

体育素养测量与评价的现状、挑战及未来

陈思同¹, 刘阳^{1, 2}, 唐炎^{1, 2}, 蔡玉军¹, 陈佩杰^{2, 3}

(1.上海体育学院 体育教育训练学院, 上海 200438; 2.上海市学生体质健康研究中心, 上海 200438;
3.上海体育学院 运动科学学院, 上海 200438)

摘 要: 系统梳理体育素养测评领域的关键问题并提出对策, 为完善该领域的方法学理论和实践应用提供借鉴。研究认为, 当前体育素养的测评存在以下主要问题: (1)概念不清晰, 导致无法确定测评的维度和内容; (2)指标的科学性存在争议, 无法提供可靠和准确的测评结果; (3)体系构建存在方法学的不规范; (4)体系应用仅停留在个人层面, 无法有效推广到大规模人群层面。研究建议, 进一步对体育素养概念进行辨析, 确定测评的维度和内容; 根据确定的测评维度和内容优先考虑成熟的测评工具构建测评体系; 采用主观和客观的方式全面选取专家, 评估测评模型; 对于体系的应用要进行大规模的人群试验, 以验证体系的可适性; 放弃传统的测评方式, 采用定量和定性测评相结合的方式, 研制以动态测评为主要方式的个性化测评体系。

关键词: 学校体育; 体育素养; 测量与评价; 指标体系

中图分类号: G807 **文献标志码:** A **文章编号:** 1006-7116(2019)05-0110-08

Current situation, challenges and future of physical literacy measurement and evaluation

CHEN Si-tong¹, LIU Yang^{1, 2}, TANG Yan^{1, 2}, CAI Yu-jun¹, CHEN Pei-jie^{2, 3}

(1.School of Physical Education and Sport Training, Shanghai University of Sport, Shanghai 200438, China;
2.Shanghai Research Centre for Physical Fitness and Health of Children and Youth, Shanghai 200438, China;
3.School of Kinesiology, Shanghai University of Sport, Shanghai 200438, China)

Abstract: The authors systematically collated key problems in the physical literacy measurement and evaluation area and put forward suggestions and countermeasures, so as to provide reference for perfecting methodological theories and practical application in this area, and concluded that current physical literacy measurement and evaluation had the following main problems: (1) unclear concepts, resulted the failure to determine measurement and evaluation dimensions and contents; (2) there was controversy about index scientificness, the indexes were unable to provide reliable and accurate measurement and evaluation results; (3) system construction was not standardized in terms of methodology; (4) system application stayed at the individual level, could not be effectively generalized to a large-scale population level. The authors put forward the following suggestions: further clarify the concept of physical literacy, determine measurement and evaluation dimensions and contents; preferably consider using a mature measurement and evaluation tool to build a measurement and evaluation system according to the determined measurement and evaluation dimensions and contents; in a subjective and objective fashion, comprehensively select experts to assess the measurement and evaluation model; large-scale population test should be done for system application, so as to verify the feasibility of system application. The authors also suggested giving up the traditional way of measurement and evaluation, and adopting the quantitative and qualitative measurement and evaluation combined way to develop an individualized measurement and evaluation system that uses dynamic measurement and evaluation as the main way.

收稿日期: 2019-03-04

基金项目: 教育部综合改革重大课题研究项目(17JGWT00013); 上海市哲学社会科学规划教育学一般项目(A1904); 上海市体育科技“雏鹰计划”项目(19C005); 上海市人类运动能力开发与保障重点实验室项目(11DZ2261100)。

作者简介: 陈思同(1991-), 男, 硕士, 研究方向: 学校体育、身体活动和体育测量。E-mail: chenst1212@163.com 通讯作者: 刘阳

Key words: school physical education; physical literacy; measurement and evaluation; index system

体育教育作为教育活动的重要形式之一,对全人发展的价值已经得到广泛认可^[1]。近年来,体育教育理论得到重要发展。其中,最为重要的就是体育素养概念(Physical Literacy, PL)的提出(注:有关PL的翻译,有学者认为应当译为“身体素养”,但是学界尚无统一表述,因此本研究暂将PL译为体育素养),这一概念极有可能对未来体育教育带来变革^[2]。国外体育素养概念创始人Whitehead将体育素养定义为“个体在一生中维持适当水平身体活动所需的动机、信心、身体能力、知识和理解”^[3]。随后,Whitehead更新体育素养定义,即个体在一生中参加身体活动的动机、信心、身体能力、需要重视的知识和理解、责任。20世纪初,我国素质教育热潮也引发体育学者“本土”体育素养概念的提出和思辨,以赖天德教授为代表的学者提出了不同理解。随后,陈思同等^[4]对国外体育素养概念进行了分析,结合我国已有研究提出了新的解释——人类在生命过程中获得的利于全人生存发展的运动要素的综合能力。尽管存在研究逻辑、语言和文化的差异,但是国内外学者都从多维度视角将体育素养理解为维持身体活动或运动的一种综合能力。围绕体育素养这一概念,国内外学者在概念辨析^[4-5]、测评工具研制^[6-7]和实践推广^[8-9]等方面进一步展开大量研究。以这些成果为基础,美国、加拿大(部分地区)进行了新一轮的体育教育改革,将“培养具备体育素养的人”(Being Physically Literate)设为体育教育的总目标。与此同时,我国国务院办公厅也于2016年下发《强化学校体育促进学生身心健康全面发展的意见》(以下简称27号文件),将“全面提高学生体育素养”作为强化学校体育的基本原则,这标志着体育素养这一概念已经逐步从学术研究领域拓展到社会教育实践层面。

