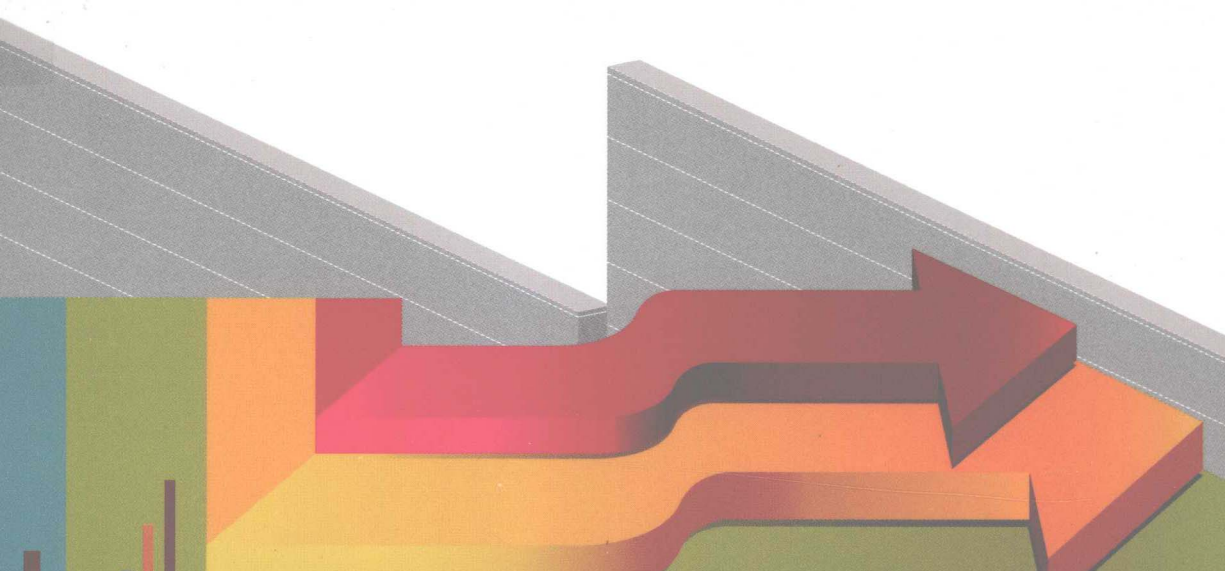
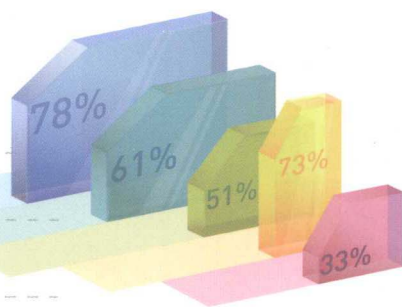




学业成就调查的 原理与方法

田慧生 孙智昌 主编

XUEYE CHENGJIU
DIAOCHA DE
YUANLI YU FANGFA

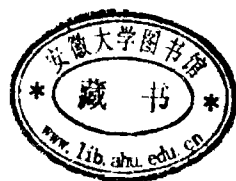


教育科学出版社

Educational Science Publishing House

学业成就调查的 原理与方法

田慧生 孙智昌 主编



教育科学出版社

· 北京 ·

出版人 所广一
责任编辑 池春燕
版式设计 杨玲玲
责任校对 贾静芳
责任印制 曲凤玲

图书在版编目 (CIP) 数据

学业成就调查的原理与方法 / 田慧生, 孙智昌主编. —
北京: 教育科学出版社, 2012. 6
ISBN 978 - 7 - 5041 - 6482 - 7

I. ①学… II. ①田… ②孙… III. ①学业评定
IV. ①G424. 75

中国版本图书馆 CIP 数据核字 (2012) 第 085845 号

学业成就调查的原理与方法

XUEYE CHENGJIU DIAOCHA DE YUANLI YU FANGFA

出版发行	教育科学出版社		
社 址	北京·朝阳区安慧北里安园甲 9 号	市场部电话	010 - 64989009
邮 编	100101	编辑部电话	010 - 64989441
传 真	010 - 64891796	网 址	http://www.esph.com.cn
经 销	各地新华书店		
制 作	北京金奥都图文制作中心		
印 刷	北京中科印刷有限公司	版 次	2012 年 6 月第 1 版
开 本	169 毫米×239 毫米 16 开	印 次	2012 年 6 月第 1 次印刷
印 张	23	印 数	1 - 5 200 册
字 数	346 千	定 价	49.80 元

