



浙江大学医学院

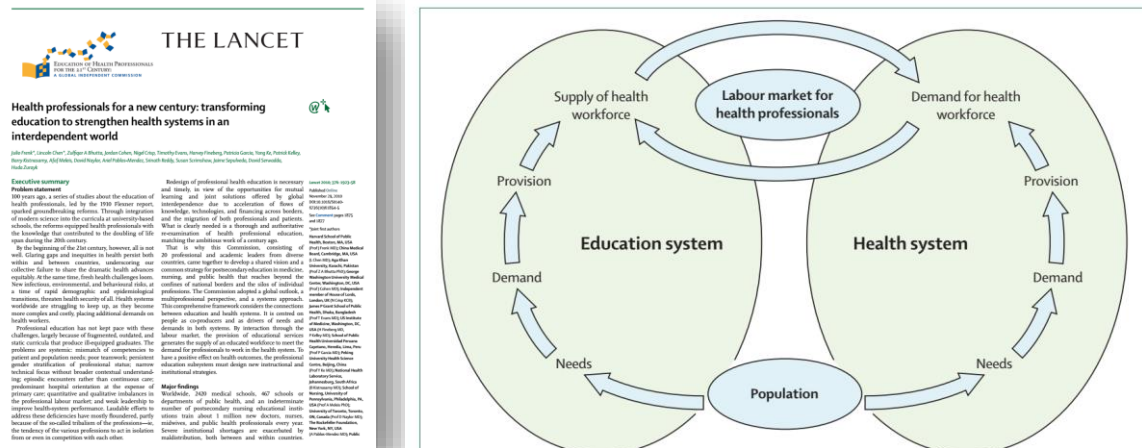
ZHEJIANG UNIVERSITY SCHOOL OF MEDICINE

基于米勒模型的 程序性评价实践与探索

报告人：韩魏

公认的第三代医学教育改革

《新世纪医学卫生人才培养：在相互依存的世界，为加强卫生系统而改革医学教育》



- 医教协同：以患者和人群为中心，以健康为中心
- 基于结果导向的教育--以胜任力为导向
 - 课程体系
 - 学习模式--转化式学习
 - 跨专业教育 (IPE)

转化式学习 (transformative)

领导特征与创新能力，培养有能力的医学变革推动者

- 向批判性思维、胜任力转化
- 向团队合作转化
- 向主动学习转化

形成式学习 (formative)

注重培养学生的价值观，目标是培养有职业素质的人才

记忆式学习 (informative)

旨在知识、技能的获取，培养专门人才

学习层次的递进

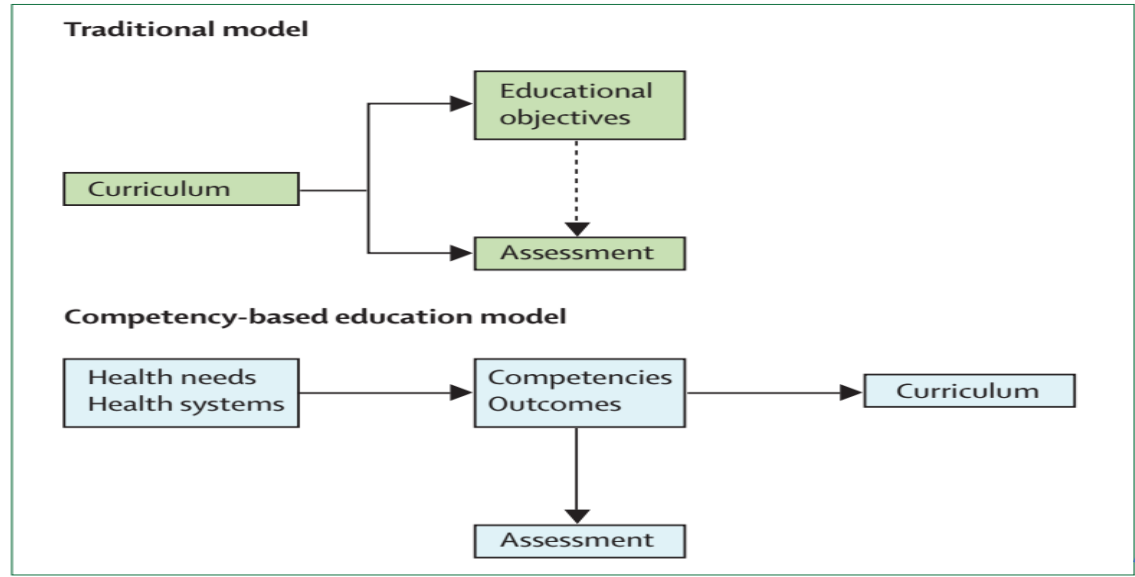
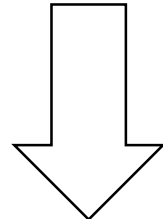


Figure 9: Competency-based education



评价：评什么、怎么评、评了怎么用 医学教育

- 结果导向的教育 (outcome-based education)
- 胜任力导向的教育 (competency-based medical education)



- **要求对医学生的能力进行合理、全面的评价**
- 形成性评价、终结性评价、**程序性评价** 理念、方法、实践
- 以评促学--->更强调形成性评价和评价的反馈--->以学生为中心学习 (student-centered learning) , 注重支持性、创造性学习

本质：是多种评价方式有机互补的联合式评价

理念

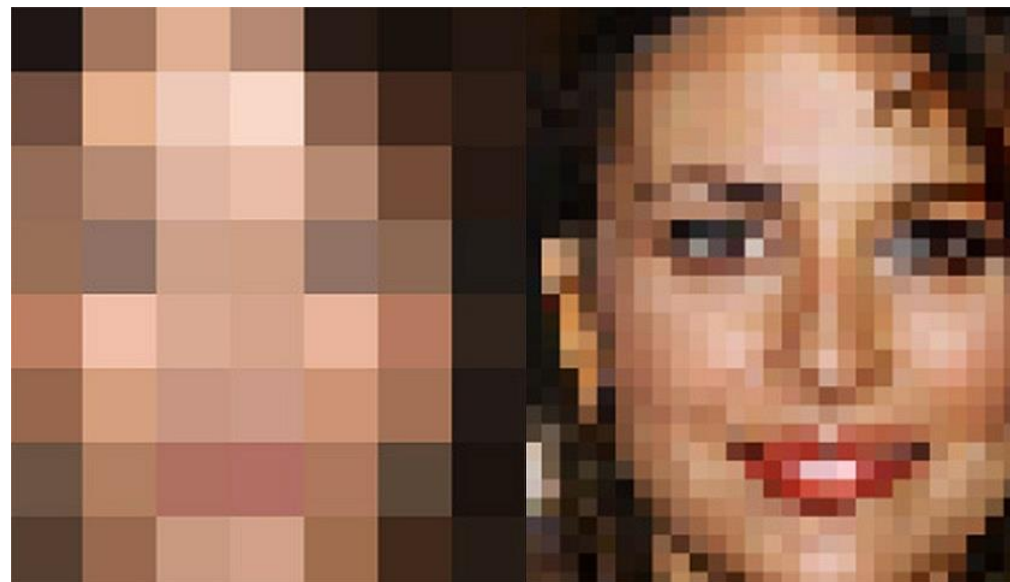
- 任何一个单项评价，都不能对能力进行全面评价
- 任何一个单项评价，都只能评价1-几个方面
- 一个像素---->众多像素的有机组合----->图

- 利用不同评价方法之间的互补性，实现 $1+1>2$ 的属性
- 将各部分评价有机汇聚，促进整体上、较为全面的评价

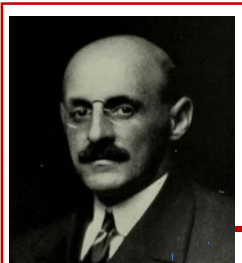
问题：

- 评价的框架--怎么有效的联合和互补，什么时候重点评价什么
- 评价的工具--评价方法的选择

高低像素对比图



浙医百年医学教育概述



Flexner's report

求是书院 1897年



韩清泉

中国人自己创办的最早的医学专门学校；1912年招医学生60人。1916年，“教育部”召开全国学校成绩展览会，以93分之成绩，居全国之首，列为特等

1912年 浙江医学专门学校

国立浙江大学 1928年

国立浙江大学 医学院 1945年

1947年 浙江省立医学院



王季午

在卫生部组织的全国部分医学院校医学系毕业生的统考中，1982年与1983年均获第二名，1984年获第一名

1952年 浙江医学院
1960年 浙江医科大学

Case Western Reserve University Reform of medical education

浙江大学医学院 1998年

以胜任力为导向的见习医生制改革 2004年

2000年 基础医学整合课程改革

2005年 招收八年制 4+4

2007年 系统/器官课程改革 ZJU-UCLA 整合课程

基础医学获批国家唯一的医学类试点学院 2011年

首先提出、首家批准 国家临床医学博士后 2015年

2016年 启动“医学+X”交叉博士研究生培养

通过临床医学专业认证 2019年

2016年 启动临床教学评价体系改革

见习医生制



浙大医学院临床医学5年制/5+3模式

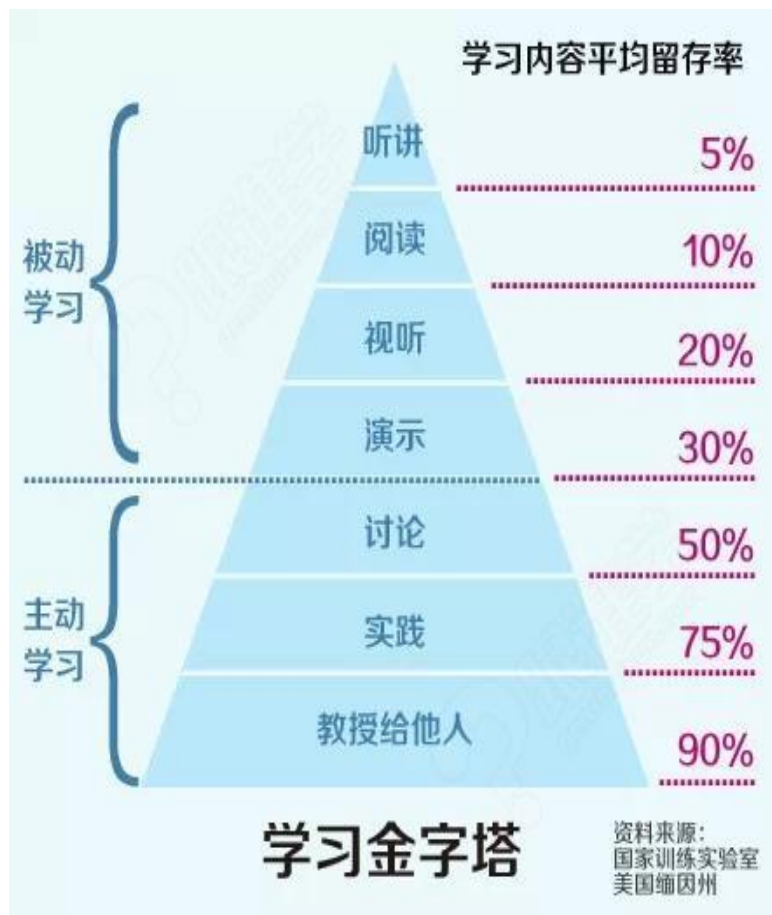
人文教育、职业道德、创新教育 **不止在课堂，更在言传身教**

基础教育阶段				桥梁课程				临床教育阶段			
第一学年		第二学年		第三学年				第四学年		第五学年	
秋冬	春夏	秋冬	春夏	秋	冬	春	夏	秋冬	春夏	秋冬	春夏
通识、大类											
	基础医学										
			预防医学		临床见习 1年		临床见习 1年		临床见习 1年		
					理论课	临床轮转		临床轮转		临床实习 1年	
				临床基础、临床见习、临床实习							

