

TIMSS

国际数学和科学趋势研究 测评的理论与思考

赵 慧○著



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE

国际数学和科学趋势研究 测评的理论与思考

赵 慧○著



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE

图书在版编目(CIP)数据

国际数学和科学趋势研究测评的理论与思考/赵慧
著. —上海:立信会计出版社,2018.4
ISBN 978-7-5429-5753-5

I. ①国… II. ①赵… III. ①数学教学—研究—
世界②科学教育学—研究—世界 IV. ①O1②G40-05

中国版本图书馆 CIP 数据核字(2018)第 064605 号

策划编辑 王艳丽
责任编辑 王艳丽
封面设计 南房间

国际数学和科学趋势研究测评的理论与思考

出版发行	立信会计出版社		
地 址	上海市中山西路 2230 号	邮政编码	200235
电 话	(021)64411389	传 真	(021)64411325
网 址	www.lixinaph.com	电子邮箱	lxaph@sh163.net
网上书店	www.shlx.net	电 话	(021)64411071
经 销	各地新华书店		

印 刷	江苏凤凰数码印务有限公司	
开 本	710 毫米×1000 毫米	1/16
印 张	17.5	
字 数	244 千字	
版 次	2018 年 4 月第 1 版	
印 次	2018 年 4 月第 1 次	
书 号	ISBN 978-7-5429-5753-5/O	
定 价	69.00 元	

如有印订差错,请与本社联系调换

前言

如何做一名好教师？什么样的教学才是有效的教学？回答这样类似常规化的问题并不那么容易，而教育教学的发展实践也告诉我们，事实上也并非诸如教师具有丰富的知识加上教学经验这么简单化的解决方式就可以万事大吉，因为问题的背后蕴涵着许多理论问题，也就是学科教育的理论问题以及将理论运用于学校教学实践的问题。在世界范围内，随着教育评价理论的日见其丰，以及改进学校教育教学、提升教育质量的呼声高涨，产生了以国际数学和科学趋势研究（the third international mathematics and science study, TIMSS）等为代表的一批国际大型教育评价项目。在经济全球化的时代，各个国家都把打造高质量的教育体系作为提升国家竞争力的保障，因此，TIMSS 等大型教育评价项目一经施测，便引来众多国家积极参与，试图通过 TIMSS 的施测和诊断，找出自身教育发展的短板，为改进和发展开出良方。这也充分说明，在国际化程度日益加深的今天，任何一个国家的发展均离不开世界，教育的发展也是如此，利用大型教育评价项目审视自身教育体系，在国际坐标系中找到自身位次和发展方向，已经成为一种发展趋势。

由于教育经历与工作的关系，作者对教育评价领域具有一定的研究基础，也参与过该领域著作的研究与编写，但之前对 TIMSS 只能算

是概貌上的了解。随着国际大型教育评价项目在我国试点测评的几次实践,国际大型教育评价项目正式走进了普通人的视野,在全国的教育研究者、教育工作者、教育决策者乃至媒体热议国际学生评估项目(programme for international student assessment, PISA)的同时,作为一名教育评价领域的研究者,作者发现,TIMSS 同样是具有国际影响力的大型教育评价项目,我们对它的关注和研究却少之又少,目前所能看到的对 TIMSS 进行较为系统介绍的中文著作仅有一本,还是出版于 2003 年,现有的期刊论文中也仅有些零星的介绍。那么,TIMSS 究竟是怎样的一个测评项目?为何能吸引众多的国家参与一轮接一轮的测评?其魅力何在?这一连串的疑问促使作者以 TIMSS 为研究对象申报了相关科研项目,并在课题研究的过程中惊喜地发现 TIMSS 具有丰富的内涵,其背后的价值理念、理论分析模型、测评工具的开发及实施、数据分析、信息收集、测评结果呈现与解释等诸多方面都展示出科学和独到之处。作者在利用 TIMSS 数据、信息做研究的基础上也尝试使用其测评工具做了一个小调查,并在研究和实践的过程中受益匪浅,也反思颇多,因此萌生将这些研究和启发梳理、编撰成书的想法,希望为相关研究工作者提供一点参考。

