

六门临床课程形成性评价实践研究

孙 敏

(南京医科大学第四临床医学院,江苏 南京 210029)

摘要:文章抽样南京医科大学六门临床课程成绩,通过 SPSS 统计软件分析形成性评价、终结性评价、综合性评价成绩的相关性和差异性。结果显示终结性评价具有较高评价效力,无论在综合性评价中占分比例大小,都与综合性评价呈显著正相关;形成性评价需要占有较大分值比例,才与综合性评价显著相关。结果提示,课程学习成绩评价需要调整和明确目标要求,不同的评价方法适用于不同的评价目标要求;不少课程现行的形成性评价方法需要进一步改进。

关键词:临床课程;形成性评价;终结性评价;综合性评价

中图分类号:G642.475

文献标识码:A

文章编号:1671-0479(2012)02-144-004

本科医学教育标准要求评定体系包括形成性评价和终结性评价。形成性评价包括测验、观察记录、查阅实习手册等,终结性评价包括课程结束考试及毕业综合考试等^[1]。2010年我校开展课程考试方法改革,要求各专业(方向)主要课程(必修课)的成绩评定均采用综合性评价,根据学科、课程特点的不同,在综合性评价中,终结性评价可占总成绩的60%~80%,形成性评价占20%~40%^[2]。两年来,全校111门必修课,71门采用了包含形成性评价分值的综合性评价,占64%,7门采用了形成性评价,占6%。形成性评价从课堂表现、课后作业、出勤率、见习考核、实验设计、答辩表现等多方面开展。

本文抽样我校六门临床课程综合性评价成绩,通过 SPSS 统计软件分析形成性评价、终结性评价、综合性评价成绩的相关性和差异性。根据统计结果,了解背后的原因,并对课程考试方法实际运用效果进行评价,探讨进一步深入开展医学课程考试方法改革的思路。

一、临床课程考核现状

目前临床课程考核多采用期末笔试这种终结性评价来检验学生的课程学习结果。终结性评价是在一个学期或者一个教程结束的时候,为了进行分等、鉴

定、进步评价,或者对课程、学程以及教育计划的效力(有效性)进行研究而进行的评价^[3]。对835名医学生的调查统计得出78.9%的医学生有“前紧后松,应付考试”情况^[4]。临床课程形成性评价研究尚在起步阶段,在中国知网搜索1979年以来“形成性评价”主题论文1475篇,采用“实证研究方法”有29篇,占2%。其中有关医学教育研究的仅有10篇,占1%。

二、样本选取与分析方法

选取30名2005级七年制临床医学(儿科方向)学生的儿童保健学成绩,45名2006级临床医学(耳鼻咽喉方向)学生的视光学成绩,32名2007级临床医学(儿科方向)学生的儿科学、小儿传染病学成绩,55名2007级临床医学(生殖方向)学生的生殖工程学、临床生殖学共六门临床课程成绩,运用形成性评价、终结性评价、综合评价得出的三类成绩分别简称“形”、“终”、“综”,具体考试方法如表1所示。运用SPSS16.0软件,采用Pearson相关分析、多配对样本Friedman检验方法对三种评价成绩的相关性、差异性进行统计分析。

三、统计分析

(一)描述性统计

六门临床课程三类评价成绩见表2。

基金项目:中华医学会医学教育分会2010年度医学教育研究课题(2010-20-07)

收稿日期:2012-03-05

作者简介:孙 敏(1981-),女,江苏盐城人,教育学硕士,助理研究员,主要从事医学教育管理与研究工作。

表1 六门临床课程考试方法

课程	考试方法
儿童保健学	团队答辩分×66.6%+平时分×3.8%+学习体会分×4.8%+期末考试分×24.8%
视光学	平时分×20%+实践操作分×30%+期末考试分×50%
儿科学	平时分×20%+期末考试分×80%
小儿传染病学	平时分×20%+期末考试分×80%
生殖工程学	平时分×5%+期末考试分×95%
临床生殖学	平时分×5%+期末考试分×95%