设计理念

做中学 (Learning by doing)

通过实践学习的留存率75%



浙医采取的模式

核心

激发学习动力

见习医生制改革

学习内容综合化

学习过程场景化

学习方式自主化

学习评价多元化

医学生学习

开始于病人 继续于病人 终止于病人

学生以床边学习为中心，学生本身就是医疗小组的成员，浸润在真实的临床环境中，并且全程参与病床管理和病人治疗，亲历医疗实践

- 足够的时间、深入地研究病人，**学生主要从病人身上学习知识，而讲课和教科书是辅助**
- 教师、学生之间充分的接触和交流
- 充满活力的学术氛围，科学探索的方法和精神



以授课为中心 **✗**

床边观察者 **✗**

知识记忆和积累为中心 **✗**

- 尽早广泛接触临床
- 提前完成医师角色定位

- 基于实践的学习，知识构建、能力培养
- 强调主动学习和自学能力
- 贯穿伦理、态度、沟通

临床见习教学安排

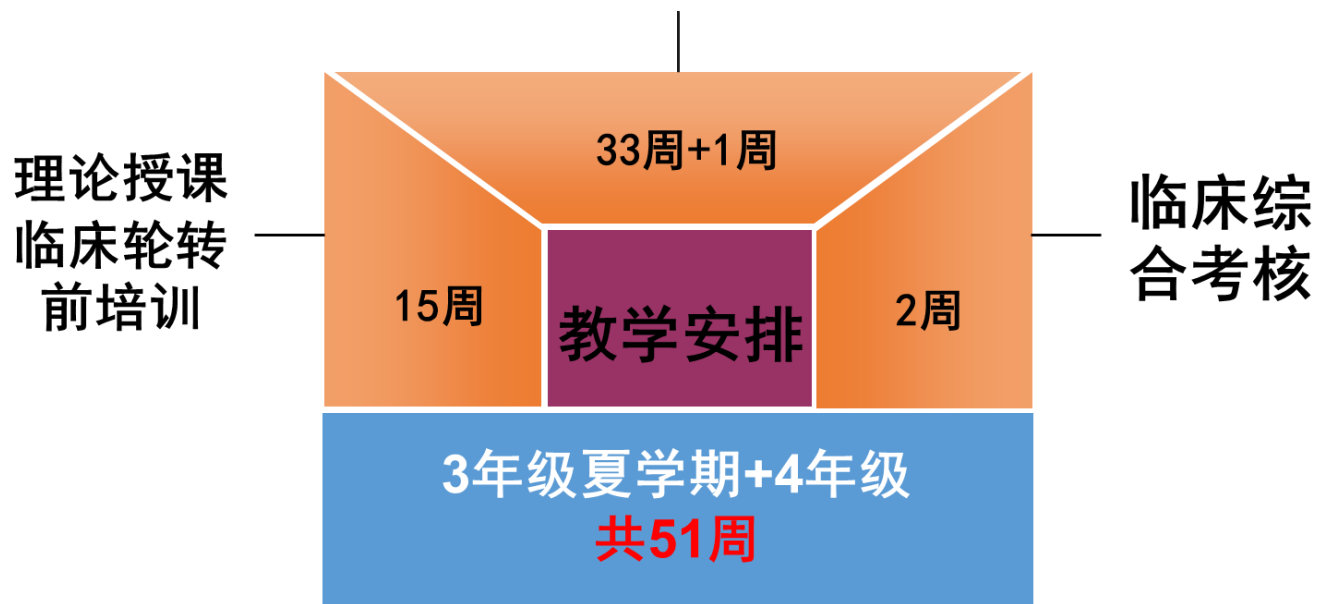
理论授课：

目的--搭建各学科的理论框架

内容--总论、典型疾病、其它核心内容等

“见习医师制”临床教学采用类似实习生在病房轮转的方式，使学生从临床见习开始就作为医疗小组成员融入临床团队，在临床工作中完成学习。“见习医师制”覆盖21个临床教学基本科室，教学形式多样，以病房轮转为主，以小讲课、思维训练、技能训练、病例讨论、自学等为补充。

临床见习轮转



- **14周**理论课（内、外、妇、儿、神经、精神、传染、眼、耳鼻咽喉、影像、皮肤、口腔、中医）
- **1周**考试（内、外、口腔、中医）

33周病房轮转安排

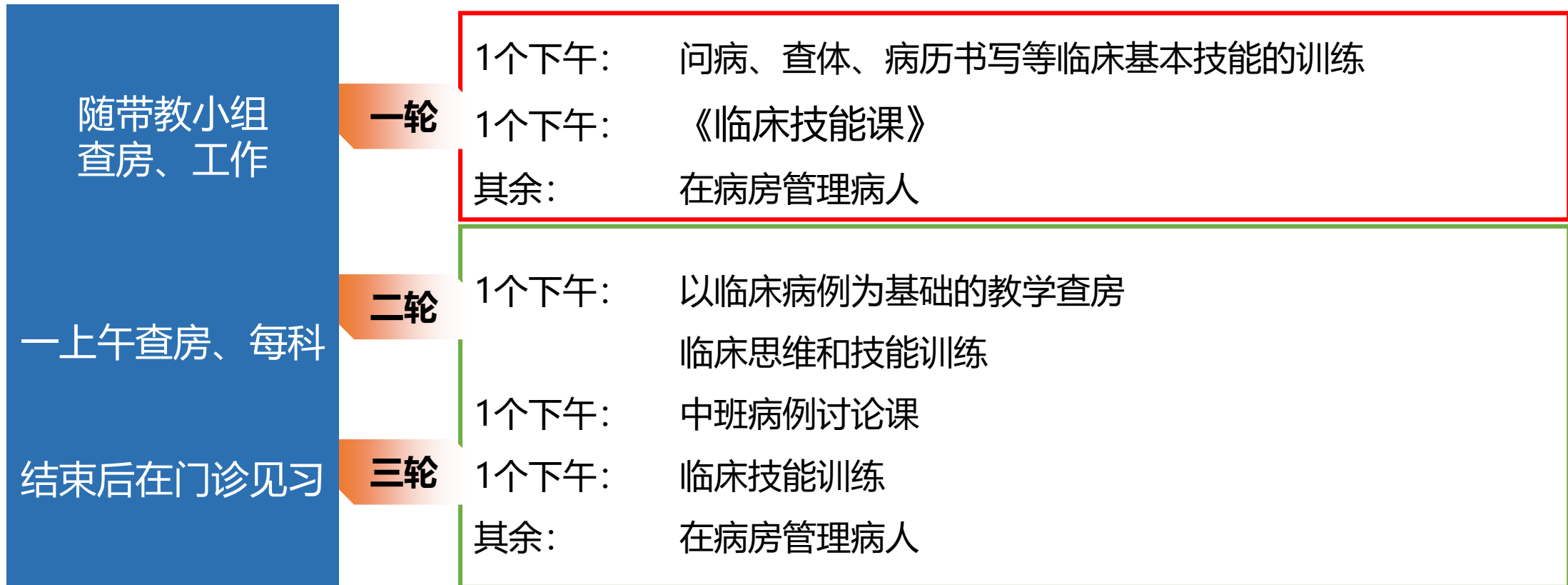
周次 轮次	1	2	3	4	5	6	7	8	9	10	11
第一轮	心内科		呼吸科		消化科		血液	内分泌	肾内	影像	考试
周次 轮次	12	13	14	15	16	17	18	19	20	21	22
第二轮	普外（肿瘤）				泌尿	骨科	胸外	脑外	眼科	耳鼻咽喉	考试
周次 轮次	23	24	25	26	27	28	29	30	31	32	33
第三轮	妇产科			传染	精神	皮肤	儿科			神经	考试

