

上海市中小学生学习质量绿色指标综合评价改革研究

汪茂华

上海市教委教研室 讲师

摘 要

在教育改革的内涵发展阶段，教育质量评价是一个关键环节。学生学业质量是教育质量的重要组成部分，以学业质量评价为切入口，2011年上海构建了中小学生学习质量绿色指标并开展了两次测试。本研究报告了2012年测试的基本结果以及评价改革取得的初步成效，同时也提出了深化评价改革的设想与建议。

随着国家、上海《中长期教育改革与发展规划纲要》的颁布与实施，上海基础教育的改革与发展进入了一个新的阶段。实现教育公平、提高教育质量，促进教育的内涵发展成为当前的重要任务。教育质量是衡量一个国家、地区社会综合实力的重要指标之一，学生学业质量是教育质量的重要组成部分，以学生学业质量的绿色评价为切入口，全面关注学生的健康成长，引导全社会树立正确的教育质量观，建立教育质量保障体系，对于促进教育的均衡发展，实现教育公平，有着重大的作用。

一、学业质量绿色指标综合评价的背景

上海市开展义务教育质量综合评价改革研究，提出学业质量绿色指标，主要有以下几方面的背景：

1. 是上海教育实施《规划纲要》的需要

《国家中长期教育改革与发展规划纲要》（2010-2020年）中明确提到“改革教育质量评价和人才评价制度。改进教育教学评价。根据培养目标和人才理念，建立科学、多样的评价标准。”，这为教育质量综合评价指明了方向，并提出了明确的要求。

2010年3月，教育部和上海市人民政府在北京签署了《教育部上海市人民政府共建国家教育综合改革试验区战略合作协议》，根据《上海市中长期教育改革和发展规划纲要（2010-2020）》和部市共建国家教育综合改革试验区战略合作协议，上海市教委向国家教育体制改革领导小组申报并被批准了27项教育体制

改革试点项目。“改革义务教育教学质量综合评价办法”即为其中之一。

2、贯彻基础教育工作会议精神的需要

2011年3月，上海市召开了基础教育工作会议。会议明确提出了“五个转型发展”的目标。一是在教育价值取向上，从过度追求现实功利，转向更加追求教育对“人”本身发展的价值；二是在教育质量评价上，从过度注重学科知识成绩，转向全面发展的评价；三是在学生培养模式上，从高度统一的标准化模式，转向更加注重基于需求导向的培养模式；四是在教师专业成长上，从注重学科知识和教学技能，转向更加注重专业能力和教育境界的提升；五是在教育管理方式上，从单纯依靠行政手段，转向更加注重思想领导和专业引领。

五个转型发展的目标中，教育质量评价的改革起着重要的保障作用。评价的科学性与有效性，有利于树立正确的教育价值观，有利于基于学生需求实施多元化、个性化教育，有利于引导教师的专业成长，有利于推动教育管理体系的升级。

3. 是上海教育促进内涵发展的需要

“十一五”以来，上海基础教育坚持“先一步、高一层”的发展定位，深化课程与教学搞革，取得了较快较大的发展，在全国基础教育领域中具有一定的优势。处在转型发展的重要历史关口的上海基础教育，要保持优势、占领教育改革的潮头，必须在阻碍教育改革发展的评价问题上有新的突破。

在创新驱动、转型发展的过程中，上海基础教育面临着新的任务、新的挑战：优先发展教育、提高教育质量已经成为转变经济发展方式、维护和保障民生的重大战略选择；人民群众已经不再满足于有学上，而是要上好学，接受高品质的教育；上海基础教育必须应对国内外教育的挑战以及自身发展瓶颈的考验。比如说上海15岁学生参加PISA2009测试，在阅读、数学、科学三个方面均取得了第一名的成绩，但上海学生每周课内外上课时间的总量（34.8小时）位居参与国（地区）第12位。

二、学业质量绿色指标体系的主要内容

学业质量绿色指标体系的研究采用了演绎与归纳相结合的技术路径。首先从学业质量的内涵分析出发，逐步演绎，不断分解核心要素；其次，从上海连续五年参加教育部基础教育课程教材发展中心“建立中小学生学习质量分析、反馈与

指导系统”项目的调查结果中提炼出与学业成绩密切相关的因素，逐步归纳，不断聚类。同时结合上海的地方特色（如上海城市的创新精神、课程领导力、立德树人），最终形成了初步的学业质量评价指标体系。

学业质量绿色指标综合评价的内容主要包括学生学业水平、学习动力、学业负担、师生关系、教师教学方式、校长课程领导力、学生社会经济背景对学业成绩的影响、学生品德行为、身心健康和跨年度进步等十大方面。

1. 学生学业水平指数

学生学业水平包含学生学业成绩的标准达成度、学生高层次思维能力以及学生学业成绩均衡度。

学业成绩的标准达成度指的是学生在各学科达到合格水平以上的人数比例。高层次思维能力主要包括知识迁移能力，预测、观察和解释能力，推理能力，问题解决能力，批判性思维和创造性思维能力等。学业成绩均衡度指的是学生之间学业成绩差异的大小。

