

孙宝志 金魁和 主编

教与学评价实用技术

PRACTICAL TECHNOLOGY OF EVALUATION OF TEACHING AND LEARNING

辽宁科学技术出版社

序

美国全国内科医学会，评价研究副主任约翰·诺西尼博士

自1949年以来，中国已经发展了适合人民卫生保健需要的完善的医学教育系统。评价工作是这个系统中的固有的一部分，也是整个教育系统中的基本环节。为了更好地教育学生，教师必须了解怎样才能有效地传输医学知识与技能，怎样才能激发学生的学习动机。与此同时，学生也必须经常得到学习上的反馈，了解自己在学习过程中的进步和弱点。这部由中国医科大学孙宝志和金魁和主编的著作，正是在这些主题上进行了有益的成功地探索。

此书的上篇集中阐述了对教师的教学质量的评价，这是提高医学教学质量非常重要的工作。目前，在中国之外，世界各国在对学生的评价方面做的研究多一些，而对教师的教学评价工作做得少一些，显然这是一个缺陷。这本书的许多贡献中的其一就是非常及时地反映了中国对教师教学评价的最新成果。本书开始介绍了教师的专业素质和如何评价一个好教师的教学质量，接着介绍了评价的理论、原则与方法，然后又阐述了在医学教学过程中不同阶段：课堂教学、实验课教学、临床实习教学中如何应用评价的理论与方法。特别是重点介绍了可以直接应用的评价技术。这些介绍都建立在他们长期的实践研究基础上，因而是非常有价值的。

这本书的下篇集中介绍了准确测量学生学习成绩的理论与方法，特别注意介绍这些方法的实际应用，如考试设计，笔试、口试，实际操作评定，直接观察，病案评定等等。还介绍了美国等国家在标准化病人、计算机模拟等方面的考试技术，这些考试的理论与方法已经囊括了目前在医学考试领域中对能力的测试方法。

总而言之，这本书提供给读者，非常有用的评价方法与技术，对提高教与学两方面质量无疑会有很大的帮助。我相信这部高质量的著作在中国出版发行，将使中国的医学教育评价活动推向一个新的台阶。

(签名) John J. Norcini, Ph. D.

美国，宾夕法尼亚州，费城

1992年7月31日

前 言

“教与学”是教学活动中的两个不同方面，“教”的主要任务是传授知识与能力，居于主导地位；“学”的主要任务是掌握知识与能力，居于被领导地位。但二者又是同一教学活动中缺一不可、相互依存的两个方面，教是学的条件，学是教的依据。正如中国古代“说文解部”中所说的“教，上所施下所效也”。从英文“教学”两个字看，learn和teach是同源派生出来的两个词，表达同样的意思。教育文献中常见的复合词句的Teaching—learning（“教”—“学”）也说明了这一点。教与学是相互交织、密不可分的。因此教学评价当然应该包括对“教”和“学”两个方面的评价。

故名思义，《教与学评价实用技术》一书的内容是介绍对教师的“教”和学生的“学”的评价理论与技术。书中介绍的评价理论与技术主要是针对高等医学院校的“教”与“学”。本文包括上篇和下篇，上篇介绍“教”的评价，下篇介绍“学”的评价。读者对象是各级各类医学院校的教师、医生、学生和管理人员。旨在提高高等医学院校教育质量，为社会主义现代化培养高质量的医学人才服务。

编写此书的背景和过程是：中国医科大学于1988年末以《高等医学院校教师教学质量评价指标体系的研究》为题，申请到卫生部科研基金，此后历时3年，由中国医科大学牵头，联合全国11所高等医学院校参加协作，它们是：（按笔画排列）上海第二医科大学、白求恩医科大学、石河子医学院、第一军医大学、沈阳医学院、首都医学院、南京医学院、重庆医科大学、湖北医学院、锦州医学院、福建医学院。从1989年至1991年，先后召开四次大型研讨会，经过评价指标体系设计、预实验、正式实验、总结分析等阶段，完成了课题研究，建立了一套评价指标体系和电子计算机软件。整个实验研究中被评教师597名，参评学生24,757名，参评教师2,355名，得到评价表27,501份。课题研究总结之后，又使用该评价指标体系在一些单位应用，取得较好的社会效益。在此基础上，协作组各院校主要研究人员在课题总负责人中国医科大学副校长金魁和主持下，于1991年9月，在石河子医学院召开会议，动议结合课题研究的实践编写《教与学评价实用技术》一书，由中国医科大学医学教育研究中心负责人孙宝志和金魁和副校长提交了全书编写总框架和编写提纲，同与会编写人员讨论，提出修改意见后，确定本书上篇的编写提纲。