随着我国教育开放的国际化程度不断提升,我国教育的变革和发展已与国际教育的发展息息相关,我国部分地区试点参加 PISA 的实践说明,我们在构建自己的教育评价体系中,充分关注并借鉴了国际上的先进做法。我国正处于全面推进学生综合素质评价过程中,需要扩大视野,以主动融入国际的心态,积极研究 TIMSS 等大型教育评价项目的跨文化、大系统和重现场(现实复杂性)的精髓,将会受益匪浅,最终构筑一个相互吸收、富有本民族特色的教育评价体系,以期实现党的十九大报告中提出的“落实立德树人根本任务、努力让每个孩子都能享有公平而有质量的教育”的目标。

本书的研究和写作得到了许多专家和同事的帮助和支持,上海市教育科学研究院张珏副院长对本书的研究框架给予了很多指导建议,上海市教育科学研究院陆璟副院长及上海市教委教研室的有关专家为

调查问卷的信息收集给予了指导并做了大量具体工作,上海市教育委员会原党委副书记长余国平、上海市教育科学研究院原党委书记江彦桥教授、上海市教育评估院原副院长陈效民教授对研究课题的设计、研究报告的撰写提出了富有针对性的意见和建议,上海市教育科学研究院高等教育研究所毛建茹同志对本书提纲的修订完善提出了具体意见,上海交通大学的陈建俞博士在数据的采集与分析方面给予了很多支持,在此一并表示感谢。也感谢立信会计出版社的王艳丽老师,她对本书进行了非常仔细的审读,并提出了许多建设性的修改意见,她负责的工作态度和良好的专业素质为本书的顺利出版提供了强有力的保障。最后,还要感谢我的家人,我的工作和研究都离不开他们的全力支持。

TIMSS 作为一项颇具国际影响力的教育评价项目,具有极为深刻而丰富的内涵,由于本人研究视野和水平有限,书中难免会有疏漏之处,敬请专家与读者批评指正,以便今后修改充实。

作 者

2018 年 3 月

目录

§ 1 大型教育评价项目的现状与趋势	1
1.1 大型教育评价项目概况	1
1.2 大型教育评价项目的发展趋势	9
§ 2 TIMSS 测评项目解析	14
2.1 TIMSS 的发展历程	14
2.2 TIMSS 的测评目的与测评模型	23
2.3 TIMSS 的价值与新变化	30
§ 3 TIMSS 2011 的设计与实施	51
3.1 评价框架与工具开发	51
3.2 样本的设计与实施	55
3.3 测评内容的翻译与校正	59
3.4 施测与质量保障	62
3.5 后续环节	65
§ 4 TIMSS 2011 测评结果与分析	71
4.1 TIMSS 2011 测评结构	71

4.2	TIMSS 2011 的测评结果与分析	89
4.3	参与国/地区的背景信息分析	114
§ 5	TIMSS 测评相关论题的探讨	124
5.1	数学和科学课程的变革	124
5.2	数学和科学教学的变化	134
5.3	非智力因素与学业成绩的关系	149
§ 6	TIMSS 测评的实践与启示	166
6.1	TIMSS 课程问卷调查实践	167
6.2	TIMSS 参与国/地区的教育改革案例分析	172
6.3	TIMSS 测评项目的启示	182
附录		192
主要参考文献		264

§ 1 大型教育评价项目的现状与趋势

高质量的教育是一个国家最宝贵的财富之一,也是国家提升竞争力和建立创新型社会的基石。基于此,世界范围内各个国家都给予了教育质量评价越来越多的关注,期望用可靠的、具有国际可比性的教育评价体系和手段实现对人才培养质量的监控,引导教育卓越发展。与此同时,随着教育评价理论与实践的不断发展,国际上逐渐形成了以 PISA、国际学生阅读素养进展研究(progress in international reading literacy study, PIRLS)、TIMSS 等为代表的大型教育评价项目。本章将简述这些大型教育评价项目的概况,并分析其发展趋势。

1.1 大型教育评价项目概况

大型教育评价项目作为教育质量监测的重要手段,具有可持续性、定期实施、循环进行等共同特征,但它们各自的评价目的、评价理念、评价对象、评价内容等不同,详述如下。

1.1.1 PISA 测评项目概况

PISA 是由经济合作与发展组织(organization for economic co-operation and development, OECD)发起与实施的国际比较研究项目,

