·医学模拟教育·

经尿道前列腺电切术实践技能模拟评价效度研究

汪成合[1]　汪晓冉[2]　王佳玉[3]　蒋光亮[1]　何竑超[1]　徐丹枫[1]

[1]上海交通大学医学院附属瑞金医院泌尿外科，上海 200025；[2]阿伯丁大学心理学院，阿伯丁，英国 AB24 3FX；[3]上海交通大学医学院附属瑞金医院医学教育处，上海 200025

通信作者：王佳玉，Email：wjy30541@rjh.com.cn

【摘要】 目的 使用经尿道前列腺电切术（transurethral resection of the prostate，TURP）关键技术评价量表对初学者实施经尿道前列腺电切术的行为表现进行打分，旨在为该量表所得成绩的诠释积累效度证据。**方法** 收集2023年1月至2024年3月期间，在上海交通大学医学院附属瑞金医院医学模拟中心参加经尿道前列腺电切术模拟培训的学员在培训前后实践技能模拟考核的成绩，各获取90份样本。本研究采用梅西克（Messick）效度框架，收集基于内容、反应过程、内部结构和其他变量的关系、测试后果5个方面的效度证据。**结果** 本研究量表的内容效度指数为0.952。评分者间一致性采用Spearman相关系数来衡量，培训前后的总分Spearman相关系数分别为0.993和0.970。量表总分的Cronbach's α系数在培训前后分别为0.987和0.947。验证性因子分析结果显示：χ^2/df=1.686、CFI=0.953、TLI=0.939、RMSEA=0.087、SRMR=0.036[CFI即比较拟合指数（comparative fit index）；TLI即塔克-刘易斯指数（Tucker-Lewis index，也称非规范拟合指数）；RMSEA即近似均方根误差（root mean square error of approximation）；SRMR即标准化均方根残差（standardized root mean square residual）]，表明模型拟合度可接受。**结论** 本研究为经尿道前列腺电切术实践技能模拟考核积累了效度证据。经尿道前列腺电切术关键技术评价量表是适合用于评价初学者经尿道前列腺电切术技能的工具。

【关键词】 经尿道前列腺切除术；　实践技能；　模拟考核；　效度验证；　毕业后医学教育

【中图分类号】 R-4

基金项目：全国医学教育发展中心医学模拟教育研究项目（2021MNYB13）

DOI: 10.3760/cma.j.cn116021-20240914-02103

A study on the validity of simulation-based assessment of practical skills in transurethral resection of the prostate

Wang Chenghe[1], Wang Xiaoran[2], Wang Jiayu[3], Jiang Guangliang[1], He Hongchao[1], Xu Danfeng[1]

[1]Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; [2]Department of Psychology, University of Aberdeen, Aberdeen, UK AB24 3FX; [3]Department of Medical Education, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

Corresponding author: Wang Jiayu, Email: wjy30541@rjh.com.cn

【Abstract】 Objective To score the behavioral performance of novices in transurethral resection of the prostate (TURP) using a TURP key skill assessment scale, and to accumulate validity evidence for the interpretation of scores derived from this scale. **Methods** Simulation-based assessment scores of practical skills were collected from 90 trainees, who participated in a simulation-based training on TURP at the Medical Simulation Center of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine from January 2023 to March 2024, before and after training. Messick's framework of validity was used to collect validity evidence for test content, response process, internal structure, relations to other variables, and

consequences. **Results** The content validity index of the scale was 0.952. The inter-rater consistency was measured using Spearman correlation coefficients, which were 0.993 and 0.970 for the total scores before and after training, respectively. The Cronbach's α coefficients for the total scale scores were 0.987 and 0.947 before and after training, respectively. The confirmatory factor analysis showed χ^2/df of 1.686, comparative fit index of 0.953, Tucker-Lewis index of 0.939, root mean square error of approximation of 0.087, and standardized root mean square residual of 0.036, indicating an acceptable model fit. **Conclusions** This study presents validity evidence for simulation-based assessment for TURP practical skills. The TURP key skill assessment scale is a suitable tool for evaluating the TURP skills of novice learners.