1991年，中国医科大学申请获得了美国中华医学基金会资助的《学生认知能力评价》研究项目。该项目的主要内容是开展对高等医学院校学生“学”的评价的研究。在美国专家的协助下，先后召开了四次专题研讨会。在此基础上，由孙宝志、金魁和两同志拟写了对“学”的评价（下篇）的编写提纲，分寄各编写人员修改。可见，本书是在一次较长时间全国性关于教学评价方面的实验研究之后，又吸收了执行受国际组织资助的关于学生评价方面项目的实践经验基础上编写的。编写人员就是课题的研究人员或项目的执行者，他们工作在全国各高等医学院校，又有实际工作经验，因此，书中的内容具有较丰富的实践基础；本书注意参考了国内外较先进的教学评价理论与技术并尽量结合我国的实际情况，力求体现先进性；本书注意了评价技术的可操作性，对一些具体操作方

法作了比较详细的介绍。

本书各部分的编写人是：前言和第一章绪论，中国医科大学孙宝志、金魁和；第二章，第一军医大学李淑芝、郑钦竟；第三章，锦州医学院李红玉、关永琛；第四章，沈阳医学院杨孟坤、张庆再，中国医科大学乔敏；第五章，重庆医科大学颜家珍、杨力；第六章，福建医学院许惠源、康元元；第七章，沈阳医学院张庆再、杨孟坤；第八章，首都医学院宋延月、兰文恒；第九章，中国医科大学孙宝志、岳巍；第十章，南京医学院谢松龄、徐国勋；第十一章，湖北医学院王永楝、方苏彬，中国医科大学于晓松、张戈；第十二章，白求恩医科大学李春昌、金祥雷，中国医科大学于晓松、孙宝志；第十三章，上海第二医科大学郑玉衡、刘克明；第十四章、第十五章，石河子医学院于明信；第十六章，中国医科大学乔敏、于洪昭；第十七章，上海第二医科大学蒯强、刘克明；最后全书由孙宝志、金魁和统稿、修改和定稿。

本书在编写过程中，参阅了国内、外专家学者资料，特别是美国中华医学基金会主席William D.Sawyer博士和全美内科学会评价研究副主任John J.Norcini博士对于下篇的编写给予指导，在此一并表示感谢。