主要评价对象是参与国(地区)即将完成义务教育的年龄为 15 岁的学生,测评他们作为社会人在多大程度上掌握了全面参与社会所需要的终身学习的能力,聚焦学生在阅读、数学、科学三个核心领域的素养。

OECD 于 20 世纪 90 年代中期着手开发 PISA 项目,并于 1997 年正式启动该项目。PISA 于 2000 年开始第一轮测评,此后每 3 年举行一轮,每一次在三个核心领域中选择一个作为本次比较测评的重点领域,其他两个作为辅助评价,每 9 年完成一个循环。迄今为止,PISA 已经举行了 6 次测评。同时,为了更全面地测评学生在现代生活中其他领域的知识与能力,PISA 还定期推出新的测评领域,如 2012 年增加了创造性问题解决能力和财经素养领域的测评,2015 年增加了协作性问题解决能力的测评。

PISA 所侧重的是学生在不同领域的技巧与能力,不一定与课程内容直接相关,PISA 测评内容考察的是学生在理解基本概念的基础上,在不同情境中运用知识和能力的程度。另外,PISA 还通过问卷调查收集有关课程实践、学生学习过程、学校背景信息、学生个人及家庭背景信息等内容,分析影响教育质量的因素,为学校教育系统的改进提供政策建议。

1. PISA 的价值取向

尽管 PISA 的评价领域与学校的学科内容紧密相关,但 PISA 与学业考试截然不同,主要在于其基于终身学习能力的知识和技能的考察。相比于学业考试内容,PISA 着眼于一个更广阔的范畴,在实际生活的场景中测评学生的习得技能和解决问题的能力。学生不可能在学校里掌握应对未来生活的每一种知识和技能,所以培养他们终身学习的能力尤为重要。简而言之,为生存和发展而学习是 PISA 测评项目的基本价值取向。

2. PISA 的研究目的

PISA 是在终身学习的动态模型基础上设计的测评,从世界范围汇集专家,经过多年实践形成了一套学生素养评价指标体系,为实现把 15 岁的学生培养成为积极、善于思考、有智慧的公民这一目标,从科学、数学、阅读素养的角度,考察各个国家的教育成效如何。PISA 作为一项权威性研究,其目的之一是开展学生成绩的国际比较,向政策制定者描

述结果。通过结果可以让政策制定者了解本国学生的知识和技能水平,看到本国取得的成绩在世界范围内的位次,学习和研究其他成绩较好国家的做法等。另一个目的是研究影响学生学业成就的因素,包括学生层面的学习策略、学习态度、参与度等,以及这些因素的综合作用;学校层面的资金投入、学校管理、问责、学校自主权、学校资源、学生入学、分组与选拔等因素;学生家庭层面的社会经济背景、移民因素等内容,全方位、立体化地对影响学生学业成就的诸多因素进行了分析,并提出了改进方法和建议。

3. PISA 的政策导向

PISA 不是全球性的学生竞赛,也非单纯的学院派研究,政策是其重要导向,正如 PISA 2000 报告中开宗明义地指出:“PISA 旨在为政策对话以及教育目标的界定与实施提供一个新的基础。”^①

OECD 历年发表的 PISA 研究报告为各国政策制定者所重视,其政策导向主要体现在四个方面^②:第一个导向是竞争力导向。PISA 给出各个国家(地区)在阅读、数学、科学等领域的成绩排名,以及达到各个精熟度水平(proficiency level,又称能力水平)的学生比例,使各国从别国的成绩中反思自己的教育,寻找更有效的教育政策。第二个导向是终身学习导向。终身学习的关键是学会学习,学会学习的关键在于具备知识技能基础、学习的内在动力(有兴趣)以及自主学习的能力(掌握学习方法和策略)。PISA 所测评的“素养”重在评价应用知识和技能的能力。PISA 还研究学生的学习策略、跨学科的问题解决能力以及他们对不同议题的兴趣。第三个导向是教育公平导向。OECD 特别重视研究不同社会经济背景的学生能否获得同样的学习机会,达到同样的能力水平。PISA 从三个方面来评价教育公平,即学习结果的平等;教育资源分配的公平,可以用学校教育资源指标和学生经济社会文化地位的相关性来衡量;克服学生背景的学习公平,指的是学生成绩受背景因

① 李刚,陈思颖. PISA 的政策影响:类型、方式及其启示[J]. 外国教育研究,2014(07).

