

# 第三方评估，你真的准备好了吗？

文/王静洁



第三方，已经成为一个越来越热门的词语：第三方支付可以让在网上素未谋面的买家和卖家，放心安全地进行交易，不用担心卖家卷款逃走，也不用担心买家“吃霸王餐”；第三方监理，可以给房子的质量与价值提供权威的鉴定，没有第三方监理的房屋甚至都不能上市销售；第三方信用评级甚至可以给一个国家的可信度评分，被标普评为“垃圾级”的俄罗斯，让一向强势自大的普京都愁眉不展。为什么越来越多的第三方不断涌来，到底什么是第三方，第三方又可以扮演怎样的社会角色呢？对于教育来讲，第三方同样不是一个陌生的新兴词汇，每一个赴美留学的非英语母语国家学生都绕不开的ETS的TOFEL考试，究其本质就是一项有独立第三方教育机构开展的评估。那么，对国人还稍陌生的教育第三方评估机构到底是什么，它又能够担当哪些社会职能？

## 一、第三方是社会文明进步的标志

“第三方评价”的概念最早起源于15世纪的欧洲，当时是为了保证物品生产的质量，后来这一概念就开始向不同领域拓展和延伸，究其本质，其实就是由第三者来对质量进行监控与评价。既然提出了第三方的概念，其中第一方和第二方分别是谁呢？从教育行业来看，最重要的两个主体是学校和教育主管部门，学校是教育实施的主体，而教育主管部门（例如，教育部，教育厅，教委、教育局等）则是教育管理的主体。学校和教育主管部门作为当事的双方，与评价对象——即教育质量——有着直接的利益关系，为了维护自己的利益或影响，第一方和第二方通常会依据自身的利益诉求实施评价活动，从而对评价结果的客观性和公正性产生负面的影响。

学校自主发展评价，是对学校管理、教学、教师发展等方面全面综合的评价，每年教育主管部门都会组织相关的专家进驻学校进行发展督导，但督导的指标却是由学校自己制定的。在真正的督导活动开始之前，学校常常会邀请相关专家对自己学校的工作进行指导，在这一来二去的过程中，专家对学校的工作就有了非常深入详细的了解，甚至会提出很多个人的调整意见，专家与学校的感情日渐深厚。等到教育评估和督导正式开展之时，学校根据自己制定的指标，逐个汇报自己被专家指导工作后获得的成效，然后

由督导专家进行评定。这就好比一个运动员参加比赛，但是比赛规则却是自己制定的，然后要求裁判根据自己制定的比赛规则进行裁定，选手的真实实力其实不得而知。

第一方和第二方评价的这一天然漏洞的存在，才催生了第三方评价的出现。教育是国家重要的公共服务事业，教育的质量关系到千家万户的利益，尤其是近年来社会也不乏对教育的质疑和拷问，上海连续两次获得PISA考试第一，为何引来教育行业内部人士如此大的反响，也正是因为第一次他们可以理直气壮地证明上海的教育质量确实领先于世界。

而在此之前，虽然经常有出入各种国际论坛会议的各种行业内专家不断宣传上海的基础教育丝毫不逊色于国外，但依然有大量家长以自己的实际行动——送孩子出国读书越来越低龄化——来宣告自己对国内教育质量的不信任。

引入第三方评估，实质就是设立真正公正的裁判，由裁判对观众（社会）负责，对行政部门进行监督和评估。从习总书记上任以来，政府简政放权不断深化，政府各条线的工作都在积极地将评估检查权交还到纳税人手中。教育改革走到现在，不能否认前人的探索和努力为教育为社会带来了非常直观根本的变化，但如何更进一步，寻找到新的方向，真正实现以社会公共利益为根本，为人民提供满意的教育，引入第三方评估是极为大胆和冒险的一招，也正是教育改革的

决心的体现。从下文列举的政策条款中可以看到，从教育部到各级教育主管部门，“都已将加快健全评估体系，将教育第三方评估纳入到教育质量监测系统中来”提上了日程。

### 政策链接：

“逐步建立健全教育行政绩效评估体系，……在做好内部评估的同时，要主动委托第三方开展全面、深入、客观的评估。”