2. 学生学习动力指数

学生学习动力主要有四个方面，分别为学生学习自信心、学习动机、学习压力和学生对学校的认同度。

学习自信心主要通过调查学生对个人学习能力的评价、尝试解决困难问题的意愿、对取得优异学习成绩和完成学习目标的预期等问题。

内部学习动机的测量包含学生对学习本身的兴趣、对于学习目的和意义的认识等问题。

学习压力主要调查学生在学习过程中产生的心理负担和焦虑。通过询问学生做作业量的多少及难易程度、考试次数的数量以及学校公布成绩、考试之前的感受来调查学生所承受的学习压力的情况。

学生对学校的认同度主要指学生对学校的认可程度，包括学生的同学关系、是否愿意参加学校集体活动、是否喜欢学校以及在学校是否会感到孤独等。

3. 学生学业负担指数

学业负担的增加并不是提高学习成绩的简单办法，学生学习时间的增加与学生学习成绩之间没有明了、简单的关系，更多的学习时间并不一定带来学生更好

的学习成绩。通过调查学生的睡眠时间、做作业时间和补课时间来反映当前学生的学业负担。

4. 师生关系指数

师生关系的调查主要包含教师是否尊重学生，是否公正、平等地对待学生，是否信任学生等。

5. 教师教学方式指数

良好的教学方式能够对学生的学业成绩起到积极的影响。学生对教师教学方式的评价是通过采集学生问卷数据，运用统计方法得到的结果。问卷内容主要包括教师是否进行情境教学、鼓励学生动手实践等问题。

6. 校长课程领导力指数

校长课程领导力的调查分析包含三个方面，分别为课程决策与计划、课程组织与实施、课程管理与评价。

7. 学生社会经济背景对学业成绩的影响指数

父母受教育程度、父母职业、家庭文化资源等综合为学生社会经济背景。学生社会经济背景与学生学业成绩结合起来，分析家庭社会经济背景对学生学业成绩的影响指数，反映学校教育的作为。

8. 学生品德行为指数

良好的品德是个人成长、终生发展的基础，更是其成为社会有用之才的重要条件。学生的品德塑造是否成功，真正反映了学校教育的成功与否。主要包括学生的理想信念、公民素质和健全人格三方面，通过热爱祖国、自尊自爱、尊重他人、有诚信和责任心、遵守公德以及拥有关怀之心、公正之心等具体指标来调查。

9. 学生身心健康指数

学生的身心是否健康，关系到民族整体素质，关系到国家的未来与兴衰。学生的身心健康水平主要通过调查学生生理、心理和情感等指标来反应。目前，身心健康重点是研究学生的体质健康，包括身体形态、生理机能、身体素质等。

10. 跨年度进步指数

对学生学业质量的评价不仅关注以上 9 大方面的横向水平现状，还关注这些

指标的纵向动态发展变化。通过对区县、学校目前的学业质量与过去的学业质量进行比较，研究区县及学校在各项指标上的发展进步。

三、评价工具的研发及结果处理技术

学业质量评价选择的试点评价年级¹小学为四年级，初中为九年级。评价指标数据的获得主要是通过学科测试和问卷调查。

（一）学科测试工具

1、学科测试框架

学业水平测试的学科，小学四年级为语文和数学两个学科，初中九年级为语文、数学、英语和科学四个学科。各学科测试框架如下表。

表 1 中小学生学习质量学科测试框架

学科	内容		能力
语文	积累与运用		读准字音；认清字形；理解词义；积累和运用常见古诗文
	阅读	获取信息	整体感知；提取信息；形成解释；合理使用信息并做出评价
		获得文学体验	
		完成任务	
写作		能根据表达的需要，选取恰当的作文材料； 能有条理地安排文章的结构；能具体明确、文从字顺地表达自己的意思	
数学	数与代数 空间与图形 统计与概率	知识技能；理解概念；运用规则；解决问题	
英语	语音 词汇 语法	听、读、写	

¹ 由于测试时间是在 10 月份，因此实际测试的学科内容是三年级和八年级。

	话题 语言功能	
科学	生命科学 物质科学 地球、宇宙和空间科学 科学探究	回忆 理解与简单应用 问题解决

2、命题原则

为了保证学科测试工具的质量，在命题时坚持了五项原则：第一，命题基于课程标准，体现课程标准的基本理念和要求。注重考查学生对学科核心知识、技能的理解和掌握，尤其是学生综合运用所学知识解决实际问题的能力、收集与分析信息的能力以及对重要学科思想方法的理解与掌握。第二，根据课程标准、统计测量方法和大规模测试数据，确定学生合格、良好和优秀标准。试卷有足够的试题鉴别学生合格与否，是否达到良好或者优秀水平。第三，控制试卷的总体难度（小学是 0.80~0.85，中学在 0.75 左右）。组卷时坚持以体现课程标准对学生基本要求的题目为主，有难度的题目主要考查学生的高层次思维能力。第四，试题形式为客观性试题和主观性试题。多使用真实的情境和任务，注重通过客观性试题考查学生高层次思维能力。第五，命题过程及程序规范、科学，实行严格的质量控制。

