大数据时代的教育测评模型及其范式构建

范涌峰、宋乃庆

**摘要：**在教育大数据时代，如何充分挖掘大数据对于教育实践、决策、评估与研究的价值及规避大数据风险，是当前教育改革和发展的重要议题。教育测评模型通过对教育现象的关键要素及其相互关系进行定量刻画和价值判断，使人们从海量数据中获取关键有效的信息，将“大数据”变成“小数据”。它是大数据时代我国教育改革和发展的战略工具，是教育决策科学化的有力支撑，也是教育研究科学化的重要突破口。教育测评模型的构建范式是教育测评模型构建的一整套规范，是教育测评模型构建的方法论依据，其主要内容是：确定教育测评模型构建的价值取向，明确教育测评对象的操作性定义，构建教育测评指标体系，确定教育测评指标权重与生成教育测评模型，验证和修正教育测评模型。应用这一范式，从宏观、中观和微观层面构建了系列教育测评模型，对相关领域的研究、决策、实践和评估起到了积极的作用。

**关键词：**教育测评  模型  范式  大数据

**作者**范涌峰，中国基础教育质量监测协同创新中心西南大学分中心、西南大学基础教育研究中心副研究员；宋乃庆，中国基础教育质量监测协同创新中心首席专家、西南大学统筹城乡教育发展研究中心教授。（重庆   400715）

20世纪90年代以来，随着信息技术和互联网技术的发展，大数据概念逐步形成并普及，人类社会逐渐进入所谓的大数据时代，至今关于大数据的应用和讨论方兴未艾，大数据正在对人类生产生活的各个领域产生深远的影响，大数据同样深刻影响着教育研究、决策与实践。**大数据是一把“双刃剑”，一方面，大数据给我国教育改革和发展带来前所未有的机遇，但大数据的价值不是自然彰显的，而是需要挖掘利用的，如果没有科学有效的工具，大数据的价值便如蒙尘之美玉、没土之金子，难以被发现；另一方面，大数据也给我国教育改革和发展带来前所未有的挑战和风险，在大数据时代仅仅凭经验认识和解决教育问题已经愈加不可能，而面对纷繁复杂的大数据，如果没有科学有效的工具，大数据给人们带来的可能更多的是碎片化的信息和惶恐、焦虑的情绪，甚至误导教育决策。**教育测评模型便是挖掘教育大数据价值和规避大数据风险的战略工具，对于提升教育研究的科学化水平，推动形成基于证据和大数据的教育决策机制，提高教育质量，促进教育公平，意义重大而深远。对于教育改革和发展而言，可以说这是个数据的时代，这也是一个模型奠基的时代，谁掌握了模型，谁便把握了大数据时代的发展方向。教育测评模型的构建需要模型思想和一定的方法论指导，教育测评模型的构建范式则可谓是模型之“模型”，是教育测评模型准确性和科学性的重要保证。研究教育测评模型构建范式，对于应对教育大数据时代具有基础性、战略性意义。



一、教育测评模型的内涵与特征

教育测评即教育测量与评价。测量，“从广义上讲，就是根据某些法则与程序，用数字对事物在量上的规定性予以确定和描述的过程。”教育测量就是对教育现象进行定量刻画的过程。而教育评价则是对教育现象作出价值判断的过程。显然，教育测量侧重定量描述，而教育评价侧重定性描述，但这并不是说教育测量没有价值判断或者教育评价没有定量刻画，格朗兰德（N.E.Gronlund）便认为，评价包括定量描述、定性描述和价值判断三个方面。教育测量与教育评价二者既有区别又有联系，测量不可能完全脱离价值判断，评价也往往离不开测量的支撑。这也是“教育测量与评价”总是被人们作为一个词或者专业领域的原因。为此，教育测评可以理解为对教育现象进行定量描述和价值判断的过程。教育测评是正确认识和评价教育现象本质和规律的重要方式，也是影响我国教育改革和发展的关键环节。

模型是人们依据特定的研究目的，在一定的假设条件下，再现原型客体的结构、功能、属性、关系、过程等本质特征的物质形式或思维形式。在这里我们更关注的是本质上作为一种思维方式的模型概念，它反映的是研究对象各变量之间的相互关系，而这种关系一般可以通过定性和定量两种方式表达出来。在教育研究领域，存在更多的是定性地表达研究对象结构关系的模型，定量模型则较为匮乏。既然是对事物本质特征的描述和反映，构建一个科学的模型就意味着有效地把握了研究对象的本质特征。在大数据时代，教育数据具有规模大、类型和来源多样、处理速度快、时效性高等特点，其中也包含依靠传统方法无法获取的教育活动实时、真实的信息，如学生在各种传感器上留下的学习、生活等活动信息及时间分配，但正是由于大数据的样本接近于总体，数据获取具有偶然性，使大数据本身参差不齐，信息价值（value）密度低，加之大数据重相关而不重因果，使得“由足够详尽的搜索和拟合产生的数据‘模式’可能是过渡拟合的结果，仅仅是随机波动的产物，并不代表所研究现象的任何本质特征”，因此，要想基于大数据把握教育改革和发展的本质与规律，首先应建立研究对象的测评模型，形成对研究现象自身的科学认识，并基于模型系统地、科学地、有目的地搜集、分析和处理大数据。从这个意义上来说，教育测评模型是教育发展应对大数据时代的战略工具。随着大数据时代的到来，各级各类教育都开始关注和构建测评模型，如义务教育阶段学生核心素养模型、职业教育均衡发展测度模型等。我们认为，教育测评模型是对反映教育现象的关键要素及其相互关系进行定量刻画和价值判断的工具。它具有以下几个方面的特征。