转轮分3轮，总共33周，覆盖了21个临床基本科室。

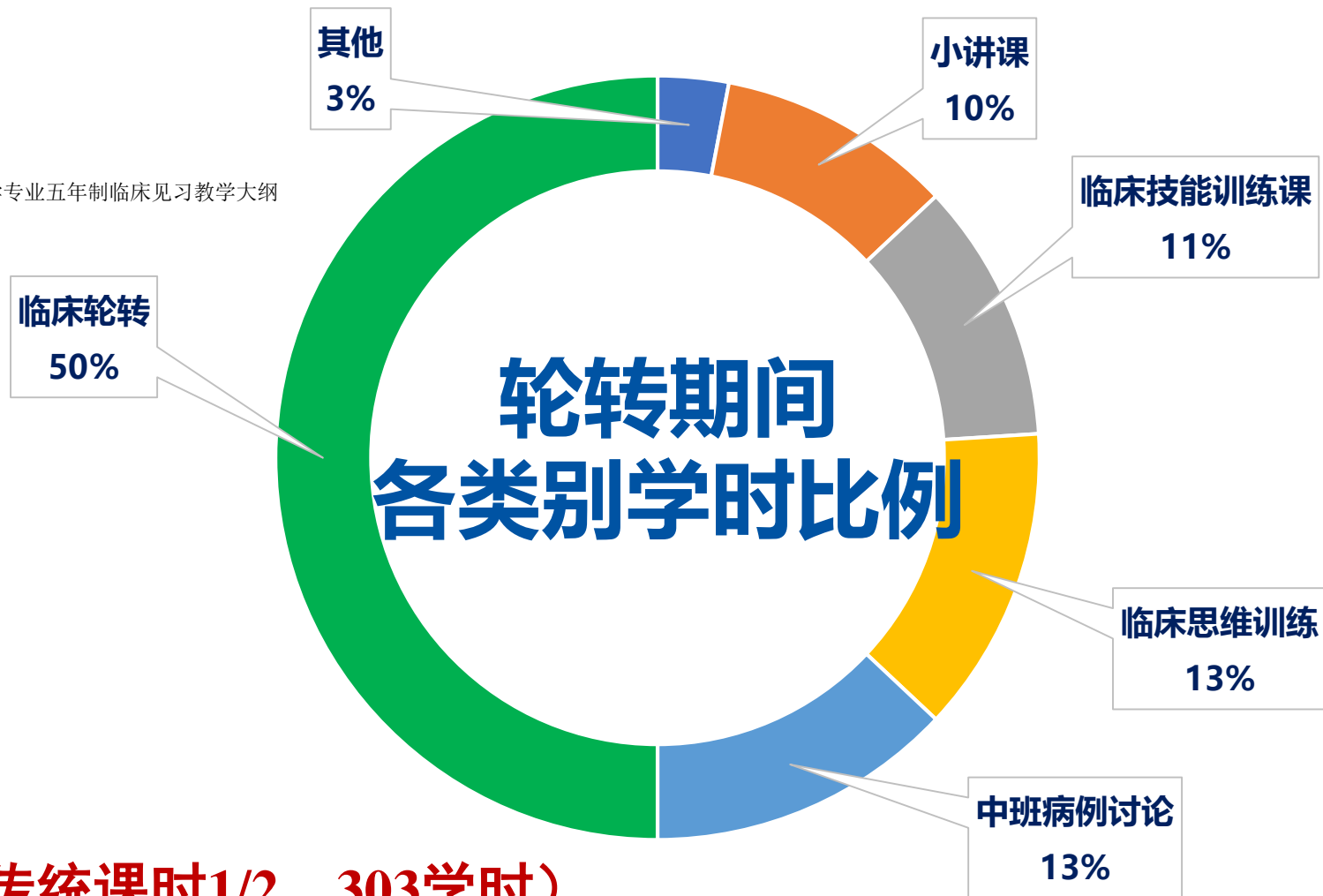
每个轮转学科配备一名“见习脱产指导教师”，每个学科每次最多6-8名学生

临床轮转期间一天典型安排

在见习轮转期间，统一设计了多种教学形式，除主要的病房轮转之外，还包括小讲课、思维训练、技能训练、中班病例讨论等



*上述数据来自2013级临床医学专业五年制临床见习教学大纲



- ✓ 压缩大班理论课程的学时数（**为传统课时1/2，303学时**）
- ✓ 改传统的小班“套餐式”见习为小组“轮转式”见习
- ✓ 脱产指导老师总体负责，临床医疗小组具体指导
- ✓ 开设病例讨论课，学生报告病例和主持讨论，老师参与引导和共同讨论

临床评价改革





3. 学业成绩考核

3.1 考核方法

- 根据不同的考核目的，采用广泛**多样的考核方法和方式**
- 积极开展考核体系与方法的研究，探索新的、有效的考试方法并加以应用

3.2 考核和学习之间的关系

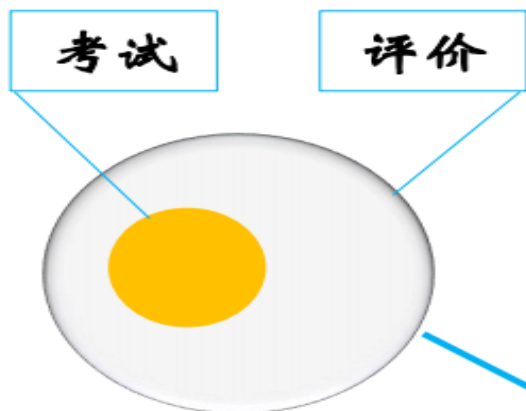
- 有利于促进学生的学习
- 做好终结性评价的同时，加强形成性评价的应用，并及时进行反馈，以便**指导学生更好地学习**

3.3 考试结果分析与反馈

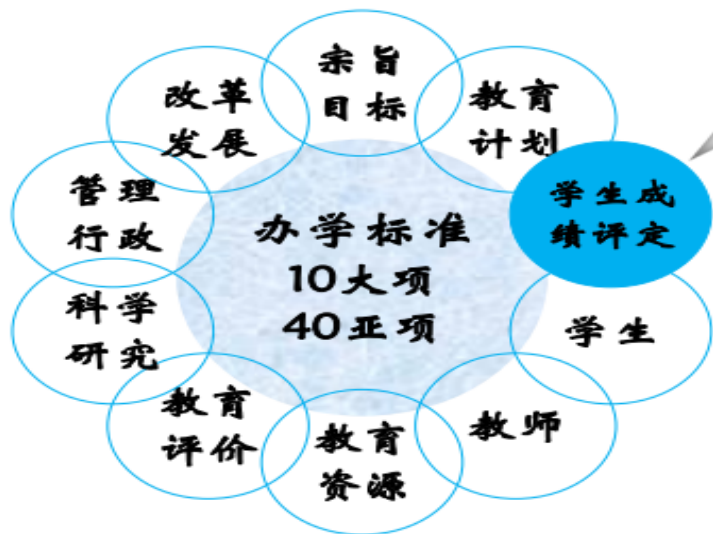
- 基于教育测量学的考试分析
- 将考试分析结果及存在的问题以适当方式反馈给学生、教师 and 教学管理人员
- 将考试分析结果用于**改进教与学**
- 加强考试的改革与研究



共性问题



- ◆ 对考试研究不足，试题结构不合理，记忆性题目权重大
- ◆ 试卷分析流域形式，反馈与改进不足
- ◆ 考试方法的研究和规范存在问题
- ◆ OSCE站点数量不足，失去客观、有序和结构化的初衷
- ◆ 评价与考核概念混淆，对形成性评价的理解偏差



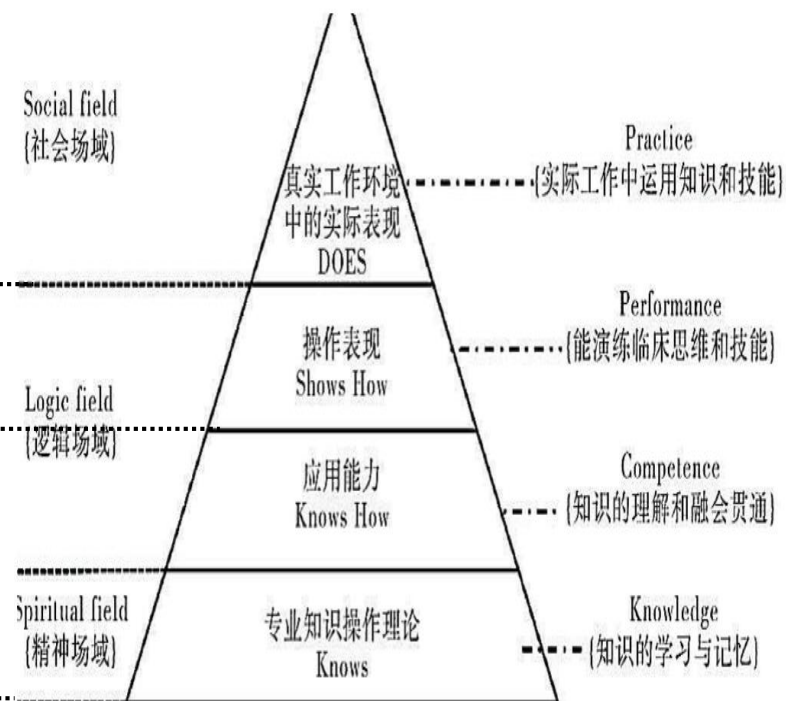
形成性评价 ≠ 平时考试

借鉴程序性评价理念

- 发挥单个评价的优势---聚焦于1个主要目的或维度
- 发挥体系设计的优势---解决单个评价方法不全面的问题

借鉴米勒模型设计框架

- 采用Mini-CEX和DOPS等，直接观察学生在实际临床中接触患者和临床操作的表现
- 通过OSCE等，以及在模型上的临床操作，来测试学生的思维过程和操作技能
- 通过笔试评价中MCQ等，对知识记忆和知识应用的评价



面向临床、面向能力评价和反馈 不同学习阶段 选取不同评价

米勒金字塔评价层次	主要评价工具	理论教学	见习轮转	实习轮转
实践（基于工作场所的评价）	Mini-CEX/ DOPS/直接观察法/MSF等		√	√
展示（OSCE和模拟测试）	OSCE/OSATS/标准化病人等	√	√	√
理解（知识应用测试）	A2型MCQ等	√	√	√
知道（事实记忆测试）	A1型MCQ/简答题/作业等	√		

评价工具选择原则（可靠性、有效性）：

- 评价目的和定位：形成性/终结性？判定及格或不及格？
- 评价反馈形式和机制
- **评价对学生的影响**--不可避免的会影响学习策略的选择
- **学生对评价的可接受度**
- **评价实施的可行性**---成本效益、教师压力

MCQ Multiple-choice questions 多选题考试	Mini-CEX Mini-Clinical Evaluation Exercise 迷你临床演练评估	OSCE Objective Structured Clinical Examination 客观结构化临床考试
MSF multisource feedback 多源反馈	Portfolio 文档与日志记录	DOPS Directly Observed Procedural Skills 操作技能直接评估
CBD case-based discussion 基于案例的讨论	Projects and theses 项目或论文	Long Case 长病例考试

1、A2型 MCQ

Eur J Vasc Endovasc Surg (2017) 54, 654–658

TOPICAL EDUCATIONAL REVIEW

How to Write a High Quality Multiple Choice Question (MCQ): A Guide for Clinicians

P.A. Coughlin^a, C.R. Featherstone^b

^aDepartment of Vascular Surgery, Addenbrookes Hospital, Cambridge, UK
^bThe Royal College of Surgeons of Edinburgh, Edinburgh, UK

Despite the variety of assessment tools available, multiple choice questions (MCQs) still play an integral part in examinations at both a national and speciality board level. MCQs have a number of methodological advantages yet their strength is related to the quality of the question posed. Specifically, there has been a move towards the MCQ testing a taxonomically higher order concept of integration-interpretation and problem solving. This paper focuses on question development and the potential pitfalls to avoid.

© 2017 European Society for Vascular Surgery. Published by Elsevier Ltd. All rights reserved.
Article history: Received 15 May 2017, Accepted 17 July 2017, Available online 1 September 2017

Keywords: Multiple choice question, Question writing

- **MCQs被视为测试事实知识的有用工具**，具有较高阶的认知过程，如**解释、应用知识和解决问题**
- 尽管有各种各样的评估工具，选择题(MCQs)在国家考试和专业考试中仍然扮演着重要的角色。**MCQs在方法论上有许多优势，但其优势与所提出问题的质量有关。**

重新认知：良好的可靠性、有效性、可行性

- 试题能评什么、测什么，关键在于内容，而不是格式
 - 开放式问题不一定能测试高级认识技能，而选择题不一定只测试知识记忆
- 暗示效应（仅仅通过识别各选项就可能找到正确答案）
 - 然而，MCQ在相反方向也有明显暗示，即可能导致考生选择错误答案。（Schuwirth et al.,1996）
 - 通过提升试题质量来避免
- 知识点覆盖广，对同一知识点在不同维度的评价
- 同一个考点可以有许许多多病例，不用担心试题枯竭

□ 有效结合信息技术的可能

- 命题易于管理、分析、反馈
- 回答和评价可以依据计算机自动完成，效率高，消除教师评分的可变性.....
- 大规模、远程进行的可能性
- 多维度数据采集、全面测量分析：成绩、考试期间的行为等
- 多种利用：课堂教学结合（TBL）、刮刮乐.....