【Key words】 Transurethral resection of the prostate; Practical skill; Simulation-based assessment; Validation; Postgraduate medical education

Fund program: Medical Simulation Education Research Project of National Center for Health Professions Education Development (2021MNYB13)

DOI: 10.3760/cma.j.cn116021-20240914-02103

良性前列腺增生(benign prostatic hyperplasia, BPH)是中老年男性常见疾病之一[1]。经尿道前列腺电切术(transurethral resection of prostate, TURP)是目前公认的外科治疗 BPH 的金标准[2]。由于 TURP 是各类经尿道前列腺手术的基础,且随着微创手术的发展及手术设备和器械的普及,TURP 已成为泌尿外科毕业后教育阶段的重点培训内容之一。准确、有效的 TURP 手术技能评价是 TURP 手术技能规范培训体系的重要组成部分。因此,本研究团队前期基于 Delphi 专家咨询法构建了包含 6 个维度、18 个评价指标的 TURP 关键技术评价量表[3]。本研究旨在利用该评价工具评估初学者的 TURP 操作技能,以积累效度证据,为其进一步使用和广泛推广奠定基础。

1　研究方法

1.1　经尿道前列腺切除术实践技能模拟考核

1.1.1　考生来源及数量

效度验证的样本量通常应为评价量表条目数的 5~10 倍。本研究所使用的量表包含 18 条评价条目。同时,鉴于单中心的学员规模,最终选取 2023 年 1 月至 2024 年 3 月期间在上海交通大学医学院附属瑞金医院医学模拟中心参加 TURP 模拟培训的学员共计 90 名,其中包括在培外科基地住院医师 70 人及在医院泌尿外科进修的医生 20 人。所有参与学员均无 TURP 手术经验(即手术例数不超过 50 例[2])。于培训前后分别对学员进行一次 TURP 实践技能模拟考核,并记录考核成绩,最终共回收 180 份考核成绩单。

1.1.2　模拟考核设置

TURP 实践技能模拟考核共设两站。第一站为器械准备和口试。器械准备考核时长为 5 min,主要考察内容为器械的选择和组装。口试考核时长同样为 5 min,主要考查内容为与 TURP 手术相关的理论知识,包括手术能量平台的选择及参数设定、冲洗液的选择、常见器械故障的处理、常见并发症的表现及其紧急处理等。第二站为 TURP 模拟器手术操作考核,考核时长为 15 min。在以往的研究中,考虑到手术学习初期的学习效果和安全因素,通常选择的病例前列腺体积为 60 mL 左右[4]。本次考核选用 VirtaMed UroS 前列腺电切手术模拟器的病例模块,相当于 60 mL 左右体积的前列腺,以两侧叶增生为主,腺体表面有部分怒张的血管。

1.1.3　评价工具及评分者培训

本研究团队于 2022 年 6 月通过 Delphi-AHP 法构建了 TURP 关键技术评价量表[3](表 1),并采用该评价工具对考生进行打分。本研究的考试由 2 位泌尿外科主治及以上职称的医生担任考官,这些考官均具备 5 年及以上 TURP 手术经验、3 年及以上 TURP 教学经验。考试前,组织所有在研究期间参与执考的考官进行了时长为 2 h 的培训。培训内容包括执考宣教、评价标准的解读,并采用框架参照法(frame-of-reference, FOR)进行打分培训,以确保所有考官在使用该评价工具时的打分趋势保持一致。

1.2　梅西克效度验证框架

效度验证是收集相关证据,为特定分数的解释提供科学依据的过程[5]。其主要任务是从多方面

收集证据，对考试各阶段提出的假设和推论进行验证，这一过程同时也验证了考试构念界定的结果及其落实情况[6]。根据梅西克（Messick）的效度验证框架，效度证据包括以下5个方面：基于测试内容的效度证据（evidence based on test content），基于反应过程的效度证据（evidence based on response processes），基于内部结构的效度证据（evidence based on internal structure），基于与其他变量关系的效度证据（evidence based on relations to other variables），以及基于测试后果的效度证据（evidence based on consequences of testing）[7]。本研究基于梅西克效度验证框架，收集TURP实践技能模拟考试的各项效度证据。