如有印装质量问题, 请到所购图书销售部门联系调换。

前言

在当代，提高教育质量、促进教育公平，已成为全世界的一种共同追求。究竟什么是教育质量，人们可能会有不同的看法，但学生的学业成就无疑是教育质量的重要组成部分。因此，要提高教育质量，必须关注学生的学业成就。教育质量的提升与教育公平的推进是相辅相成的。通过学业成就调查来了解学生、学校及相关地区的学业成就现状，已成为提高学生学业水平、改善教师教学效能、提升学校办学质量、促进地区教育均衡发展的重要手段。

但是，学生学业成就调查是一项专业性、技术性和政策性都很强的工作，需要有专业的调查人员、调查团队和专门的技术予以支撑，需要有明确的制度和政策予以保障。这其中隐含着一系列的原理与方法需要深入研究。

正是在这样的背景下，我们于 2006 年承担了国家社科基金“十一五”规划（教育学科）国家重点课题“中小学生学业成就调查研究”（课题编号：AHA060008，课题主持人：田慧生），在进行实证调查的同时，深入探讨了学业成就调查的概念、原理、方法等核心问题。本书就是这一课题研究的重要成果。

全书由学业成就调查概论、学业成就调查的国际比较与借鉴、学业成就调查的理论体系与实践策略组成。

概论部分梳理了学业成就调查的历史发展及其基本框架，是全书的一个粗略图景，既引导了全书的写作，又便于读者的整体把握。

学业成就调查的国际比较与借鉴部分，通过大量的国内外研究数据、案例，详细比较了国际组织和美国、日本等国的学业成就调查经验，系统总结了我国学业成就调查的成果和存在的问题，详细阐述了国外学业成就调查体系的特点和发展趋势。由于现代学生学业成就调查起源于西方，其经验也为它们所创造，所以对这些成果的充分吸收，为本书提供了重要的理论资源 and 实践策略，对我国学业成就调查的本土化发展具有重要的引领价值。

学业成就调查的理论体系与实践策略是本书的重点。这部分开创性地探讨了学业成就调查的哲学基础——价值论基础和认识论基础，具有填补空白的重要意义。对学业成就调查的教育心理学基础和测量理论的研究，为这项工作的实际操作提供了直接的理论依据，同时，对有关 SOLO 分类理论的介绍，将学业成就调查的实施推向教育心理学的最前沿。关于学业成就调查的目标体系与命题技术、学业成就调查的问卷设计和组织实施，直接回答了这项工作的实际操作过程，特别是对目标体系和命题技术的研究，具有重要的创新性，并提出了进一步研究的问题。最后，该书回答了学业成就调查与教育教学改进的教育政策问题和教育教学实践问题，直接服务于教育质量的提高和教育公平的推进。

本书是团队合作研究的成果。尽管具体章节的执笔人各有分工，但每一章都体现着团队的集体智慧。各章写作的具体分工如下：前言：田慧生、孙智昌；第一章：单志艳；第二章：张布和；第三章：任春荣、杨希洁；第四章：史亚娟、孙智昌；第五章：孙智昌；第六章：孙智昌；第七章：单志艳、卿素兰；第八章：单志艳；第九章：张雨强；第十章：单志艳；第十一章：左晓梅、杨希洁、任春荣；第十二章：孙智昌。全书由田慧生、孙智昌负责统稿。由于学生学业成就调查是一项难度很大的研究课题，我们水平有限，书中错漏和不足之处在所难免，敬请各位专家批评指正。