题目的质量是关键

□ 试题质量保证

• 核心：考试大纲

--基于胜任力需要和课程内容重新修订

• 支撑：临床骨干教师队伍

--涵盖38个三级学科，共200名附属医院临床教师

--加强培训，宽松考核，以共同理念、目标、愿景而汇聚在一起

□ 学以致用--要求学生出题

--教师点评、反馈；学生互评

--用于学生自测、TBL的iRAT、tRAT环节

□ 基于计算机的考核、分析反馈系统

--加密远程考试

--加强结果反馈

该题干考核阵发性室上性心动过速的治疗原则，难度较易。PSVT很少会出现黑矇，持续时间太长或引起血流动力学障碍时才有脑供血不足表现。删去“黑矇”。
谢小洁 17-09-23 04:04

编号:	Re心
等级:	<请选
主题词:	阵发
说明:	金华 换压

试题

患者，女性60岁，因
性心动过速。查体：

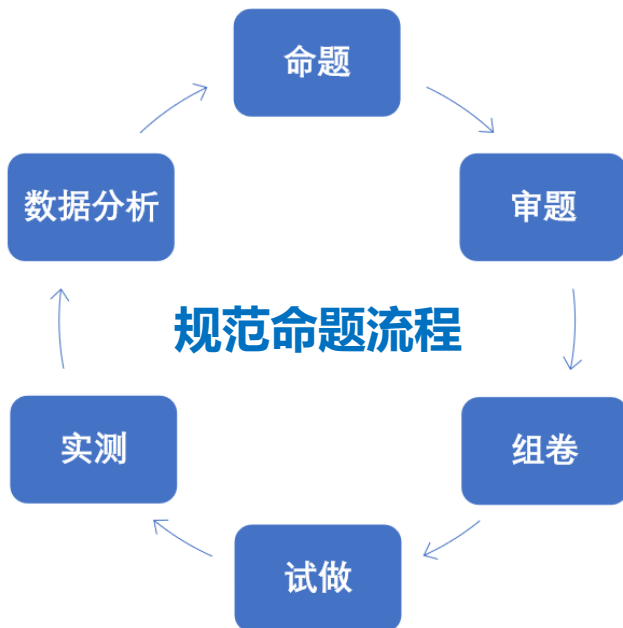
A. 按摩颈静脉窦

B. 注射维拉帕米

C. 口服普萘洛尔

D. 经食道调搏

E. 同步电复律



13门临床课程的试题平均区分度



一般：≤0.2不良；0.2-0.3尚可；0.3~良好

题库中正式试题（不含学生出题）

已有20门课程3万余试题，其中临床课程13门，基础课程7门

命题管理 - 试题 - 试卷 - 基本资源 - 系统管理 -

METESP™



试题

新增

修改

删除

添加到

Word导入

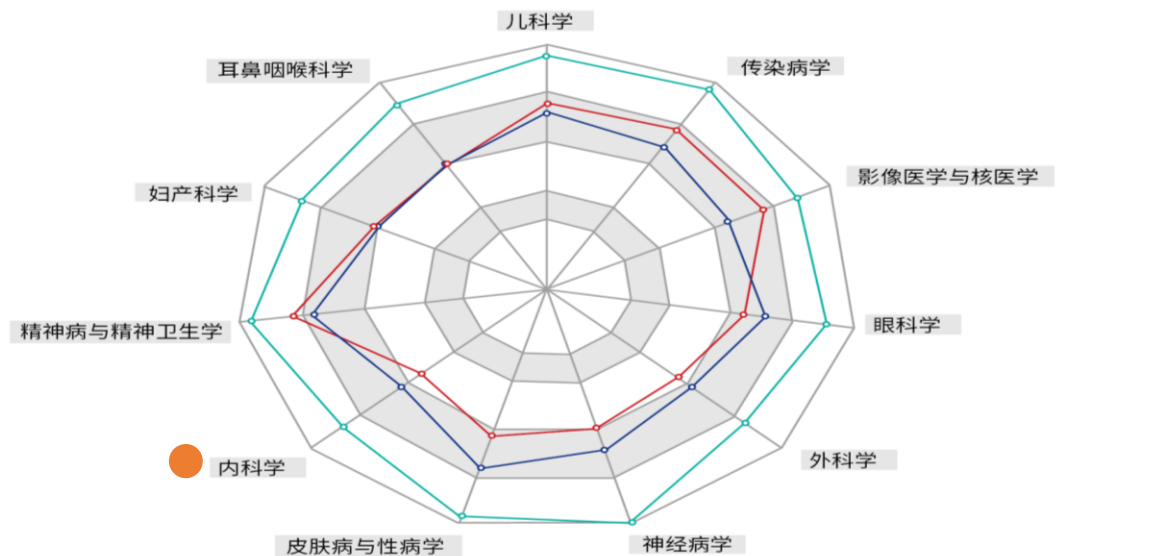
细目表

第 331-360 项, 共 34571 项



<input type="checkbox"/>	题型	编号	说明	等级	计数	最近处理人	命题人	更新时间
<input type="checkbox"/>	A2	2017级神经病学-浙一-余丽华-036	患者反复发呆，发作持续时间短暂...	-未定级-	1	梁辉	余丽华	2020-10-14 18:39
<input type="checkbox"/>	A2	2017级神经病学-浙一-余丽华-050	时间窗内的急性脑梗死，首选静脉rt...	-未定级-	1	梁辉	余丽华	2020-10-14 18:38
<input type="checkbox"/>	A2	2017级神经病学-浙一-余丽华-001	症状、既往史与MRI检查都支持脑...	-未定级-	1	梁辉	余丽华	2020-10-14 18:38
<input type="checkbox"/>	A2	2017级神经病学-浙一-余丽华-011	患者深感觉障碍，感觉性共济失调...	-未定级-	1	梁辉	余丽华	2020-10-14 18:37
<input type="checkbox"/>	A2	内分泌-备用4-013	2014级实习命题（第11-12周）-秦...	2级	1	张哲	秦锡怡	2020-10-14 18:36
<input type="checkbox"/>	A2	2017级神经病学-浙一-余丽华-037		-未定级-	1	梁辉	余丽华	2020-10-14 18:33
<input type="checkbox"/>	A2	2017级神经病学-浙一-余丽华-014	缓慢进展，踝阵挛阳性，Hoffman...	-未定级-	1	梁辉	余丽华	2020-10-14 18:30
<input type="checkbox"/>	A2	内分泌-备用4-017	2014级实习命题（第5-6周）-黄天...	1级	1	张哲	黄天皓	2020-10-14 18:14
<input type="checkbox"/>	A2	内分泌-备用4-015	2014级实习命题（第17-18周）-钱...	1级	1	张哲	钱聪	2020-10-14 16:59
<input type="checkbox"/>	A2	2017级神经病学-浙一-余丽华-024		4级	1	梁辉	余丽华	2020-10-14 16:20

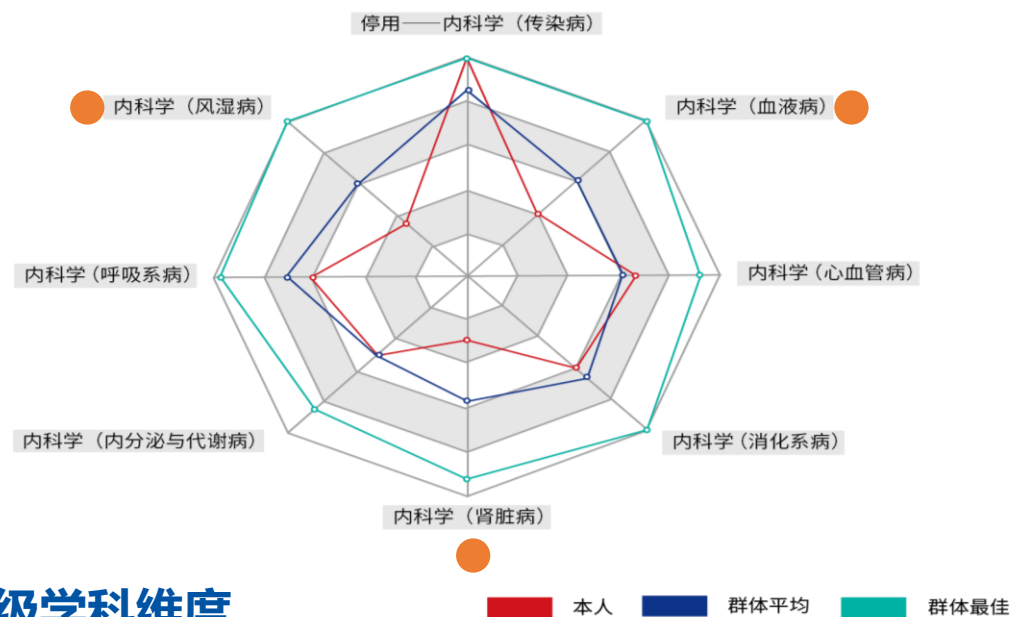
例：反馈给考生（手机端--微信小程序推送）



二级学科维度

■ 本人 ■ 群体平均 ■ 群体最佳

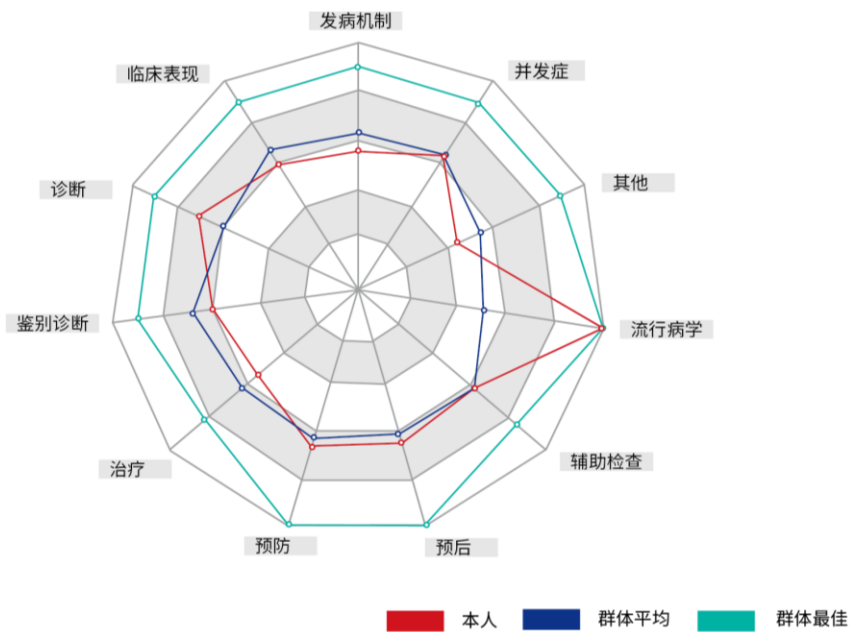
维度	本人掌握率	百分等级	群体平均	群体最佳
儿科学	74.19	58	71.86	94.62
耳鼻咽喉科学	60.00	51	60.28	90.00
妇产科学	61.00	55	59.87	87.00
精神病与精神卫生学	83.33	82	76.88	96.67
内科学	53.77	28	61.46	87.74
皮肤病与性病	63.33	15	76.43	96.67
神经病学	58.33	23	68.62	98.33
外科学	55.25	24	62.50	83.76
眼科学	65.00	29	72.30	91.67
影像医学与核医学	77.00	92	63.82	89.00
传染病学	76.79	79	69.15	96.43