在实际教育情境下,作为体育教育的目标和学校体育的重要原则,体育素养测量与评价的重要性不言而喻。国内外学者对于体育素养的测评研究已经有近20年的历史并取得一定成果,推动了该领域的发展。以加拿大学者为代表,Tremblay^[10]和Longmuir^[11]带领的科研团队历经近10年的时间开发出加拿大体育素养测评工具(Canadian assessment of Physical Literacy, CAPL),并于2018年完成改良。此外,Lodewyk^[12]和Cairney等^[13]也相继开发适合不同年龄儿童青少年的体育素养测评工具。美国和英国一些教育组织或协会也开发了针对儿童青少年的体育素养测评工具,并在实际中得到应用^[14]。我国学者对体育素养的测评也有多年研究历史。2002年,陈琦^[15]对学生体育素养的测评

进行了初步探索,并构建了7个指标的体系。随后,于秀等^[7]带领的科研团队对学生体育素养的测评工具进行研究,开发了相对完整的测评工具并对我国部分城市的中小學生进行了调查。近年来,一些国内研究者对国外研究进行介绍,主要集中于对加拿大研究者开发的测评工具的介绍和分析。比如,陈思同和刘阳^[6]介绍并分析了CAPL的特点,赵雅萍等^[16-17]先后报道加拿大青少年体育素养测评体系(Passport for Life, P4L)和加拿大体育素养测评体系(Physical Literacy Assessment for Youth, PLAY)。上述3种已在国内公开报道的测评工具为今后研究提供了研制思路和方法范式,但这些测评体系仍存在一些问题,并引发学者的探讨^[6]。比如,测评体系的测量可靠程度、反映概念内涵的全面性以及实际应用的可适性。然而,有关当前体育素养测评的问题和不足国内研究鲜有报道,且大多数研究局限于儿童青少年人群,未拓展到其他年龄段人群,甚至是特殊人群。

尽管体育素养测评面临一些问题和挑战,但从教育需要和政策执行的角度来说,体育素养的测评至关重要。首先,测评体育素养是全面评价体育教育成效的重要方式,这不仅反映了体育教育工作的效果,也反映了学生体育学习的效果及质量。其次,体育素养的测评更是调整和改进未来体育教育工作的实证基础。此外,体育素养作为一个综合性概念,评价方式突破了过去体育教育采用单一评价方式(如单以体质健康测试为评价方式)的弊端,为教师、学生、家长和政策制定者提供了全面的体育教育评估措施。在当前我国教育综合改革的背景下,研究体育素养的测评有利于高考综合改革的实施和推进,为建立综合素质评价制度提供可靠的评估工具。最后,从长远角度来看体育素养作为制定公共健康政策的依据势在必行,因为其测评结果为政策的“有效制定”和“精准落实”提供了重要的决策依据。因此研究有必要解决当前体育素养测评中的问题,开发出科学的测评工具。综上,体育素养测评研究领域尚处于“萌芽阶段”^[18],仍有许多问题需要回答。有鉴于此,本研究旨在系统梳理国内外体育素养测评研究的基本现状和关键问题,并提出建议和对策,以进一步完善体育素养测评的方法学基础,指导体育素养测评的实践和应用,以期为未来研究提供启示。

1 体育素养测评的现状

到目前为止,世界上关于体育素养测评研究仍处

于起步阶段^[14, 18], 但存在很多争议问题。在可考证的学术文献和公开资料(测评手册、操作协议等)中, 加拿大研究人员是该领域的先锋。我国部分学者已经对加拿大开发的 3 种体育素养测评工具进行了介绍, 分别为 CAPL^[6]、P4L^[16]和 PLAY^[17]。CAPL 由加拿大学者 Tremblay 和 Longmuir 及其团队开发, 针对 8~12 岁儿童青少年。基于常模参照和标准参照评价标准, CAPL 能够给每位测试者提供全面评价。尽管如此, 关于 CAPL 的质疑仍然值得关注, 如其测评时间成本较高以及测量准确性较低。P4L 是针对幼儿至 12 年级青少年的测评工具, 每个年龄段的测评指标和方式存在不同。作为一个过程性评价工具, P4L 为体育素养的发展提供了很好的参考标准。但是, 关于 P4L 的争议, 比如非综合性测评、可靠性不足也需要引起研究者重视。PLAY 系统是一套针对青少年的体育素养测评工具, 可以从不同视角对受测对象进行评估。其测评结构由 PLAY-fun、PLAY-basic、PLAY-self、PLAY-parent、PLAY-coach 和 PLAY-inventory 构成, 具有关注变化趋势、形成“立体环绕式”评估的优势。但是, 针对 PLAY 系统不足的争论也一直存在, 比如测量的可信度、项目过于繁琐、分数不易理解。

