

PISA 是一项大型系统工程,首先是设计测评框架;接着根据框架设计测评工具,包括试题本和问卷,以及包括项目管理、翻译修订、抽样、数据管理、考务、评分等各方面的实施手册,用精细化管理确保测试达到 PISA 技术标准的要求;之后组织试测,对试测结果进行技术分析,用于筛选和修订试题;其后再组织正式测试;最后得出研究结果和政策启示;根据研究结果再改进下一轮的研究。

上述过程中的各个环节是环环相扣的整体设计,一开始就考虑到后面的需要。例如设计认知测试的时候不仅要全面均衡地反映测评框架各个维度的要求,还要考虑下一个环节的试题分析技术方法,问卷的变量设计也要充分考虑多层统计的需要。又如 2007 年试测准备时不仅要确定 2008 年试测的时间,也要同时确定 2009 年正式测试的时间,以避免试测学生和正式测试学生年龄交叉重复,等等。

先进的理念、科学的方法、系统的设计都需要严格的过程管理来落实。PISA 2009 是有 60 多个国家和地区参与的大型研究,如果没有严格的程序规范,实施过程就会产生多方面的不一致性,影响数据的可比性。因此,PISA 不仅制订了严格的技术标准,还编制了一系列操作手册(有 40 本左右),并定期组织培训和工作会议,定期要求上报在线表格,以规范实施过程中的具体做法。

总之,PISA 是一项具有前瞻性的评价项目,我们希望通过参与 PISA,从国际的视角来审视上海学生是否为终身学习做好了准备;PISA 是一项在理论、技术、管理上具有先进性的系统研究,我们希望能借鉴 PISA 的理论、技术和管理方法,结合上海的实际情况,建立科学的质量监测体系;PISA 是一项以改进教育政策为目的的实证研究,我们希望借鉴这一视角,加强教育决策的科学性。

我们从 PISA 测评中学到了什么

张民选(上海市教委副主任、上海 PISA 项目负责人、上海师范大学教授)

2009 年国际学生评估项目(PISA 2009)测试,有很多单位参与了相关工作,包括教委的 5 个处室、4 个直属单位,19 个区县的教育局招办,以及 150 多所学校,还有很多评卷专家和质量监测专家,其中测试当天参加考务工作的领导和管理人员就达 375 人。为了一个科研项目,投入如此多的人力,在上海历史上可以说是很少见的。PISA 是一个要坚持很多年的项目,从 2006 年底上海市教委决定要参与 PISA,到 2007 年、2008 年做准备,整理和翻译大量资料,进行试测,2009 年正式测试,2010 年要出报告,之后还要为 PISA2012 的数学(主项)测试做准备,到 2015 年进行科学(主项)的测试。用这么长的时间、投入这么多人力去参与一项测试,是一件很不容易的事。

或许有人会心存疑问:用这么多的心血、投入这么多钱、花这么长时间开展 PISA 测评,究竟值不值得?关于这个问题,我想说的是:中国大陆,只有上海

正式参与了 PISA。PISA2009 共有 68 个国家和经济体参加,这些国家和经济体代表了世界 GDP 总量的 90%。虽然全世界大大小小的国家有近 200 个,联合国成员有 192 个,但参与 PISA 测评的国家和地区数量却很有限,几乎所有大国都参与其中。这里要思考的是,我们究竟可以从 PISA 中学到什么?我觉得从开始参与 PISA 到现在,我们确实学到了很多。

首先,一项测评可以完成多个目标。这在以前是没有的。之前很多测评都是所有人共同参加的,只为完成同一个目标,就是看学生学得好不好、这所学校所谓的质量高不高。一项测评解决一个问题,大家都很辛苦,而这次 PISA 测评同时要完成多项任务,包括评估学生的质量、教育的公平、学生的参与、政府的投入等。所以,如果一项测评组织得好,就可以同时实现多项目标。我们之前的想法是,如果是一个高相关的选拔性测试,就一定不能完成水平性的测试;而一个水平性的测试,就一定不能完成选拔