“坚持以学校为主体、以学生发展为本位，健全多元化评价标准，积极采用现代化评价方法和技术，保证教育评价的科学性、规范性、独立性，切实发挥教育评价的诊断、导向和激励作用。……鼓励成立教育评价的行业组织，发挥其在评价机构的资格准入、业务指导、监督管理等方面的作用。”

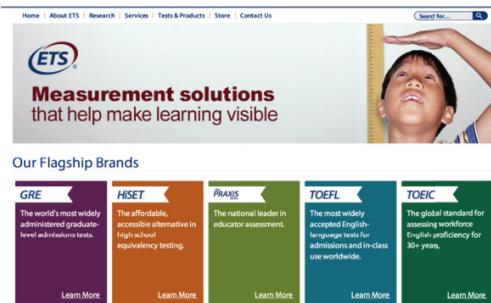
——教育部《关于深入推进教育管办评分离 促进政府职能转变的若干意见》

## 二、国内第三方评估尚处萌芽

目前国内承担教育第三方评估的机构主要有三种类型：

（1）教育主管行政部门的业务下属机构，类似于教育研究院、评估院、督导中心等。此类机构一般负责对该地区的学校办学质量、学生学业水平、教师专家发展等教育科研和教育评估工作的开展。参与工作的人员也往往是该地区教育领域的专家，他们以全职或兼职的形式负责对该地区的学校和教师进行专业指导与监督。当然，他们本身也可能就是来自于这些被指导与监督的学校。因此，这类机构虽然不是教育行政主管部门，但是其行动基本是遵循教育行政主管部门的工作要求的，其人员又大多来自于基层学校。所以他们的评估虽然在业内会获得比较大的认可与重视，但社会对“专家”固有印象已经不会为这些单位的评估结果买单了。

（2）高校教育研究机构，师范类院校或者是具备教育领域强势学科的高校，也会经常开展相关的教育评价工作。在某种意义上高校教育研究机构可以称之为第三方机构，因为一般高校都是独立的法人单位，并不直接受该地区的基础教育行政机构管理，但这类机构开展的教育评价往往出于两种主要的目的，第一种即为自己本身科研需求，那么这类教育评价必然受到研究目的的限制，在评价内容和评价时效上都有许多局限；第二种则为政府委托开展的教育评价，那么其性质其实与第一类机构并无本质区别。因此，这类机构在社会公信力上相较第一类机构并没有很大的优势，而基层学校往往又会因为高校多从理论研究出发，以不接地气等理由否定评估的结果。



（3）独立的社会机构，主要为民非机构或企业，这类机构不隶属于任何行政主管部门，评估工作也多以项目委托制的形式开展，即根据评估的目的与需求，单独制定针对性的评估方案，独立开展评估并递交评估结果。目前，成都、江苏、上海等地的教育主管部门也开始尝试和引入这一类型的社会机构，进行评估与调研。从第三方的定义来看，这一类机构是最接近第三方本质的机构，但其数量却是少之又少。目前，国内被邀请进行此类第三方评估的机构，大多为综合性的管理咨询公司，他们借助管理咨询通用的调研方法进行调研与评估。此类调研方法可以满足于类似于社会满意度等评估的需求，但若真需要深入教育核心内容，行业门外汉的评估与建议就只能隔靴搔痒了。

## 三、国内第三方评估的发展困境

中国是在人类文明历史上最早开始进行教育评价的。早在公元606年的隋朝，中国已经出现了科举制度，这是教育评价活动大规模实践的标志。但近代教育评价学科的发展，中国的参与却并不多。上个世纪四十年代，教育评价学科奠基人，美国教育家泰勒发表具有里程碑意义的“八年研究”时，我们正处于浴血抗战中，教育评价研究水平至此与世界拉开了距离。历经七十年的发展，以美国为代表的欧美国家教育第三方评估体系已经非常完备与成熟。文章开头提到的PISA考试就是由经济与合作发展组织（Organization for Economic Co-operation and Development, OECD）对其数十个成员国开展的基础教育质量的第三方评估，它抽取各个国家15岁的学生进行统一的学科素养考试，考试的结果即代表该国家（或地区）基础教育质量的水平。PISA考试的命题、实施与结果评估，并不依附于任何国家的教育体系，换句话说，PISA考试的内容完全独立于任何国家的教学大纲，而是从科学、阅读和数学素养的角度，由OECD组织的教育专家进行独立命题开展评估。那么，为何各个国家都依然非常信服PISA考试的结果呢，这便是第三方评估的奥秘所在。