目前美国没有公开的体育素养测评工具, 但是美国学者正积极推广体育素养理念。虽然暂无专门体育素养测评工具, 但是一些研究学者利用现有测评工具作为“代替办法”测评儿童青少年的体育素养。比如, PE Metrics(一种体育教育评价综合性工具)就被采用^[19], 作为评估儿童青少年体育素养的工具。但是, 这种测评无法反映儿童青少年的体育素养水平。与美国类似, 英国也未开发综合的体育素养测评工具。部分体育和教育协会, 比如英格兰青少年体育基金会(Youth Sport Trust)、体育威尔士(Sport Wales)、北爱尔兰年轻人体育(Young People & Sport)和体育苏格兰(Sports Scotland)的计划——运动吧! 智者(Better Movers and Thinkers)都在 Whitehead 的学术理念影响下开展了一系列测评研究, 但效果并不理想。这些计划中, 有些侧重于运动技能, 但忽视了心理和动机维度; 有些侧重于运动能力, 但无法体现体育素养概念的全面性。

近年来, 澳大利亚学者对体育素养的相关研究也得到了学界关注。澳大利亚体育教育专家 Keegan 领衔的团队围绕体育素养的概念辨析、测评工具、教育标准和政策制定开展了深入研究^[20]。但是, 澳大利亚学者也未能开发出一个相对合适的体育素养测评工具。他们认为当前体育素养测评研究总体处于无法反映概念本质的模糊状态, 这可能与概念本身还未辨析清楚有关。

陈琦^[15]是较早涉及该领域的中国学者, 他于 2002 年就学生体育素养测评进行初步探索, 并提出 7 个指标的体系。但是, 该测评指标体系是否具有可行性尚需论证。于秀等^[7]对学生体育素养测评工具进行一些研究, 并开发相对完整的测评工具, 对我国部分城市的中小学生进行了调查。然而, 在这一测评体系中部分指标的测量难度较大, 如体育道德, 因为至今仍未有公认的行为学或心理学测量工具; 该测评工具并未研制出体育素养水平的总评指标得分, 无法反映学生体育素养的总体水平。此外, 个别研究就大学生体育素养测评进行了分析, 但过于简单, 类似于“拼凑”测评工具。随着研究的深入发展, 一些学者开始介绍国外研究^[6, 16-17], 这也一定程度推动了我国相关研究的进步。但是, 针对中国人群较为认可的体育素养测评工具仍然处于“空白”状态。

尽管过去我国的研究存在局限, 但目前少数关于体育素养测评工具的研究已经取得了一定突破。比如, 上海体育学院牵头进行的学生体育素养测评体系研究在国内外相关专家的支持下已经初步完成了体系构建, 该测评体系在综合国内外相关研究的基础上对研制方法进行了改进, 一定程度能够减少过去国内相关研究的局限。

2 体育素养测评的挑战

虽然一些国家的学者对体育素养测评进行了较为深入的探讨^[14, 18, 21], 但是当前体育素养测评研究的局限性和争议性仍是学界面临的挑战。受制于过去研究的视角, 我国并无系统分析这些挑战的相关研究, 以下从概念的差异、指标的科学性、体系的构建、体系的应用以及其他问题 5 个关键方面进行详细分析。

2.1 概念的差异

现代实证科学的逻辑起点就是明确概念, 提出清晰的理论定义, 进而对理论定义进行操作化, 形成操作定义, 最后发展测评工具。因此, 测量体育素养的首要前提则是明确概念的理论和操作定义。操作化是科学研究中必不可缺的一个环节。基于经验观察, 通过操作化过程, 将抽象概念转化成一套可以观察或者测量的指标体系或工具(如具体的测量问卷、量表)。事实上, 形成合适的观察或者测量指标体系或工具是开发体育素养测评工具的最终目标。因此, 有关体育素养理论定义和操作定义的关键问题, 就是最终形成的测评体系中应该包含哪些内容。