性的要求。换句话说,对质量的评估都要切成两块来考虑。现在我就觉得不一定如此,因为 PISA 告诉我们,不仅质量问题可以在一个测试项目中解决,有很多要素也都可以设计在里面,关键是有能力去设计。

第二,测评不一定会加重学生负担。如果一项考试要求所有学生都参加,就有可能从低利害变成高利害,还可能造成学生过重的心理负担。但 PISA 告诉我们,通过抽样性的测评,同样能够得到学校、区县、市政府乃至国家政府所希望得到的数据、达到所希望达到的效果。而且抽样不要求所有人都参加,这样就可以减轻学生的负担。这就给我们一个启示:以后开展质量监测,也不必要求全部学生参加。

第三,PISA 测评安排得非常精致。68 个国家和经济体共同参与 PISA,必然对信度和效度有很高的要求,所以测评就一定要设计得非常精致、非常公平。OECD 有那么多措施、标准、手册和培训,并且统一了思想,每一个环节都设计得非常精细,出差错的可能性就变小了。例如,一个学校参与 PISA 测试的只有 35 个学生,但 35 个学生中考卷就有 13 套。如果一个极端偶然现象出现,比如有 10 多个学生都缺席,学校就要重考。所以,每一个环节都要设计得很细致,尽量避免出差错。我们现在也天天在喊工作要精细,但却没有真正做到,因此我们还可以做很多事情。

第四,教育方面还可以有许多创新。我们目前的各种考试中,女孩子的成绩都比较突出,而男孩子的成绩就不怎么有优势(所以有人提出要“救救优秀小男孩”)。另外,诸如学生学习负担重、教学中出现的各种问题等都迫切需要得到解决。我们的语文测试中也许从来没有像 PISA 那样关注各种不同的情境。PISA 的试题情境可以分为个人的、公众的、教育性的、参加社会事务的等等,在我们即将要做的下一轮测试中,很可能就会有解决问题领域的测试。我们常说“要解决问题”,但首先要知道“解决问题”到底包含多少环节、有多少要素构成一个人解决问题的能力。所以,教育的创新,不管是测试、评价,还是教

育决策方面,都可以从 PISA 命题、考务组织、试卷批改与评分等方面学到很多东西。

邓小平早在上世纪 80 年代初就提出教育“要面向世界”。既然我们要走现代化道路,世界也已经为我们开辟了现代化道路,我们就要努力去学习。只有在学习的基础上,才能不断吸收世界各国先进的思想、理念与做法。中国人也创造了很多好的理念和做法,但我们没有经过工业的现代化,没有精细地做好每一个过程,因此,在“怎么做”方面我们经常犯错误。在中小学,经常听到“一切为了孩子”、“为了孩子的一切”、“为了一切孩子”等类似口号,理念很好,口号也提得很漂亮,但学校所做的却不一定是“一切为了孩子”,也没有“为了孩子的一切”。我们往往觉得,讲过了就好像做过了,想到达彼岸好像就到了彼岸,但很少有人真正去“做”。PISA 就给了我们“如何做”的启示。

最后,参与 PISA 测评到底是为了什么?总的说来,我们希望通过这个项目,学习到世界各国的先进经验,做好我们自己的基础教育改革,使我们的基础教育能够促进学生的发展,使学生能在更幸福、更愉快的过程中得到发展,使测评能够真正帮助我们的教育决策、教育改革和教育监测。所以,我们最直接的目的之一就是通过参与 PISA 测评和研究,同时通过参加全国性的测试,建立我们自己的基础教育质量监测系统,使这个系统能够反映中国特色,能够与国际的质量要求相一致,能够同时兼顾学生的学习能力、知识学习的掌握程度、知识的应用程度、促进学生的终身学习,使所有学生真正在公平和均衡的状态下都得到发展。世界已经向我们提出了“公平而卓越”的新挑战,我们的目标是要将学生培养成为一个中国的公民,一个走向世界的公民,一个负责任大国的公民,一个建立和谐世界的公民。这任重而道远。同时我也希望,上海的研究与实践可以为全国积累一点经验,为世界宣扬一些中国的经验,更希望通过这个项目使我们的教育能切切实实发生一些变化。