评估工具;②测评内容侧重学习环境、学生的参与度和表现;③用于复盘后对其行为有效性的质量评分
OSAD	①评估时间短;②维度丰富;③信度、内容效度及收敛效度良好;④有电子版本,评估便捷,易于管理数据	①仅应用于欧美人群,尚未在亚洲人群中应用;②样本量较少,难以准确反映量表的测量特性;③儿科 OSAD 评分者间信度中等,有待改进	①适用于外科和儿科复盘教学活动;②测评内容侧重评估复盘教学全过程
SHORT	①评估便捷;②信度良好;③针对性强,解决问题的效率高	①量表未经过跨文化验证,泛化能力受限;②未进行效度检验;③样本量少	①适用于紧急情况发生后的复盘;②侧重评估复盘活动的环境条件、结构化程度、实际效果、分析能力和成果转化能力
DART	①内部一致性佳,工具可靠性强;②易于使用,仅需少量培训即可掌握工具的操作;③定量评估工具,客观性强	①工具的研制过程欠严谨性;②调查对象少,限制评估工具的普遍性和外推性	适用于评估不同培训水平的跨学科学习者的复盘质量
PADI	①可用于不同医疗学科,适用性广泛;②信度佳,测量工具可靠	①评估时间长,不适用于快速反馈的情况;②同行评估可能受到人际关系和主观偏见的影响;③样本量较少,限制了工具的可推广性	适用于不同学科,用于复盘后的自我评估和同行评估
DES	①信度和效度均优秀;②跨文化适用性强;③已经完成汉化版本的验证	①仅专注于复盘的体验,没有触及其他领域,应用面狭窄;②汉化版本的样本量有限,需要进一步的数据收集和验证以确保其准确性和可靠性	①适用于护理专业复盘教学活动;②重点评估学习者在复盘时的体验
FCR	信效度良好	模拟导师进行自我评估时可能存在主观性,导致评估结果的偏差	①适用于护理专业;②用于模拟导师自我评估其在模拟教学的准备、预复盘、引导、复盘和评估 5 个阶段的能力
DMLSES	①信效度佳;②能够促进护理计划的质量改进,实用性强	仅评估有意义学习复盘应该存在的行为的有无,不评价这些行为的完成质量	①适用于护理专业;②评估有意义学习复盘中的应有行为是否存在

3.3 对评价工具未来研究的启示 未来在开发复盘教学评估工具时,应严格遵循标准化的开发流程。研究者应依据理论和规范理论为指导,建议在开发量表的过程中将 Delphi 法、访谈法以及焦点小组等多元方法有机结合,并充分参考学生意见,确保评估工具研制过程的质量与规范性,以确保工具能够更好地满足学习者的实际需求。此外,在对工具进行汉化时,应充分考虑我国的医学教育背景和复盘教学实际情况,对现存工具继续进行大样本、多中心的完善验证,丰富量表的内涵,并进一步探索其心理测量学属性,同时考察其对于护理教学的效果。此外,还可结合互联网与 5G 技术,利用手机 App、微信小程序等评估教学活动中的复盘质量,实现实时数据传输和反馈,增强教学的互动性和个性化。

4 小结

近年来,复盘教学方式虽然在我国快速发展,但其评估工具在我国护理领域应用较少。针对复盘的评估工具的研究处于起步阶段、发展缓慢,研究者可参考国外成熟的评估工具,并结合我国护理模拟教学

的特点和需求,通过科学的信效度检验,构建或引入具有我国文化特色的高质量、特异性和规范化的评估工具,有助于评估复盘者技能的掌握情况、是否实施正确的复盘方法等,以提升复盘实施过程和结果的质量,进而推动护理模拟教学的科学化发展。本研究未纳入灰色文献,且仅考虑了中英文资料,因此未来研究应拓展文献来源,以提高评价的客观性。

参考文献:

[1] 葛敏静,王巍,黄华兴,等. 医学模拟教育课程的构建与实践[J]. 南京医科大学学报(社会科学版),2020,20(4): 384-388.

[2] Committee I S. INACSL standards of best practice:simulationSM,simulation glossary[J]. Clin Simul Nurs,2016(12): 39-47.

[3] 童斌,朱滨海,占伊扬,等. 医学模拟教学中 Debriefing 的认识与实践[J]. 中国毕业后医学教育,2020,4(3):258-261.

[4] 迪力努尔·波拉提,刘婷,虞彬. 结构化引导性反馈在心内科护生临床实习中的应用[J]. 中华医学教育探索杂志,2022,21(6):753-756.