② 陆璟. PISA 测评的理论与实践[M]. 上海:华东师范大学出版社,2013:8-9.

素的影响程度。第四个导向是效益或效率导向。指的是投入产出的有效性,包括教育经费的分配和使用效益,校内外投入学习的时间和学生表现相比较所体现的教育有效性。在这四个政策导向中,教育公平是优先关注点,并通过公平导向的政策扩展了公平的范畴。

PISA 凭借其先进的评价理念与技术获得了巨大的国际影响力,被越来越多的国家和地区认可并引入,国家(地区)参与数从 2000 年的 43 个发展到 2015 年的 72 个。2009 年和 2012 年,中国上海地区作为试点参与,2015 年,北京、上海、江苏、广东组成的中国部分地区联合体参加了 PISA 测评。

1.1.2 PIRLS 测评项目概况

阅读是学生获得信息的重要渠道,良好的阅读素养是学生拥有终身学习能力的必要条件,因此有关阅读的研究成为国际教育研究的热点。PIRLS 是由国际教育成就评价协会(international association for the evaluation of educational achievement, IEA)主持的国际性学生阅读素养比较研究。PIRLS 以小学四年级学生(年龄 9~10 岁)的阅读素养为评价对象。之所以选择小学四年级的学生来评价,是因为这个学段处在阅读的关键期,学生由学习阅读转型为通过阅读来学习。PIRLS 测评项目始于 2001 年,每 5 年进行一次,迄今一共举行了 4 次,最近一次的 PIRLS 2016 测评项目的参与国(地区)达到 50 个。

1. PIRLS 测评项目的核心

PIRLS 旨在评估儿童阅读素养发展趋势,通过国际比较,发现造成学生阅读素养差异的经济、社会和教育因素,从而进一步为政府改善自身的母语教育提供必要的国际基准和数据,为各国改善阅读教学及促进儿童阅读能力提升提供参考。基于此项研究的目的和测评对象的特点,PIRLS 测评项目的核心是对阅读素养(reading literacy)进行了界定。PIRLS 2001 将阅读素养定义为:理解和运用社会需要的或个人认为有价值的书面语言形式的能力,儿童阅读者能够从各种文章中建构意义,他们通过阅读来进行学习,参与阅读者群体,并进行娱乐。

PIRLS 2006 对阅读素养又做出进一步的描述:理解和运用社会需要的或个人认为有价值的书面语言形式的能力,年轻的阅读者能够从各种文章中建构意义,他们通过阅读来进行学习、参与学校中和日常生活中的阅读者群体,并进行娱乐。这一改动突出了学生在学校和日常生活中进行阅读的重要性,强调了阅读活动发生的不同情境。

由此界定可以看出,PIRLS 测评至少有两个鲜明的特征:一是非常关注阅读的目的与情境,即阅读者通过阅读要做什么,阅读活动发生有其特定的目的,如为了娱乐而进行阅读、为了文学体验而进行阅读、为了获取信息而进行阅读、为了学习知识而进行阅读、为了完成某项任务而进行阅读……不同的阅读目的在阅读活动的具体实践中体现为不同的阅读情境。二是将阅读与生活相联系。在 PIRLS 阅读素养评价中,阅读能力已经不再被局限于学科范围内,单纯地作为一种特定的学科能力,而是与生活紧密相连,融入现实生活的各个层面,如娱乐、学习、工作等。

2. PIRLS 的测评框架

PIRLS 测评框架主要包括两个部分:阅读测评和调查问卷。

基于对“阅读素养”的界定,PIRLS 确定了“阅读素养”需要测评的三个方面:理解过程、阅读目的、阅读行为和态度。理解过程和阅读目的是阅读测评的评价内容,阅读行为和态度则是通过调查问卷来评价。

PIRLS 评价框架确定了与被测对象相关的两个主要的阅读目的:为学习体验而进行阅读;为获取和使用信息而进行阅读。另外,框架还确定了被测对象在两种阅读目的中需要使用的四个理解过程:①关注并提取明确陈述的信息;②进行简单推论;③解释并整合观点与信息;④检视并评价文章内容和表达方法。以上每一个理解过程在两种阅读目的中都将受到测评,但不同阅读目的及阅读过程在评价中所占的权重不同。对理解过程、阅读目的、阅读行为和态度的评价构成了对学生阅读素养的实际情况的评价。