本书的出版，得到教育科学出版社总编辑李东的大力支持，教师教育编辑部主任刘灿、责任编辑池春燕的高度责任心，使本书生色不少，在此一并表示感谢！

Contents 目 录

第一编 学业成就调查概论 / 1

第一章 学业成就调查的历史发展 / 2

第一节 学业成就调查的产生与形成 / 2

第二节 学业成就调查的发展 / 9

第二章 学业成就调查概述 / 15

第一节 学业成就调查的目的 / 15

第二节 学业成就调查的内容与技术考量 / 20

第三节 学业成就调查的主体 / 26

第四节 学业成就调查结果的分析与使用 / 30

第二编 学业成就调查的国际比较与借鉴 / 37

第三章 国际组织的学业成就调查体系 / 38

第一节 PISA 的学业成就调查体系 / 38

第二节 IEA 的学业成就调查体系 / 76

第四章 国外的学业成就调查体系 / 107

第一节 NAEP 的学业成就调查体系 / 107

第二节 日本的学力调查体系 / 138

第五章 中外学业成就调查体系的比较与借鉴 / 161

第一节 国外学业成就调查体系的特点与发展趋势 / 161

第二节 我国学业成就调查体系的现状与问题 / 182

第三编 学业成就调查的理论体系与实践策略 / 191

第六章 学业成就调查的哲学基础 / 192

第一节 学业成就调查的价值论基础 / 192

第二节 学业成就调查的认识论基础 / 204

第七章 学业成就调查的教育心理学基础 / 213

第一节 布卢姆的教育目标分类体系 / 213

第二节 SOLO 分类理论 / 228

第八章 学业成就调查的测量学基础 / 247

第一节 经典测验理论 / 247

第二节 项目反应理论 / 253

第三节 概化理论 / 257

第九章 学业成就调查的目标体系与命题技术 / 264

第一节 学业成就调查的目标体系 / 264

第二节 学业成就调查的命题技术 / 276

第十章 学业成就调查的问卷设计 / 293

第一节 学生问卷的设计 / 293

第二节 学校问卷的设计 / 302

第十一章 学业成就调查的组织实施 / 310

第一节 学业成就调查的组织测试系统 / 310

第二节 学业成就调查结果的分析与反馈 / 320

第三节 学业成就问卷调查结果的分析与反馈 / 327

第十二章 学业成就调查与教育教学的改进 / 340

第一节 学业成就调查与宏观教育政策调整 / 340

第二节 学业成就调查与区域、学校教育教学改革 / 350

参考文献 / 360

第一编

学业成就调查概论

学业成就调查的历史发展

学业成就是学生在学习课程的结果，是教育质量的重要内容。学业成就调查是人们关注和保障教育质量的重要措施。自有学校教育以来，就有对学生学业成就的考核，这种考核也是一种广义上的学业成就调查。但是，在现代教育的意义上，学业成就调查却是一个特殊的范畴。这种特殊性体现在它总是与人们对教育质量的关注，对教育公平的推进，对教育政策的调整，以及对教育教学的干预等紧密结合在一起。

关注教育质量、推进教育公平的学业成就调查并不是自古以来就有的，它有一个萌芽、产生、形成和发展的历史过程。深入了解学业成就调查的历史，对认识它的本质，进而提升教育质量、促进教育公平发展具有非常重要的意义。

第一节 学业成就调查的产生与形成

一、学业成就调查的产生

学业成就调查的历史可谓源远流长。就广义的学业成就调查而言，早在

公元前 2000 年, 中国官员的录取就采用了考试的形式, 可以说它是最早的学业成就调查。公元前 5 世纪, 雅典的教师在其教学方法中使用了评价性问题, 它也是学业成就调查的一种形式。^① 至于产生于我国隋朝的科举制度, 更是一种典型的古代学业成就调查, 后来演变为西方的文官制度, 对世界历史的发展有重要影响。

现代意义上的学业成就调查萌芽于美国。1845 年, 以霍勒斯·曼 (Horace Mann) 为首的美国波士顿学校委员会进行了后来被称为“波士顿调查”的学业成就调查活动。他们使用印刷好的试题开展大范围的测试, 试题主要集中于定义、语法、书写、地理、本国史、自然哲学、天文学、美术。调查连续进行了两年, 用以评估波士顿学校的教育质量。虽然有关的调查信息并没有用于学生学习的改进, 但它毕竟开了学业成就调查之先河。

从 1887 年到 1898 年, 美国的约瑟夫·赖斯 (Joseph Rice) 在一个较大的城市学校系统中, 调查了 33 000 名学生的拼写能力。根据调查结果, 他认为, 虽然人们很强调拼写教学, 但学生在学习上还是没有取得实质性进步。这一举措被认为是世界上第一次正式实施的教育项目评价。^② 同时, 该调查是在一种统一的标准化的情境下获知真实的学生学业成就信息, 至此, 标准化的测试已初见端倪。

学业成就调查与心理测量运动有着不解之缘, 心理测量、教育测量是它的源头之一。1895 年, 法国的比奈 (A. Binet) 和亨利 (V. Henri) 设计了一套智力测验。1905 年, 美国编制了著名的比奈—西蒙智力量表。此后, 以心理测量原理为基础的学业成就调查得到很大发展。1909 年, 桑代克运用统计学的“等距原理”发明了编制量表的单位, 编成了《书法量表》《拼字量表》《作文量表》等标准测量工具。这些都使学业成就调查在定量化、客观化和标准化方面取得了重要进展。此后, 各种学业成就调查的标准化测试工具应运而生, 为开展大规模学业成就调查提供了重要条件。

① 胡森, 波斯尔斯韦特. 教育大百科全书: 第 1 卷 [M]. 重庆: 西南师范大学出版社, 海口: 海南出版社, 2006: 567.