[5] 杨娜娜,曾慧,王燕,等. 3C 引导性反馈下情景模拟教学

- 对本科护生批判性思维能力的影 响[J]. 护理学杂志, 2024,39(2):75-78.
- [6] 龙晓焕,王西富,关天悦,等. 医学模拟教育中复盘教学方法的发展和 应用综述[J]. 中华医学教育杂志,2022,42(3):241-245.
- [7] Husebø S E, Reiersen I Å, Hansen A, et al. Post-simulation debriefing as a stepping stone to self-reflection and increased awareness: a qualitative study[J]. *Adv Simul (Lond)*, 2024,9(1):33.
- [8] Lockwood C, Dos Santos K B, Pap R. Practical guidance for knowledge synthesis: scoping review methods[J]. *Asian Nurs Res (Korean Soc Nurs Sci)*, 2019,13(5):287-294.
- [9] Brett-Fleegler M, Rudolph J, Eppich W, et al. Debriefing assessment for simulation in healthcare: development and psychometric properties[J]. *Simul Healthc*, 2012,7(5):288-294.
- [10] Olvera-Cortés H E, Hernández-Gutiérrez L S, Scherer-Castanedo E, et al. Who can debriefing assessment for simulation in healthcare? different perspectives on teledebriefing assessments[J]. *Simul Healthc*, 2022,17(5):308-312.
- [11] Arora S, Ahmed M, Paige J, et al. Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery[J]. *Ann Surg*, 2012,256(6):982-988.
- [12] Abegglen S, Krieg A, Eigenmann H, et al. Objective Structured Assessment of Debriefing (OSAD) in simulation-based medical education: translation and validation of the German version[J]. *PLoS One*, 2020,15(12):e0244816.
- [13] Runnacles J, Thomas L, Sevdalis N, et al. Development of a tool to improve performance debriefing and learning: the paediatric Objective Structured Assessment of Debriefing (OSAD) tool[J]. *Postgrad Med J*, 2014,90(1069):613-621.
- [14] Runnacles J, Thomas L, Korndorffer J, et al. Validation evidence of the paediatric Objective Structured Assessment of Debriefing (OSAD) tool[J]. *BMJ Simul Technol Enhanc Learn*, 2016,2(3):61-67.
- [15] Zamjahn J B, Baroni de Carvalho R, Bronson M H, et al. eAssessment: development of an electronic version of the Objective Structured Assessment of Debriefing tool to streamline evaluation of video recorded debriefings[J]. *J Am Med Inform Assoc*. 2018,25(10):1284-1291.
- [16] Rivière E, Aubin E, Tremblay S L, et al. A new tool for assessing short debriefings after immersive simulation: validity of the SHORT scale[J]. *BMC Med Educ*, 2019,19(1):1-7.
- [17] Baliga K, Coggins A, Warburton S, et al. Pilot study of the DART tool: an objective healthcare simulation debriefing assessment instrument[J]. *BMC Med Educ*, 2022,22(1):636-636.
- [18] Baliga K, Halamek L P, Warburton S, et al. The Debriefing Assessment in Real Time (DART) tool for simulation-based medical education [J]. *Adv Simul (Lond)*, 2023,8(1):9.
- [19] Saylor J L, Wainwright S F, Herge E A, et al. Peer-Assessment Debriefing Instrument (PADI): assessing faculty effectiveness in simulation education[J]. *J Allied Health*, 2016,45(3):e27-e30.
- [20] Saylor J L, Wainwright S F, Herge E A, et al. Development of an instrument to assess the clinical effectiveness of the debriefer in simulation education[J]. *J Allied Health*, 2016,45(3):191-198.
- [21] Reed S J. Debriefing Experience Scale: development of a tool to evaluate the student learning experience in debriefing[J]. *Clin Simul Nurs*, 2012,8(6):e211-e217.
- [22] Tosterud R, Petzæll K, Wangensteen S, et al. Cross-cultural validation and psychometric testing of the questionnaire: Debriefing Experience Scale[J]. *Clin Simul Nurs*, 2015,11(1):27-34.
- [23] Almeida R G, Mazzo A, Martins J C, et al. Validation to portuguese of the Debriefing Experience Scale[J]. *Rev Bras Enferm*, 2016,69(4):705-711.
- [24] Farrés-Tarafa M, Julian D B, Lorenzo-Seva U, et al. Cultural adaptation, translation and validation of the Spanish version Debriefing Experience Scale[J]. *PLoS One*, 2022,17(5):e0267956.
- [25] Xie Y D, Li X Y, Liu Q, et al. Cross-cultural validation and psychometric testing of the Debriefing Experience Scale (DES): a cross-sectional study [J]. *BMC Med Educ*, 2022,22(1):272.
- [26] Leighton K, Mudra V, Gilbert G E. Development and psychometric evaluation of the Facilitator Competency Rubric[J]. *Nurs Educ Perspect*, 2018,39(6):E3-E9.
- [27] Forbrig T A, Gellert P, Biniok M, Gräske J. Facilitator competency rubric in nursing simulations: transcultural adaptation and validation of the German version[J]. *BMC Nurs*, 2023,22(1):139.
- [28] Bradley C S, Dreifuerst K T. Pilot testing the debriefing for meaningful learning evaluation scale[J]. *Clin Simul Nurs*, 2016,12(7):277-280.
- [29] Bradley C S. Confirmatory factor analysis of the debriefing for meaningful learning inventory[®] [J]. *Clin Simul Nurs*, 2018,14:15-20.
- [30] Arul N, Ahmad I, Hamilton J, et al. Lessons learned from a collaborative to develop a sustainable simulation-based training program in neonatal resuscitation: simulating success[J]. *Children (Basel)*, 2021,8(1):39-39.
- [31] Dreifuerst K T. Getting started with debriefing for meaningful learning[J]. *Clin Simul Nurs*, 2015,11(5):268-275.