

国家学生体质健康监测效果的真实性评价设计

于红妍¹, 张亚平², 常冬青³

(1. 上海交通大学 体育系, 上海 200030; 2. 新加坡楷博高等教育学院, 上海 200050;
3. 上海体育局 水上运动中心, 上海 310000)

摘要:从教育评价的视角分析,我国学生体质健康监测是一项工作,更是一项教育活动。当前由于忽略了它的教育功能,导致对监测效果的评价出现了评价主体单一、评价内容单一、测试工具单一等问题。运用文献分析法和逻辑归纳法,以真实性评价理论为依据,对我国学生体质健康监测如何开展和应用真实性评价进行研究。发现,我国学生体质健康监测具有评价任务的真实性、评价内容的真实性以及评价方式的真实性等特征。据此提出了学生体质健康监测实施真实性评价的五个步骤,并设计了四种类型的学生体质健康监测效果评价方案,为拓展我国学生体质健康监测的教育功能、提高体质健康监测效果提供方法上的支持。

关键词:国家学生体质健康监测;真实性评价;监测效果;评价设计

中图分类号:G804.49

文献标志码:A

文章编号:1008-3596(2015)02-0006-05

我国十分重视学生体质的健康发展。1952年我国出台了新中国第一个体质测试达标标准——《准备劳动与保卫祖国体育制度》(简称《劳卫制》,1952—1963年),伴随着《国家体育锻炼标准》(1975—2002年)、《(大、中、小学)学生体育合格标准》(1990—2001年)、《学生体质健康标准》(2002—2006年)和《国家学生体质健康标准》(2007年至今)等学生体质健康评价标准的出台,我国学生体质健康监测已走过了60余年的发展历程,目前已成为学校体育教育中的一项系统工程。然而从教育评价学的视角分析,我国的学生体质健康监测存在着较突出的问题。如,评价主体较单一,学校或教师是评价的主体,师生之间的关系是评价和被评价的关系,忽视了学生、学生家长及社区等评价主体多元多向的参与。再如,测评工具较单一,目前使用的唯一的评价标准是《国家学生体质健康标准》,由于每种评价工具都有一定的优缺点,较单一的评价工具难免会导致评价结果的片面性。还有,评价内容较单一,评价内容主要是对学生的身体形态、机能和运动素质进行评定,缺少对学生认知态度、参与表现、测试技能掌握等内容的评价^[1]。究其原因,笔者认为可能是我们将学生体质健康监测仅仅视为一项工作而不是一项教育活动,忽略了它教育的本质功能。正在实施的《国家学生体质健康标准》明确说明了学生体质健康监测“是促进学生体质健康发展、激励学生积极进行身体锻炼的教育手段。”基于以上原因,本研究尝试采用教育评价中真实性评价理论和方

法,对学生体质健康监测如何开展真实性评价进行研究,以期拓展我国学生体质健康监测的教育功能、提高体质健康监测效果提供借鉴。

1 真实性评价的内涵

真实性评价(authentic assessment)不是与“假”评价相对的。一方面,它作为一种评价方法最早由美国的学习评价和学校结构顾问委员会执行主任(Executive Director of Consultant on Learning, Assessment and School Structure) Grant Wiggins于1989年提出,被定义为:学生运用自己所学的知识和掌握的技能解决生活中的或与现实情境相似的真实性任务,以便通过自己的创造性活动,培养、展示和证明自己的知识、才能以及解决问题过程中的策略^[2]。它由真实性任务和评价量规两个部分构成^[2-3]。其中,真实性任务是指现实生活中或模拟现实生活中的一件任务,学生可以用他们所学的知识和技能去解决。如操作一项实验、表演或运动技能、展示作品、角色扮演等等,藉此考察学生在真实性任务中能够做什么,而不是知道什么。量规是一种评分工具,它有评价标准、评价分数或等级以及评语三个方面,用来判断学生在完成一项真实性任务的过程中是否达到了所提出的要求。概括来说,作为一种具体的评价方法,真实性评价就是让学生完成一个操作性的任务,用以考察学生的知识与技能在解决实践问题过程中表现的交流合作、批判性思考等多种复杂能力的发展状况。

收稿日期:2014-10-18

基金项目:2013年教育部人文社科规划项目基金项目(13YJC890042)