② 同上。

但是，现代学业成就调查自产生以来，就存在两个严重缺陷：一是一直到了泰勒创立“教育评价”概念为止，它所测试的只是学生对教材内容的记诵，而不是对高级智慧技能的测量；二是测量量表建立的理论基础是正态分布或高斯分布理论，其目的是要对学生的学业成就做出最大限度的区分，^①即它主要是为甄别和选拔学生服务的，而不是为了促进学生的学习和发展。从这两方面来说，它们还都不是严格意义上的学生学业成就调查，而只是其萌芽。

泰勒创立的“教育评价”概念，把教育评价与教育测量区分开来，不仅是教育评价发展的里程碑，而且也是学业成就调查的里程碑。如上所述，泰勒之前的学业成就调查，所测量的仅仅是学生对知识的掌握，没有测试学生的高级智慧技能。但泰勒证明，对高级智慧技能的测量和对知识的测量不是一回事。如果高级智慧技能是教育目标的话，那么，这些技能须直接加以测量。同时，泰勒认为，在20世纪初发展起来的所谓常模参照测验（Norm-Referenced Tests），一般并不能提供评价教育目标所需要的信息。起源于智力测验的常模参照测验是以区分考生为目的的，它对了解学生的学习进展并无多大价值。而与此相对的目标参照测验（Criterion-Referenced Tests），则可为判断学生学业的进展提供很多帮助。^②因此，泰勒不仅创立了学业成就调查的“目标为本”模式，而且还将学业成就调查与选拔性考试区分开来，从而确立了学业成就调查的基本原理。“目标为本”的测验至今仍在学业成就调查中占有重要地位。

二、学业成就调查的形成

学业成就调查作为保障和提升教育质量的措施，是在一系列条件的支持下才得以形成的。首先，随着初等教育特别是义务教育的普及，世界各国越

① 胡森，波斯尔斯韦特．教育大百科全书：第9卷[M]．重庆：西南师范大学出版社，海口：海南出版社，2006：709．

② 陈玉琨．教育评价学[M]．北京：人民教育出版社，1999：62．

来越关注教育质量，将质量作为教育存亡攸关的重大问题来看待。其次，教育评价理论和技术的发展，能够支持这项活动的开展。另外，现代教育的主体——政府，采取了有关的措施，以实际行动来保障教育质量。没有这些条件的支持，也就不可能形成学业成就调查这一历史现象。这些条件在 20 世纪 50—60 年代全都具备，学生学业成就调查的形成也就成为历史的必然。

1. 对教育质量的关注

20 世纪 50 年代以来，关注教育质量成了世界历史性现象。1957 年，苏联发射世界上第一颗人造卫星，震惊美国朝野。在深刻的反思中，美国人认为是本国的教育质量出了问题。1958 年，美国总统亲自批准颁布了《国防教育法》（*National Defence Education Act*），声明要加强普通学校的自然科学、数学和现代外语教学，要求大力更新教学内容，设置实验室、视听设备、计算机等现代教学手段，充实教学参考资料；加快外语教学中心发展，提高师资质量；强调“天才教育”，鼓励有才能的学生完成中等教育后攻读高等教育课程，培养拔尖人才；拨付 8 亿多美元的经费，支持教育发展，全力提升教育质量。

继《国防教育法》之后，美国掀起了大规模课程改革，即著名教育家布鲁纳发起的影响了全世界的结构主义课程改革。它强调早期教育，发掘儿童智力潜力；强调逐级下放科学技术课程，缩小高级知识与初级知识之间的差距；强调课程的结构性和学生的探究发现学习。其核心是提高教育质量。

1965 年，美国国会又批准了《初等和中等教育法案》（*Elementary and Secondary Education Act*，简称 ESEA），它与实用主义教育思想完全不同，与《国防教育法》一脉相承，仍致力于提高教育质量，强调小学的目标是加强普通科学文化知识教育，为将来接受专业教育打好基础；中学教育目标是培养未来的专家学者打好基础，学习科研方法，为高等学校输送合格生源。同时，它还强调白人和黑人学生合校，采取措施帮助处境不利儿童的学习，迈出了教育公平发展的重要步伐。

