

欧洲医学教育学会

教育指南

(第 1 3 、 1 8 号文件)

Association for Medical Education in Europe
Education Guide

翻译 万学红 邓洪



人民卫生出版社

欧洲医学教育学会 教育指南

(第 13、18 号文件)

翻译: 万学红 邓 洪

主审: 周同甫 李晓松

人民卫生出版社

图书在版编目(CIP)数据

欧洲医学教育学会教育指南/万学红,邓洪翻译.
北京:人民卫生出版社,2002

ISBN 7-117-05145-0

I. 欧... II. ①万...②邓... III. 医学教育—欧洲
IV. R-4

中国版本图书馆 CIP 数据核字(2002)第 072715 号

欧洲医学教育学会教育指南
(第 13、18 号文件)

翻 译: 万学红 邓洪

出版发行: 人民卫生出版社(中继线 67616688)

地 址: (100078) 北京市丰台区方庄芳群园 3 区 3 号楼

网 址: <http://www.pmph.com>

E-mail: pmph@pmph.com

印 刷: 三河市潮河印刷厂

经 销: 新华书店

开 本: 850×1168 1/32 印张: 4.5

字 数: 79 千字

版 次: 2002 年 11 月第 1 版 2002 年 11 月第 1 版第 1 次印刷

标准书号: ISBN 7-117-05145-0/R·5146

定 价: 8.50 元

著作权所有,请勿擅自用本书制作各类出版物,违者必究
(凡属质量问题请与本社发行部联系退换)

前 言

为了推动正在我国广泛深入开展的医学教育评估的研究与改革,为了吸取发达国家的先进经验和研究成果,我们从国外大量的相关文献中选择一些对我国医学教育改革和医学考试有重要借鉴价值的文献,译成中文供国内同道参考。本小册子是源自欧洲医学教育学会出版的医学教育指南第 13 号和 18 号文件,由纽约中华医学基金会(China Medical Board of New York Inc. USA)主席 R. Schwarz 博士和国际医学教育专门委员会(Institute for International Medical Education, IIME)主任 Andrzej Wojtczak 教授的推荐,并得到作者同意和支持,在中国出版。

对医学生操作表现(performance)的评估历来是医学教育评估中的一个难点,也是近年国内外研究的热点。特别是我国正在引进国际医学教育专门委员会制定的全球医学教育最基本要求(Global Minimum Essential Requirements in Medical Education)之际,有必要认真研究国外在医学生操作表现评估方面的成功经验。

对医学生的操作表现的评估,与笔试大不一样,它需要观察学生与病人之间的互动。由于病人的应用对考试医学生能力具有高度的有效性,因此被用于医学考试中操作表现的评估。但是由于其对考生不能始终提供一致的情况等原因,对大量的考生而言,存在考试的不公正性等问题,因此早在 1963 年的美国医师执照考试中就停止使用。目前,除针对较少数应考者的考试外,已不可能用真实病人进行全校医学毕业生或全国性医师

执照考试等较大规模的考试。因此很有必要深入研究,建立一些至少在某些方面或一定程度上能代替真实病人的考试方法。因此,各种各样的模拟技术应运而生,并越来越显示出其独特的优势,其研究与发展也非常迅速。

医学教育指南第 13 号文件对模拟病人、模拟工具以及真实病人的使用以及他们各自的优缺点作了深入的分析,对我们很有启发意义,也有助于我们把握和应用。影像、音像制品、塑料或其他材料的各种模型,以及计算机模拟等都可作为病人替代品,尤其是计算机模拟值得我们深入研究和发 展。我国部分医学院校,如四川大学(原华西医科大学)、浙江大学(原浙江医科大学)、九江医学专科学校和中国医科大学等从 90 年代初就开始积极探索标准化病人在医学生临床技能评估中的应用,及时地引进了发达国家在考试方面的先进技术,取得一定的成绩,积累了经验,为我国各医学院校进一步研究和发 展这一考试技术提供了一个样板。国内医学院校在其它模拟技术方面也作了大量的研究,有些已试验性的应用于学生的临床考试中。吸取国外先进经验,结合国内学者的研究成果,探索并建立符合我国实际情况的各种模拟技术,是当今医学教育,特别是医学考试工作者们的重要课题。

第 18 号文件则对一个非常重要而我们常常没有给予高度重视的问题作了科学的阐述,并介绍了具体的方法供读者参考,这就是学生评估标准的设定。如我们常遇到的问题,要区分开具备能力和不具备能力的人,60% 的分数线最好吗? 这种决定应该以什么为基础? 多少分才合格? 我们对标准设定的程序及其结果有多少信心呢?