**第一，定量描述与价值判断相结合。**构建教育测评模型的主要目的就是通过关键要素及其相互间关系来呈现复杂的教育现象，将复杂问题简洁化、计量化，这其中，有些教育要素，如入学率、教育经费、教师数、学生数等是容易以数字呈现的，而由于教育问题具有价值性、复杂性、生成性的特点，诸多教育要素，如教师幸福感、学生心理健康水平、校长胜任力等，即便是在互联网技术和数据科学如此发达的今天，也难以直接获取定量数据。在教育测评模型构建中，对于这类难以计量的教育要素要最大程度地进行量化，便于相关要素上大数据的挖掘，从而使教育测评对象的认识更加精确、科学，也使测评结果更具说服力、可比性和参照性，人工智能、机器学习、机器阅读等方法的迅猛发展，也为对传统上难以定量分析的非结构化、半结构化数据进行量化分析提供了更多可能。不过，常为人们所忽视的是，追求量化的教育测评模型并不是完全脱离价值的冷冰冰的符号和数据，绝对价值中立的教育测评模型是不存在的，教育测评模型的构建首先是建立在一定的理论假设和价值立场基础上，对于为什么要构建一个测评模型，怎样构建这个测评模型，构建怎样的测评模型等问题，都必然地包含研究者的价值评判，也可以说，当决定构建一个测评模型的时候，研究者的价值已经“入场”。如上述“教育均衡发展测评模型”的构建基于一定的教育公平观，学生核心素养测评模型的构建基于人的全面发展观。不同的研究者基于不同的立场和价值观念，对于同一教育现象，可能构建出不同的测评模型。因此，教育测评模型是定量描述与价值判断的统一体。

**第二，精确性与局限性并存。**大数据给教育改革和发展带来更多的不确定性，在不确定性的世界中找寻相对确定的方面，这便是教育研究者、实践者和决策者在大数据时代的重要使命之一。教育测评模型可以通过定量刻画实现对教育现象的精确描述和再现，把握教育现象的本质规律。如《中国教育现代化2035》强调“到2035年，总体实现教育现代化”，但是这一目标较为宏观和抽象，教育现代化的表征为何？如何衡量教育现代化水平？通过构建教育现代化测评模型，便可以明确教育现代化的各项指标，对教育现代化总体及各分项指标作出精确的规定，并定量刻画我国教育现代化水平，从而克服传统上仅仅通过演绎、推理等方式进行思辨研究和主观决策的缺陷。不过，这种精确性是相对的，教育测评能够反映教育现象的关键特征，但却不可能绝对完整地呈现教育现象的全貌，加之对一些指标尤其是主观指标的定量刻画不可避免地存在误差，因此教育测评模型也必然存在一定的局限性。规模大、类型多、时效性强的大数据使人们可以应用教育测评模型更加精确地认识教育现象，但是大数据同样难以克服数据价值密度低、样本偏差及在主观指标反映上的局限性，因而在大数据时代教育测评模型依然是精确性与局限性并存，最大限度地超越局限，实现最大程度的精确是教育测评模型的基本追求。

**第三，综合性与精简性相结合。**教育测评模型对教育现象的定量刻画与价值判断应该是相对全面的、综合的，也就是说能够尽可能全面地包含反映教育现象的教育要素，全面地再现教育现象，同时，测评模型应该是普适而超越具体的、个别的教育现象的，如学生核心素养测评模型应该能够适用于区域乃至全国的学生，而不是某一个或部分学生，只有这样，教育测评模型的外延价值才能彰显。然而，反映教育现象的指标要素是众多的，在教育测评模型中，只能提取关键指标和数据，从而将复杂的教育现象简单化，将承载其中的大数据变成小数据，即教育测评模型应该力求精简，而不能包含太多的指标，否则便失去了构建测评模型的基本价值。大数据强调通过在海量数据中找寻相关关系来洞察和透析教育现象，强调数据的体量和效率，而不一定强调数据的精准，强调相关关系而不注重因果关系，而教育测评模型恰恰强调通过“降维”和“提纯”以提高数据的价值密度，因此，大数据有利于提高教育测评模型的综合性，而教育测评模型强调“精简性”有利于提高大数据服务教育实践和决策的价值，两者相得益彰。

**第四，稳定性与发展性相结合。**教育测评模型中所包含的指标和数据是研究者根据实际需要和基于一定的价值取向选取的，这些指标和数据往往都只是反映了一定时期内教育现象的基本状况，随着时间的推移，经济社会发展状况和教育政策发生改变，其中一些指标和数据可能会失去效用，而需要以新的指标和数据替代。因此，要保证教育测评模型的效度，既要使其在一段时间内保持一定的稳定性，又需要对测评内容进行追踪，根据实际需要，对测评模型予以调整和修订。需要指出的是，由于教育现象自身的复杂性等方面原因，尽管在大数据时代往往强调基于实时数据变动而更新模型，但教育测评模型则更多地是因“势”而动而非因“数”而动，即基于教育实践和教育政策的变化与需要更新模型。