在美国教育改革的推动下，世界各国都掀起了以提高教育质量为核心的教育改革。如联邦德国对范例教学的强调；日本在 1960 年的《国民收入倍增

计划》中强调普及高中教育，充实科学技术教育；苏联 1966 年的教育改革强调提高普通知识水平和技术知识的重要性，提出要把提高专家培养质量当做头等任务抓紧抓好，等等。

对教育质量的世界性关注，为学业成就调查的形成提供了前提条件。

2. 教育评价理论和技术的基本成熟

20 世纪 30 年代，泰勒“教育评价”概念的提出和“目标为本”评价原理的创立，为学业成就调查的形成奠定了重要的教育评价理论和技术基础。此后，布卢姆教育目标分类学的问世、迈克尔·斯克里文（M. Scriven）形成性评价概念的提出以及项目反应理论（Item Response Theory，简称 IRT）的逐步成熟，共同为学生学业成就调查的形成提供了重要的理论和技术支持。

1956 年，泰勒的学生、著名教育家布卢姆等人发表了《教育目标分类学（第一分册）：认知领域》，1964 年，他与哈罗（A. J. Harrow）又发表了《教育目标分类学（第二分册）：情感领域》，辛普森（E. J. Simpson）和哈罗发表了关于动作技能领域的研究成果（1965 年、1972 年）。布卢姆等人的教育目标分类学将泰勒的思想进一步深化和具体化，如他们将人的认知过程分为知识、领会、运用、分析、综合、评价，将情感学习过程分为接受、反应、赋予价值、组织、性格化等，每个大类之下又分出亚类，如将领会分为转化、解释和推断等。这样，就为学生学业成就调查和考试评价的工具编制提供了基本依据，大大推进了学业成就调查的科学化。

学业成就调查之所以能够形成，还与教育评价领域将总结性评价与形成性评价予以区分有密切联系。1967 年，研究教育评价的又一重要人物斯克里文，正式提出形成性评价和总结性评价的概念。他认为，形成性评价是对教育教学过程进行诊断和改进的评价，而总结性评价是用于筛选和晋升的，其性质有很大不同。^①同时，他还提出了“目标游离评价模式”，强调评价不能被目标所钳制，要关注教育实际发生的效果，特别是非预期效应的产生。^②

^① 胡森，波斯尔斯特·教育大百科全书：第 1 卷 [M]．重庆：西南师范大学出版社，海口：海南出版社，2006：585．

^② 陈玉琨．教育评价学 [M]．北京：人民教育出版社，1999：65．

此后,李·约瑟夫·克龙巴赫(L. J. Cronbach)还将评价定义为为决策者收集和提供相关信息的行为。^①由此,不仅带来了教育评价的大发展,而且也必将为学业成就调查从教育评价中区分出来创造了条件。

20世纪60年代,项目反应理论逐渐成熟,并逐步运用于实践。它假定学生对项目(测试题目)的反应是学生在连续体中的位置和随机误差的一个函数。它能够诊断学生的学习情况提供技术支持,因为它能够将个人的特质(如能力)和项目都置于同一个维度上,有清晰的关于量表间距或然性的解释。^②这些都与形成性评价的思想相一致,从而解决了学业成就调查的测量和统计技术问题。

3. 政府保障教育质量的行动

现代教育作为国家主权,如果没有政府的责任行为,将是不可能的。同样,学业成就调查作为保障和提升教育质量的行动,如果没有政府的支持和参与,也是不可能形成和壮大的。

从历史来看,学业成就调查萌芽于个人行为,如美国的赖斯和桑代克等人开展的调查,日本的久保顺一于1951年开展的小学生数学学力调查等。^③但是,无论是从教育规模来看,还是从学业成就调查的庞大工作量与高度技术性来说,任何个人都不可能承担如此纷繁复杂的工作。因此,只有它成为一种政府行为时,才有形成的可能。

正是在这样的意义上,20世纪50—60年代形成了真正的学业成就调查领域,其标志是美国国家教育进展评价(National Assessment of Educational Progress,简称NAEP)的产生和国际教育成就评价协会(the International Association for the Evaluation of Educational Achievement,简称IEA)的成立。

NAEP产生于20世纪60年代。当时,联邦政府迫切需要了解国家基础

① 胡森,波斯尔斯韦特.教育大百科全书:第1卷[M].重庆:西南师范大学出版社,海口:海南出版社,2006:597.