以前,标准设定的方法大都普遍使用在多选题和其它形式的笔试。在过去的十年中,随着操作表现的评估方法越来越普

遍和特定的专业认证考试的出现,有必要重新审视现存的标准设定的方法是否还能适应对操作表现评估的复杂性。目前,对书面和操作表现的考试,都有多种多样的标准的设定方法。本指南将着重介绍操作表现的评估,并用实例说明设定标准的程序。设定评估操作表现的标准,相对而言是一个比较新的研究领域,因此,迄今还没有一个广泛接受的方法。本指南首先对考试理论中的几个重要概念,如常模参照与标准参照、补偿标准与合取标准等作了深刻阐述;再介绍标准设定方法的框架、原则和一些应该考虑的实际问题,使读者在为考试需要而选用最恰当的标准设定方法时做出理性的选择;最后,以客观结构化临床考试为例,应用改良 Angoff 法制定其学生评估的标准,让读者看到了改良 Angoff 法这一最常用的标准设定方法的每一个实践步骤,具有很好的指导意义。

英国 Dundee 大学,医学教育中心 J. P. Collins、R. M. Harden 和 M Friedman Ben - David 教授都是医学生能力评价方面的著名专家,著作颇丰。

J. P. Collins 博士 新西兰奥克兰大学医学教育中心的主任,外科副教授,奥克兰墨的默医院学术部主任。他长期致力于医学大学本科生和医学毕业生的教育。

R. M. Harden 教授 英国 Dundee 大学医学教育中心主任, Dundee 大学医学、口腔和护理系教务长、英国 Dundee 教育发展署(苏格兰临床医学和口腔医学毕业后教育理事会)主任。Harden 教授是著名的医学教育和考试专家,他于 1979 年在全世界首次报道了客观结构化临床考试。

M Friedman Ben - David 博士是美国外国医学毕业生培养委员会(ECFMG)临床技能证书颁发项目的前副主任,目前在英国 Dundee 大学医学教育中心做访问教授。她主要的研究领域是

在校教育及继续教育中医生(医学生)能力的评价。

四川大学副校长张肇达教授是本书的总策划者,对本书的翻译工作给予了极大的关注和指导;在编译过程中,我们还得到了四川大学外语学院邱望生教授、雷丽敏副教授的大力支持,在此表示诚挚的谢意。对纽约中华医学基金会主席 R. Schwarz 博士和国际医学教育专门委员会主任 Andrzej Wojtczak 教授的大力支持表示最衷心的感谢。

由于文中知识面广而编译者水平有限,在文中难免有不妥甚至错误之处,望读者斧正。

(万学红 邓洪)

目 录

真实病人、模拟病人和模拟工具在临床考试中的应用	1
欧洲医学教育协会教育指南第 13 号文件	1
概述	1
临床能力考试	3
临床考试中“病人”的使用方法	5
“真实”病人	7
模拟病人和标准化病人	9
病人替代物	16
“病人”的选择	19
对病人角色的预期	19
标准化水平的要求	22
准备事项	23
评估的场景	25
现实性或真实性要求	25
在临床考试中使用病人的建议	26
涉及真实病人或者模拟病人的事项	26
仅涉及真实病人的事项	27
仅涉及模拟病人的事项	27
结论	28
学生评估标准的设定	31
欧洲医学教育学会教育指南第 18 号文件	31
概述	31

为什么要使用标准设定程序	31
本指南的目的	32
关键概念	33
常模参照与标准参照	33
补偿标准与合取标准	35
标准设定的方法	37
以考试为中心的模式	37
以应试者为中心的模式	39
改良 Angoff 法	40
专家的选择	42
书面标准与操作表现标准	44
标准设定对教育的好处	45
教师教学水平的提高	45
考试素材的质量控制	46
第一步——总体介绍	46
第二步——对“操作”考站的说明	47
第三步——临界考生特点	49
第四步——专家小组成员评分	51
评估	55
结束语	57
英文原文	59