首先, 在概念的理解上国内外学者都提出了多种且不同的辨析^[4-5], 这导致了体育素养的概念定义存在定义不清晰和构成元素不同的情况, 进而导致不同的

测评维度和内容。举例来说,如采用 Higgs^[22]的“在不同身体活动、律动运动和竞技运动情况下,儿童自信、可控制地参与活动的基本运动能力和运动技能”的定义,按照此思路体育素养概念的构成元素就是环境(对应不同情况)、心理情感(对应自信)、身体能力和运动技能(对应基本运动能力和运动技能)。因此,上述内容应该作为测评的具体内容。事实上,尽管不同学者对体育素养概念的理解存在差异,但是大部分学者都认同体育素养是一个综合性概念^[3-5],包含行为、情感、认知和效果等多个维度,而并非一个仅仅涉及身体的概念。因此, Higgs^[22]的概念缺少认知这一构成元素。再如,采用 Whitehead^[3]的概念,就应该包括日常行为、身体能力、知识理解和情感动机,这在 CAPL 中得到体现。又如,在 P4L 中生存技能作为一个重要方面被纳入测评。关于“生存技能”这一部分是否是体育素养的维度,理论层面上一直没有得到很好解答。因为从概念上分析,生存技能是比体育素养更为广泛的上位概念。因此,概念不清晰和差异造成了当前研究无法明确概念的构成元素,导致研究者无法明确哪些元素需要测评,进而对未来的概念操作化造成困难。由此可见,不同概念定义对于测评维度和内容造成了差异。那么,哪些构成元素是体育素养概念的构成,进而被纳入测评体系进行操作化呢?当前研究似乎没有一致认同。

其次,关于体育素养概念的定义都存在不同程度的抽象性^[5]。尽管一些定性测量遵从体育素养概念的理论基础——身心一统的一元论^[23],但是在具体测量时如何反映这个概念的整体性仍存在问题。以 Whitehead 的定义为分析样本(国外大部分学者都参照 Whitehead 的定义),国外体育素养概念本身就是根据一元论、存在主义和现象学等抽象的哲学理论构建的,这使得体育素养概念的特性和内涵存在不同程度的抽象性。在 Whitehead^[3]的概念中提及了“个体阅读环境”和“与外界环境进行交互式互动”类似的特性,试问如何理解“阅读环境”和“与外界环境交互”,这些构成元素如何被定义、如何被操作化等需要进一步阐明。赖天德^[24]提出的“体育的文化水平”,冯古首等^[25]提出的“人的各种体育精神要素及其品质的综合”,余智^[26]提出的“综合体育素质与修养”,杨献南等^[27]提出“对身体和精神文化内在追求的综合性文化塑造”,这些包含“文化”“精神”“修养”和“内在追求”字眼的体育素养概念内涵特性,抽象甚至“形而上学”,那么这些概念构成可以被定义么?进而被操作化进行测评么?再如,我国一项关于城市学生体育素养研究提出了包含“体育道德”这一指标在内的测评体系^[7]。然而,到目

前为止有关儿童青少年体育道德的测量与评价,由于概念不明确、工具不完善等问题仍有很多争议^[28]。因此,在测评个体的体育素养过程中,测量个体的体育道德是否合理或者能够被测评仍需要探讨。

美国著名体育学研究者 Corbin^[29]提出,应该就体育素养的各个构成元素开发特定具体的测评工具。然而,Whitehead 提出的概念指出体育素养概念的构成元素是彼此交互的,且随着环境和文化背景而改变。也就是说,在明确概念构成元素基础上构成元素之间的关系也需要测评。如果采用 Corbin 的理念,“拼凑”一套测评体系无法体现构成元素之间的关系,似乎有“1+1 < 2”之嫌,更无法体现概念的整体哲学思想。因此,当前测评工具能否完全反映体育素养概念的整体抽象性及其构成元素的交互关系,学界存在很大争议。

2.2 指标的科学性

即便已经确定体育素养的测评维度和内容,其测评指标都有具体操作化条件,而体育素养的测评指标仍有一些问题需要考虑。按照评估理论,测评指标需要从可行性、信度、效度和敏感度4个方面进行评估。Longmuir^[18]指出当前研究中,体育素养概念的部分维度和内容还未有较好的测量方式,比如信心、动机和知识在目前测量领域内还无法被直接测量。再如,测量个体身体活动参与的责任和价值判断也无可行工具。也就是说,目前这些维度和内容的测评似乎并没有一套可行的方式。此外,Whitehead 指出体育素养测评应该贯穿个体一生并反映“体育素养的历程”,同时放弃传统的常模参照和标准参照,但是目前的研究未能在这方面有所突破,也未能开发出可行的测评工具。

在测评工具的信度和效度上,一些问题仍需要研究者重视。关于体育素养的体质水平维度,我国一些学者采用《国家学生体质健康标准》(或者该标准的2014修订版,以下统一简称《体质测试标准》)作为测评工具。《体质测试标准》作为国家推广的测评体系,能够保证测量工具的统一性,具有高度的可行性,但是这一标准也存在争议。以心肺耐力测评为例,《体质测试标准》采用男生1000 m和女生800 m作为具体测评方法。心肺耐力是指在锻炼过程中向肌肉运用氧气并利用氧气产生能量以支持肌肉活动的的能力,所体现的重要特性就是个体在长时间过程中肌肉继续工作的能力,而男生1000 m和女生800 m的测评不符合上述概念。因为,个体1000 m和800 m的测评以时间作为评价基准,无法体现个体的长时间做功能力,导致测评效度存在争议。在信度方面,Edwards 等^[21]指出定性测量的效果受到测评人员经验的制约,倘若经验不足的测试人员,那么测量结果的可信度更难