本书在出版过程中得到辽宁科技出版社王连汉同志的大力协助；中国医科大学张训巍、时瑾、娇娜等同志也作了很多具体工作，在此表示衷心感谢。

限于作者的经验和水平，本书在框架上的编排、文字的表达、观点和方法的叙述等方面，一定存在很多不妥之处，热切希望各位专家和读者批评指正，我们将不胜感激。

孙宝志 金魁和

于中国医科大学

1992年8月1日

目 录

上篇 对教的评价

第一章 绪论——医学院校教师教学质量评价的研究与实践	3
第一节 医学院校教师教学质量评价研究的国内外进展概况	3
第二节 医学院校教师教学评价指导思想的转变	5
第三节 医学院校教师教学质量评价的影响因素及相互关系	6
第二章 评价对象——医学院校教师的基本素质要求	15
第一节 思想政治素质	15
第二节 业务知识素质	17
第三节 能力素质	19
第四节 身体心理素质	21
第三章 评价依据——教师应掌握的现代教学理论	22
第一节 现代教学理论发展概况	22
第二节 教学目的与任务	27
第三节 教学过程	30
第四节 现代教学方法	40
第四章 评价方法——医学院校教师教学质量评价方法概述	55
第一节 评价指标体系的设计原则与方法	55
第二节 收集评价信息的方法技术	64
第三节 处理评价信息的方法技术	67
第四节 评价结果的分析与反馈方法	74
第五节 电子计算机在教师教学质量评价中的应用	84
第五章 评价组织——医学院校教师教学质量评价的组织管理	95
第一节 教师教学质量评价的组织机构	95
第二节 教师教学质量评价过程的组织管理	97
第三节 教学质量评价的心理及调控	100
第六章 课堂教评——医学院校教师课堂教学质量评价方法	104
第一节 建立课堂教学质量评价指标体系的依据与内容	104
第二节 课堂教学质量评价具体实施方法	115
第三节 医学院校课堂教学质量评价应注意的问题	124
第七章 实验教评——医学院校教师实验教学质量评价方法	129
第一节 建立实验课教学质量评价指标体系的依据与内容	129
第二节 实验课教学质量评价具体实施方法	134
第三节 医学院校实验课教学质量评价应注意的问题	143
第八章 临床教评——医学院校教师临床实习教学质量评价方法	145

第一节	建立临床实习教学评价指标体系的依据与内容	145
第二节	临床实习教学评价的实施方案	146
第三节	临床实习评价量表的统计分析工作	147
第九章	国外介绍——美国医学院校教师教学质量评价方法	156
第一节	美国医学院校教师教学质量评价概况	156
第二节	美国医学院校教师教学质量评价的实施方案	157
第三节	美国医学院校教师教学质量评价的具体实例	185

下篇 对学的评价

第十章	现代学力观与学力评价	193
第一节	现代学力理论	193
第二节	医学生学力结构	197
第三节	学力评价的方法与特点	201
第十一章	医学生学业成绩评价方法的概述	205
第一节	教育目标分类学简介	205
第二节	医学生学业成绩评价的设计	208
第三节	教育目标分类学在医学考试中的应用	212
第四节	学业评价的可靠性与有效性	215
第十二章	笔试方法	222
第一节	自由应答题型	222
第二节	固定应答题型(是非题、填充题)	226
第三节	固定应答题型(多项选择题)	229
第四节	考试分数评定及统计分析	237
第五节	笔试评价特点	243
第十三章	口试方法	248
第一节	医学口试概况	248
第二节	口试方法介绍	250
第三节	口试评价的特点	254
第十四章	现场考试与直接观察评定	258
第一节	概述	258
第二节	现场考试方法	259
第三节	直接观察评定	267
第十五章	病案记录评定方法	273
第一节	病案的概念与作用	273
第二节	病案记录的种类与要求	274
第三节	病案记录评定方法	281
第四节	病案评定要求和注意事项	284
第十六章	国外介绍——美国医学考试方法	285

第一节	结构化客观临床考试.....	285
第二节	标准化病人考试.....	291
第三节	处理病人问题考试.....	295
第四节	计算机模拟考试.....	298
第十七章	国外介绍——法国医学考试方法及特色.....	302
第一节	法国医科各教学阶段的特别考试.....	302
第二节	法国医科教学考试方法特色.....	303
	参考文献.....	307

上篇 对教的评价

第一章 绪论—医学院校教师教学质量 评价的研究与实践

教育评价是对教育目标及其价值的判断的系统调查,为教育决策提供依据的过程。

教师的教学质量评价是根据一定的教育目标的标准,制定科学的评价指标体系,通过系统地收集教师的教学活动信息,准确地了解实际情况,进行科学分析,对教学工作质量作出价值判断,为改进教学工作提供真实可靠的依据。

从系统论观点看:高等学校教育评估是一个系统工程,宏观上包括学校合格评估,办学水平评估等,微观上包括学校内部评估。学校内部评估包括专业评估,教学评估等等。由于教学是高等学校的中心任务,因此教学评估是高校内部评估的主要的评估。教学评估又包括对教学管理、课程、教师、学生、教学条件等方面评估。由于教师是教学的主导,教师教学的质量如何直接关系到学校为社会培养人才的质量,如果说产品质量是企业的生命,同样地,教学质量也应该是学校的生命,是学校能否办得兴旺,能否办出特色,能否办出水平,能否办出声望的基本条件的保证。因此面对我国的高校现状,尊重高等教育的客观规律,研究我国高等院校教师教学质量评价的指标体系及应用具有重要的意义。