真实病人、模拟病人和 模拟工具在临床考试中的应用

欧洲医学教育协会 教育指南第 13 号文件

J. P. Collins 博士 新西兰 Auckland 大学

R. M. Harden 教授 英 国 Dundee 大学

概 述

在临床能力的评估中,观察应试者与病人之间的互动很重要。此时,该病人的角色将随预期的学生与病人之间的互动的程度而变化,无论病人有无体征。被用于考试的病人可能是真实的病人,或许是接受过这方面训练的模拟病人,后者能够重复提供一种特定的场景。模型、模拟工具、录像带、录音带以及电脑模拟都可以被用来替代病人。

从没受过训练的真实病人到受过系统训练的模拟病人是一个连续体:

1. 临床实践中“真实”病人;
2. “真实”病人同意参加临床考试,但是没有事前的排演;
3. “真实”病人已经就可能发生的情况进行了排演;
4. 为了考试目的,“真实”病人被要求在病史陈述或病情介绍方面做出一定改变;

5. 在考试中,“真实”病人自身的医学经历是他病情介绍的基础,但他也常常为了考试而对自己的表演做些实质性的改变;

6. 模拟病人则常常得到一张提纲,规定了对他们的要求;

7. 模拟病人要对于给予的病情摘要或方案很熟悉,除此之外,他们可按自己的意愿回答问题;

8. 在考试之前,模拟病人对考试的内容要有全面的了解,并就此进行充分准备。简单和先进的模拟工具可以被用来评估查体的技巧和实际操作的程序。

只有在少数情况下,考试中对病人表现的选择才被局限于某一方法。比如,某些体征不可能被模拟,必须要用真实的病人;在非常微妙或情绪表现难度较大时使用模拟病人;在使用真实病人不合适的情况下,如心肺复苏等。然而,在许多情况下,没有哪一种方法是绝对正确的。方法的采用将由当时的具体情况以及考试的需要来决定。

在考试中,影响病人选择的因素包括:

1. 对什么进行评估,包括所要求的异常程度和与病人互动的程度。

2. 所要求的标准化水平,重点强调高级别的国家级考试所需的标准化要求。

3. 准备事项,包括真实病人和受过训练的模拟病人的可获得性及其成本。

4. 场景条件,例如基于实践的或正式的客观结构化临床考试。

5. 所要求的现实性或真实性。

无论病人是真实的还是模拟的,在临床考试中都可采用切实可行的步骤来获得最大程度的好处。

临床能力考试

在最近的一次国际医学教育会议上,一位参会者问了这样一个问题,“在客观结构化临床考试中是否有可能使用真实的病人?”这个问题真是让人吃惊不小。因为许多医学中心在客观结构化临床考试中只使用真实病人,也有一些既使用真实病人又使用模拟病人。实际上,从最初对客观结构化临床考试的描述(Harden et al. 1975),到随后的美国医学工程师学会(ASME)所发表的医学教育小册子(Harden & Gleeson 1979),都涉及到利用真实的病人。然而,提出这个问题仍然是可以理解的。在许多医学中心,利用标准化或模拟病人来评估考生的临床能力已成为客观结构化临床考试的同义词。人们很少注意到什么时候使用真实的病人,什么时候使用标准化的模拟病人最为合适这个问题。

对医学生临床能力的评估包括诸多不同但又相互关联的方面,如与病人沟通交流和体格检查的技能等。多年以来,以病人为中心的床旁考试一直是评估医学生是否掌握这些技能的卓越方法。它在最终的考试中的重要性得到公认,医学生在临床能力的考查上是否能过关成为考生是否能最终毕业的一个必要条件(Stokes 1974)。然而,也有人质疑这种传统的临床考试是否具有可靠性。

大家已经认识到不同的病人所提供的考试难度具有很大的差别,考官在判定过程中的客观性会受到挑战,这些可能导致严重的心理测量缺陷以及对考试可靠性的质疑(McGuire 1966, Pokorny & Frazier 1966, Wilson et al. 1969, Foster et al. 1969)。这些因素使得一些主考者和考试委员会在六十年代疏远了以病人为评估基础的考试(Hubbard et al. 1965),而去寻求其它的替