作者简介:于红妍(1976—),女,山东莱阳人,副教授,博士,研究方向为学生体质健康监测与评价。

另一方面,随着国内外学者如 Resnick (1987)、Savery (1995)、Hammond (2000)、Kathleen Montgomery (2001)、Judith (2009) 以及 姐媛媛 (2000)、梁恒 (2013) 等^[4-10]对真实性评价标准和评价功能等理论研究的拓展和深入,真实性评价的涵义已超出了仅仅作为评价方法的局限。它通过将观察性和理解性的评价、学生自身的评价以及成长档案袋评价等质性评价方法与标准化测验相结合的方式,突出了评价的发展性功能、强调了评价的开放性和过程性,做到了对学生的全面、真实、有效、动态的评价,因此,真实性评价作为一种评价理念,成为多元性、过程性和发展性评价的典型代表。

2 学生体质健康监测中真实性评价的体现

2.1 学生体质健康监测任务的“真实性”

任务的真实性是真实性评价最重要的特点。这里的“真实性”是指任务是源自于现实生活的,故此学生的学习行为和学习成果能够用具体的行为或行动方案表现。完成真实性的任务就是要求学生用所学到的知识和技能在实际的任务中进行链接和运用,以此来证明学生能够做什么,而不是知道什么。学生体质健康监测通过这样“真实性”的任务,要求学生不仅学会体质健康监测的知识,更重要的是掌握体质健康监测的操作性技术和方法,并能够运用这些掌握的知识和技能解决现实生活中的问题,如体质健康水平的自我测量、根据体质健康评价结果给自己制定出有针对性的体育健身计划等。

2.2 学生体质健康监测内容的真实性

我国学生体质健康评价标准从 1952 年实行《劳卫制》到现行的《国家学生体质健康标准》一共经过了五次调整,尽管每一个新标准的出台都有着不同的历史背景或原因,但是几次在测试项目上的调整都呈现出了一种趋势,那就是,在测试内容上已由军事技能类和体能主导类项目转向与健康相关的体质健康项目的测试。如去掉了射击、手榴弹、举重、爬杆(绳)等项目,增加了肺活量(检测肺功能)、体成分(测量体脂含量)等项目,而且正在积极开发心理健康测试项目。调整后的测试内容与人们健康发展的观念相符合,真正考虑了学生现实生活和身体健康发展的需求,体现出人的“全”健康的评价理念,因此,对学生的全面发展来说是真实的和可实现的^[11]。

2.3 评价方式的真实性

真实性评价中典型的评价方式包括了档案袋评价、表现性评价以及基于日常行为观察的评价等。这些评价方式体现了真实性评价的发展性、个性化、开放性及多元性等过程性评价的特征,这也是将真实性评价应用于学生体质健康监测、解决当前只对监测结果进行一次终结性评价问题的原因。如应用档案袋评价方法,收集学生在体质健康监测过程中或结束后的一段时间内参加各类体育活动的记录、体育竞赛取得的成绩、体育课成绩,以及体质健康监测登记卡、参

与体育锻炼的感想等,通过这些信息的收集和分析,真实反映他们所处的学习阶段、过程中取得的进步或成长。再如观察性评价是基于对学生日常体育行为进行完整的连续的观察,做出的准确具体和详尽的记录,也可以邀请家长、专家以及同学参与评价的全过程,全面了解学生学习过程和学习表现,对学生做出真实而有意义的反馈。目前在学生体质健康监测中使用的《国家学生体质健康标准登记卡》以及建设中的学生体质健康测试数据的信息系统就是建立和管理档案袋评价的方式。这些评价方式是培养学生良好的体育锻炼习惯和体育生活方式的一种监督和鼓励,也是学生体质健康监测目标真实体现的途径。

3 学生体质健康监测真实性评价的设计流程

开展真实性评价并没有严格的程序,可根据不同的评价重点和要求灵活地实施。本文在这里设计了学生体质健康监测真实性评价的实施流程,该流程包括了实施真实性评价的五个基本步骤,为实施体质健康监测真实性评价提供操作性方法。

3.1 确定评价目标

评价始于对目标的确认,止于对目标达成程度的判断(Lirm&Groulund, 2003)^[12]。确定学生体质健康监测的目标就是回答学生参与了体质健康监测后应该具备什么样的知识和能力的问题。这些知识和能力将成为真实性评价的标准(standards)。目标可分为总体目标和具体操作性目标,无论是哪类目标都要求目标的提出要清晰明确,这是对学生体质健康监测的具体要求。例如“学生们通过参加体质健康测试,能够了解体质健康相关知识,掌握体质测评的技能和方法,并能够独立完成对自身体质健康状况的全面评价。”