② 胡森,波斯尔斯韦特.教育大百科全书:第9卷[M].重庆:西南师范大学出版社,海口:海南出版社,2006:710,713.

③ 田中耕治.教育评价[M].高峡,田辉,项纯,译.北京:北京师范大学出版社,2011:212.

教育的质量,并加大对教育的干预和支持力度。1964年,凯普尔(F. Keppel)成立了教育进展评估调查委员会(Exploratory Committee on Assessing the Progress of Education,简称ECAPE),通过调查实施全国性评价。1969年春,在卡耐基基金的资助下,ECAPE实施了科学、公民学以及写作三个领域的评估。同年7月,全国性评价的管理工作交给了各州教育委员会(the Education Committee of the States,简称ECS),美国联邦教育署资助ECS 200万美元支持该计划。从此,该评价体系得以确认,被命名为国家教育进展评估(NAEP)^①。

IEA原名国际教育成就评估委员会,后来改称国际教育成就评价协会,成立于1959年。它致力于大规模的学生学业成就跨国调查,是一种政府间组织。20世纪60年代以后,IEA展开了大规模的调查研究,获取了大量有关学生学业成就、态度及影响因素的比较数据,如1964年和1980年分别对12个和18个国家数学成就的国际研究,1970年和1971年分别对8个和19个国家的科学教育、阅读理解、文学教育、公民教育、英语及法语6门科目的调查等。^②

到20世纪70年代中期或之后,经济合作与发展组织(Organization for Economic Co-operation and Development,简称OECD)大多数成员国已经制定了关于学业成就的国家监测系统或一种新的制度。这些国家和地区有英格兰和威尔士、法国、荷兰、西班牙、苏格兰、瑞典等。^③ 尽管这些监测制度或体系还不是很成熟,但它们却正式宣布了学业成就调查领域的真正形成。

由上可见,学业成就调查作为教育评价范畴之一,是与世界各国对教育质量的高度关注、教育评价理论和技术的成熟以及政府直接参与教育质量保障运动密不可分的。可以说,正是在以上三股力量的合力作用下,20世纪50—60年代形成了真正的学业成就调查范畴。

^① Vinovskis. Overseeing the Nation's Report Card: the Creation and Evolution of the National Assessment Governing Board [M]. Washington, DC: U. S. Department of Education, 1998: 8.

^② 胡森,波斯尔斯韦特. 教育大百科全书: 第1卷 [M]. 重庆: 西南师范大学出版社,海口: 海南出版社, 2006: 627.

^③ 同上,第635页。

第二节 学业成就调查的发展

20 世纪 80 年代以来,随着面向 21 世纪的世界教育改革的兴起,学业成就调查呈现出前所未有的发展格局:调查的层面越来越多,规律性越来越明显,监测和提升教育质量、促进教育均衡发展、以公平和卓越为标准的新的教育质量观明确显现。全世界的学业成就调查正在迈向新的高度。

一、学业成就调查发展的规模特征

学业成就调查自 20 世纪 50—60 年代形成以来,经过 70 年代的发展,到了 80 年代,呈现出明显的大规模发展态势。一些国际组织所开展的学业成就调查已经形成规模和规律;一些国家的学业成就调查体系也逐步建立起来,开始规律性地开展调查;与此同时,学业成就调查与教育改革相互促进,成为世界教育发展领域令人瞩目的重要现象。

20 世纪 80 年代以来,IEA 所开展的学生学业成就调查,越来越引起世界各国的关注和参与。1977—1981 年,IEA 组织了第二届国际数学研究,1982—1986 年,又组织了国际科学教育研究。此外,IEA 还进行了作文写作研究(1983—1988)、文学阅读研究(1985—1994)、教室中使用电脑的研究(1987—1993),以及班级环境研究和学前教育研究等。自 1995 年以来,IEA 连续开展国际数学和科学评测趋势研究(The Trends in International Mathematics and Science Study,简称 TIMSS),每 4 年一次,主要考查世界各国四年级和八年级(分别以 9 岁和 13 岁的学生为主)学生的数学和科学学业成就,以及影响其学业成就的教育政策和教育教学因素。同时,它还开展了国际阅读素养进展研究(The Progress in International Reading Literacy Study,简称 PIRLS),每 5 年进行一次全球范围的阅读素养调查,用来监控儿童阅读能力的发展。现在,参与 IEA 调查的国家越来越多,如 2001 年就有 35 个国家和地区参与 PIRLS 阅读素养调查。