第一节 医学院校教师教学质量评价研究的国内外进展概况

一、国外进展概况

教育评估在发达国家起步很早,以美国为例,有近百年历史。从19世纪末期开始,经历了萌芽阶段,30年代测验时期,40年代“泰勒”时期,50年代低谷,60年代勃兴,70年代形成专业化。美国现有117所大学开设教育评估专业可授予硕士、博士学衔。1962年美国医学院校协会召开首届教育研究年会,开始开展高等医学教育评估的研究。美国医学教育联络委员会(LCME)负责医学教育评估工作,主要是对美国医学院校进行鉴定合格的认可性评估,在申报认可评估之前,各学校首先进行自我评估,然后提交认可评估申请报告。据文献资料,美国高等医学院校内部教学评估已经形成制度,主要采用学生评、同行评、自评三种方法,各校已经制定各种评价指标体系,如1988年华盛顿大学医学院组织专家通过听课,调查了解学生在临床上遇到最好的教师的教学实证,制定一份学生教学评价量表,包括7个方面44条指标,这7方面是:知识性、组织性、教学热情、讲授技能、临床管理技能、临床医疗能力及职业道德。评价表按7分制评分。经过正式评价之后,应用逐步回归分析方法得出结论,以下四种特征最能显著评价教师的教学质量(1)教师使我专注于学习过程($R^2=0.779$)。(2)教师使我明确掌握,临床实习要求($R^2=0.846$)。(3)教师激发我的学习兴趣($R^2=0.864$)。(4)教师处理病人技术熟练($R^2=0.876$)。这些资料为科学制定评价量表提供依据。

美国医学院校对教师教学质量评价影响因素和数据处理研究很深入,研究成果有研

究学生评价的可靠性即内部一致性,不同时期评价的稳定性,学生评价的有效性,学生个性特点及学习成绩对评价的影响。教师的年龄、职称、性别、教学负担、教学经历、科研成果对评价的影响。学生评、同行评、自评三者关系等等。对这些研究方向的题目,以前我国高等医学院校基本没有涉猎,当然美国没有像我们这样全国十二所高校统一制定评价指标,使用相同量表,统一利用几万名学生去评价几百名教师。这也是我国评价课题研究的长处。

二、国内进展概况

我国的教育评价起步较晚,1985年中共中央公布了《关于教育体制改革的决定》后,对我国的高等教育评估起到巨大推动作用。在国家教委的领导下,边研究边试点开展了大量的活动。1990年12月国家教委发布了《普通高等学校教育评估暂行规定》,使我国的高校教育评估的研究与实践,进入一个新的阶段。

由于,在我国高等学校开展评价活动是一项开创性的工作,无论是评价理论还是方法都很薄弱。近年来,国内一些医学院校也开展了教师教学评价指标体系的制定,大部分由教务行政部门,自己制定量表,缺乏对量表本身的反复实验和科学论证。

对评价结果的分析处理也没有达到高级阶段,缺乏对测量本身可靠性,有效分析。对学生评、自评、他评三者的关系,缺乏定性定量分析。对评价结果的解释,对评价各种影响因素及相互关系都缺乏科学研究,等等。

中国医科大学1988年末获得的三年6万元经费的卫生部基金课题,从1989—1991年由医大牵头联合全国十一所不同层次的医学院校协同研究,课题由总负责人金魁和领导了主要研究人员近40余名开展工作。

课题最初定的题目和目的是研究制定教师教学质量评价的指标体系。随着研究工作的开展,全国参与学校12所,涉及面广、资料丰富、数据库庞大。因此,研究题目从指标体系发展到研究评价方案与方法,研究电子计算机处理数据的软件系统,研究如何对评价结果科学的解释,研究评价结果反馈信息的利用,研究自评、他评、学生评之间的关系及各种影响因素等等。现在课题经过三年研究一年推广,已经全部完成。最终形成的主论文的题目是:“高等医学院校教师教学质量评价的研究与实践”。

由中国医大等12所医学院校开展的此项研究的特色和创新之处是:

1. 研究教学评价的指导思想端正,坚持研究的社会主义方向。重视形成性评价为学校教育改革服务。力图把教学评价真正作为促进教学质量动力。
2. 在国内外研究基础上创新制定评价指标体系和评价方案并经过科学论证和反复实践,力争达到先进性,导向性和可行性。
3. 国内首次不同层次12所高等医学院校统一步调、统一方案进行评价。实验范围广、时间长、数据丰富。被教师597名,参评学生24,757人次,参评专家2,355人次;涉及医学专业33个学科。
4. 创新完成电子计算机数据处理软件系统,定性定量地分析各种因素关系。
5. 从评价结果中找出目前我国高等医学院校在教学质量方面的现状与问题,在全国会议上交流,并为上级行政部门决策提供咨询。

目前已取得的成果

1. 制定出比较科学的“高等医学院校教师教学质量评价指标体系”九套评价量表及较实用的评价方案。

2. 制作出电子计算机对评价数据处理的软件系统。

3. 从结果分析中找出目前存在问题,为教学改革决策提供咨询。

4. 已经公开发表十余篇研究论文。

5. 出版发行《教与学评价实用技术》专著40万字。

6. 评价方案已经在全国部分院校推广应用。

第二节 医学院校教师教学评价指导思想的转变

评价指导思想是教学评价成败的关键。这是十年来我们进行教学评价实践中,体会最深,教训最深刻之处。

以中国医科大学为例。教学评价在中国医大起步很早,1982年开始试点搞教师教学评价,当时的出发点是:在高校的职称评定过程中“两文”(论文、外文)是硬指标、教学质量是软指标,只好借助于教学评价为手段,把教学质量评出高低,评出数量值,便于鉴别排队和作为晋升职称依据。这种指导思想实质是40年代美国教育家泰勒的“目标模式”,即评价就是确定教育目标实际实现程度的过程。它强调的是对价值判断、鉴别、验证和检查对象的工作,称之为总结性评价。

当时对中国医大基础部某一教研室14名教师进行课堂教学质量评价,主要是单渠道收集信息,学生填评价表、计算分值,最后搞排队。这次评价结果有一定差异,但是不太显著。特别是有两名教师和平时掌握有出入,评价结果不便于公开。由于评价结果勉强能排出队来,因此也写了一篇总结文章,认为方法基本可行。

1983年在基础部和临床医学系大面积推广时,遇到了教师的阻力。被评教师50余名,评价结果没有显著差异,平均在85—90分之间。仅有2—3名教师评价分数较低。费了大量人力、物力、财力,评价结果却是不显著,收效甚微。

1984年,1985年,医学教育界又出现单纯按照学生成绩评定教学质量的倾向,即按照学生考试分数,衡量教学质量高低,甚至按分数把学校排队,学校再按分数把教研室排队,教研室按分数又把教师排队,结果使广大教师产生抵触情绪。

1989年我们承担卫生部基金课题以后,认真回顾了几年来教学评价实践的经验教训,总结出以下几个问题:

(1) 以往评价的出发点有片面性,偏重于总结性评价而忽视了形成性评价。

(2) 以往评价信息渠道偏重于学生评价,忽视了全渠道评价,即教师自评,专家组他评价等。

(3) 评价方法过于简单,偏重1—2次抽样调查,忽视全过程评价。

(4) 评价者与被评者之间关系不协调。评价者居高临下,把被评价者视为检查对象,忽视对他们的指导帮助,造成情感对立等等。

按照传统的评价模式即泰勒模式,没有突出形成性功能和为决策提供信息的功能。价值判断唯一性,带来的只有高低优劣的横向比较,会引起评价对象的不安,使评价受到阻力,而不是改进教学的动力。

按照党的实事求是的路线来研究教育,认识到教育评价用在教育竞争机制中要持慎重态度。教育不同于经济领域及其它领域,硬套优胜劣汰原则对人才问题是不合适的。教师之间的竞争主要不在于战胜他人,而在竞争中发展自我,找出问题,改进教学,善于发现别人优点,以所长带所短、完善自我。教育现象有其本身的复杂性、层次性、超前性,不能过份强调平等,不能急于应用绝对比较的量标,要对具体问题具体分析等。