二、教育测评模型的战略意义

我国自20世纪20年代便开始教育统计与测评的研究和实践，但百年来的探索并未使其充分支撑教育研究、决策和实践，教育研究的科学化水平落后于发达国家，教育研究者面对教育大数据常感有心无力，基于证据和大数据的教育决策机制尚未形成，大数据对于教育质量和教育公平的推动作用也显不足。这源于我国教育领域数据挖掘、分析的意识和技术不足，尤其是模型思想不足，没有充分认识到教育测评模型的重要意义。在大数据时代，教育测评模型的战略价值更需重新认识。

（一）教育改革和发展的战略工具

公平与质量是我国教育改革和发展的主旋律。而模型思想对于促进教育公平和提高教育质量是不可或缺的。教育测评模型对于我国教育改革和发展具有十分重要的意义和作用，这主要体现在以下方面：**第一，反映和描述教育现象与问题。**教育大数据意味着教育数据的爆炸式增长和数据模式的高度复杂化，仅凭经验准确地认识教育现象是不可能的，教育测评模型的意义在于它集中了教育现象最关键、最典型的特征，并加以量化，从而将教育现象较为准确地呈现出来，易于人们了解和把握。**第二，监测与评估教育过程。**以基于大数据的质量监测作为评估手段，是教育评估领域最前沿的变革，这种评估方法扭转了以往“唯分数”、“唯升学”的不科学的教育评价导向，而科学的教育测评模型是教育评估有效性的关键保证，通过教育测评模型的应用，可以实现对教育过程的质量监测，保证教育目标的达成和教育事业的健康发展。**第三，预测与规划教育发展。**基于现有的数据，通过构建一定的教育测评模型，在对过去和现实状况分析的基础上，可以对测评模型中的指标变化规律，对教育发展的走向作出预测，并根据预测结果作出教育规划。大数据为基于测评模型的教育发展预测提供了可能，大数据带来的不确定性也要求对未来教育的发展作出更为有效的预测，以增强教育改革成功的可能性。**第四，导向教育价值。**教育测评模型包含人们对教育现象的系统思考，是工具性和价值性的统一体，教育测评可以导向人们对测评对象是什么、为什么、怎么做以及怎么评的价值判断，如“义务教育质量测评模型”实际深层次融入了研究者的义务教育质量观，这种质量观通过测评结果将会引领人们形成与研究者趋同的教育质量观。传统的教育测评中，为研究某一教育现象在总体中的情况，往往通过随机抽取有代表性的样本，通过统计分析得到样本中不同变量之间的关系，并依此推论总体的情况，而基于大数据的教育测评模型应用在理论上可以将总体全部纳入分析框架，而无需参数检验，这为更加全面、系统地反映和描述教育现象与问题、监测与评估教育过程、预测与规划教育发展以及导向教育价值提供了更大的可能，为国家和地方提供更加全面的反映教育现实水平的客观数据，为加快教育现代化、办好人民满意的教育提供重要基础。

（二）教育决策科学化的有力支撑

当前我国教育决策中凭经验主观决策的现象仍然较为普遍，教育决策的科学化程度有待提高。这一方面源于教育决策机制不够健全，另一方面则源于我国教育智库建设不完善，教育决策的科学支撑不足。根据《全球智库报告2018》，目前全球共有8162家智库，中国2018年共有507家智库，仅次于美国（1871家）、印度（509家），位居世界第三。尽管如此，在教育领域，教育智库建设及其对教育决策的作用则仍显不足。美国詹姆斯·科尔曼1966年完成的《关于教育机会平等》报告、1983年美国高等教育质量委员会完成的《国家处在危险之中：教育改革势在必行》报告等对美国的国家教育改革和发展产生了深远的影响，而在我国则难有此类报告。这与我国教育领域长期以来注重定性研究而忽视定量研究的传统有关系。基于大数据分析，可以对教育现象和教育问题进行定量刻画，并作出更为准确和更具说服力的判断和预测，但是，正如舍恩伯格和库克耶提出的，大数据“侧重于分析相关关系，而不再寻求每个预测背后的原因”,大数据分析可以告诉我们教育现象在统计上的相关性，但是却往往不能检验逻辑上的因果关系。如果没有科学的模型支撑，大数据有可能给人们呈现一些“统计欺骗”。也正因此，大数据分析在一些领域带来了巨大的商业价值，并体现了蓬勃的应用前景，但是在教育领域，却鲜见基于大数据应用的高水平学术成果和教育决策案例。教育测评模型的作用便在于能够将大数据变成“小数据”，从不确定性中寻找确定性，增强大数据对教育决策的价值，而降低大数据带来的决策风险。因此，教育测评模型的构建是未来教育决策科学化的战略需要和支撑，将催生教育智库发展，并推动教育智库发挥更大的决策咨询作用。