### 案例链接：

全球最大的第三方教育评估机构——ETS

如果你想进入世界上大部分英语国家甚至是部分非英语国家的高等院校进行学习，TOEFL考试都是第一块敲门砖。TOEFL考试成绩关系千千万万出国留学、移民者的命运，是由谁组织命题和实施的？答案就是ETS（美国教育考试服务中心 Educational Testing Service），在如此官方的名字之下，其身份却是一家非盈利教育考试评估机构，也就是说它并不是政府下属的行政单位，仅仅只是为政府、学校提供第三方评估的独立机构。

ETS最著名的评估与考试有：

- TOEFL非母语者英语能力鉴定考试
- GRE美国研究生入学考试
- HiSET美国高中同等学力考试
- PRAXIS美国基础教育教师资格考试
- NAEP美国国家教育进步评价
- SAT学术能力评估考试（类似于美国高考）

从ETS官网上可以看到已经比较成熟的系统评估和考试共有50余项，其中不乏美国的高考、研究生入学考试、教师资格认定、甚至是学校综合绩效评估等带有非常权威官方形象的考试与评估。那么是什么让美国的政府和学校心甘情愿地出让如此重要的评估权，由一所第三方独立机构来评判大学申请者的水平，来评判谁可以成为美国的教育者，来评判怎样的学校是教育质量更好的学校呢？

ETS成立于1947年，由其他三个同样是非盈利的教育组织：美国教育委员会、卡内基教育进步基金和大学入学考试局共同在美国注册成立。ETS最初的目的就是为上述三个机构组织举办考试，包括国家教师考试、研究生入学考试、学术能力评估考试等。为何这些机构不单独自行组织原有的评估与考试，而要组合成立一个单独的考试机构呢？在ETS众多的考试业务背后，更加庞大的是其对考试和评估开展的学术研究。接受TOEFL考试的同学们可能都遇到过这样的情况，原定只需要完成3篇阅读理解，但是测试系统却偏偏要求完成4篇，并提示只有3篇文章会列入成绩的计分。多出来的一篇就是为供ETS机构研究所用，TOEFL考试的命题并不像我们的高考、中考一样，考完就结束了，所有的考试结果都会被进行二次数据分析和挖掘，以不断完善考试的科学性和准确性。这就是为什么我们的高考常常有大小年，而国外学校录取学生时只会说我们需要TOEFL分数超过80分的学生，并不需要指定是某一次TOEFL考试的结果。

ETS的学术研究主题：

- 残障人士的评估
- 英语学习的研究
- 提升教学质量的研究
- K-12创造力的研究
- 政策研究
- 评分技术研究
- 统计与心理测量的研究



了解了ETS的运作机制，出国申请者大概就不会再为1500元一次的TOEFL考试报名费而叫冤了，高额的报名费才支撑了全球最顶尖的教育研究者提供公正、科学、全面、反映真实能力的考试和评估。也正是由于第三方评估机构强大的研究实力，欧美国家大部分教育相关的考试与评估都已经委托第三方评估机构开展。

回望国内，虽然政策东风已经吹起，但是各类第三方评估机构依然凤毛麟角不成气候，真正具备官方信任，又有民间影响力的第三方独立评估机构还未横空出世。从目前国内开始出现的第三方独立评估项目来看，依然暴露出非常多的矛盾与困难。