代方式。以多选题为基础的笔试具有极大的可靠性,具有容易打分和覆盖考试内容面积大等优点;其它采用的笔试方法包括改良的问答题(Knox 1975)和病人处理问题(Harden 1983)。

客观结构化临床考试(OSCE)的出现和发展提供了一种不同的选择,它强调病人的参与,旨在达到评估的可靠性和有效性。Barrows 和 Abrahamson 于 1964 年提出了模拟或程控病人这一概念,他们的目的是促进临床技能的学习。这一概念后来发展到被用来评估临床技能(Barrows 1968, Lamont and Hennen 1972, Stillman 1976)。

使用客观结构化临床考试和模拟病人在八十年代得到发展(Newble & Swanson 1988, Harden 1990, Stillman et al 1986, Stillman & Swanson 1987, Barrows et al. 1987)。观察考生和病人的互动成为评估有效性的必要条件,在这一点上大家已经达成了共识。在 90 年代,真实的或以操作表现为基础的评估进一步得到强调,病人在这一评估程序中扮演了一个中心的角色。渥太华会议(Hart et al. 1986, Hart & Harden 1987, Bender et al. 1990, Hart et al. 1992, Harden et al. 1992)反映出通过利用病人对学生临床能力进行评估的兴趣正方兴未艾。

在临床考试中存在着三个变量,即学生、考官和病人(见图 1)。应该达到的目标是将主考人和病人标准化,那么考生的表现就可以被看成是测量他或她临床能力的依据。这本小册子为怎样在临床评估中使用病人提供了指南。它规定了在评估过程中病人参与的不同方式,其中包括了真实病人、模拟病人或病人替代品。本指南还描述了在选择要使用的病人时,哪些因素应加以考虑。

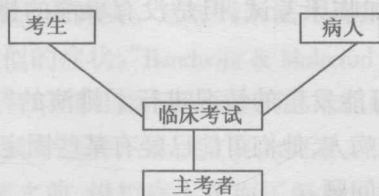


图1 临床评估中的三个变量

临床考试中“病人”的使用方法

在临床考试中,有三种使用病人的方法:

1. 在医疗实践中所遇到的“真实”病人;
2. 模拟病人,即通过训练而扮演病人角色的个人;
3. 病人替代品,包括音像制品、塑料或其它模型以及计算机模拟。

真实和模拟病人之间的区别可能并不像文献中描述的那样明确清晰。实际上,从真实的病人(即完全没有预演成分的病人)到模拟病人(即广泛地接受过培训,其每一反应都经过周密的设计,见图2)是一个渐变的连续体,它包括以下几个方面:

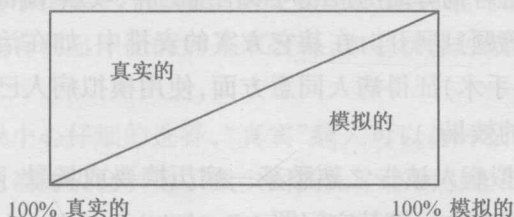


图2 在临床评估中使用从真实病人到模拟病人的连续体

1. 临床实践中未经预先排演准备的真实病人。如通科实践中的会诊,学生表现的评估判定是通过录制下来的会诊技能记录进行分析后而获得(Campbell et al. 1993)。

2. 同意参加临床考试,但是没有事前的排演的“真实”病人。

3. 已经就可能发生的情况进行了排演的“真实”病人。可能有这样的一组病人,他们可能已经有某些固定的体检发现,如心脏杂音或关节问题。

4. 为了考试的目的,被要求在病史陈述或病情介绍方面做出某些改变的“真实”病人。例如,病人可能被要求强调其病史中的某个特别的方面或体检中某一发现。他们可能得到训练,以变更考试的难易水平。

5. 在考试中,“真实”病人自身的医学经历是他病情介绍的基础,但也会常常为了考试而对自己的表演做些实质性的改变。他们可能被要求模拟关节炎中关节的僵硬、神经疾患中肌肉的僵直或者在胆道疾病中右上腹压痛等体征。