此外,PIRLS 还对影响学生阅读素养的因素进行了调查,即阅读学习环境调查,PIRLS 评价以问卷调查的方式展开有关阅读素养发展背

景信息的收集。问卷调查面向参加测评的学生以及他们的父母、教师 and 校长,收集的信息是有关学生在家庭和学校中发展阅读素养的经历与体验,国家教育部门相关人员所要完成的调查问卷是关于教育体制和阅读课程政策等信息的,调查要点包括以下五方面:①国家和社区环境调查,即国家对读写能力和语言学习的重视程度,人口统计学呈现的状况和资源,教育体制的组织和结构,小学各年级的阅读课程设置等。②家庭环境调查,即家庭的经济、社会和教育资源,父母对读写能力发展的重视,父母的阅读习惯与态度等。③学校环境调查,即学校的特点,学校有关教学的组织形态,学校的学习氛围,学校的人力、物力资源,父母对课程的参与等。④课堂环境调查,即教师的专业教育和发展状况,教师的个性特点、对教育教学的态度,教室的环境、结构特点和氛围,教材和教学技术,教学策略和教学活动以及评价方式等。⑤学生特征与态度调查,主要关注学生良好的阅读行为、积极的阅读态度和学习阅读的态度对阅读学业成绩的影响。^①

PIRLS 综合以上两方面的研究结果对学生“阅读素养”的进展情况
进行监控和分析。

1.1.3 其他的大型测评项目

需要指出的是,TIMSS 也是一项具有国际影响力的大型教育评价项目,由于其是本书的主要研究对象,在后续章节中将作重点分析探讨,因此本节不再赘述。除了以上大型国际教育评价项目以外,还有一些非国际性的教育评价项目也富有影响力,主要有美国国家教育进展评估(national assessment of educational progress, NAEP),澳大利亚在 2008 年推出的全国性评价项目——澳大利亚教育进展评价(national assessment program, NAP),英国对 7 岁、11 岁、14 岁学生实施的标准成绩考试(standard attainment tests, SATs)以及对 16 岁学生实

^① 郑飞艺,童志斌. PIRLS 评价内容及其对小学阅读课程变革的启示[J]. 外国中小学教育,2013(03).

施的中等教育证书考试(general certificate of secondary education, GCSE)等。本节将着重概述 NAEP 和 NAP。

1. NAEP 概况

NAEP 是美国国内的一项全国性学生学业成就调查,也被称为“国家成绩单”,它是了解美国学生知识和能力的一个大型国家级代表性评估。在《不让一个孩子掉队法案》通过后,它成为衡量美国全国、各州及越来越多的城区教育进展的重要工具。NAEP 由联邦统计机构——美国国家教育统计中心(national center for education statistics, NCES)来管理,国家教育统计中心是美国教育部的一个研究部门。同时,NAEP 也受国家评估管理委员会(national assessment governing board, NAGB)的管理,该委员会是一个由普通公民、专家和政治家组成的委员会,负责监督评估计划的走向。NAEP 实施的首要目标是向美国公众报告中小学的教育状况,促进教育质量和学生学业成就的不断提高。

NAEP 有两条主线^①:一是长期趋势评估;二是 NAEP 主测评,由国家级、州级和城区级测评组成。就长期趋势评估而言,从 20 世纪 70 年代初开始,美国大约每 4 年在全国范围内举行一次国家教育进展评估长期趋势评估,现在其评估重点是 9 岁、13 岁和 17 岁学生的数学和阅读水平,对应的学段是 4 年级、8 年级和 12 年级。长期趋势评估的主要优点在于,它从一开始就采用长时间跟踪的形式,并一直保持评估框架的稳定性,在对两个主要科目进行评估时能够获得学生一段时间内表现的连续性观测,但保持评估框架的更新与趋势性数据之间的平衡是其需要面对的一个重要问题。

NAEP 的特点主要有两个:①NAEP 的长期趋势评估是按照年龄取样进行的,主要是对来自 4 年级、8 年级和 12 年级学生取样。以科目测评成绩、教师的教学经验以及学校的学习环境等为依据收集数据。除此之外,与政策相关的因素,如性别、种族(民族)和收入等,也会影响

① 施耐德·马克.美国国家教育进展评估[J].考试研究,2011(03).