表1　TURP关键技术评价量表

一级指标	二级指标	分值
器械准备	选择合适的器械并熟练组装	5
	选择合适的能量平台并设定工作参数	5
	选择合适的冲洗液	5
组织保护	准确地在内镜下辨认重要结构：尿道外括约肌、精阜、膀胱颈、输尿管开口	5
	准确地在内镜下辨认前列腺中叶、侧叶、尖部、前叶、外科包膜	5
	操作过程中未发生可能导致严重后果的损伤：尿道外括约肌、直肠、输尿管开口、膀胱壁等	5
	操作过程中未发生可能导致一般后果的损伤：尿道、前列腺包膜、精阜等	5
	操作过程中未不当使用暴力	5
	在合适的时机止血，保证操作全过程视野清晰	5
切除效率	熟练使用器械完成切除、止血操作	5
	眼-脑-手-脚协调统一	5
手术流程	熟悉整个手术步骤的每个细节	5
	有计划地完成手术：完整观察—切除腺体—彻底止血—收集标本	5
基本知识	掌握常见器械故障的处理（光源昏暗、切割无力、视野偏斜、出水不畅等）	5
	掌握常见并发症的表现和紧急处理（水中毒、尿道损伤、括约肌损伤、包膜穿孔、输尿管口损伤、膀胱损伤、静脉窦开放大出血等）	5
团队协作	合理指挥助手配合完成手术	5
	与手术室护士、麻醉科医生沟通合作	5
	必要时寻求上级医师指导	5

1.3　内容效度

邀请7位泌尿外科领域的临床专家（这7位专家未曾参与量表开发工作）对TURP关键技术评价量表条目与构念的相关性进行评价。评价采用4级评分标准：1分表示“不相关”，2分表示“弱相关”，3分表示“相关”，4分表示“强相关”。评价标准为条目水平的内容效度指数（item-level content validity index，I-CVI）≥0.71，以及量表水平的内容效度指数（scale-level content validity index，S-CVI）≥0.90，达到这些标准表示量表的内容效度较好[8]。

1.4　统计学方法

使用Excel建立受测者成绩数据库，使用JASP 0.17.2.1软件对所得数据进行统计处理，使用Jamovi 2.6.23.0软件进行验证性因子分析。计量资料以均数±标准差表示，计数资料以频数和百分率[例（%）]表示。两评分者间信度采用Spearman等级相关分析，Spearman相关系数取值在-1~1之间，$P<0.05$表示两评分者的评分结果有相关性，此时Spearman相关系数越大，表示两专家评分的等级相关性越好，结果越趋向于一致[9]。使用Cronbach's α系数来表明本考试研究的内部一致性信度，通常其值在0和1之间，达到0.7~0.8时表示量表具有相当的信度，达到0.8~0.9时说明量表信度非常好[10]。采用极端分组法测定考核的难度及区分度特性。当项目的难度指数P值增大时，意味着该考核整体倾向于较为简单；相反，若难度指数P值较低，则反映考核具有较高的挑战性。在考察区分度时，区分度高的项目具有卓越的能力去区分受测者的实际技能水平，区分度较低的项目则难以有效辨别受测者之间的能力差距。本研究运用Pearson相关分析检验考核内容间的关联性，相关系数范围在-1~1之间，负值表示负相关，正值表示正相关，零表示无相关。绝对值越接近1，关联性越强；越接近0，则关联性越弱。一般而言，绝对值0.2~0.4为弱相关，0.4~0.6为中度相关，大于0.6为强相关[11]。

2　结果

2.1　基于测试内容的效度证据

统计结果显示，I-CVI分布区间为0.857~1.000，S-CVI值为0.952。

2.2　基于反应过程的效度证据

2.2.1　受测者测试过程

部分受测者在培训前未能在考核规定的时间内完成模拟手术操作。经过培训后，所有受测者均能在考核规定的时间内完成考核。

2.2.2　评分用时及评分者反馈

先后有4位考官担任过模拟考核的评判工作，所有考官均能在考生完成考核时及时完成评分。部分考官表示，简洁的评价量表条目描述有助于他

们快速掌握评分要点，明确在考核中的观察重点并及时打分。所有考官均表示，考核前针对打分趋势进行的评分者培训十分必要，其有助于他们在实际评分过程中规范打分行为。有部分考官提出，“团队协作”相关指标在本研究所设置的模拟环境下较难准确打分，这些指标更适用于在真实手术环境下的直接观察。

2.2.3　评分者间一致性

经过对培训前考核成绩的分析，结果显示总维度的Spearman相关系数为0.993，各维度分布区间为0.953~0.989，P<0.01。在培训后考核成绩的研究中，总维度的Spearman相关系数为0.970，各维度分布区间为0.888~0.981，P<0.01。