国外教育界60年代提出了“评价是为了进行决策提供信息的过程”的思想,首次指出了教育评价具有改善和发展一种正在进行的活动的功能即形成性评价功能,使评价成为改善自己工作的帮手,成为反馈、调控作用的杠杆。

按照以上思路,中国医大等12所院校在课题设计时指导思想由总结性评价真正转变为形成性评价模式,由传统的教育价值观转变为社会主义教育价值观,发挥教育评价的目标导向作用;激励续能作用;反馈矫正作用;区别分等作用等等。

第三节 医学院校教师教学质量评价的影响因素及相互关系

医学院校教师教学质量评价的理论方法包括很多方面,将在以下各章分别阐述。因为评价影响因素及相互关系的内容在各论没有机会阐述,特在绪论这一节说明。

虽然近些年,我国的医学院校已开始使用学生、同行和教师自我评价方法来评价教师的教学质量,并且已研究设计了许多评价表,但是许多与教学评价有关的问题研究甚少,例如,学生评价、同行评价和教师自评之间存在什么样的关系?当解释评价结果时应当注意些什么?学生、课程及教师因素是否对学生评价产生影响?应当使用什么样的评价方法?怎样使用评价结果?这些问题对教师教学质量评价来说是非常重要的。当评价结果是用于教师的晋级、任用等管理目的,对评价结果的解释更应当慎重。评价方法的可靠性和有效性也显得更加重要,然而,实际中确有一些影响学生及教师评价的因素可能误归于评价的有效性,还有一些教师控制不了的因素错误解释为教师的教学效果,这样可能伤害一些教师的感情,影响教师教学质量评价的正常开展,很多人已经认识到这些问题的重要性,但由于缺乏研究证据,也无法解释这些问题。

我们对由学生、同行和教师自我使用的课堂教学、实验教学和临床教学评价表的可靠性和有效性进行了检验;对学生、同行和自我三种评价方法的关系做了分析;对学生评价、同行评价的内部一致性进行了检验;对可能影响学生评价的学生、课程和教师因素进行了研究,学生和课程因素包括学生的年级、班级规模、学科等,教师因素包括教师的职称、年龄、教学负担、科研成就、教学经历等,对影响教师自评因素也进行了分析,并提出了建议。另外,对关于如何使用结果以及最有效、最可靠和最可行的评价方法也进行了研究。下面介绍一下这方面的研究成果。

一、学生评、同行评、自评间的关系

(一) 研究结果

表1—1是学生、同行和自我评价的相关系数和评价均值,如表所示,学生评价与同行评价的相关系数已在0.01水平上有显著性,然而,自评和学生评分别与同行评价的相关系数很低。学生评价的均值有显著意义低于同行和教师自我评价,同行评和自评的均值基本一致。

表 1—1 6452 名学生 (1) 683 名同行、(2) 160 名自我、
(3) 评价课堂教学的 Pearson 相关系数和均值

关系 a-b	被评教师 数量	Pearson 相关系数	均 值				u
			Xa	SD	Xb	SD	
1-2	160	0.20**	78.2	9.2	82.6	10	-4.25**
1-3	148	0.14	78.2	9.4	82.2	8.7	-3.38**
2-3	148	0.12	82.8	9.9	82.2	9.6	-0.55

* 在 0.05 水平上有显著性

** 在 0.01 水平上有显著性

表 1—2 和表 1—3 所示分别是实验课教学和临床教学学生、同行和自我评价的均值和相关分析结果。

对于实验课教学, 学生评价和同行评价的相关系数非常显著, 比学生评与自评, 同行评与自评的相关系数要高, 另外, 同行评价均值有显著意义低于学生评均值 ($u = 4.69$, $P < 0.01$) 和自我评价均值 ($u = -3.13$, $P < 0.01$)。

表 1—2 实验课教学 4892 名学生评价 (1) 429 名同行评价
(2) 90 名自我评价 (3) 的关系

关系 a-b	被评教 师数量	Pearson 相关系数	均 值				u
			Xa	SD	Xb	SD	
1-2	95	0.30**	90.9	6.5	85.2	9.8	4.66**
1-3	80	0.25*	90.8	6.6	89.7	7.9	1.00
2-3	80	0.11	86.1	10.1	89.7	7.9	-3.13**