三级学科维度

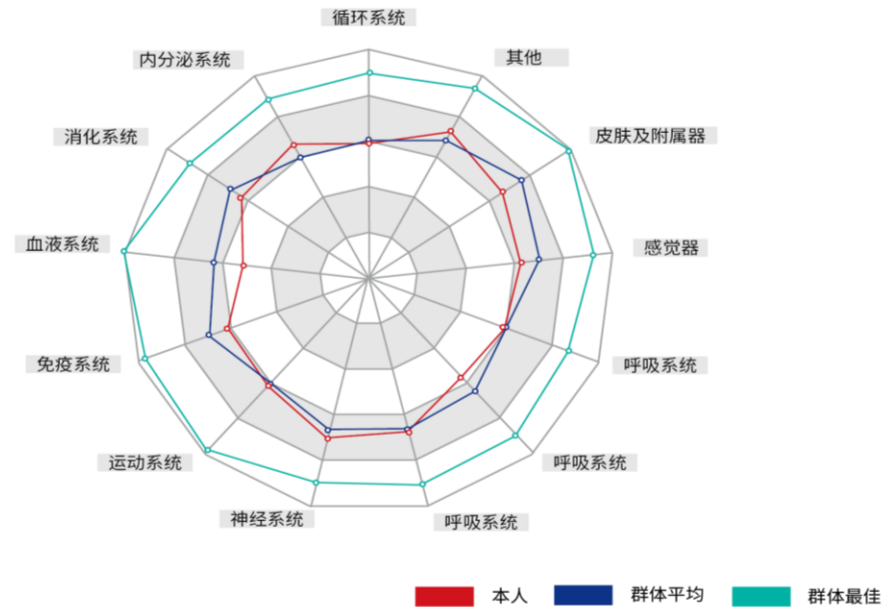
■ 本人 ■ 群体平均 ■ 群体最佳

维度	本人掌握率	百分等级	群体平均	群体最佳
停用——内科学（传染病）	100.00	100	83.96	100.00
内科学（风湿病）	33.33	11	61.19	100.00
内科学（呼吸系病）	59.46	23	69.18	97.62
内科学（内分泌与代谢病）	50.00	52	50.11	86.36
内科学（肾脏病）	30.00	7	56.50	90.91
内科学（消化系病）	59.46	25	66.99	100.00
内科学（心血管病）	65.12	72	58.58	90.70
内科学（血液病）	38.46	16	58.42	100.00



维度	本人掌握率	百分等级	群体平均	群体最佳
发病机制	56.00	25	62.96	88.31
临床表现	60.00	30	66.18	88.89
诊断	71.13	52	69.92	90.82
鉴别诊断	59.83	23	69.10	90.48
治疗	55.16	27	61.12	81.73
预防	66.67	64	64.20	100.00
预后	66.67	69	61.43	100.00
辅助检查	63.33	54	62.64	88.04
流行病学	100.00	100	50.95	100.00
其他	45.45	31	55.29	91.67
并发症	63.33	52	63.06	93.10

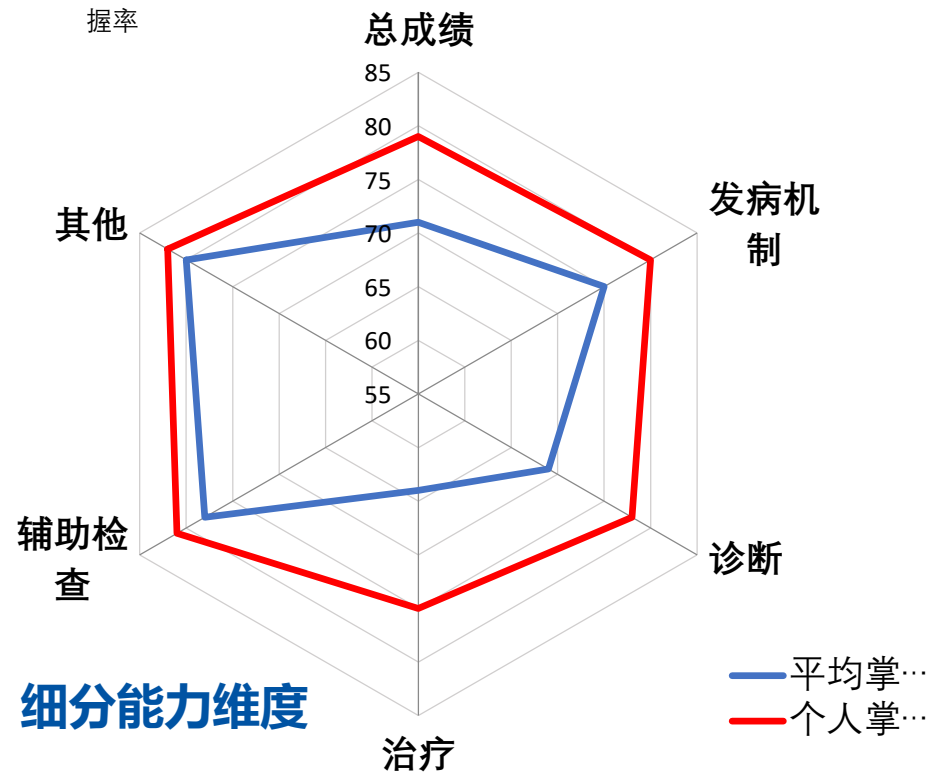
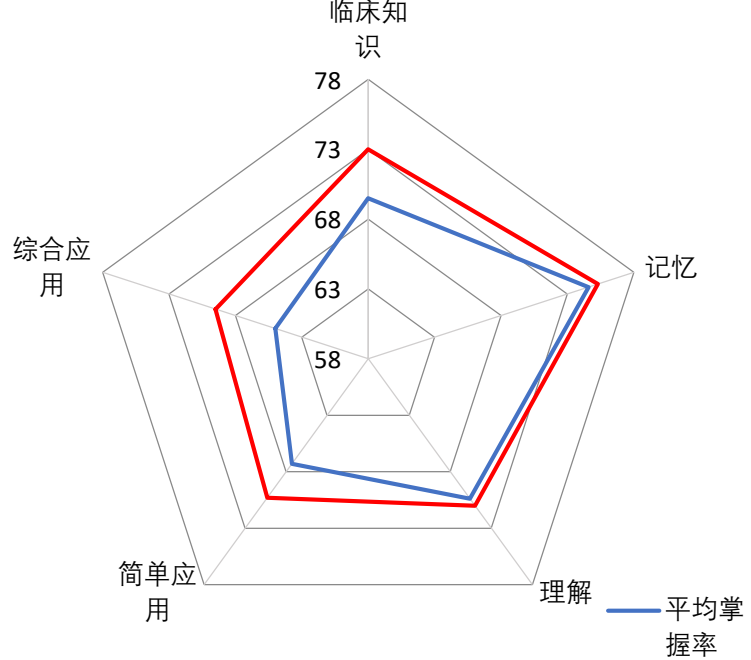
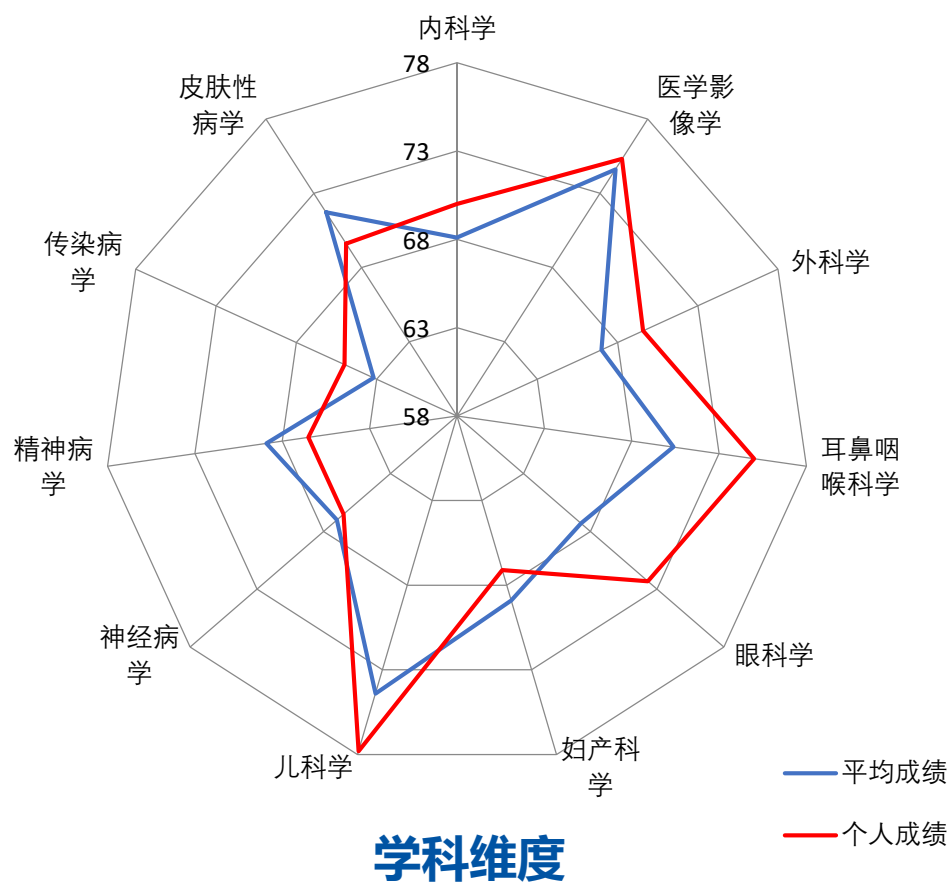
细分能力维度