6. 模拟病人则常常得到一张提纲,规定其应该做出的表现。在考试中,这些模拟病人的使用是当病人和该考生的互动需求最小的时候,例如在使用检眼镜进行常规检查时,或就糖尿病、哮喘、高血压、造瘘术的护理、免疫和避孕对病人进行教育的时候。来自咨询方面的挑战可以围绕肥胖、吸烟、酗酒和恶性肿瘤的防治等题目展开。在其它方案的安排中,如在治疗程序上(包括外科手术)征得病人同意方面,使用模拟病人已经取得了令人满意的效果。

7. 模拟病人被告之熟悉某一病历摘要或场景,除此之外,他们可以随意地回答其它问题。Baerheim 和 Malterud 在 1995 年就已经描述了对模拟病人的使用,他们鼓励模拟病人根据个人的背景来调整自身的角色。除了根据自己的选择对个人的经历做一些变更外,他们还受训被安排排演某些有限的角色。那些模拟病人报告说这样的表现容易和自然得多。他们中有人说

“这使人感到相当的真实。这对我来说尤其容易,因为我以前有过这样一些类似的症状。”Baerheim & Malterud 在 1995 年描述说模拟病人的短暂培训可能不足以达到标准化病人的表现。他们认为在许多考试中这并不是很重要的因素。

8. 在考试之前,模拟病人全面了解考试的内容,并就此进行充分的准备。这样的病人可能受训模拟体格检查中的某些异常,比如咳嗽、腹部压痛或其它更为复杂的症状和体征。很多情况都可以被模拟出来,如异常的个性、敌对的情绪、不愿交流或渴望帮助的病人或病员家属,其中还包括精神障碍,诸如抑郁症(有或者没有自杀的倾向)等等。药物滥用、精神混乱、精神崩溃或严重的外伤后遗症全都可以被模拟出来(Collins 1992)。

使用真实病人、模拟病人和病人替代物在本指南中有更为详尽的描述。

“真实”病人

在临床评估中,“真实”病人的传统使用方法包括长期和短期病例表现(Stokes 1974, Newble 1991)。作为一个长期的病情,病人被要求提供一个明确的病史和足够的体征。而短期病情常常由那些有明显体征的病人来表现,如甲状腺肿或风湿性关节炎病人。

如果小心仔细的选择,“真实”病人可以为评估学生临床技能提供适当的机会。为了减少因病人个体而引起的变化,可以招聘一小群病人,他们都具有同样的病情和相同的体征,例如:慢性阻塞性气道疾病、多囊肾或主动脉瓣疾病(Newble 1992)。

在许多医学中心里,真实病人,而非模拟病人被用于客观结构化临床考试。病人对于他们能参与重大的临床考试感觉良好(Sharma et al 1994, Persaud & Meux 1990)。Sharma 等报道说,有

三分之二的病人感觉考试令人愉快、有帮助和非常重要。大多数病人说他们愿意再次参加,很多人还把它当成参加社会活动来期盼。

在过去,使用将接受外科手术的麻醉病人来做直肠和阴道检查的评估常常没有得到病人的许可。然而,这些做法在伦理道德上不能接受,合法性也受到质疑,而且教育价值也不确定。气管插管技巧的教学和评估也使用麻醉病人和刚死去的病人,虽然在有协议的情况下,尸体上插管获得推荐(Hinchley、Tyrrell et al 1992),但是这些做法仍受到谴责(Tonks 1992)。Brattebo 等在挪威的一次民意调查(1993)中发现大多数人同意利用最近去世的病人来训练和评估插管技巧。

使用“真实”病人的长处:

- 他们可以清楚地显示病变,比如甲状腺肿、心杂音、高血压和怀孕;
- 他们在许多情况下容易获得;
- 他们不需额外的资金并且只需最低限度的组织上的支持;
- 除了提供门诊病人的旅差费外,他们不需要更多的成本。
- 对于医疗人员和学生来说,他们有很高的可接受性(Newble 1991);
- 如果将他们和模拟病人一并包括在考试中,可以使整个评估具有更高的可信度。如果在一个考试里,只使用模拟病人,其可信度就会大打折扣,而使用真实的病人则可大大降低可质疑度。

使用“真实”病人的不足:

- 在某些情形下不太好找真实病人;