（三）教育研究科学化的重要突破口

“科学界普遍把学科定量分析的水平当作研究科学性的重要的、不可或缺的指标”，如果以此作为衡量教育研究科学化水平高低的标准，那么我国教育研究科学化水平还十分不尽如人意，与教育发达国家的教育研究科学化水平还有很大的距离。有关研究表明，近年来我国教育研究领域定量研究的学术成果的占比还不到15%。尽管如此，对于教育研究科学化问题还存在诸多争论，不少学者对教育学注重实证、关注数据的倾向持质疑、否定态度，认为实证取向的教育研究盲目迷信各种证据，盲目迷恋大数据，而缺乏对教育学的价值和人文观照，实际上，保持对教育研究实证化倾向的理性审视是必要的，然而如果在大数据时代漠视教育大数据，则无疑是固步自封的做法。纵观社会科学发展历程，“量化是促进包括教育学在内的社会科学取得突破性进展的关键”，在我国教育学仍然徘徊于人文主义和科学主义的十字路口之时，心理学、经济学、社会学等社会科学的量化研究范式已经走向成熟，大大提升了学科的科学化水平，并与国际接轨，即便是在历史学领域，计量史学的发展也为历史研究注入了新鲜血液和活力，而对教育学科学性和学科地位的质疑却从未间断。因此，在大数据时代教育学者应该认识到的是，以量化研究为主要特征的教育研究科学化是教育研究的必然趋势，数据既非洪水猛兽，也非至上真理，理性地寻找基于数据的教育研究科学化路径，抓住大数据机遇增强教育研究的科学性才是我国教育研究者的重要学术使命。教育测评模型构建强调对教育现象的定量刻画，它从教育发展的现实问题出发，通过科学的研究方法尤其是定量研究方法，将复杂的、质性的教育问题最大限度地定量刻画，增强教育研究的科学性，是大数据时代教育研究科学化的重要突破口和前沿课题。同时，教育测评模型探索教育研究与计算机、互联网、人工智能和大数据的结合，为“计算教育学”等前沿领域发展奠定了基础，也为我国教育研究与国际接轨提供了新的路径和机遇。



三、教育测评模型的构建范式

托马斯·库恩（Thomas S. Kuhn）认为，“范式是一个特定共同体成员所共有的信念、价值、技术等等构成的整体”，包括符号概括、共同体成员共同承诺的信念、价值、范例等要素，

是任何一个科学领域在发展中达到成熟的标志。尽管库恩作为物理学家和科学哲学家、科学史家，其提出“范式”更多是自然科学意义上的，但是“范式”一词广泛应用于人文社科研究领域，如李吉林提出“构建择美构境，境美生情，以情启智，把情感活动与认知活动结合起来，引导儿童在情境中学、思、行、冶的儿童情境学习范式”，唐亚林提出建构“当代中国大都市治理的范式” 等。这些范式内容迥异，但都基本指向了某一研究领域共同遵循的价值取向、科研信念、研究领域、研究方法和研究程度等。

在库恩看来，凡是具有以下两个特征，便可称之为“范式”：“成就空前地吸引一批坚定的拥护者，使他们脱离科学活动的其他竞争模式。同时，这些成就又足以无限制地为重新组成的一批实践者留下有待解决的问题。”依此判断，我们开拓了教育测评模型这一新的研究领域，构建了义务教育资源配置均衡发展测评模型、学校特色发展测评模型、课业负担测评模型、教材难易度测评模型、青少年动商测评模型等系列测评模型，这些模型在理论和实践上已经产生较为广泛的影响，并且基于大量的教育测评模型研究、实践与反思，探索出教育测评模型构建的基本价值取向、规范、内容、方法和程序等，基本符合库恩所谓的范式特征，因而概可以将其称为教育测评模型的构建范式。“范式”创造一种一贯的传统，为特定的连贯的科学研究的传统提供模型，为此，教育测评模型构建范式则是为教育测评模型研究提供模型，也可以说，教育测评模型构建范式是教育测评模型之模型。它包括教育测评模型构建的一整套规范，这绝非本文寥寥数字能够阐明，笔者仅是对其核心内容加以阐释。

需要指出的是，由于教育现象的高度复杂性、生成性、价值性等特点，以及对教育测评结果解释性和预测性的高精确度要求，教育测评模型构建强调实践导向的模型构建，而非大数据科学强调的数据驱动的模型构建，即教育测评模型构建的逻辑起点在于教育改革和发展的现实需要，而非教育大数据本身。**因此，所谓教育测评模型是应对教育大数据时代的战略工具，其主要意蕴在于，教育测评模型通过从复杂的教育现象中提取关键要素，并予以定量刻画，从而将复杂的教育现象简单化，将纷繁复杂的“大数据”变成“小数据”，即通过应用成熟模型把握教育大数据世界中的关键要素和信息，解决教育改革和发展中的实际问题。**而教育测评模型对“大数据”的回应和价值主要体现在模型的应用中，教育测评模型构建本身则借助更多的依然是“小数据”，为此教育测评模型构建相对于大数据思维和方法而言，依然是遵循较为传统的研究逻辑和范式，教育测评模型构建根据需要充分发挥“小数据”优势，而教育测评模型应用则充分利用“大数据”优势，两者相辅相成。

（一）确定教育测评模型构建的价值取向

教育测评模型是价值性和工具性的统一体，价值取向在某种程度上决定了教育测评模型构建的目的、理论框架、思路和方法，它引领教育测评模型的整个过程。如果没有理清价值取向的问题，教育测评模型有可能沦为冷冰冰的、无用的工具。正如约翰斯顿（Denis F. Johnston）认为的，“如果没有理论，我们为继续收集和分析不断增多的统计资料所作的努力，可能降为一种‘任意的实证主义’（casual empiricism），只是提供无意义的事实和无重点的指标，而且研究所得的结论，其造成的问题可能远较所解答的为多”。如对区域教育测评来说，公平取向和效率取向的测评模型不同；对于学校测评来说，工具取向的测评和人文取向的测评不同；对于学生测评来说，分数取向的测评和素质取向的测评不同；等等。教育测评模型构建的价值取向的确定应主要考虑以下因素：**第一，测评对象的本质属性。**如果对测评对象缺乏足够的认识，便没有正确的价值取向可言，为此，可以从历史的视角、国际比较的视角、实证分析的视角等多重视角分析测评对象的本质属性。如中国学生核心素养测评框架的构建便体现了多视角观照。**第二，测评对象的教育及经济社会基础。**一是考虑教育发展的现实基础和需求。二是考虑经济社会发展包括全球经济社会发展对测评对象的影响。教育测评模型的构建要充分考虑其是否是教育和经济社会发展的现实所需，以及是否具备教育和经济社会发展的基础条件。脱离了现实基础和需要，教育测评模型构建便没有可操作性。**第三，教育科学理论支撑。**社会学者普遍认为，没有理论框架的任何实证资料都是无效的。教育测评模型是通过严格的科学研究过程研制的，这必然要求教育测评模型基于科学的理论支撑。