以保证。因此,采用定性工具测量对体育素养进行评估仍需持谨慎态度。

在测评的敏感度上大部分研究还未涉及,仅有 CAPL-2(CAPL 第二版)对这些问题进行了研究^[30]。测评敏感度可以分为测评模型理论和实际敏感度。早先,CAPL-1 分为日常行为、身体能力、动机和信心以及知识和理解 4 个维度,共计 25 个指标。但是,后续研究发现部分指标对体育素养测评的理论模型并无太大贡献,进一步导致部分指标在反映其所在维度时存在敏感度较低的问题。因此,研究人员对 CAPL 进行了新一轮的因子分析,形成了新的 CAPL-2^[31]。此外,测评工具敏感度在实际应用中具体表现为测评能否有效区分个体在某方面的差异。比如,在体育素养的知识测评中测评工具能否反映个体间在知识水平上的细微差异。倘若测评工具无法反映个体在某方面存在的明显差异,那么这样的工具在测评敏感度上便存在明显缺陷,无法进行准确测评。因此,当前关于体育素养测评的研究在这些方面需要改进,以进一步检视测评工具的科学性。

2.3 体系的构建

在当前体育学研究中,德尔菲法是常用的构建完善测评体系的科学方法。比如,CAPL 就采用德尔菲法构建其测评的理论模型^[1]。CAPL 研究团队选取来自 6 个领域(儿童青少年身体活动、健康体适能、身体活动和体适能测评、运动技能发展、认知和残疾人适应身体活动)专家作为征询对象,对体育素养测评体系的构建进行评判。从整体上来看,6 个领域专家基本能够涵盖体育素养概念的内涵,保证测量工具的完整性。但是 CAPL 研究并未报告测评理论模型的稳定性^[1],所以无法得知专家对理论模型的贡献大小以及模型质量。我国也有学者采用德尔菲法选择专家构建学生体育素养的测评体系,但是一些问题依旧需要商榷。比如,一项研究选择教育学、心理学和体育学领域的专家构建测评体系^[7]。选择这 3 个方面的专家固然正确,但是在当今科学研究精细化的背景下,有可能出现上述 3 个领域的专家显得过于宽泛,因此对体育素养概念会出现不了解的情况。比如,研究运动生物力学的专家很有可能不太了解体育素养概念,即便了解也不能保证其达到研究需要的水平。不可否认,构建测评工具的研究者也会认可上述假设并实施了相应的问题规避。倘若研究者能更加精细化地区分专家的研究领域,选择与体育素养内涵最为相关的高水平专家并就理论模型的稳定系数予以适当报告,不仅有利于提高所构建的测评体系的质量,也有利于未来研究的进一步改进。就专家的选择方式而言,如何选择高水平专

家也关系到构建测评体系的质量。理论上讲,一方面主观选择专家容易造成一定的“臆断”,极有可能选择的专家不了解体育素养概念及其内涵;另一方面主观选择专家可能会遗漏体育素养领域及其相关领域的专家,无形中降低了专家库的“质量”,影响所构建体系的质量。

除了内容不同,构建体育素养测量方法也存在不规范的地方。陈思同等^[6]已经对我国一些体育素养测量工具提出了意见和建议,发现一些测评工具在研制过程中未采用严格的方法学程序,因此所提出的体育素养测量的构成内容有待论证。在国外研究中,除了一些现有的体育素养测量内容与当前公认的体育素养构成内容的一致性较低外,方法学上的问题依然存在。比如,PLAY 和 P4L 的研制方法就未能像 CAPL 一样,公开具体的方法学步骤。因此,学界无法判断这些未公开方法学步骤的体育素养测评构成内容的科学性。因此,考虑到体育素养内涵的广泛性,选择哪些领域的专家能够较好涵盖概念的内涵范围,通过什么样的方法选择专家才能保证专家质量的客观性,同时以上述问题为基础衍生出的新问题,如“根据专家的研究背景如何请专家进行判定,如何根据专家意见整合测评体系的测量指标,如何根据专家意见进行测评体系的修改等问题”,都是未来研究者需要考虑的。除与德尔菲法相关的技术问题,更多问题亟待方法学知识提供解决方案。

2.4 体系的应用

一般来说,能够在实践中服务科研目的的测量工具必须具有较强的可操作性。遗憾的是,现有体育素养测评工具的可操作性并未得到研究者重视。

在 CAPL-2 中,研究者对原有知识和理解部分的测评问卷进行修订,认为该问卷的可操作性较强^[32]。而其他研究并未报告测评工具的可操作性,而测评工具的可操作性需要在实践中予以确认。诸如测评体系的时间、空间和实施经济成本,这些有可能影响测评工具的应用以及测评质量。在时间上,测评时间太长可能会造成测试对象的消极情绪,进而“抵触”测评过程,降低测评的有效性;在空间上,开展测评的空间过大可能会导致无法有效监控测评过程,潜在威胁测评质量;在经济成本上,所需花费过大可能会提高整个研究项目的难度,影响后续研究的开展,最终造成研究项目的“流产”。以 CAPL 为具体案例,在配备 5 名测试者的情况下完成对每 20~25 位个体的测评(除行为部分中的身体活动测评)共计花费 1.5 小时。实际上,这还不包括行为部分中的身体活动测评,因为该指标采用的是“在连续 7 天中对个体的步行数(Walk