(1) 第三方数据触动神经

目前最为人知的教育第三方评估莫过于各种各样的学校排行榜，从“世界大学排行榜”、到“中国优秀高中排行榜”，甚至连幼儿园都有“中国最顶尖的100所幼儿园排行榜”。这类内容往往是民众喜闻乐见的，以往高高在上的学校竟然被排名了，别人家的娃读的好学校其实不过世界排名200多。几乎每个家庭都有孩子正在上学，所以教育与每个家庭都息息相关，大家都会对教育质量特别关注。而在中国，相对来讲优质的教育资源依然稀缺，但父母望子成龙的心情却又是如此迫切，对自己能否占据相对较好的优质资源是非常敏感的。而在这些看似热闹的排行榜之后，其实普通民众并不知道公布这些排行榜的机构的专业性和可靠度。

案例链接：

作为全球公认最具权威性与影响力的三大世界性大学排名机构之一的“QS世界大学排名”（Quacquarelli Symonds World University Ranking，其他两大为“泰晤士高等教育世界大学排名”以及“世界大学学术排名”），它对大学的排名有非常严格透明的评估框架：

- 学术领域的同行评价（Academic Peer Review），占40%；
- 全球雇主评价（Global Employer Review），占10%；
- 单位教职的论文引用数（Citations per faculty），占20%；
- 教师/学生比例（Faculty student ratio），占20%；
- 国际学生比例（International Student Ratio），占5%；
- 国际教师比例（International Faculty Ratio），占5%。

其中每一项内容的评分，又有非常庞大的指标体系支撑，同时有非常严格标准的数据采集与分析的程序，其调研评估过程是相对独立而客观公正的。

另一个对各类学校排行榜趋之若鹜的则是校长。我们经常能够看到校长们在朋友圈转发各种排行榜，顺便加上自己的抱怨。先且不论这些排行榜的专业性和权威性，校长们抱怨背后的心态极有可能将未来正要出现的真正独立公正的第三方评估扼杀在了摇篮里。校长们的抱怨，很大程度上是源于自己学校的排名并没有预期的那么高，当然他们一定不会承认是自己的心态问题，他们一定能找到评估指标中“不科学”的地方，比如：“为什么要把学校面积作为评分指标，面积并不是我能够控制的啊？”“越有追求的校长对老师的要求当然越高，把教师

满意度作为对学校的评价，太不公平！”

“我们学校的家长工作是特色，可是这个排行榜对这块的评分占比这么低，其实我们比许多排在前面的学校都要好。”“怎么能绝对地从升学率来评价一所学校的好坏呢？”一旦被评价，校长们作为利益相关方，自然格外在乎排名，排名总有先后，于是就质疑指标的科学性，总觉得人家的优势占了便宜，自己的劣势吃了亏。

其实，第三方评估的初衷和存在的基础就是从完全独立的外部标尺来对质量进行评估。它没有必要，也不可能从每一个被评估者的立场去回答问题。如果过分的考虑个性化的因素，那就成了第一方或者第二方评估了。

第三方评估从某种程度上有着导向的作用，它立足于整个社会的需求，对教育提出考量标准，这些标准也许你内部人士认同的“专业”并不完全一致。但恰恰是这种立场差异产生的不一致，才是第三方评估重大的价值——把在“专业”路上走得太远，从而忘记自己为何要出发的教育者呼唤回来。

所以，神经已经过敏的学校和家长们，在面对第三方评估的到来时，首先要做的是摆正心态，第三方评估的结果，只是一种评估的视角，它能够发现问题，当然也可能忽略了一些信息。如果第三方评估机构能够在国内蓬勃发展，我们完全可以邀请不同类型的第三方机构共同进行评估，然后再综合不同的结果，形成全面的评估。

(2) 第三方评估机构的独立性无法保证

第三方评估机构存在的根本点在于其独立性。只有完全独立于第一方和第二方，第三方评估的结果才真正真实可信。而目前大部分第三方评估项目，是由政府委托的。由于第三方机构行业发展还较为弱势，评估内容和评估技术等还没有形成非常固定的实施方案。因此，行政主管部门的领导常常在委托第三方评估的同时，会对评估内容、评估方法提出个人的要求，甚至会细致到对某一个具体调研问题提出个人的修改意见。