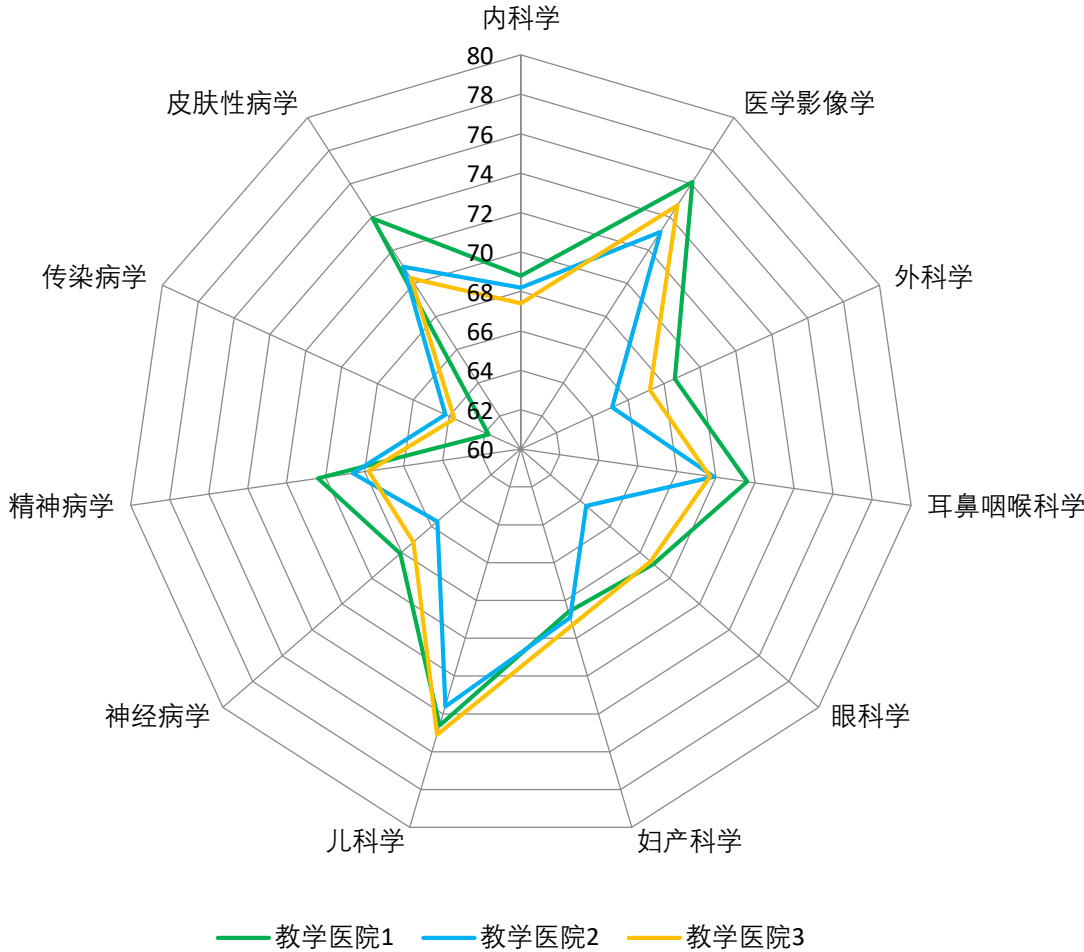
维度	本人掌握率	百分等级	群体平均	群体最佳
循环系统	59.22	51	60.36	89.25
内分泌系统	65.38	64	59.83	87.76
消化系统	63.01	32	67.16	88.71
血液系统	51.52	25	63.07	100.00
免疫系统	60.98	25	69.20	97.56
运动系统	61.63	52	60.42	100.00
神经系统	71.04	56	69.17	90.16
泌尿系统	57.14	27	65.08	91.43
生殖系统	59.13	48	59.66	87.07
感觉器	62.22	24	70.09	92.21
皮肤及附属器	65.52	19	76.14	98.28
其他	71.43	63	68.18	93.06

器官系统维度

例：反馈给考生 (PC端)

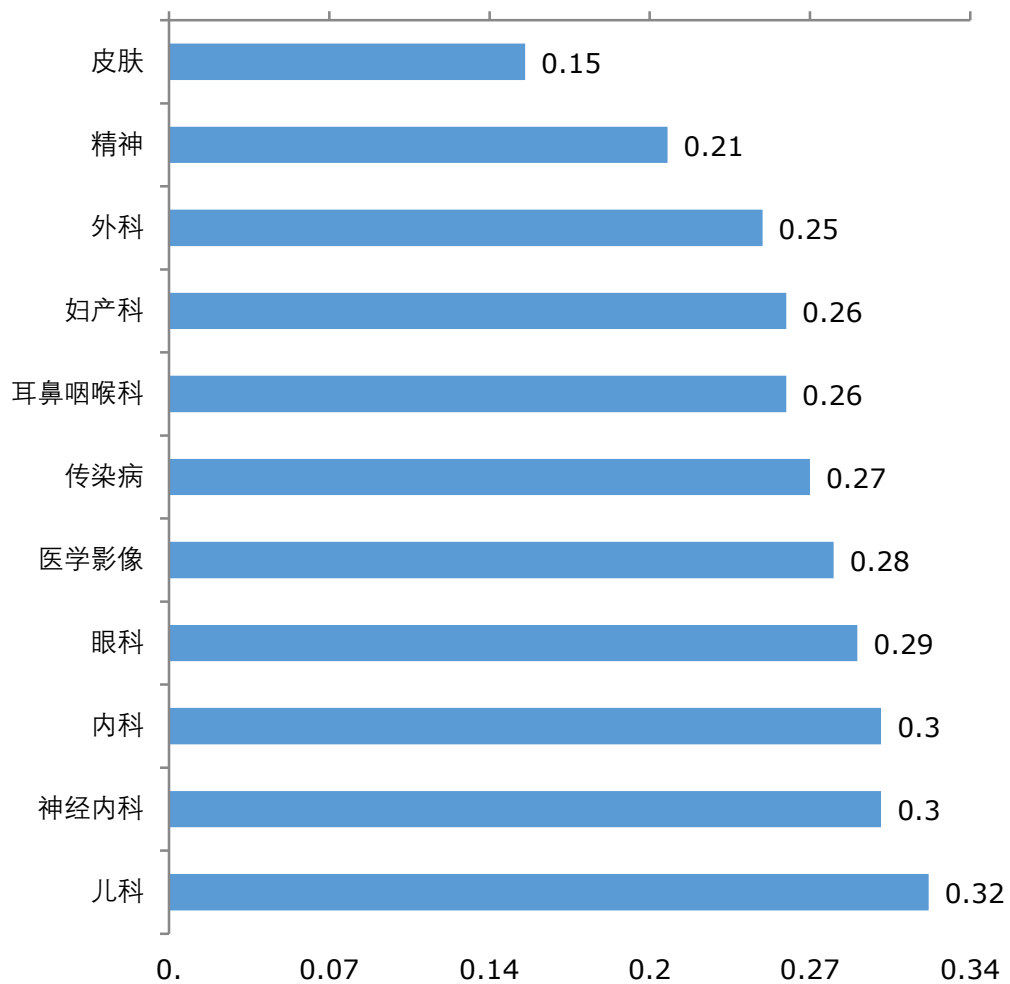


例：反馈给教师、管理人员或科室

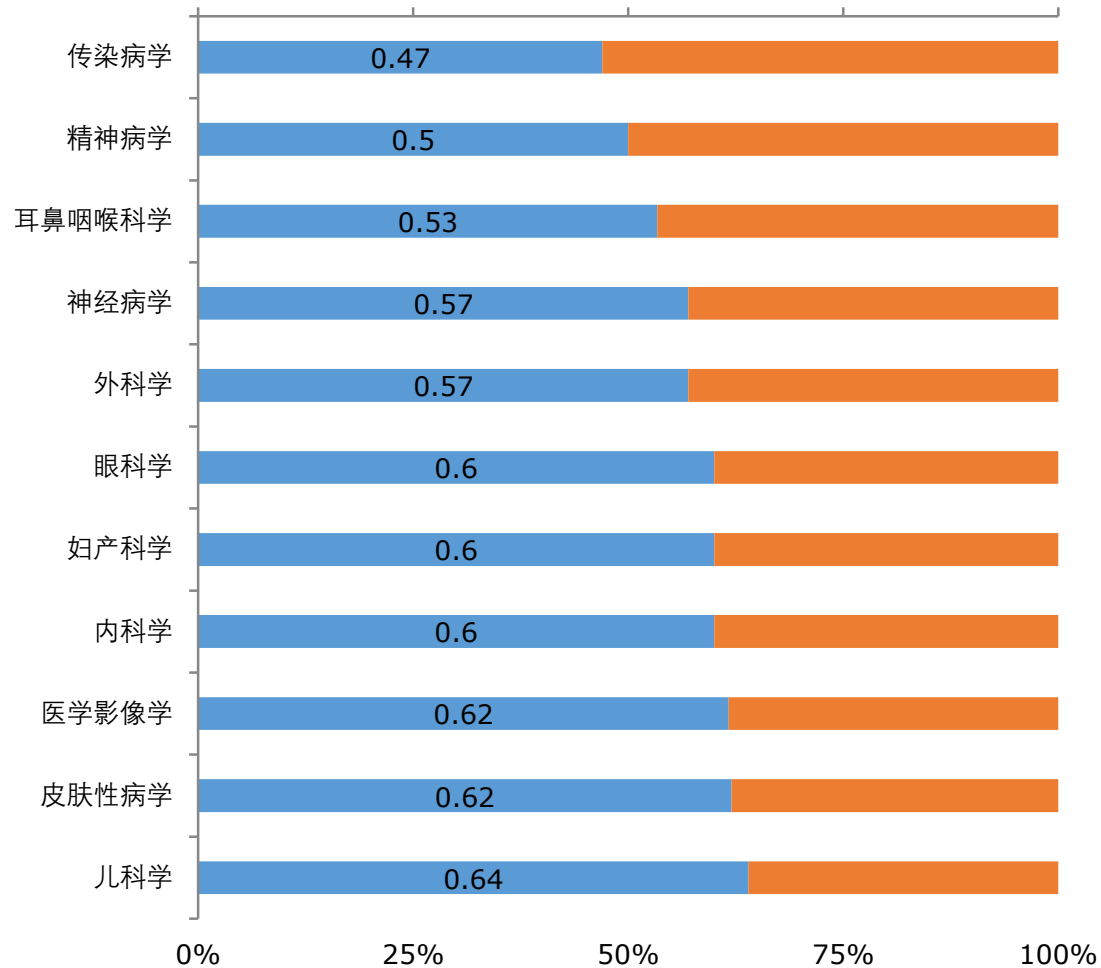


结合学生对带教科室的评价，往往可以发现，考试成绩较低的科室，其带教满意度也相对较低。

例：反馈给教师、管理人员的试卷、试题整体质量情况



试卷平均区分度



■ 区分度优良试题 (D ≥ 0.2)
■ 区分度不良试题 (D < 0.2)

例：反馈给教师的试题实测参数（考试后）

试题

1. 男性，27岁，右上腹钝器伤后1小时急诊来院，心率95次/分，血压103/70mmHg，体格检查提示右上腹压痛，无反跳痛，肠鸣音4-5次/分，急诊血常规提示白细胞 $10.5 \times 10^9/L$ ，中性粒细胞百分比76%，血红蛋白104g/L，腹部立位平片提示腹膜后花斑样改变。该患者最有可能的腹部脏器损伤为

- A. 肝脏
- B. 结肠
- C. 十二指肠
- D. 空肠
- E. 回肠

35.9649%

7.0175%

53.5087%

1.7543%

1.7543%

各选项学生的
选择率

临床医学（五年制/八年制/留学生）/见习阶段（4年级）/临床见习轮转考试

难易度： 0.5351

区分度： 0.6619

题库系统的拓展应用——理论大课的TBL教学

(学生端)

(教师端)

17级A1/A2型试题...