* 在 0.05 级平上有显著性

** 在 0.01 水平上有显著性

对于临床教学, 如表 1—3 所示, 学生评价与同行评价的相关系数最显著, 学生、同行与自评的相关系数不显著, 同行的评价均值有显著意义低于学生评价和自我评价均值。

表 1—3 临床教学 1005 名学生评价 (1) 356 名同行评价 (2)
88 名自我评价 (3)

关系 a-b	被评教 师数量	Pearson 相关系数	均 值				u
			Xa	SD	Xb	SD	
1-2	81	0.31**	90.7	9.5	87.1	7.7	2.59**
1-3	73	0.01	89.8	9.6	90.5	7.5	-0.18
2-3	73	0.13	87.6	7.9	90.6	7.5	-2.26*

* 在 0.05 水平上有显著性

** 在 0.01 水平上有显著性

表 1—4 是学生和同行使用同一种课堂评价表在评价指标的 5 个方面评价均值和相

表 1—4 学生和同行使用同一种评价表对教师课堂教学
在总体教学效果和 5 个方面的 Pearson 相关系数和均值

教学范畴	平均相关系数		均 值	
	学生*同行		学生	同行
	被评教师(30)	同行(30)	被评教师 (30)	(30)
1. 各课	0.35*		90	94
2. 知识传授	0.42*		85	88
3. 能力培养和 解决问题	0.52**		81	85
4. 思想教育	0.47**		78	83
5. 教学方法	0.43*		85	91
总体效果	0.45**		83	88

* 在 0.05 水平上有显著性

** 在 0.01 水平上有显著性

关系数，由学生和同行教师总体教学效果和 5 个方面的评价有显著意义的相关关系，学生评价均值在总体教学效果的 5 个方面都低于同行评价，但只有两个方面有显著差异。

(二) 学生评、同行评、自评三者的关系讨论

许多研究对单一评价方法进行了分析，然而，对不同评价方法之间关系的研究甚少，其原因可能是由学生、同行和自我对教学质量进行评价较难同时进行，但它是一个重要的问题，它涉及到学生、同行、自我或其他评价渠道对教师教学质量的想法是否一致，这种一致有两种意义：由一种评价渠道评价好的教师，其他评价渠道是否也评价他好，他们对同一教师的评价是否在同一水平上。例如，学生认为好的教师，同行是否也认为好，对他们之间关系的分析可以分为教师的晋级、任用或提高教学质量合理使用学生、同行和自我评价提供依据。

在该研究中，分析结果表明，不论在何种教学形式下即课堂教学、实验课教学、临床教学中学生评价和同行评价都非常显著相关，即学生认为好的教师，同行他评也认为好，说明这是客观标准评价。学生评价与自评、同行评价与自评的相关系数都很低（P 值无意义），缺少统计学意义的结果提示我们，教师对自身教学效果的评价与学生、同行的意见不一致，总的来说，教师自评属于主观评价，这个渠道不能忽视，不能为学生评、他评渠道代替，在运用评价结果时，要具体问题具体分析，参照教师本人意见。

进一步分析表明，学生、同行和自我评价的均值不在同一水平，学生评价均值有显著意义低于同行和自我评价（如表 1—3 所示）。另外，学生评价的均值在评价表学生表二指标体系的 5 个不同方面都低于同行评价，其中两个方面有显著性（如表 1—2 所示）。评价结果看来似乎学生评价比同行和教师自我评价严格，这也反映了学生对教师的教学效果不满意，教师对自己的评价偏高，同行评价较高可能是由于多数同行都来自被评教师同一教研室，因此评价较宽容。