另一方面，委托者对评估的结果也往往会有自己的预设结果，引入第三方评估机构的调研与评估，常常是想验证自己的经验与结论。当第三方评估结果与自己的

经验产生差异时，行政领导和被评估单位领导往往会去评估的内容、方式和数据分析的观点提出质疑，进而否定第三方评估，这在一定程度上也压制了第三方评估机构的中立性和客观性。



全社会各行各业都在积极倡导引入第三方评估机制，是为了让我们的行政部门和基层单位避免自行其是，客观地了解与发现自身存在的问题。虽然国内的第三方评估才刚刚起步，在技术和研究实力上确实还比较稚嫩，但如果没有好的培育土壤与宽容的态度，那么国内的第三方评估行业就无法健康快速地发展，真正的第三方评估就无法到来。

(3) 真正的第三方评估委托机制尚未形成

如前文提到，目前国内开展第三方评估的机构主要包含教育行政主管部门下属的业务单位、高校附属研究机构以及社会机构。虽然政策已经大力鼓励政府采购第三方评估机构的专业服务，但并未对第三方评估机构有明确的界定，同时在服务引进过程中对采购流程的政策也并未有明确的指向和支持。

因此，政府在选择第三方评估服务的时候，依然会倾向于选择自己系统内部的业务部门或者高校来进行第三方评估，以此来回避繁复的审批流程以及潜在的舆论风险。与此同时，正因为目前真正来自于社会的第三方评估机构的稀缺，一些政府公开招标的第三方评估项目，常常会因为没有明确的招标资质要求或者投标机构数量过少，而造成无法立项或者流标的情况出现。无法获得这些第三方评估项目，这些社会机构要么很难长期存活，要么就转为开展其他业务。

根据第三方评估机构的定义与初衷，其性质应该是中立的机构，并能够在提供第三方评估服务的同时，开展长期、科学的评估研究，为其评估服务提供数据支

持，以保持其科学性与客观性。显然行政主管部门的下属机构其实并不满足完全中立的定位，而高校研究机构又无法提供系统性的、全面持续的评估。因此，如果能对目前政策中提及的第三方评估机构的性质、资质要求有明确的界定。那么，社会第三方评估机构的生存状态应该会有较大的改善。

成熟的第三方委托机制是第三方评估的前提，更是基础。从国外成熟的教育第三方评估行业来看，第三方评估机构的直接委托方并不全是政府行政主管部门，而是由政府提供资金成立的独立组织，例如，基金会、非营利组织等。由他们设立一定的准入机制，去挑选适合的第三方评估机构开展评估工作。

同样最终都是政府出资，为什么要绕一个圈子才引入第三方评估服务？因为如果直接由政府行政主管部门采购第三方评估，其实在一定程度上依然无法避免两者的利益牵连。在委托关系中，甲方具有可以挑选乙方的天然属性，很难做到完全不去干涉作为乙方的第三方评估机构工作；第三方评估机构作为乙方，有时也会迫于要获得符合甲方预期的项目评价目的，而无法保证完全呈现真实的评估结果。

因此，由政府设立相应的预算，由社会和政府共同组建独立的机构，再由机构去组织选择最适合的第三方评估机构，由第三方评估机构开展独立调研，并对独立机构负责。虽然委托机制更复杂，但对第三方评估机构的发展和评估开展却有了更实际有效的保证。

政府对第三方评估机构的支持，是社会细胞自我反应能力的体现，也是社会文明进步的直观反映。许多创新的行政主管部门，从内在意识上已经认识到第三方评估机构的重要性，但面对评估结果时依然需要不断更新自己的理念，调整心态与过去经验判断的惯性做斗争。

当改革进入深水区，我们需要更多的是壮士断腕的决心，而接受第三方评估的结果有时也需要直面淋漓鲜血的勇气。直面教育第三方评估，政府需要打开心胸倾听不一样的意见，学校需要打开大门展示自己真实的情况，社会需要重拾信心相信评估的公正公平！