（二）明确教育测评对象的操作性定义

操作性定义是相对抽象定义而言，抽象定义是对测评对象本质内涵的概括，而操作性定义是用可感知、可度量的事物、事件、现象和方法对测评对象作出具体的界定、说明。它的最大特征就是可操作性。由于教育现象具有复杂性、生成性、价值性等特点，许多教育现象是内隐性的、潜在的、难以观测的，因此，要以测评模型对教育现象进行定量刻画，通过操作性定义将教育现象进行可观测、可度量的界定，便是其中的基础和关键环节。它直接关系到测评模型的信效度。因此，最早提出操作性定义的美国学者布里奇曼认为，要想避免概念混淆不清，最好能用测量它的操作方法来界定。而英国学者迈克尔·卡利甚至认为，社会指标的建构过程实质上就是将抽象的、难以测量的社会概念翻译成可以考察、分析的操作性术语。一般而言，操作性定义有以下三种方法：**第一，动态特征描述法，顾名思义就是通过描述测评对象的动态特征对其予以定义，如对有效教学进行定义，便需描述有效教学活动中的具体教学行为与表现；第二，静态特征描述法，即通过描述测评对象的静态特征对其予以定义，如对教师胜任力进行定义，可以通过描述教师胜任教育教学工作的具体静态特征；第三，方法和程序描述法，是通过描述使测评的教育现象发生的方法和程序对教育现象进行定义的方法。**如对学业负担进行定义，便可以通过描述学业负担是如何形成的来对其进行定义。

（三）构建教育测评指标体系

对教育现象的测评最终需要依托系列指标为载体，因此，指标是测评模型的基础要素。操作性定义中的每一项具体内容都可以成为测评指标，一组具有内在联系的、综合反映测评对象状况的指标便构成指标体系。如前所述，教育测评模型是大数据时代捕捉教育现象关键要素的战略工具，指标体系便是这些关键要素。指标体系应尽可能全面反映操作性定义的内容，但这并不是说指标体系应该是“大而全”的，精准简洁是指标体系构建应该遵循的基本原则。从指标的展开方式上分，指标体系的构建可以有两种方式：横向展开方式和纵向展开方式。以横向展开的方式构建指标体系意味着指标之间在逻辑上是相对平行的关系，如从办学条件、教师队伍、教育经费测评教育均衡发展水平；而另一种以纵向展开方式构建的指标之间具有一定的时序逻辑，往往体现为“投入（input）—过程（process）—产出（output）”的模式，如通过教育人财物投入、教育过程管理、办学质量来测评教育均衡发展水平。这种指标展开方式为人们所广泛使用，世界银行、联合国教科文组织(UNESCO)、经合组织(OECD)的教育指标体系都以此种方式构建，如OECD《教育要览》中的指标体系包括教育人口背景、教育财政与人力投入、公民的教育参与进步、学校环境与学校组织、教育的个人与社会产出、学生的学业成就等。而从指标构建的逻辑顺序来说，指标体系构建有三种方式：第一种，自上而下的方式。由研究者根据一定的理论基础和文献研究的结果初步构建指标和框架，再深入实践征求专家和有关人士的意见和建议，对初步指标和框架进行修订和完善。UNESCO便是根据这一思路提出了七个维度的核心素养。第二种，自下而上的方式。由研究者在实践中广泛征求专家和有关人士的意见和建议，并提炼出指标体系。第三种，混合式。即综合自上而下和自下而上两种思路，在开展理论研究的同时，广泛征求专家和有关人士的意见和建议。不论以哪种方式构建指标，指标体系的构建不是随意的，而是从操作性定义出发遵循一定的逻辑架构而生长的，好的指标体系应该实现价值框架、逻辑框架和结构框架三个框架的结构统一、指标融合和目标一致。

（四）确定教育测评指标权重与生成教育测评模型

从指标体系到测评模型的关键步骤是指标权重的确立，只有确立了指标权重，才能将各指标的值整合为一个综合值，将分项指标合成测评模型。显然，某一因素的权重的变化将会影响整个测评结果。因此，选择和正确使用权重确定方法显得十分重要。权重的确定方法有数十种之多，根据计算权重系数时原始数据来源和计算过程的差异，可以将权重计算方法分为主观赋权法、客观赋权法和主客观综合赋权法。其中，主观赋权法包括层次分析法（AHP）、德尔菲法（Delphi）等；客观赋权法包括最大熵技术法、主成分分析法、因子分析法、变异系数法等；主客观综合赋权法即综合运用主观赋权法和客观赋权法，从而对两种赋权法扬长避短。由于篇幅所限，对于上述方法的使用在此不再赘述，在指标体系和测评模型的构建过程中，需要根据实际需要合理地选择权重确定方法。