3.2 设计真实性任务

为了评价学生是否达到了学生体质健康监测评价目标,评价者就要设计或选择相关的真实性任务,让学生通过执行和完成任务来证明他们具备和掌握了相关的知识和技能、达到了评价预定的目标。如调查并分析本校学生体育锻炼行为与其体质健康水平的关系,并写出评价报告;跟踪测量本人的身高体重指数,并预测未来可能的发展趋势等。

3.3 建立评价标准

如何判断学生在完成任务过程中的表现或完成真实性任务应该具备哪些标准?真实性评价标准是服从和服务于评价目标的,是对学生在某一方面应该达到的知识和技能的简短陈述。它必须是可观察、可测量的,而且公开和透明。本文设计了学生自评样表(表1),将体质健康监测的评价过程划分为三个阶段,每个阶段都列出了不同的评价项目。这样,评价者就可以根据被评价者对项目的完成情况做出等级或评分的评定。另外,为了解决评价主体单一的问题,这里的评价者可以是学生本人、小组成员、教师和家长,他们应各占一定的评分比重来提高评价的信效度。

3.4 制定量规 (assessment rubric)

根据任务表现标准,为了将不同学生的表现区分开,并与所确定的评价标准相对应,需要建立评价量规。量规也叫评分规则,主要由表现标准(criteria)、表现等级(performance level)和评语三个部分组成^[12]。量规是一个非常灵活的工具,评价者可以根据需要自行开发量规形式或分配

分值的权重^[13]。但不管是哪种形式的量规,它们或者是集中测量一定的表现、行为或品质目标,即表现标准的陈述,或者是用一系列评分(等级)来评价表现,即在每一个等级标准下都列有具体的表现特征,用以解释这个等级的表现应达到的程度。量规是给学生提供反馈的具体实物,在实际使用中要根据反馈和需要适当地调整和修改量规。

表1 学生体质健康监测效果评价表

学生姓名_____ 年级_____ 班级_____ 评价者_____		评价等级/评分			评语	权重
评价项目		A	B	C		
		3分	2分	1分		
体质监测前的准备	是否能正确理解监测工作的意义和作用?					
	参加体质监测的目标和动机是否明确正确?					
	在参加体质测试过程中是否能够尽自己的努力完成整个测试?					
体质监测参与过程	你的体质健康自测中采用的测量方法是否正确?					
	根据测量结果,你使用的评价标准是否正确?					
	在评价过程中与他人配合是否默契?					
监测评价后的应用过程	通过分析评价结果,你是否掌握了自己的健康状况?					
	此次评价后,你是否能够对自己的健康状况再次测评?					
	根据评价结果,你是否能够制定一份符合自己的健身计划?					
	你是否能够在测试后坚持体育锻炼?					

3.5 价值判断与反馈

价值判断的过程是对教育活动满足社会与个体需要的程度做出判断的过程^[14],它是实现教育决策科学化的重要环节。美国学者格朗兰德(Gronlund, N. E., 1971)曾用一个简单的公式描述了教育评价的特点:评价=测量(量的记录)或非测量(质的记录)+价值判断^[15]。因此,学生体质健康监测的评价公式=体质测量+真实性评价+价值判断,可见,体质测量与真实性评价后开展价值判断是不可缺少的一个环节。最后,再将价值判断的结果向监测组织或学生反馈,即完成了学生体质健康监测真实性评价的整个过程。由此来看,真实性评价贯穿于体质健康监测、评价和反馈的每个环节中,是三者联结的锁链。下面的这种闭合式的循环设计使评价不断趋向成熟和完善(图1)。

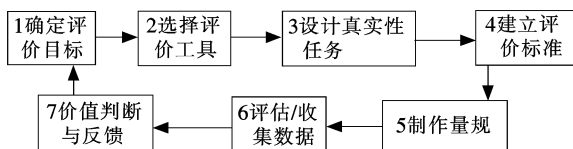


图1 学生体质健康监测真实性评价流程图

4 评价学生体质健康监测效果的几种典型设计

效果评价是监测活动整体构建中的一个重要部分^[16]。科学的评价设计能从总体上对监测过程进行有效控制,优化监测活动管理,提高监测的实效性。本文提出的几种典型的设计方案是按照体质健康监测开展的时间顺序,将真