Steps)进行测评”。可以看到,CAPL的测评时间花费较长,那么这些潜在的问题是否会影响测评。此外,现有测评工具的可操作性是以个人的测评时间为评估手段,假设将这些测评工具应用到大规模的人群测量中,比如类似我国学生体质健康测试,还会有哪些问题呢?现有研究尚无法给出明确答案,这也提示未来研究者需要对测评工具的可操作性进行分析,尤其是在特定的不同环境和不同人群中。

2.5 其他问题

除上述问题以外,体育素养测评仍存在一些其他不容忽视的问题尚未得到较好回答。如当前研究主要针对12岁以下人群^[21],可能这一阶段是发展体育素养的关键时期,而针对幼儿人群的测评鲜有报道。此外,对于老年人群的测评研究还处于理论模型构建阶段^[9]。这也说明针对不同年龄段人群的测评研究之间存在许多“真空状态”。按照Whitehead的观点,体育素养是贯穿全生命周期的概念,应当对所有人群的体育素养测评予以足够重视。因此,未来应该加强对不同年龄段人群的测评研究。

Whitehead将体育素养的发展阶段分为学龄前、儿童、青少年、青年人、中年人和老年人6个阶段,而健康个体生物发展轨迹基本遵循这6个阶段的变化。但是有必要指出的是,每个人在不同阶段的身体、心理和所处环境等特点都存在不同,同时体育素养概念又表明身体、心理和环境等这些属性存在交互关系,那么标准化的测评工具尤其是定量测量工具是否能有效地对个体的体育素养进行测评。此外,在独立的体育素养发展阶段,比如儿童阶段(一般指5~12岁),其身体和心理特点随着年龄的变化也呈现显著变化,那么标准化工具是否能准确反映出这些变化也值得商榷。同时,当前主要在体育教育环境还是身体活动环境下的测量也存在争议^[29, 31]。而测量不同环境下不同活动的心理学特性,采用同样的工具是否合适也尚无定论。在测评体育素养时,一个特性随时间的“变异性”也需要研究者予以考虑。比如,短时间内知识水平随着时间的变化不会产生太多变动,而动机和信心部分随着时间以及环境的影响会产生显著变化。也就是说,个体的一些体育素养特性是“短暂的”,其状态会发生明显变化。当前传统的测评工具采用标准化方式,几乎无法观察个体一些特性的变化。因此,在测评体育素养时这些问题也需要慎重考虑。

体育素养的一些维度已经有成熟的测评工具。比如体能部分,美国FitnessGram针对美国儿童青少年体能水平开发的标准参照测评工具,儿童青少年能够通过此工具了解自己的体适能是否属于健康水平。但

是,体育素养的总体评价采用这样的方式是否适合还不得而知。尽管CAPL采用这样的方式,但是按照Whitehead的观点,个体的体育素养不应该采用比较方式进行评估^[4],因此传统的常模参照和标准参照似乎不适合评价个体的体育素养水平,那么对体育素养的评估是否能按照以往经验采用常模参照或标准参照还不得而知。Corbin^[29]提出,体育素养的测评可以采用自评、个性化评价和人群评价,但是他的观点仅仅处于构想阶段,要落实到具体操作还需要更多研究。在这3种评价类型中,人群评价似乎有悖于体育素养个性化的理念。因此,体育素养是否需要进行人群评价值得探讨。还有一个值得注意的地方,一些测评工具将不同维度进行了组合从而进行评价。这种组合方式看似是符合体育素养概念的定义,但是如何解释这一组合概念的分数呢?Corbin^[29]对此提出质疑。在实际的现场测评中,测评体育素养的成本也是一个重要因素。此外,采用过程性评价还是总结性评价又是另一个值得思考的问题。现有的一些测评工具仅能有效测评体育素养概念的部分,比如体能和技能。

3 体育素养测评的未来

第一,在确定体育素养概念的定义上优先考虑综合性概念。Edwards等^[21]研究表明,当前大部分研究都采用Whitehead的定义。因此,建议未来研究可以采用她的定义作为测评的起点,确定哪些构成元素纳入测评范围,这也有利于未来研究进行横向比较和分析。由于Whitehead的概念扎根西方文化不一定适合我国研究,因此选择一个兼具“国际性”和“民族性”的概念则是一个可取的办法,陈思同等^[4]提出的概念定义则为这种尝试提供了依据。

第二,在确定概念后对于概念的操作化,优先考虑可以被操作化的构成元素,且这些构成元素已经具备成熟的测评工具。当然,这些暂时无法测评的构成元素,比如体育道德、构成元素之间的交互性以及概念的整体哲学性等,随着研究的发展未来一定能得到令人信服的回答。

第三,在体系构建方面,所选择的指标应该符合体育素养概念构成元素的范畴,确定初步的体育素养测评模型,所选择的评判专家应该来自更为广泛的领域。因此,专家的研究背景是否符合体系构建的需求需要进一步斟酌。以现有的研究技术,人工选择和工具辅助相结合的方式则是可行的办法。比如,研制开始前研究者可以依据经验选定一些专家,作为后备人选,然后,采用CiteSpace等工具进行再次遴选,根据该软件提供的数据辅以实证依据的参考,以提高专家