教育测评模型实质上是定量刻画教育指标体系及其相互之间的关系。在指标权重确定以后，根据各指标的权数和指标之间的实际关系抽象、构建出数学模型Y=f(X1,X2,X3,…,Xn)，从而清晰、简练地表达指标之间的关系。如Y=a1X1+a2X2+a3X3+…+anXn，其中X1,X2,X3,…,Xn为构成模型的若干指标，a1，a2，a3,…,an为各个指标相对应的权数。需要指出的是，从指标体系构建到教育测评模型生成之间常伴随测评工具的研制，通过测评工具搜集各指标上的数据，从而对模型进行应用。这其中通常有三种路线：**第一种，从测评工具到指标体系再到测评模型。**一些研究在初步构建指标体系后，构建测评工具，之后经过反复探索分析，对量表的项目进行增加或删减，常常将因素负荷过低、明显独立于其他项目或者与几个公共因子高负荷相关的项目删除，在此过程中，不断调整项目，直到量表的各项指标都趋于良好，最终确定指标体系，并通过因子负荷确定指标权重，构建测评模型。如“学习自我效能感”研究。**第二种，从指标体系到测评工具再到测评模型。**一些研究先通过德尔菲法等方法确立指标体系后，在指标不变的情况下开发指标的测评工具，通过工具的应用搜集指标数据，最后基于数据统计分析确定指标权重，构建测评模型。如“学校特色发展测评模型”研究。**第三种，从指标体系到测评模型再到测评工具。**一些研究在确立指标体系后，通过层次分析法等方法构建测评模型，再开发一定的测评工具对相应指标予以测量和数据搜集。学科教育领域常遵循这样一种逻辑路线。“一项研究的好坏往往有赖于所得数据的质量”，指标数据的质量会直接影响模型的准确性。传统的指标数据的获取主要有两种方式，一种是从现有的材料中搜集有用的数据，另一种是通过调查或实验获取新的数据，包括使用问卷法、访谈法、观察法、实验法等。在大数据时代，指标数据的获取变得更加多样，互联网发展及一些新兴技术、设备的发展，使以往难以获取的数据成为可能，如通过智能设备记录的学生每天活动轨迹、运动健康数据、学习数据等，这些大数据将使模型建构更加精准，也使模型应用具有更大的空间。

（五）验证和修正教育测评模型

为了保证教育测评模型的信效度，需要对其进行验证。这种验证一般有三种方式：**第一种，基于数据的验证。**即通过一定的统计分析方法检验模型的拟合度。如通过验证性因素分析来验证假设模型的效度，检验模型各项拟合指数（含X2/df、GFI、CFI、TLI、NFI、IFI等）是否符合统计学要求。**第二种，基于专家的验证。**即在模型构建后由相关领域权威专家对模型的效度作出评价。**第三种，基于实践应用的验证。**即在应用过程中检验模型的优劣，测评模型构建后，往往通过在小范围试验检验其在实际教育情境中，对于解决教育大数据时代实际教育问题的效用，如对学校发展质量的测评，可以对比依据模型测评的结果与依据经验评判的结果、专家评判的结果之间是否存在明显的差异。此外，通过个案应用与深入分析，也可以检验测评模型的效度，并弥补以团体数据构建和验证模型的不足。实际上，教育测评模型验证是一个系统收集各种证据以证明测评模型有效性的过程。因此，教育测评模型验证往往不是采用单一的验证方法，而是综合利用多种验证方法。此外，由于教育现象的复杂性、动态性特点，教育测评模型尤其注重在实践应用中检验其科学性和有效性，而且教育测评模型的检验和修正可能会是持续性的、动态性的、周期性较长的过程。教育测评模型具有广泛的应用领域，包括教育投入测评、教育过程运行监测、教育质量监测等，通过测评模型应用可以发现模型本身存在的问题，并根据实际情况对模型作出修正。



四、教育测评模型构建范式的应用

通过教育测评模型构建范式的应用范例，可以进一步让研究共同体理解教育测评模型构建范式所包含的符号概括、信念、价值等。应用教育测评模型构建范式可以构建多种多样的教育测评模型。如前所述，我们在长期的探索实践中，从宏观、中观和微观层面构建了系列教育测评模型，并在实践中得到有效推广应用，对于相关领域的研究、决策、实践和评估起到了积极的作用。这些测评模型不一而足，仅从宏观、中观、微观教育领域三个层面举要分析。

（一）宏观测评：义务教育资源配置均衡发展测评模型构建

**在宏观教育领域，义务教育是整个教育体系中的基础之基础，义务教育均衡发展是我国教育改革和发展的战略性任务，但相关研究多数仅从理论和经验上探究教育均衡或者教育差距，而缺乏衡量义务教育均衡发展科学有效的工具，尤其是缺乏对义务教育均衡发展核心要素义务教育资源配置均衡发展的测度工具，因此构建义务教育资源配置均衡发展测评模型具有重要的现实意义。**为此，综合经济学、教育学以及教育管理学和教育经济学等相关理论，通过文献研究、访谈调查和编码分析等方法，研究了义务教育资源配置的理论框架，选取了义务教育资源配置均衡发展的测评指标，通过探索性因素分析和验证性因素分析的方法确定模型的结构，通过两轮德尔菲法确定该模型的权重，最终构建了义务教育资源配置均衡发展测评模型，并基于样本区县的实际测算数据对义务教育资源配置均衡发展水平进行初步划分。最后，通过对义务教育资源配置均衡发展测评模型进行小范围测试对模型予以验证。最终建构的义务教育资源配置均衡发展测评模型为：Y=0.28A1+0.41A2+0.31A3，其中A1是办学条件指标、A2是教育经费指标、A3是教师质量。办学条件A1=0.55B1+0.45B2（B1为学校布局，B2为设备设施）；教育经费A2=0.67B3+0.33B4（B3为经费投入，B4为经费使用）；教师质量A3=0.82 B5+0.18B6（B5为结构分布, B6为交流发展）。通过内部一致性检验、验证性因素分析及专家效度交互检验表明，模型具有良好的信效度。