学生评价和自评在实验和临床教学都显著高于同行评价，这种情形与课堂教学不同，教师在实验和临床教学能较多参与学生的学习活动可能是造成较高学生评价的原因

之一，实验和临床教学中，学生的数量较少，教师与学生有个别接触的机会，这样，增加了师生间的相互了解。

二、影响学生评价的学生自身因素和课程特点

(一) 研究结果

学生年级：表1—5是不同年级学生对教师的评价均值，虽然一年级的学生评价最低，但与其他三组学生没有显著性差异 ($F = 2.23, P > 0.05$)。

学生班级人数：表1—6 (1) 表1—6 (2) 和表1—6 (3) 分别是不同规模班级学生对教师评价均值和方差分析情况，方差分析表明，不同学生人数的班级对教师的评价结果有显著意义的差别 ($F = 2.93, P < 0.05$)。对评价结果又进一步做了两两比较分析，结果发现下述学生班级间存在显著性差异：小于40名学生的班级与100—160名学生的班级 ($t = 2.19, P < 0.05$)。小于40名学生的班级与160—200名学生的班级 ($t = 2.64, P < 0.01$)；40—100名学生的班级与160—200名学生的班级 ($t = 2.03, P < 0.05$)。方差齐性检验证明方差齐 ($P < 0.05$)。

学科：表1—7 (1) 是学生在正式评价和预评价对不同学科教师 (普通基础，医学基础，临床和预防医学) 的评价值，学习临床课程的学生不论在正式评价和预评价阶段对教师的评价值都最高。在正式评价阶段，不同学科的评价结果方差分析无显著差异，然而，对于预评价，确存在显著性差异 ($F = 5.08, P < 0.01$)。表1—7 (2) 和表1—7 (3) 分别是方差分析和两两比较分析结果。

学习临床课程和学习医学基础课程的学生对教师的评价高于学习普通基础课程的学

表1—5 不同年级学生对教师课堂教学的评价均值

学生年级	被评教师数	评价均值
1	32	74.8
2	22	80.9
3	20	76.8
4	37	78.4

表1—6 (1) 课堂教学不同规模班级学生对教师评价均值

组号	单位班级学生数	涉及各类班级数	均值
1	10—	16	83.5
2	40—	29	80.3
3	100—	25	76.4
4	160—200	12	73.5

表1—6 (2) 课堂教学不同规模班级学生对教师评价的方差分析

评价来源	平方和	均方	自由度	F值
组间	905.2	301.7	3	2.93*
组内	9077	103.1	88	

表 1—6 (3) 课堂教学不同规模班级学生对教师评价的两两比较

比较组 a-b	$X_a - X_b$	t
1-2	3.24	1.07
1-3	7.10	2.19*
1-4	9.92	2.64**
2-3	3.86	1.48
2-4	6.68	2.03*
3-4	2.82	0.80

* 在0.05水平上有显著意义

** 在0.01水平上有显著意义

表 1—7 (1) 正式评价和预评价阶段不同学科对教师课堂教学的评价均值

学 科	正式评价		预评价	
	被评教师数	均值	被评教师数	均值
普通基础 (BS)	19	78.4	17	77.3
医学基础 (BMS)	30	77.7	22	81.8
临床 (CS)	46	81.4	26	83.8
预防医学 (PM)	12	76.7		

表 1—7 (2) 预评价阶段不同学科学生对教师课堂教学评价的方差分析

评价来源	平方和	均方	自由度	F值
组间	403.4	215.2	2	5.08**
组内	2585.1	43.2	61	

表 1—7 (3) 预评价阶段不同学科学生对教师课堂教学评价的两两比较

比较组 a-b	$X_a - X_b$	t
临床·医学基础	2.0	1.05
临床·普通基础	6.5	3.33**
医学基础·普通基础	4.5	2.25

* 在0.05水平上有显著意义

** 在0.01水平上有显著意义

生评价, 统计学上有显著意义 ($t = 3.33$, $P < 0.01$ 和 $t = 2.25$, $P < 0.05$), 方差齐 ($P > 0.05$)。

(二) 学生自身因素对评价的影响

虽然学生评价总的来说是可靠的, 但学生和课程方面许多因素会影响学生评价。可能产生影响的学生因素包括学生年龄、性别、年级、学业能力、平均成绩、对课程期待的分数、选课的动机以及个性因素, 课程特点包括必修课类型 (主修、必修及选修)、学科领域、班级规模和教学方法等, 在该研究中对上述部分因素进行了方差分析、相关分析和T检验等统计学分析。