数据收集。②NAEP并不为学生个人或者个别学校提供分数报告,而是参照全国范围建立的学生成绩标准(分为基本水平、熟练水平、优秀水平三个级别),反映美国学生必须掌握的有关知识和技能,这是其最根本的特点。

2. NAP 概况

NAP 作为澳大利亚的一项全国性教育评价,其目的在于通过这个项目使得澳大利亚政府、教育部门和学校能够判断澳大利亚青少年是否达到了一些重要的教育预期。该项目启动于 2008 年,由联邦政府采用招标的方式委托澳大利亚教育研究理事会(Australian Council for Educational Reserch, ACER)和澳大利亚考试中心(Educational Assessment Australia, EAA)等机构承担。NAP 对中小學生进行持续的跟踪评估,以对全国各地的教育水平和发展作出定期的、系统的评价,包括澳大利亚教育、就业、培训以及青少年事务部(ministerial council on education, employment, training and youth affairs, MCEETYA)旗下的所有考试,其中比较著名的是全国读写算数统考(national assessment program—literacy and numeracy, NAPLAN)。

NAP 的题型有选择题、简答题和问答题。科学科目还有两道实验题,采取学生先集体分班分组做实验、收集数据,然后再单独回答的方式,考试时间为 45 分钟。NAP 结果报告有公共报告和技术报告两种,主要提供给政府和学校。

NAP 的主要特点可以概括为五个方面:①满足学生的特殊需求。理论上所有的 3 年级、5 年级、7 年级、9 年级的学生都要参加考试,但学生如有特殊需要,家长可与教师或校长商谈,视情况确定是否参加考试。②所有的政府和非政府机构都有义务发展 NAPLAN 的考试内容。③与家长紧密联系。④尊重学生的隐私,每个学生的测评报告都是严格保密的。⑤用途广泛。学生和家长可利用测评报告与教师讨论学生的学习成果与学习进展;教师可知道哪些学生需要更多的支持,而哪些学生学有余力,可增加学习内容;学校可利用测评报告判断学校在教学方面的优缺点,以此设立读写能力及数学方面的教学目标。

1.2 大型教育评价项目的发展趋势

随着教育评价水平的提升,判断一个教育体系是否先进的标准也在发生变化,越来越多的教育体系主动以世界范围内表现最为卓越的教育评价体系做参照。近年来,我国在制定国家教育标准、探索“管办评”分离的同时,也在学习和借鉴国际基础教育评价项目的经验。因此,分析和研究国际基础教育评价的发展趋势,具有重大的现实意义。

1.2.1 对国家教育决策的影响力逐步增强

国际基础教育评价由于科学性、规范性、广泛性等特点,吸引了越来越多的教育体系参与。许多参与评估的教育体系以评估结果为参考依据,进而解决自身的教育问题。对于宏观层面的教育决策者和中观层面的学校管理者而言,这些评估结果有利于其增进对本领域教育情况的了解,从而制定更为合理的教育政策。可以说,评估结果会影响教育政策的制定、实施,从而推动一个教育体系的不断完善,进而顺应本国乃至国际教育的发展趋势。许多国家和地区积极借鉴国际教育评价坚持的理念和采用的技术,将其纳入本国教育政策的制定当中。比如,在德国 2004 年进行的教育改革中,其小学、初中、高中水平考试的州际比较在某种意义上就仿照了国际大规模教育评估,即借鉴了国际教育评估的比较方式。虽然国家和政府仍是行为主体,但是在涉及世界事务,尤其是文化、教育等方面,社会各方可以参与并结成伙伴关系,参与全球治理。教育领域跨国、跨区域的评估,特别是诸如 PISA 等大型国际教育评估项目就是在全球治理的背景下实现的。在联合国教科文组织、OECD 等机构的牵头下,各个国家积极推动教育国际化发展,在多边合作中制定方针和协议。要了解教育治理的效果如何,各个国家就需要通过国际教育评价来了解自身教育情况,从而获得适合自身发展的教育模式和路径。

1.2.2 更加关注真实的社会情景

国际教育评价的主要目的在于帮助学生提升解决问题的能力。