2.3　基于内部结构的效度证据

2.3.1　内部一致性信度

针对培训前的90份成绩，计算得出总量表的Cronbach's α系数为0.987，各维度的Cronbach's α系数分布区间为0.937~0.969；针对培训后的90份成绩，计算所得总体Cronbach's α系数为0.947，各维度的Cronbach's α系数分布区间为0.798~0.914。

2.3.2　难度及区分度

总结本次考核的研究结果，整体来看，考核的难度适中，区分度表现出明显的高效性，能够有效地分辨出受测者之间的技能层次差异（表2）。

2.3.3　验证性因子分析

由于本研究评价量表的结构有参考来源，因此使用验证性因子分析来检验评价量表的结构效度。选用培训后的成绩数据，采用最大似然法进行验证性因子分析，以检验模型的拟合情况。结果显示，模型拟合度可接受[12]，具体结果见表3。

表2　TURP关键技术评价量表难度及区分度

维度	基于培训前成绩（n=90）		基于培训后成绩（n=90）	
	难度指数 P值	区分度 D值	难度指数 P值	区分度 D值
器械准备	0.655	0.747	0.929	0.661
组织保护	0.608	0.723	0.904	0.538
切除效率	0.587	0.681	0.896	0.493
手术流程	0.558	0.722	0.872	0.513
基础知识	0.536	0.731	0.894	0.739
团队协作	0.579	0.674	0.895	0.513
考核总分	0.595	0.704	0.900	0.561

表3　验证性因子分析模型的拟合优度

检验量	统计值	判定标准
χ^2/df	1.686	<5可接受，<3良好
RMSEA	0.087	<0.1可接受，<0.05良好
SRMR	0.036	<0.1可接受，<0.05良好
CFI	0.953	>0.9，接近0.9基本可接受
TLI	0.939	>0.9，接近0.9基本可接受

注：P<0.001；RMSEA即近似均方根误差（root mean square error of approximation）；SRMR即标准化均方根残差（standardized root mean square residual）；CFI即比较拟合指数（comparative fit index）；TLI即塔克-刘易斯指数（Tucker-Lewis index，也称非规范拟合指数）。

2.4　基于与其他变量关系的效度证据

基于培训前成绩，考核各维度间的相关性检验结果均达到P<0.01的显著水平，各维度间呈现出显著的强相关性，因此可用总分来综合反映考生在各维度的能力水平（表4）。

基于培训后的成绩，考核各维度间相关性检验的结果均达到P<0.05的显著水平。大部分维度间呈现出中至强的相关性，这说明对于培训后的受测者而言，不同维度之间既相互关联又有所区别。因此，在评估受测者的能力时，需要同时关注其在各个维度上的水平（表5）。

表4　TURP关键技术评价量表考核维度间相关分析（基于培训前数据，n=90）

项目		器械准备	组织保护	切除效率	手术流程	基础知识	团队协作	总分
器械准备	r值	1.000	0.880	0.812	0.825	0.823	0.816	0.934
	P值		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
组织保护	r值		1.000	0.826	0.840	0.842	0.811	0.961
	P值			<0.001	<0.001	<0.001	<0.001	<0.001
切除效率	r值			1.000	0.882	0.696	0.836	0.900
	P值				<0.001	<0.001	<0.001	<0.001
手术流程	r值				1.000	0.786	0.868	0.925
	P值					<0.001	<0.001	<0.001
基础知识	r值					1.000	0.772	0.886
	P值						<0.001	<0.001
团队协作	r值						1.000	0.913
	P值							<0.001

表 5　TURP关键技术评价量表考核维度间相关分析（基于培训后数据，n=90）

项目		器械准备	组织保护	切除效率	手术流程	基础知识	团队协作	总分
器械准备	r值	1.000	0.504	0.487	0.259	0.225	0.558	0.641
	P值		<0.001	<0.001	0.014	0.033	<0.001	<0.001
组织保护	r值		1.000	0.597	0.672	0.573	0.557	0.900
	P值			<0.001	<0.001	<0.001	<0.001	<0.001
切除效率	r值			1.000	0.448	0.363	0.606	0.745
	P值				<0.001	<0.001	<0.001	<0.001
手术流程	r值				1.000	0.523	0.503	0.759
	P值					<0.001	<0.001	<0.001
基础知识	r值					1.000	0.439	0.692
	P值						<0.001	<0.001
团队协作	r值						1.000	0.787
	P值							<0.001