（二）中观测评：学校特色发展测评模型构建

**在中观教育领域，主要涉及学校管理问题。在义务教育基本均衡逐步实现的基础上，如何办出学校特色、为学生提供适合的教育，是当前基础教育改革和发展的核心议题。**然而人们对于特色发展普遍认识不清，实践混乱，更无法科学评判学校和区域特色发展水平，此种情况下，以测评为切入点，构建学校特色发展测评模型，可以引领人们对学校特色发展形成相对统一的认识，并形成学校特色发展的基本理论框架及测评框架和方法。为此，将定性研究方法与定量研究方法相结合，综合使用文献研究法、问卷调查法、访谈法、案例研究法和探索性因素分析、结构方程模型、方差分析、回归分析等统计分析方法，尤其是先后调查访谈全国范围内专家146人次，调查4726个中小学校管理干部和教师样本。**第一，构建了学校特色发展的操作性定义。第二，构建了学校特色发展测评指标体系。**结合前期学校特色发展内涵研制过程中的理论研究和专家调查，初步构建学校特色发展测评框架，之后将学校特色发展指标体系广泛征求专家意见，最后确定了三维度、七指标的学校特色发展测评指标体系。**第三，构建了学校特色发展测评量表和模型。**自主开发学校特色发展测评量表，基于量表大样本施测，通过因子分析法确定学校特色发展测评指标权重，并形成学校特色发展测评模型，这一模型近似地用以下数学线性表达式表示为：S=0.33A+0.34C+0.33Q；A=0.50P+0.50D；C=0.60C1+0.40M；Q=0.36S1+0.33F+0.31S2。其中，S表示学校特色发展水平，A表示学校特色发展理念体系的适切度，C表示学校特色实践与理念体系的一致度，Q表示学校特色发展效果的优质度；P表示办学理念，D表示方向目标；C1表示课程体系，M表示组织管理；S1表示学生发展，F表示教师发展，S2表示学校发展。**第四，对学校特色发展测评模型进行试验和验证。**通过在八省市应用，一是整体分析中小学校特色发展水平及其在若干变量上的组群差异。二是选取有代表性的小学、初中、高中进行个案学校特色发展状况诊断。三是初步分析了学校特色发展水平与学生学习成绩之间的相互关系。

（三）微观测评：学生与学科测评模型构建

**在微观教育领域，学生发展是教育发展的根本目的，“减负提质”是当前基础教育改革和发展的中心工作，围绕减负，构建了学生课业负担测评模型，而围绕提质，由于课程是促进学生发展的关键载体，构建了系列学科测评模型。**

学生课业负担测评模型研究在构建课业负担操作性定义的基础上，通过对4省市、1350名被试的抽样调查，初步构建了学生课业负担的测评模型。其中，675名被试用于探索性因素分析，得到的课业负担测评模型包括成绩压力、课业难度、精力消耗和课内学习任务4个维度，合计14个指标；675名被试用于验证性因素分析，结果显示：课业负担测评模型拟合良好；课业负担各维度与睡眠时间、学习成绩有轻微的负相关，与厌学倾向有中等程度的正相关，说明该模型具有良好的测量学性能，可作为学生课业负担的测评工具。

而在学科测评模型方面，构建了数学教科书难度测评模型、数学学习兴趣测评模型、数学符号意识测评模型、数学几何直观测评模型、学生阅读素养测评模型、学生统计思维测评模型、青少年动商测评模型、学生综合素质测评模型、学生信息技术素养测评模型等。如小学低段数学符号意识测评模型研究探索从“小学低段数学符号意识的含义→表现形式→操作性定义→内涵维度的确定→测评指标的筛选→测评模型的构建→测评模型的检验”的测评模型构建的研究思路。通过实证研究陈述了小学低段数学符号意识的存在性，首次探析了小学低段数学符号意识的含义、表现形式和操作性定义，初步尝试构建小学低段数学意识的测评指标（3个内涵指标和7个行为因素指标）和测评模型Y=0.3X1+0.3X2+0.4X3，且模型的准确率为70.56%，其中，X1、X2、X3分别为数学符号的抽象、数学符号的识记和数学符号的应用。