实性评价方法应用其中,通过设计真实性评价在监测活动中实施的时机和实验组与对照组的比较,计量出监测效果的定量值或是定性分析结果,实现对体质健康监测效果的检验。

4.1 单次评价设计

单次监测活动评价设计是指在体质健康监测开始后,将真实性评价方法应用于监测活动整个过程中,在监测活动结束后进行一次性的评价数据的收集(图2)。这种方法的优点在于简单易行,但是由于它在监测之前没有收集任何资料,许多难以控制的因素会影响评价的结果或效度,其缺点也很突出。

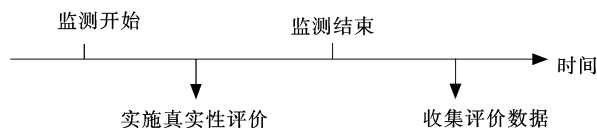


图2 单次评价设计

4.2 同一小组实施监测前后对比的评价设计

在监测活动实施的前后各进行一次真实性评价数据的收集,两者的差距即为监测活动实施的效果。这种设计要比单次的监测活动设计优化了一步,将监测实施前与监测实施后的评价指标进行比较,就可以确定改进的情况。这种方法较为科学,操作也简单,但如果监测活动开展的时间较长,监测过程中没有对相关因素进行控制的话,不能完全保证监测效果或是监测中的某一些改进是通过实施监测活动得来的

(图3)。

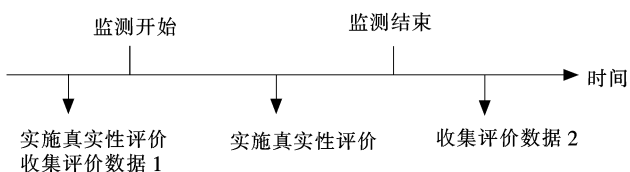


图3 实施监测活动前后对比的评价设计

4.3 同一小组多重评价设计

多重评价设计就是在监测前后各进行多次的评价数据收集(图4)。这是一种较为精确的评价设计。实验组是自己的对照组。在监测之前实施多重监测会避免因不用对照组而产生的问题。而事后采取重复性评价,不仅可以对原始结果进行比较,也可以对监测项目的长期效果进行评价。但是该评价设计需要投入大量的时间和费用,需要专人专门研究,实现起来比较困难。

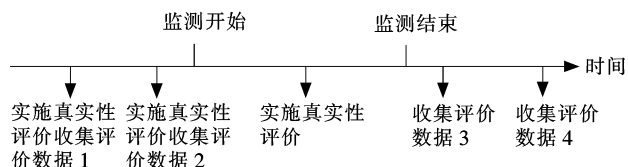


图4 同一小组多重评价设计

4.4 有对照组的评价设计

监测活动实施前从同一学生群体中随机选取两组,进行真实性评价数据的收集,然后一组实施体质健康监测,称为实验组,另一组学生不进行体质健康监测,为对照组,实施监测活动结束后两组再次进行真实性评价数据的收集,与对照组比较,实验组的结果表明了监测活动的实施效果。对照组的另一个功能是排查监测活动的实施过程中影响监测效果的一些不利因素。这种设计是目前较有效的评价设计之一,但是两组能否做到随机抽样是该方案实施的关键,在实践中如果采用这种设计方案,则需要详细说明随机抽样的办法或操作过程。

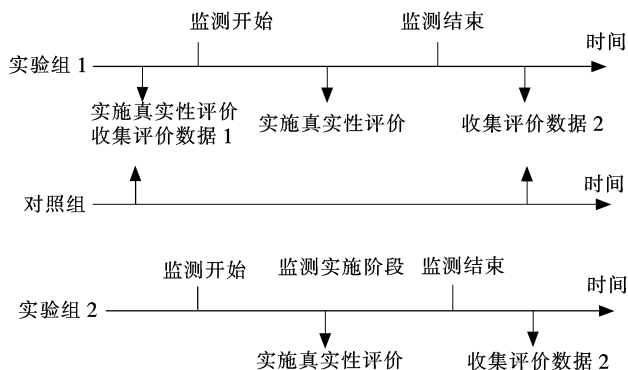


图5 有对照组的多重评价设计

另外,有对照组的最理想的设计是在上面的设计方案中

再增加一个实验组2。与实验组1比较,实验组2只是没有前测,用于排除前测对监测结果的影响。这种设计把监测以外的干扰因素尽可能地消除,在实施设计中,这种设计是最好的。然而,从实际来看,要得到三个随机选择的抽样小组有一些困难,另外,这种操作所用的时间长、费用高等缺点也较突出。