2.5　基于测试后果的效度证据

本研究在实施考核时并未设定合格线，通过计算受测者培训前后的成绩得出：培训前平均分为53.53 ± 17.31，培训后平均分为 81.05 ± 6.32。培训后的平均分相较于培训前显著提升（培训前–培训后，t=−21.68，P<0.001）。培训前的中位成绩为49.25分，上四分位成绩为63.38分，下四分位成绩为39.13分；培训后的中位成绩为81.00分，上四分位成绩为84.75分，下四分位成绩为78.00分。

3　讨论

3.1　TURP关键技术评价量表是适合用于评价初学者TURP技能的评价工具

通过本研究的效度验证结果，研究者发现本研究的I-CVI分布区间为0.857~1.000，这体现了量表条目的内容效度在专家评判下的较高一致性与适宜性。S-CVI值为0.952，这一数值充分表明该量表在评价TURP关键技术方面具有较高的内容效度。在培训前后应用该量表所得到的成绩显示，培训前的Spearman相关系数分布区间为0.953~0.993，培训后的分布区间为0.888~0.981。这表示在采用TURP关键技术评价量表时，不同评定者间的评分具有极高的一致性；培训前后，总量表的Cronbach's α系数分别为0.987和0.947，这表明量表具有高度的内在一致性。考核难度中等、区分度高，评价各维度间的相关性在低水平受测者中呈现高度相关，而在相对高水平受测者中则既相关又相互区别。验证性因子分析结果显示：χ^2/df=1.686、CFI=0.953、TLI=0.939、RMSEA=0.087、SRMR=0.036，模型拟合度可接受。由此可以认为，在TURP初学者中，不论其处于模拟培训的任何阶段，该考核都是适合用于评价其TURP关键技术的理想工具。

3.2　模拟培训可改善初学者TURP关键技术行为表现

本研究中，受测者培训后的考核成绩较培训前明显提升。观察受测者培训前后的成绩标准差及四分位成绩间距，发现培训后的考核成绩区间较培训前有所减小。这说明经过TURP模拟培训能够缩小不同基础受测者在行为表现上的差距，以此提示开展TURP模拟培训对于初学者习得TURP关键技术是有效的，这也符合既往相关研究的结果[13-14]。此外，基于培训后成绩所得出的评价维度间相关性降低的现象，可能是由于培训前受测者在各维度上的能力水平均较低。经过培训后，他们在不同维度上的能力呈非同步提升。因此，通过分析受测者在不同维度的成绩情况，可以提示受测者进一步训练的重点方向。未来，可进一步研究模拟培训的有效性，并探索TURP关键技术的学习曲线，以获得模拟器训练的最佳实践方案。

3.3　开发可应用于手术直接观察的TURP技术评价工具的必要性

泌尿外科手术技能教学的目的在于通过规范的手术技能培训，使毕业后阶段的泌尿外科住院医师掌握泌尿外科手术的特点，并将这些技能应用到临床实践中[15]。因此，开发适用于手术室直接观察的手术技能评价工具显得十分必要。在本研究评

价量表的使用过程中，发现部分指标更适用于手术室直接观察的应用场景。未来，研究团队计划将本评价量表与毕业后教育的“监督—置信—允许”五阶段模型[16]进行整合，以开发针对手术室TURP关键技术的评价量表。在符合医学伦理的前提下，该量表将在已通过模拟准出考核的住院医师的临床手术中应用，并开展相关的效度研究。

3.4 研究的局限性

本研究目前仅在瑞金医学模拟中心进行了单中心应用，接下来需进一步扩大研究范围，纳入全国各地更多的中心，以获得更加充足的效度证据，从而论证该评价工具在不同社会背景和医疗环境下的分数解释效力。本研究尚未对评价工具的分数线进行划定，接下来将在后果证据部分进一步探索，结合模拟器训练的最佳实践，以研究并划定模拟培训的合格线。

4 结语

综上所述，本研究结果表明，在TURP模拟实践考核中参照《TURP关键技术评价量表》的成绩来评价初学者的TURP关键技术水平是可行的。未来需要进一步扩大该评价量表的应用范围，以获得更加充分的效度证据。本研究所提供的效度证据，为使用该评价量表所得分数作为评价初学者TURP关键技术水平的依据提供了初步支持。