由于篇幅所限，以上仅简要呈现了一些教育测评模型构建范式应用的结果。上述模型的构建对于宏观层面的教育决策、中观层面的学校管理决策及微观层面的教学决策及学生的学习改进都将提供强有力的支撑，同时为相关领域研究提供了数据、工具支持和研究参考。以学校特色发展测评模型为例，学校特色发展是一个内隐性概念，以往区域、学校对于自身特色发展水平到底如何，更多是基于一种主观的感受，区域之间、学校之间也难以比较，而通过以测评这一“牵一发而动全身”的研究，实际涉及学校特色发展的内涵与操作性定义、指标体系、测评工具、测评模型等方面的系列研究，为学校特色发展的实践提供系统理论框架和系列工具，并通过全国大数据采集与分析，可以数据化和可视化全国、区域和中小学特色发展水平，诊断学校特色发展现实水平和存在问题，为中小学校特色发展提供理论指导和诊断工具。与此同时，基于模型应用数据，可以通过多元回归分析等方法探索学校特色发展的影响因素，而通过中介效应、调节效应等方法可探索学校特色发展对其他相关变量的影响，如校长领导力对学校特色发展水平的影响或学校特色发展对学生学业水平、学生核心素养等方面的作用，从而探索学校特色发展的驱动因素和作用机理；又如就当前而言，教育大数据主要体现在学生大数据上，学业负担测评模型的应用过程中，可以借用信息技术手段包括可穿戴设备搜集学生在学习时间、作业时间、运动时间、睡眠时间及学习压力等方面的大数据，精准刻画全国中小学生学业负担状况，为督导评估各级政府落实减负提质情况及因地制宜地制定“减负”政策提供专业支撑。尽管在大数据时代教育测评模型具有广阔的应用前景，但是，教育测评模型的应用应从实际出发，发挥教育测评模型的引领和诊断等积极功能，避免走入新的“机械化”、“形式化”、“功利化”误区，而成为教育发展新的“枷锁”。



结语

教育测评模型构建是我们近年来探索创新的教育研究新兴领域，由于这一领域的研究时间毕竟不长，加之教育测评模型构建本身是一个极为复杂的工作，教育测评模型构建范式还有诸多问题有待在今后的研究中进一步完善，未来对于教育测评模型及其构建范式还将从以下几个方面进一步探索。

**第一，构建教育测评模型体系。**今后，应用教育测评模型构建范式，围绕我国教育改革与发展的重点、热点、难点领域，从宏观、中观、微观层面还可以继续构建系列模型，如在宏观领域，可构建教育现代化测评模型、城乡教育一体化测评模型、义务教育优质均衡发展测评模型、县域义务教育质量测评模型等模型；在中观领域，可构建学校教育质量测评模型、课程质量测评模型、课堂教学质量测评模型、校长领导力测评模型等模型；在微观领域，可构建学生发展质量测评模型、学生学业水平发展测评模型、学生核心素养测评模型、教师专业发展测评模型等模型，每一个模型中又可有若干子模型，从而构建大数据时代的教育测评模型体系。

**第二，探索教育测评模型、教育大数据和信息技术的深度融合。**如前所述，当前我们对于教育测评模型的构建更多地是借助“小数据”，尽管“小数据”具有自身独特的优势，但面对新的时代变革，教育测评模型构建必须探索新的方法和路径：一是探索基于大数据分析的模型建构方法。在教育测评模型过程中，充分运用大数据的思维和方法，如人工智能、机器阅读、机器识别、机器学习等，综合小数据和大数据优势，探索更加科学的教育测评模型构建方法。二是探索基于教育测评模型的大数据应用路径。模型是教育大数据应用科学化和精准化的基础，借助信息技术手段和基于教育测评模型的大数据应用，构建教育测评云平台，实现测评对象的全样本化、测评信息的多元化和丰富化（包括文本、图片、音频、视频、动作、表情等）、测评过程的智能化、情境化和生活化，为教育实践与决策提供即时性、个性化、精准化的改进路线。最终通过将教育测评模型系统、大数据收集与分析系统、信息技术系统三大系统整合，形成相对完整的大数据时代教育决策系统。

**第三，探索非线性测评模型的构建方法和范式。**本研究中的教育测评模型构建范式从测评对象的操作性定义入手，这一范式具有独特优势，有利于形成对教育现象客观、统一的认识，提高研究结果的准确性、可比性和可检验性。但任何模型都不具有绝对普遍性，而是局限于一定的适用对象，本研究中的教育测评模型构建范式更多地适用于线性模型的构建，其中最基本的假设是，X和Y总体上具有线性关系，同时Y服从正态分布。但如果X和Y是非线性的变量关系，则需要构建非线性测评模型。而由于篇幅所限，对于非线性测评模型的构建范式将另文研究。

**第四，保持教育测评模型的工具价值与教育现象的人文属性之间合理的张力。**对于教育测评合理性的争议由来已久，科学主义者认为，科学知识以经验事实为基，人文社会科学的问题同自然科学的问题一样可以用自然科学的方法进行研究，因此，教育现象是可测可评的。但存在主义者认为，教育的人文属性决定了教育现象是复杂而又无法测量的，如对学生品德发展水平的测量，由于“个体从其内在的道德信念和道德动机向外在言行过渡的过程中，会受到诸多因素的影响，因此，从外显的指标去推断一个人真实的道德状况，不可避免地带有不同程度的失真情况”。教育研究中的科学主义和人文主义各有其合理性，两者的争论不会停休，对于教育测评模型及其构建范式而言，最重要的是，在构建和应用教育测评模型时，不陷入对数据的盲目迷信和乐观，保持对教育人文属性的敬畏，同时，对于尤其难以定量测评的教育现象，以科学精神和科学的测评方法客观予以认识，最大程度地寻找反映教育现象本质的指标，最大限度地突破教育测评的人文限度。这也是大数据时代教育测评模型及其范式构建面临的重要方法论难题。

随着大数据在教育领域的深度应用，模型思维将愈来愈重要，教育测评模型的价值将愈来愈凸显，尽管人们对于教育测评尚存争议，尽管教育测评模型及其范式还未必成熟，但相信这些对教育测评模型及其范式的初步思考对我国教育改革和发展如何应对教育大数据时代具有启迪意义，希望它能引起学界对教育测评模型及其范式问题更多的关注和争鸣，促进教育测评模型及其构建范式不断走向完善。