选择的合理性。

第四,有关体系应用上,应重视对测评体系应用效果的评估。体育素养测评体系的应用是一个动态改良过程,需要结合环境和人群特点以及测评效果进行调整和优化,以求用最小成本获得最可靠准确的测评结果。

除以上之外,一些学者提出新测评思路有助于解决本研究无法回答的问题。由于传统方法的局限,采用简单、线性的方法无法有效测评体育素养概念,因此,Green等^[14]、Longmuir^[18]建议采用图示法代替传统的测量和评价方法。理由如下:

体育素养是个人且无法与他人进行比较的特性。Whitehead对传统测评带来的“横向比较”表示担忧,因为这不利于体育素养的个性化发展。此外,体育素养是一个涉及到个体天赋、所处环境和适应文化的概念,在此背景下每个人都是独特的,无法进行比较。因此,采用图示法这样的方法适合记录个人的体育素养发展。另一方面,体育素养是贯穿个体一生发展的动态且非线性的过程,传统测量方法无法反映过程和结果^[14]。采用时间历史的方法是记录体育素养发展的一个方式。Whitehead认为体育素养的发展没有终点,而当前经典的测评方式基本采用常模和标准参照对个体进行评估。

如何从新的视角构建体育素养的测评呢?Green等^[14]提出了一些构想。他们以体育素养的属性是交互的,并以整体视角这一观点切入,提出体育素养概念包含身体、情感、认识,体育素养的发展是非线性过程,因此提出了一些具体措施。在测评的本质,体育素养测评要以行为变化为主要观测点入手,与其相关的认知、心理和情感部分也需要进行测评,这些维度的权重相同;在测评过程中需要考虑文化和环境的特点;测评形式以自我比较为主要形式,避免传统的常模参照和标准参照比较;测评目的以观察增长为主,促进个体进一步提高;测评数据收集和分析通过观察个体体育素养的增长,能够反映个体的体育素养发展过程,通过不同的呈现方式,比如视频、图片和反思性文本来观察个体的变化。

也就是说,未来的体育素养测评要以整体观(身心一统的一元论)为测评的核心思想,放弃传统的测评方式,建立以自我比较为基础的测评体系,采用定性和定量相结合的方式,对个体的体育素养进行个性化的动态描述^[14, 21]。体育素养作为一个隐喻(形而上学)的学术概念,其测量并非是将其构成部分的逐一测量与再次整合^[14]。尽管一些研究提供上述思路,但目前研究还未提出相关理论和依据。相信随着未来研究的

深入,体育素养的测评问题终将会被学者们一一攻克。

体育素养研究目前仍处于“问题多于答案的幼年期”,就如何测量体育素养这一命题而言,由于不同研究视角尚存的争议,概念的综合和复杂性以及操作过程中涉及到各方面问题,体育素养的测评研究还有很长路要走。本研究的目的,一方面旨在介绍当前关于体育素养测评的一些理论和实际问题,并提出一些尝试性的解决对策甚至是假设。另一方面,也试图“抛砖引玉”,希望更多学者参与到这一研究命题讨论中,为体育素养测评研究的推动提供更多智慧。有必要说明的是,本研究中的部分问题目前仍无较好的解决方案,是当前以及未来研究面临的挑战,因此本研究对这些问题仅做一些建议性回答,为未来研究提供参考,同时期待着更多学者能够参与到体育素养测评的研究中,为我国体育素养测评提供更多智慧。

参考文献:

- [1] 邵伟德,邵天逸,李启迪. 学校体育价值论[J]. 北京体育大学学报, 2016, 39(7): 83-88.
- [2] CHEN A, SUN H. A great leap of faith: Editorial for JSHS special issue on physical literacy[J]. Journal of Sport and Health Science, 2015, 4(2): 105-107.
- [3] WHITEHEAD M. The concept of physical literacy[J]. European Journal of Physical Education, 2001, 6(2): 127-138.
- [4] 陈思同,刘阳,唐炎,等. 对我国体育素养概念的理解——基于对 Physical Literacy 的解读[J]. 体育科学, 2017, 37(6): 41-51.
- [5] EDWARDS L, BRYANT A, KEEGAN R, et al. Definitions, foundations and associations of physical literacy: A systematic review[J]. Sports Medicine, 2017, 47(1): 113-126.
- [6] 陈思同,刘阳. 加拿大体育素养测评研究及启示[J]. 体育科学, 2016, 36(3): 44-51.
- [7] 于秀,孙夕鹭,刘海燕. 城市学生体育素养评价指标的研究[J]. 沈阳体育学院学报, 2012, 31(5): 1-5.
- [8] 阳艺武. Physical Literacy: 内涵解读、中外对比及教学启示[J]. 上海体育学院学报, 2016, 40(4): 73-78.
- [9] JONES G, STATHOKOSTAS L, YOUNG B W, et al. Development of a physical literacy model for older adults: A consensus process by the collaborative working group on physical literacy for older Canadians[J]. BMC Geriatrics, 2018, 18(1): 13-24.
- [10] LONGMUIR P, BOYER C, LLOYD M, et al. The