利益冲突　所有作者声明不存在利益冲突

作者贡献声明　汪成合：研究实施、数据收集、撰写论文；汪晓冉：统计处理数据、统计分析、撰写论文；王佳玉：研究设计、研究指导、论文构思、撰写论文；蒋光亮：研究实施、数据收集；何竑超：研究实施、论文修改；徐丹枫：行政支持

参考文献

[1] Cornu JN, Gacci M, Hashim H, et al. EAU guidelines on management of non-neurogenic male lower urinary tract symptoms (LUTS), incl. benign prostatic obstruction (BPO)[M]. EAU Guidelines Office, Arnhem, the Netherlands: Edition Presented at the Annual EAU Congress, 2023.

[2] 孙国锋，张弋，于澄钒，等. 前列腺电切模拟机表面效度和内容效度研究[J]. 中华外科杂志，2012，50(2)：157-160. DOI: 10.3760/cma.j.issn.0529-5815.2012.02.015.

[3] 王佳玉，何竑超，江哲涵，等. 基于Delphi-AHP法构建经尿道前列腺切除手术关键技术评价量表[J]. 中国毕业后医学教育，2023，7(7)：551-554，570. DOI: 10.3969/j.issn.2096-4293.2023.07.010.

[4] 左超，杨昆霖，李志存，等. “双沟双环法”经尿道铥激光前列腺剜除术单一术者学习曲线分析[J]. 中华医学杂志，2023，103(20)：1563-1567. DOI: 10.3760/cma.j.cn112137-20230212-00198.

[5] 张洁，李亚男.《国际汉语教师证书》考试的效度研究[J]. 华文教学与研究，2020(2)：64-70. DOI: 10.16131/j.cnki.cn44-1669/g4.2020.02.009.

[6] 刘庆思. 考试的构念界定及验证探析[J]. 中国考试，2018(10)：30-34. DOI: 10.19360/j.cnki.11-3303/g4.2018.10.007.

[7] AERA, APA, NCME. Standards for educational and psychological testing [M]. Washington, DC: AERA, 2014: 26-31.

[8] 史静琤，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)：49-52. DOI: 10.3969/j.issn.1672-7347.2012.02.007.

[9] 苏丽敏，何慧爽. 基于区间数的Spearman秩相关系数的多属性决策方法[J]. 统计与决策，2019，35(6)：51-53. DOI: 10.13546/j.cnki.tjyjc.2019.06.011.

[10] 王佳玉，叶庭均，黎健，等. 膝关节镜实践技能模拟器考核的效度研究：来自三家三甲医院的多中心研究[J]. 中国毕业后医学教育，2022，6(6)：610-613，619. DOI: 10.3969/j.issn.2096-4293.2022.06.024.

[11] Chen P, Li F, Wu C. Research on intrusion detection method based on pearson correlation coefficient feature selection algorithm [J]. J Phys Conf Ser, 2021, 1757(1): 012054. DOI: 10.1088/1742-6596/1757/1/012054.

[12] Xu L, Li J, Yin L, et al. Examining the structure of difficulties in emotion regulation scale with Chinese population: a bifactor approach [J]. Int J Environ Res Public Health, 2021, 18(8): 4208. DOI: 10.3390/ijerph18084208.

[13] 张弋，朱鹤柳，金顺，等. 虚拟模拟器在经尿道前列腺切除术培训中的有效性评价[J]. 中华泌尿外科杂志，2011，32(7)：486-489. DOI: 10.3760/cma.j.issn.1000-6702.2011.07.015.

[14] 邱智，张小东. 前列腺电切模拟机在泌尿外科教学的应用和效果评价[J]. 中国病案，2014，15(11)：71-72. DOI: 10.3969/j.issn.1672-2566.2014.11.035.

[15] 马琳，纪志刚. 从传统手术教学到泌尿外科手术技能三阶段培训体系的建立[J]. 国际外科学杂志，2021，48(9)：583-586. DOI: 10.3760/cma.j.cn115396-20210915-00365.

[16] Holmboe ES, Durning SJ, Hawkins E. 临床胜任力评价实用指南[M]. 张抒扬，主译. 北京：北京大学医学出版社，2021：12-13.

（收稿日期：2024-09-14）

（本文编辑：邬小梅）