标准差反映了一个数据集的离散程度,标准差大,表示学生成绩的离散度大,提示评价方法的评价效力和区分度高。六门课程中,儿童保健学、视光学、小儿传染病学、生殖工程学、临床生殖学五门课程终结性评价成绩的标准差大于5分,形成性评价成绩

的标准差在视光学、生殖工程学、临床生殖学三门课程大于5分。结果说明终结性评价相对形成性评价有效性和区分度更高。

(二)三类评价成绩相关分析

采用 Pearson 相关分析。Pearson 相关系数通常用 r 或 ρ 表示,是用来度量两个变量之间的相互关系(线性相关)的,取值范围为 $-1 \leq r \leq 1$ ^[5]。对于相关系数的价值如何解释,统计学家的见解尚不一致,但通常作如下的解释: $0.1 < |r| \leq 0.3$,表示相关程度低,可称之为微相关; $0.3 < |r| \leq 0.5$,表示确实有关系,可称之为切实相关; $0.5 < |r| \leq 0.8$,表示相关程度比较高,可称之为显著相关或密切相关; $0.8 < |r| \leq 1.0$,表示相关程度很高,可称之为高度相关^[6]。

表2 六门临床课程成绩描述性统计

(分)

课程	人数	形成性评价	终结性评价	综合性评价
		最小值~最大值 (均值 ± 标准差)	最小值~最大值 (均值 ± 标准差)	最小值~最大值 (均值 ± 标准差)
儿童保健学	30	77.00~92.70 (84.86 ± 3.47)	59.60~100.00 (83.96 ± 9.62)	76.00~94.00 (84.70 ± 4.48)
视光学	45	63.00~92.00 (82.80 ± 6.88)	63.00~97.00 (86.04 ± 7.88)	64.50~93.00 (84.53 ± 5.30)
儿科学	32	75.00~89.00 (80.38 ± 3.28)	80.00~94.00 (89.03 ± 3.28)	79.00~92.00 (86.81 ± 2.73)
小儿传染病学	32	75.00~89.00 (80.38 ± 3.28)	63.00~93.00 (82.38 ± 6.55)	65.50~92.00 (81.95 ± 5.42)
生殖工程学	55	60.00~90.00 (74.70 ± 7.70)	65.00~98.00 (89.36 ± 6.39)	65.00~97.00 (88.65 ± 6.25)
临床生殖学	55	49.00~90.00 (70.74 ± 9.67)	75.00~100.00 (90.64 ± 5.81)	74.00~100.00 (89.65 ± 5.78)

表3显示了三类评价成绩两两之间的相关系数。儿童保健学、生殖工程学、临床生殖学三门课程的“形”与“终”显著相关($P < 0.01$)。“形”与“终”之间应该有一定的相关性,如果相关性过低是违反学习规律的,说明其中一种评价方法是不可靠的。但另一方面,相关程度亦不应过高,因为“形”与“终”的评价目标要求应有所不同,否则就失去了“形”评价方法存在的必要性。因此,“形”与“终”评价成绩的相关性在切实相关与显著相关之间为宜。

六门课程“形”与“综”都呈正相关($P < 0.01$),这是由于“形”是“综”的一部分。但不同的课程要用的“形”和“综”的评价方法不同,“形”与“综”相关的程度不同。儿童保健学“形”与“综”高度相关,视光学、生殖工程学、临床生殖学“形”与“综”显著相关,儿科学、小儿传染病学“形”与“综”切实相关。

从表3还可以看出,儿童保健学“形”占“综”的比例最高,“形”与“综”也呈高度相关;生殖工程学和临床生殖学“形”占“综”的比例很小,仅5%,对“综”的成绩形成影响甚微,但却与“综”显著相关;而儿科学与小儿传染病学“形”占“综”20%,却与“综”仅有切实相关。说明“形”占“综”的比例不是影响其间相关程度的决定因素。关键因素是“形”与“终”的相关程度,因为“综”是由“形”和“终”构成的,“形”与“终”相关程度很低,“形”与“综”的相关程度就不能提高,儿科学与小儿传染病学就是如此。