- Canadian assessment of physical literacy: Methods for children in grades 4 to 6 (8 to 12 years)[J]. *BMC Public Health*, 2015, 15(1): 767-778.
- [11] FRANCIS C, LONGMUIR P, BOYER C, et al. The Canadian assessment of physical literacy: Development of a model of children's capacity for a healthy, active lifestyle through a delphi process[J]. *Journal of Physical Activity and Health*, 2016, 13(2): 214-222.
- [12] LODEWYK K, MANDIGO J. Early validation evidence of a Canadian practitioner-based assessment of physical literacy in physical education: Passport for life[J]. *Physical Educator*, 2017, 74(6): 441-475.
- [13] CAIRNEY J, CLARK H J, JAMES M E, et al. The preschool physical literacy assessment tool: Testing a new physical literacy tool for the early years[J]. *Frontiers in Pediatrics*, 2018, 6(9): 138.
- [14] GREEN N, ROBERTS W, SHEEHAN D, et al. Charting physical literacy journeys within physical education settings[J]. *Journal of Teaching in Physical Education*, 2018, 37(3): 272-279.
- [15] 陈琦. 学生体育素养的评价[J]. *体育学刊*, 2002, 16(6): 12-14.
- [16] 赵雅萍, 孙晋海. 加拿大青少年体育素养测评体系“生命护照”研究及启示[J]. *成都体育学院学报*, 2018, 44(4): 92-97+102.
- [17] 赵雅萍, 孙晋海. 加拿大青少年体育素养测评体系 PLAY 解读及启示[J]. *首都体育学院学报*, 2018, 30(2): 10-15.
- [18] LONGMUIR P, TREMBLAY M. Top 10 research questions related to physical literacy[J]. *Research Quarterly for Exercise and Sport*, 2016, 87(1): 28-35.
- [19] ZHU W, RINK J, PLACEK J, et al. PE metrics: Background, testing theory, and methods[J]. *Measurement in Physical Education & Exercise Science*, 2011, 15(2): 87-99.
- [20] KEEGAN R. A sceptic's introduction to physical literacy: Is it new? Is it science? And is it worthy of study?[J]. *Journal of Science and Medicine in Sport*, 2014, 18(Supplement 1): e101.
- [21] EDWARDS L, BRYANT A, KEEGAN R, et al. 'Measuring' physical literacy and related constructs: A systematic review of empirical findings[J]. *Sports Medicine*, 2018, 48(3): 659-682.
- [22] HIGGS C. Physical literacy—Two approaches, one concept[J]. *Physical and Health Education Journal*, 2010, 76(1): 6-10.
- [23] WHITEHEAD M. Physical literacy: Philosophical considerations in relation to developing a sense of self, universality and propositional knowledge[J]. *Sports Ethics and Philosophy*, 2007, 1(3): 281-298.
- [24] 赖天德. 试论素质教育与学校体育改革(上)[J]. *中国学校体育*, 1998, 18(11): 64-66.
- [25] 冯古首, 王勇慧. 体育素养教育模式研究[J]. *山西师大体育学院学报*, 1999, 6(2): 24-26.
- [26] 余智. 体育素养概念研究[J]. *浙江体育科学*, 2005, 27(1): 73-76.
- [27] 杨献南, 鹿志海. 形式逻辑视角下的体育素养概念辨析[J]. *南京体育学院学报(社会科学版)*, 2015, 14(2): 89-92.
- [28] 祝大鹏. 运动员体育道德: 概念、测量、影响因素与展望[J]. *武汉体育学院学报*, 2013, 47(7): 64-70.
- [29] CORBIN C. Implications of physical literacy for research and practice: A commentary[J]. *Research Quarterly for Exercise and Sport*, 2016, 87(1): 14-27.
- [30] GUNNELL K, LONGMUIR P, BARNES J, et al. Refining the Canadian assessment of physical literacy based on theory and factor analyses[J]. *BMC Public Health*, 2018, 18(2): 1044-1052.
- [31] LONGMUIR P, GUNNELL K, BARNES J, et al. Canadian assessment of physical literacy second edition: A streamlined assessment of the capacity for physical activity among children 8 to 12 years of age[J]. *BMC Public Health*, 2018, 18(2): 1057-1068.
- [32] LONGMUIR P, WOODRUFF S, BOYER C, et al. Physical literacy knowledge questionnaire: Feasibility, validity, and reliability for Canadian children aged 8 to 12 years[J]. *BMC Public Health*, 2018, 18(2): 1035-1043.
- [33] ROBINSON B, RANDALL L. Marking physical literacy or missing the mark on physical literacy? A conceptual critique of Canada's physical literacy assessment instruments[J]. *Measurement in Physical Education and Exercise Science*, 2017, 21(1): 40-55.