总之,不同的设计方案所用的人力、物力、财力、消耗的时间不同,能够获得的信息量也不同,因此,在实践运用中要根据具体情况对这些设计方案做恰当的选择。

5 结论与建议

我国学生体质健康监测是一项工作,更是一项教育活动。从真实性评价的理论和方法分析,学生体质健康监测具备了监测任务的真实性、监测内容的真实性和评价方式的真实性等特征,因此,学生体质健康监测应用真实性评价方法是可行的,也是现实所需的。

学生体质健康监测实施真实性评价有五个基本步骤:确定评价目标—设计真实性任务—建立评价标准—制定量规—价值判断与反馈。由这五个步骤构成的闭合式的循环结构,保证了评价过程的逐步成熟和完善。

本文根据真实性评价在体质健康监测中的应用时机设计了四种效果评价方案,并分析了每个方案的优缺点。建议在我国学生体质健康监测的评价中采用真实性评价的方法,这是对提升我国学生体质健康监测效果的有益尝试。

参考文献:

- [1] 冯建新. 现代教育评价与测量学[M]. 北京:中国社会科学出版社,2005.
- [2] Wiggins G. A true test: toward more authentic and equitable assessment [J]. Phi Delta Kappan, 1989 (20): 703-713.
- [3] 威金斯. 教育性评价[M]. 北京:中国轻工业出版社,2005.
- [4] Resnick L B. Learning in school and out[J]. Educational Leadership, 1987, 16(9): 13-20.
- [5] Savery J, Duffy T. Problem based learning: an instructional model and its constructivist framework[J]. Educational Technology, 1995(35): 31-38.
- [6] Darling H, Snyder J. Authentic assessment in teaching in context[J]. Teaching and Teacher Education, 2000 (16): 523-545.
- [7] Kathleen M. Authentic Assessment: A guide for elementary teachers[M]. Essex: Addison Wesley Longman, Inc, 2001.
- [8] Judith T M, Theo J B, Paul A K. 真实评价设计的五维框架[J]. 过程教育杂志, 2009(4): 62-66.
- [9] 组媛媛. 真实性学生评价研究[D]. 上海:华东师范大学, 2007.
- [10] 梁恒. 综合实践活动中真实性评价的探索与应用[J]. 基础教育课程, 2014(3): 52-56.

- [11] 孟庆光,胡国鹏.大学生体质健康动态分析[J].河北体育学院学报,2012(1):86-89.
- [12] 张继玺.真实性评价:理论与实践[J].教育发展研究,2007(01B):23-27.
- [13] 张向众,李永琚.真实性评价中的评分规则[J].外国教育研究,2004(12):35-38.
- [14] 王笃勤.真实性评价:从理论到实践[M].北京:外语教学与研究出版社,2007.
- [15] Gronlund N E. Measurement and evaluation in teaching[M]. New York: Macmillan Publishing Co., Inc. and Collier,1971.
- [16] 唐晓义,甄志平,刘铁军.我国研究生体质测试现状与分析[J].河北体育学院学报,2012(1):73-75.

Authentic Evaluation Design of National Student Physical Fitness Test in China

YU Hong-yan¹, ZHANG Ya-ping², CHANG Dong-qing³

(1. Department of Physical Education, Shanghai JiaoTong University, Shanghai 200030, China;

2. Kaplan Higher Education Academy of Singapore, Shanghai 200050, China;

3. Aquatics Center, Shanghai Sports Bureau, Shanghai 310000, China)

Abstract: From the perspective of educational evaluation, the National Students' Physical Fitness Test (NSPFT) is task, and more importantly an educational activity. Currently, neglect of its educational function leads to a series of problems in the evaluation of the testing effectiveness, such as, the evaluation subject is single; the evaluation items are not rich; and the testing tools are not diversified. This paper applies methods of literature analysis and logical induction to research how to carry out NSPFT and its authentic assessment based on authentic assessment theory. It finds out that our NSPFT appears some features with the authentic tasks, the authentic contents and authentic evaluation methods. On this basis, it proposes five steps for the implementation of NSPFT authentic evaluation and four kinds of typical NSPFT evaluation design proposal, aiming to expand NSPFT'S educational function and to provide methodological support for the promotion of the evaluation effectiveness.

Key words: National Student's Physical Fitness Test; authentic assessment; test effectiveness; evaluation design