再看“终”与“综”的关系,表3显示除视光学显著相关外,五门课程“终”与“综”都呈高度相关。不管“终”占“综”的比例大小,特别是儿童保健学、视光学“终”占“综”的比例只有24.8%和50%,却也能达到显著相关。说明“终”占“综”的比例无论大小都能与“综”达到显著相关。

表 3 六门临床课程三类考核成绩相关统计

课程	相关系数			比例关系(%)	
	“形”与“终”	“形”与“综”	“终”与“综”	“形”与“综”	“终”与“综”
儿童保健学	0.558*	0.891*	0.870*	75.2	24.8
视光学	0.029	0.659*	0.765*	50.0	50.0
儿科学	0.010	0.451*	0.873*	20.0	80.0
小儿传染病学	0.033	0.461*	0.878*	20.0	80.0
生殖工程学	0.554*	0.609*	0.996*	5.0	95.0
临床生殖学	0.515	0.557*	0.998*	5.0	95.0

*: 相关系数的假设检验显示 $P < 0.01$, 可认为两者间存在相关关系。

(三) 三类评价成绩差异分析

在不确定总体是否呈正态分布的情况下, 需用非参数检验方法, 由于三类评价成绩来自同一个体, 是配对样本, 采用多配对样本 Friedman 检验。结果显示, 儿科学、小儿传染病学、生殖工程学、临床生殖学四门课程的三类评价成绩来自的总体成绩分布有明显差异 ($P < 0.05$), 儿童保健学、视光学这两门课程三类评价成绩来自的三个总体成绩分布无显著差异 ($P > 0.05$)。

结合考试方法构成来看, 儿童保健学、视光学这两门课程期末考试比例较小, 分别占 24.8% 和 50.0% (表 3), 而形成性评价的要求较高, 需要团队答辩、学习体会、平时表现考核、实践操作考核, 不是笼统的平时分 (表 1)。而儿科学、小儿传染病学、生殖工程学、临床生殖学四门课程三种评价方法得出的成绩总体分布差异明显, 结合这四门课程的形成性评价方法来看, 都采用“平时分”的方式, 说明这几门课程的“平时分”考核不够严谨, 不能反映学生的实际学习状况。在接下来的课程考试中需要进一步改造并合理规范操作“平时分”考核。

四、结论与思考

(一) 终结性评价效力高, 应该坚持运用

从以上分析可以看出, 终结性评价方法虽为传统评价方法, 但同时也是经典的、最成熟的评价方法, 其区分度高, 评价效力高, 不管占综合性评价比例大小, 都能反映出学生学习水平。

因此, 终结性评价依然是评价学生学习效果的有效方法。但目前的终结性评价的主要方式是期末闭卷考试, 主要目标在于测试学生对基本知识的记忆、理解和简单应用, 无法考察学生的分析问题、沟通交流、依靠原有知识技术获取未知答案等能力。对测试学生在学习后形成的真正能力 (运用知识解决实际问题的能力) 没有作用。因此要积极推行终结性评价的考试方法, 增加测试能力的有效手段如临床技能操作考试、病例分析考试等, 使终结性评价

既评知识又评能力。

(二) 形成性评价方法需要改进, 并得到正确运用

从相关分析来看, 儿童保健学、视光学的形成性评价比例分别占 75.2% 和 50%, “形”与“综”、“终”与“综”的相关性都很显著。不管终结性评价占综合性评价比例大小, 都与综合性评价呈显著相关, 说明综合性评价是否客观, 不取决于终结性评价, 而是取决于形成性评价。只有形成性评价客观有效, 才能与综合性评价显著相关。如果随意操作形成性评价, 不能考核学生真实的学习状态和水平, 会对综合性评价形成干扰, 使综合性评价成绩发生偏移, 不能很好地解释学生平时学习水平。

从差异分析来看, 高形成性评价比例的儿童保健学、视光学这两门课程三类评价成绩来自的三个总体成绩分布无显著差异。其他四门形成性评价比例在 20% 以下的, 三类评价成绩有显著差异。说明形成性评价分比例偏小的课程, 在综合性评价中不能反映一个学生的平时学习水平。所以形成性评价占高比例, 有利于从综合性评价中体现出学生的平时学习水平。