患者，男，66岁，因左手腕摔伤1小时入院，疼痛剧烈，活动困难，行左手平片检查（如图）。该患者诊断为

第3题 单选/1.00分



点击查看大图（共 1 张）

左手尺骨远端骨折

左手桡骨远端骨折

12 /// 浙四-28047 患者为老年男性，因左手腕摔伤入院，平片可见左侧桡骨远端透亮骨折线，骨折远端向外后方移位，周围软组织肿胀，故诊断左手桡骨远端骨折。

左手腕骨骨折

左手指骨骨折

左侧肱骨骨折

剩余时间：00:39:34

题号 下一题

2017级肝胆胰练习

成绩单

名次	姓名	学号	得分
第1名	徐欣	3170105566	10分
第2名	杨博文	3170104486	8分
第2名	林雨欣	3170104225	8分
第4名	胡佳慧	3170105573	7分
第5名	白晨昊	3170101602	6分
第5名	徐绿弯	3170103814	6分
第5名	王欣润	3170103954	6分
第5名	郑艺越	3170103964	6分
第5名	俞一凡	3170103941	6分
第5名	裘颖寅	3170103930	6分
第5名	林伟博	3170103982	6分

确认

2017级肝胆胰练习

实测情况

题号	掌握率	平均耗时
第1题	56.67%	46秒
第2题	61.67%	28秒
第3题	40%	52秒
第4题	31.67%	43秒
第5题	10%	30秒
第6题	31.67%	55秒
第7题	33.33%	31秒
第8题	11.67%	37秒
第9题	43.33%	34秒
第10题	45%	32秒

确认

2017级肝胆胰练习

患者，女，32岁，1月余前无明显诱因下出现右上腹痛，为持续性刺痛，偶感胸闷，无气急，无恶心呕吐，无畏寒发热等不适，查肝胆胰脾彩超提示：一囊袋样结构伴其内强光点。最有可能的诊断是：

第2题 单选/1.00分

掌握率：61.67% 平均耗时：28s

6.67% 肝囊肿

61.67% 胆囊结石

3.33% 胆囊癌

点评时间：2020/08/08 18:08:16

题号 作答路径 返回

2017级肝胆胰练习

3.33% 胆囊癌

0% 肾结石

0% 肝癌

选项 1

刘孝龙

施柯熠

张瑞洁

金梦蕾

点评时间：2020/08/08 18:08:16

题号 作答路径 返回

2017级肝胆胰练习

3.33% 胆囊癌

0% 肾结石

0% 肝癌

试卷解析 - 第 2 题

秦家盈: B

陈梦佳: B

施柯熠: A B

赵思凡: B

徐皖北:

选择路径

客观结构化临床评价

Objective Structure Clinical Examination

指学生在一系列限定时间的“站点”中进行轮转的训练或评价项目，通过每一个站点内和SP的接触，要求所有学员完成一致的特定性任务。学员的表现通过预先设计好的评价标准进行标准化评价

原则

direct observation
objectivity
structured format

2004年

美国将OSCE纳入STEP2

2001年

ASPE成立（标准化病人教育者协会）

1991年

浙江大学、华西医学院、九江医学院引入SP，并逐步实施OSCE

1990年

加拿大第一个将OSCE纳入执医考核

1975年

OSCE的起源

Harden建立了一种模拟多个临床场景的多站点考核方式

- MCQ因其可靠性高、知识评价更全面而被引入，但MCQ并不能评价临床技能
- OSCEs是基础、高级临床技能（形成性评价、终结性评价）评估的重要教学工具，从形式和内容等方面模拟医生-病人实际接触情况进行评价

SP 1991年开始

OSCE成功的关键所在

使用SP的过程本身就是对生物-心理-社会医学模式加深理解的过程，是对医学生交流能力、健康教育、职业精神等胜任力深入理解的过程，有助于医学人文教育

SP75人，其中全英文23人

★ SP案例达**170**多个

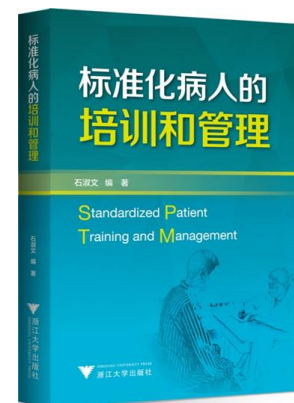
★ **2019年**接待学生或考生**9000**多人次，教学、考试、教师培训等

教师	退休人员28人 其他在职或自由职业37人	
10	65	
大专及以下	本科27人 研究生13	
35	40	
<50岁: 42人	50 ~ 60岁: 22人	>60岁
64		11
男性	女性	
22	53	

SP师资队伍近30人，涵盖各临床专科



**尊重
规范
发展**



主创、演唱--SP志愿者之歌

2019第二届国际SP医学教育论坛



作词: 科促会SP社区志愿者团队
浙江大学医学院SP团队
作曲: 刘熬宁 创作指导: 刘熬宁
合唱: 浙江大学医学院SP团队

简要说明各站点与所评价胜任力之间的关系

【注释】

- 临床医学课程包括诊断学、内科学（含神经病学、传染病学等）、外科学（含外科学总论、麻醉学等）、妇产科学、儿科学、精神病学、眼科学、耳鼻咽喉与头颈外科学、皮肤性病学、口腔科学、中医学或其他民族医学、全科医学等核心课程；以及急诊医学、康复医学、老年医学、肿瘤学、舒缓医学、物理治疗、放射治疗学、临床药学（含抗生素合理使用）等拓展课程。临床医学课程也可以整合的形式呈现。核心课程与拓展课程的含义见2.4 生物学课程注释。
- 临床技能包括病史采集、体格检查、沟通技能、辅助检查、诊断与鉴别诊断、制定和执行诊疗计划、临床基本操作等。
- 职业能力包括患者处置能力、团队协作与交流能力、领导力、跨学科/专业合作能力等。

每个站点都可以对医学生的多个胜任力进行评价★

确保能覆盖所需评价的各胜任力★
确保每个胜任力能被多次考核★

重要问题：

- 是否反映典型的临床场景/实践
- 是否反映课程/轮转内容
- 是否覆盖所有需评价的胜任力
- 案例是否多样性？性别、年龄、民族、文化背景等

站点/ 案例	内容	胜任力/技能								适用对象	备注
		病史采集	体格检查	沟通技能	辅助检查	诊断与鉴别诊断	制定诊疗计划	临床基本操作	人文关怀		
内科SP	胸痛	√	√	√	√	√	√		√	实习结束	15'+5',男 35
妇科SP	腹痛	√	√	√	√	√	√		√	见习结束	15'+5',男 65

引入**CBD** (基于案例的讨论, 由英国Healthcare Assessment and Training (HCAT) 所设计,15'+5') :

- 引入了由来自不同学科的教师组成考官组的形式, 来实施多角度的评价
- 每次评估由学生自己挑选病例个案, 事先送交考官
- 病例讨论内容涉及诊断、治疗、跨学科知识、医学伦理等内容, 然后评分反馈
- **评价表** (平衡: 条目越多---可靠性、有效性上升; 评价准确性下降)
 - 根据考试大纲制订**结构化评分表**
 - **建立原则**
 - 具体行为性条目评价**
 - 整体等级评价** (综合评价更能反映他们的能力)
 - 采用分级评价** (明显缺陷、低于预期、符合预期、高于预期、十分出色)
- **反馈**
 - **基于胜任力的、多站点的纵向评价反馈** (多个站点可以对同一胜任力进行评价, 利用信息反馈平台)
 - **实施OSCE考后的即时点评反馈**

在反馈之后, 尤其是对于不合格的学生, 实施个性化指导、制订补习计划

Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference 2020年渥太华会议关于评价共识

Recommendations for OSCE

1. Design OSCEs as part of a system of assessment (评价系统--将OSCE作为评价系统(程序性评价)的一部分)

--The inclusion of oral discussions, MCQs, or Short Answer Questions (SAQs) in an OSCE is inappropriate and poses a serious threat to the validity of an OSCE, these are more appropriately tested using paper or computer based formats

2. Adhere to validity framework criteria (评价效度框架--Kane)

3. Define purpose of the OSCE and make purpose explicit to stakeholders (评价目标明确并告知相关人员)

4. Blueprint to learning objectives/outcomes and assess clinical interactions (考核大纲)

5. Plan adequate sampling with sufficient stations and testing time (足够的站点和测试时间)

--OSCEs consisting of less than 12 stations and/or less than 150 min of testing time will not yield reliable results

6. Design marking schemes to align with clinical task and clinical thinking (根据临床任务和临床思维设计评分方案)

7. Use criterion referenced standard setting (Borderline Regression method) (科学进行标准设定)

8. Generate metrics for OSCEs to examine reliability and station level issues (评价OSCE考试的质量)

9. Value examiner diversity and focus examiner training on conduct, behaviours and bias (Embrace examiner variability by ensuring sufficient numbers of examiners, rather than trying to standardize their judgements. 考官)

10. Handle test security through task design and circuit design to group stations

11. Triangulate data from OSCE performance with other assessments or outcomes, to inform decision making



3、Mini-CEX

2010年

引进中国修订为中国简体版的min-CEX

2005年

陈伟德教授在获得ABIM授权同意翻译成中文版引进台湾。

2001年

ABIM重新修订评量项目后成为目前大家所熟悉的七项。

1995年

美国内科医学会（ABIM）建立的一套兼具教学与评量的工具，能针对学员在一次面对一位真实病人中，使用较短的时间来评估住院医师某些特地被要求的工作能力。

迷你临床演练评估

mini clinical evaluation exercise

由一位指导老师、一位学员和一位共同的诊治的患者，选择合适的时间和地点，在15-20分钟内进行重点诊疗行为，而指导教师在直接观察学员与患者的互动后，给予评分，并做5-10分钟的反馈。

观察七大项目

医疗面谈、体格检查、人文关怀、临床判断、临床咨询、组织效能、整体表现

优点和缺点

优点：直接观察真实；实用性高，不受空间限制；可行性高、短时间完成；弹性大，适用对象广；立即反馈；有效且可信的评量标准

缺点：评量部分临床能力；评分者主观误差；病人接受度；临床情境变化大

临床操作技能评价

Direct Observation of Procedural Skill

英国皇家内科学院应用在Foundation Programme中的评估工具，其方法是用一套清楚定义该操作技能的每个步骤的标准清单，或是总体评估计分表，或是两者一起用，而对学员的临床操作技能进行直接观察评估。

优点和缺点

优点：评估在病人身上进行临床操作技能实际操作细节；据大部分的临床操作性技能皆可纳入进行；能结合操作与回顾，教育意义高；具有良好的信效度；经过妥善安排，可使评量时间短，临床运用可行性良好；临床教师与学员较能接受DOPS的评量方式。