形成性评价应真正反映学生的学习过程, 对学生的学习态度、学习能力、学习兴趣、阶段性成果予以客观评价。形成性评价也要测试学生运用知识探索未知、解决实际问题的能力。形成性评价不能过于简单, 要努力避免区分度小, 与终结性评价过多重复的问题, 真正提高过程性评价的质量。

(三) 学习成绩评价方法的评价与改革

从六门医学课程的实际评价运用效果来看, 我校目前形成性评价实际运用还不够成熟, 评价效力尚待提高。需要扭转长期形成的单纯知识点考核模式, 结合培养医学生临床运用能力的教学改革, 如 TBL、PBL 教学改革, 评价医学生的团队合作能力、以问题为核心的临床诊断思维等。期末考试这样的终结性评价之所以被长期采用与它的客观性、准确性、区别性有关, 但它也有不足之处, 不能反映出学生的学习情感、学习态度、实践操作能力, 需要注重

过程性的形成性评价的补充,所以结合形成性评价和终结性评价的综合性评价是实际需要的。

要说明的一点是,形成性评价可以理解为“平时分”,终结性评价则不能理解为闭卷试题考试模式,主要的区别在于评价目标不同。终结性评价的目标在于评价学生学习的成果,包括对知识的掌握、运用、职业态度与情感的养成、职业能力的塑造等,而形成性评价的目标在于评价学生的学习过程,是否参与其中、学习态度、学习兴趣、所花时间和精力、阶段性成果等。因此,终结性评价的方法不能限于一张考卷,要拓展其方式方法,客观评价其能力、态度、情感形成的结果;形成性评价也要多种形式、多种手段,客观详细地反映学生的学习情况;再把两者结合起来进行综合评价,才能建立起真正客观、有效的学习评价体系。

要坚持和深化考试方法改革,需要实际了解教师是怎么具体操作形成性评价的,他们是否能将包含形成性评价的综合性评价坚持下去。三种评价方式的差别和优势还有待时间和学生的后期学习效果去考证,要找几门有相互关系的课程,如小儿传染病

学的基础是免疫学、微生物学,一二年级的免疫学、微生物学的三种评价成绩与三年级的小儿传染病学成绩是否有相关性?如呈高度相关,才能说明这种评价方法是有效的。

参考文献

- [1] 本科医学教育标准 [EB-OL]. [2008-09-16]. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohbgt/pw10812/200811/38226.htm>
- [2] 南京医科大学学生成绩评定与考试管理工作实施指导意见 [EB-OL]. [2011-10-11]. <http://jwc.njmu.edu.cn/s/34/t/79/43/16/info17174.htm>
- [3] Bloom BS, Hastings JT, Madaus GF. Handbook on formative and summative evaluation of student learning [M]. New York: MacGraw-Hill, 1971: 48
- [4] 祖勤. 医学生学习动因的调查分析 [J]. 医学教育探索, 2010, 10(9): 1312
- [5] 王孝玲. 教育统计学 [M]. 上海: 华东师范大学出版社, 1999: 251
- [6] 王汉澜. 教育统计学 [M]. 北京: 教育科学出版社, 1986: 118

Research on the practices of formative assessment of six clinical courses

SUN Min

(The Fourth School of Clinical Medicine, Nanjing Medical University, Nanjing 210029, China)

Abstract: The article analyzed the correlations and differences among formative, summative and comprehensive assessment of six clinical courses with SPSS. The results showed that summative assessment was more effective and had a significant correlation with comprehensive assessment no matter the proportion, and the correlation between formative and comprehensive assessment was related to the big proportion of formative assessment. These results suggested that it was necessary to adjust and clear the goal and requirement of course assessment, different methods of assessment should be applied to different goals and requirements of assessment, and many methods of formative assessment for existing courses should be further improved.

Key words: clinical course; formative assessment; summative assessment; comprehensive assessment