缺点：学员可能因紧张而有失常表现；个案难易度差异过大可能会降低评量的信度；评分者间的差异；仅能针对受测者的操作部分加以评量，无法评估受测者的整体表现。

临床见习期间Mini-CEX/DOPS安排

见习轮转期间各科室问诊、体格检查、操作项目要求表

分类	科室项目	心内	呼吸	消化	内分泌	肾内	血液	普外	泌尿外	骨科	神外	心胸外	妇产科	儿科	神内	精神	传染	眼科	耳鼻咽喉	皮肤	影像		
必须完成的能力评价	Mini-CEX	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		
	DOPS	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
	医学生轮转表现评价	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
指导学生问诊、体格检查和病历书写	问诊	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		
	书写病历	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
	心脏检查、血管检查	●										●											
	肺部检查		●																				
	腹部检查			●					●														
	甲状腺检查				●																		
	肝脾触诊							●															
	浅表淋巴结、胸骨压痛检查							●															
	泌尿系统检查						●			●													
	脊柱和四肢运动系统检查										●												
	神经系统体格检查											●				●							
	自定义体格检查项目													●	●		●	●	●	●	●		
指导学生临床操作	骨穿						●																
	胸穿		●																				
	腰穿														●								
	腹穿			●																			
	插胃管			●																			
	心电图	●																					
	拆线换药							●		●													
	导尿								●														
	切开缝合							●															
	消毒铺巾穿脱手术衣							●															
	自定义操作项目											●	●	●	●		●	●	●	●	●	●	●

每位学生见习期间,要完成19次Mini-CEX和19次DOPS。

20个科室出科时,需要进行临床轮转表现的评价和反馈。

以上评价均用手机小程序进行。学生的反馈也通过手机微信自动推送

教师手机评价界面



学生收到评价结果界面

2019.11-2020.10期间，共完成
Mini-CEX 5500人次
DOPS 5100人次

Mini-CEX

姓名: 临床医学二系-严 (316010) 评价人: 附属儿童医院-叶贝 时间: 2020/6/8 16:05:00

PS-病例概况

■ 病例概况

- 患者情况
 - 地点: 病房
 - 患者简要概述 (50字以内): 咳嗽3天, 伴喘息1天
 - 疾病诊断 (200字以内): 急性支气管炎
 - 年龄: 2
 - 性别: 男
 - 病情复杂程度: 中
 - 评估重点: 病史采集

PS-Mini-CEX 评估表 (数值版)

■ 评估约在15分钟左右完成

- 1~3有待加强; 4~6达到预期; 7~9超过预期; **0指未评估
- 医疗面谈: 5
- 体格检查: 6
- 人文关怀: 6
- 临床判断: 6
- 临床咨询: 6
- 组织效能: 5
- 整体表现: 6
- 观察时长 (分钟): 15

PS-评语

■ 点评

- 教师评语
 - 优点: 问诊比较全面, 对主要症状、伴随症状能详细询问, 诊治经过、一般情况都有问及。既往史、个人史及家族史都有询问。
 - 不足: 对喘息相关的既往史、个人史及家族史还需要有重点地挖掘 (这方面对实习同学可能要求过高)
 - 建议和计划: 总体不错, 以后临床工作中多实践, 把握询问的主动性, 可以做到有序全面并从容!
- 反馈时长 (分钟): 5

DOPS

姓名: 临床医学二系-麻焯(3160103721) 评价人: 附属一院-俞国栋 时间: 2020/6/8 17:12:00

PS-PS-DOPS 首页

■ 概况

- 患者情况
 - 类型: 模型
 - 性别: 男
 - 年龄: 35
 - 病例号 (模型时填“无”): 无
 - 操作地点: 病房
- 执行项目
 - 名称: 胸腔穿刺
 - 技术复杂程度: ★☆☆
 - 操作经验
 - 作为助手 (第几次): 2
 - 独立操作 (第几次): 0

PS-DOPS 评估表 (数值版)

■ 请注意计时

- 1~2有待加强; 3~4达到预期; 5~6超过预期; 0指未评估
- 知晓操作适应症和禁忌症: 5
- 核对患者信息并取得同意: 5
- 准备工作: 5
- 操作步骤正确、规范: 4
- 操作手法准确、熟练: 4
- 操作后处理: 4
- 具有良好的无菌观念: 5
- 应对及沟通技巧: 5
- 医学人文关怀: 5
- 整体表现: 5
- 操作时长 (分钟): 15

PS-评语

■ 点评

- 教师评语
 - 优点: 耐心仔细
 - 不足: 不够熟悉
 - 建议和计划: 多练习, 多提问
- 反馈时长 (分钟): 5

命题管理 - 试题 - 试卷 - 基本资源 - 系统管理 -

METESP™

评价表

还原 反馈 导出

适用范围	名称
<input type="checkbox"/> /其他	课堂教学质量评价表 (学生用)
<input type="checkbox"/> /其他	中班病例讨论评价表
<input type="checkbox"/> /其他	DOPS
<input checked="" type="checkbox"/> /其他	Mini-CEX
<input type="checkbox"/> /其他	临床轮转表现评价表

评价结果

日期: [选择] 第 1-30 项, 共 18

日期	回收总数	已反馈数量
2020-10-20	5	5
2020-10-19	18	18
2020-10-18	11	11
2020-10-17	1	1
2020-10-16	154	154
2020-10-15	90	90
2020-10-14	49	49
2020-10-13	26	26
2020-10-12	23	23
2020-10-10	5	5
2020-10-09	118	118
2020-10-08	27	27
2020-09-30	14	14
2020-09-29	29	29
2020-09-28	22	22
2020-09-27	10	10
2020-09-26	7	7
2020-09-25	178	178
2020-09-24	53	53
2020-09-23	30	30
2020-09-22	39	39
2020-09-21	20	20
2020-09-20	2	2
2020-09-19	4	4
2020-09-18	110	110
2020-09-17	39	39
2020-09-16	32	32
2020-09-15	21	21
2020-09-14	17	17
2020-09-13	13	13

临床期间的评价经历

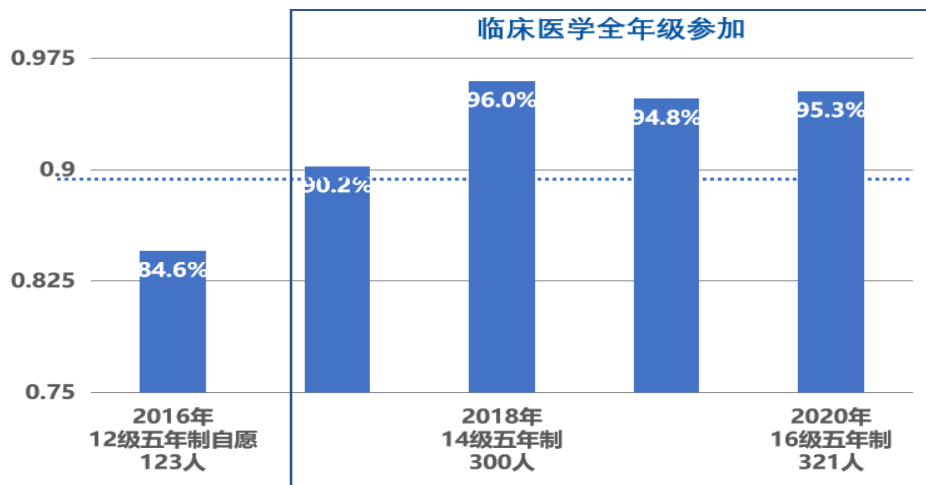


图1 浙大医学院执业医师分阶段第一阶段考试通过率

医综III
(MCQ+osce)
paper

命题专家

- 29个学科近200名临床骨干教师
- 18位教师成为国家医学考试中心命题专家

□ 反馈-反思-指导-改变（循环）

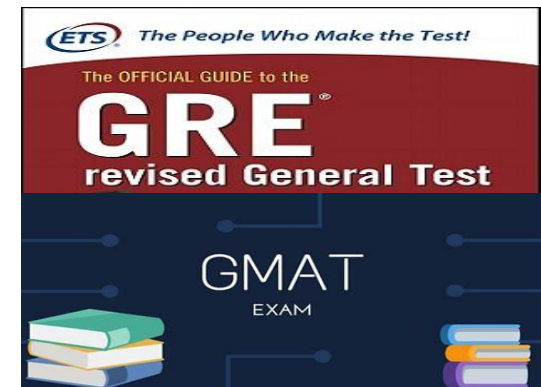
- 基于工作场所的评价（WBPA）越来越**聚焦于督导学习和及时、具体、建设性反馈，以最大限度地发挥这些工具对学习者的教育潜力。**应考虑评价策略、目标和定位，利用不同方法、使教育者可以在这些评价工具的过程中得到训练和指导
- **有效反馈：合适的时间、促进学生反思、给予学习指导、改变学习行为**
- **学生学习行为改变：调研正在进行中**

□ 分数等值

多轮考试之间的等值处理（样本量小）

□ 基于IRT（项目反应理论）的计算机适应性测验(CAT)

- 由计算机根据学生的回答，在后续自动选择最适宜的试题，以达到对学生能力作出恰当评价的方法
- 优点：
 - ✓ 题目针对性强，**较少题目--较精确地评估**
 - ✓ 因人而异地选题--**被试可自由选择测验时间**，不必因露题而统一测验时间
 - ✓ **基于信息技术的优点**

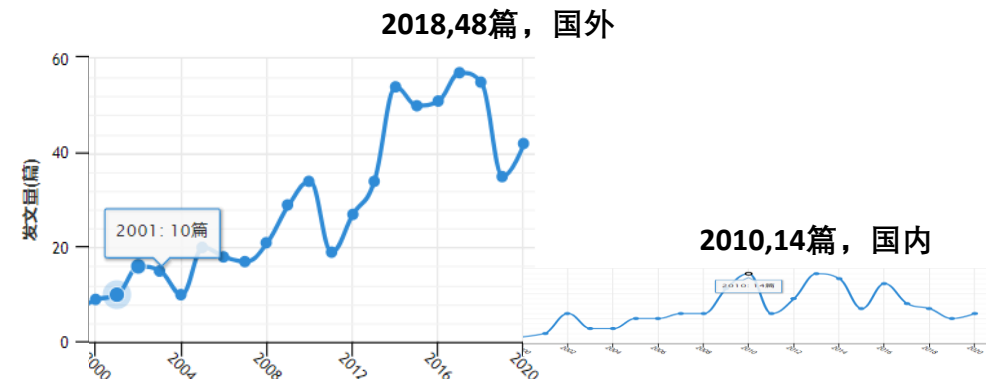


□ 电子档案袋评价：构建学生能力成长轨迹，发挥评价的价值

档案袋与结果导向的教育和胜任力导向的学习相匹配

档案袋成功的关键因素确保可行性以及让学习者立刻受益

电子档案袋减轻工作量，在档案袋中纳入反馈循环





感谢垂听!

THANKS FOR LISTENING