3、评价试题样例

例 1 如图，在 34×12 的竖式中，箭头所指的这一步表示的是（ ）。

$$\begin{array}{r}
 34 \\
 \times 12 \\
 \hline
 68 \\
 \underline{340} \quad \leftarrow \\
 408
 \end{array}$$

- A. 10 个 34 的和 B. 12 个 34 的和
- C. 1 个 34 的和 D. 2 个 34 的和

本题考查的内容是数与代数，考查的能力是理解概念，主要是测试小学三年级学生是否能理解两位数乘两位数的竖式中每一步的含义。

例 2 夏天买回的豆腐，存放时间稍长就会变质。有人说，在水里加入适量的盐，再将豆腐浸泡在其中，就能延长保鲜时间。请设计一个实验方案检验这种说法？

表 2 评分标准

编号	回答
	较高水平的回答
20	<p>所设计的对照实验方案，考虑了空气中存放、清水中浸泡和盐水中浸泡三种情况，且有控制各对照组实验样品用量、放置的环境条件等因素相同的意图和措施。</p> <p>例 1：将一块豆腐分成相等的三小块，一小块放在空容器中，第二小块浸泡在清水中，第三小块浸泡在盐水中。三者放在同一地方，定期观察这些豆腐变味的情况(可用鼻子闻豆腐是否有异味)。</p> <p>例 2：将一块豆腐分成相等的三小块，一小块浸泡在清水中，第二小块浸泡在淡盐水中，第三小块浸泡在浓盐水。三者放在同一地方，定期观察这些豆腐变味的情况。</p> <p>例 3：将一块豆腐分成相等的四小块，一小块放在空容器中，第二小块浸泡在清水中，第三小块浸泡在淡盐水中，第四小块浸泡在浓盐水中（水量一致）。都放在同一地方，定期观察这些豆腐变味的情况。</p>
29	其他较高水平的回答
	一般水平的回答
10	<p>所设计的实验方案，考虑了空气中存放、清水中浸泡和盐水中浸泡三种情况中的两种，且有控制变量的意图和措施。</p> <p>例1：将一块豆腐分成相等的两小块，一小块放在空碗中，第二小块浸泡在盐水中。两者都放在同一地方，定期观察两小块豆腐变味的情况。</p> <p>例 2：将一块豆腐分成相等的两小块，一小块浸泡在清水中，第二小块浸泡在盐水中。两者都放在同一地方，定期观察两小块豆腐变味的情况。</p>
19	其他一般水平的回答。
	错误回答
70	<p>没有对照实验。</p> <p>例：将豆腐放在盐水中，看它是否变味。</p>
71	<p>有对照实验，但明显地没有控制变量。</p> <p>例 1：两块豆腐，一块放水箱，一块放盐水中，看谁先变味。</p> <p>例 2：买一些豆腐，放在普通环境下，看多久变味；再买一些豆腐，</p>

	放在盐水中，看是不是更长时间才变味。
79	其他错误回答（包括不切题、字迹不清等）
	不回答
99	空白

本题主要是在科学探究的视野下，考查初中九年级学生设计科学实验的能力。在科学探究的视野下考查科学实验能力，重点应定位在设计和实施科学实验。实验仪器工具的使用、实验技术技巧细节对于保证实验的顺利进行和成功虽然重要，但在学业测评中，如果命题重点落在这样的操作性细节上，测评目标就容易被误读。基于这样的思想，科学实验能力考查命题的重点主要放在了科学实验的设计上。特别是考虑到控制变量在设计实验中具有举足轻重的地位，因此将控制变量设计实验作为九年级学生设计实验能力考查的主要方面。

（二）调查问卷

1、调查问卷的框架

调查问卷共分为学生问卷、教师问卷和校长问卷三种。各类问卷的调查框架如表 3。

表 3 各类问卷的调查框架

问卷类型	主要维度
学生问卷	学生基本情况（父母受教育程度、父母职业、家庭文化资源）；学习自信心；学习动机；学习压力；学生对学校的认同度；学业负担（睡眠时间、做作业时间、补课时间）；师生关系；对教师教学方式的评价；品德行为（理想信念、公民素质和健全人格）
教师问卷	教师基本情况（学历、任职经历、职称）；教师对课程的看法；对学校管理的看法；面临的困难；教育观和教师专业发展；教师教学方式（因材施教、互动教学和探究与发展能力）
校长问卷	学校基本情况；校长办学理念和教育观；校长的学校管理方法；课程领导力（课程决策与计划、课程组织

2、调查问卷编制过程

问卷编制严格按照心理测量学调查问卷的编制程序进行开发。首先详细分析、界定了各个概念的具体内涵。广泛收集国内外与研究内容相关的各种问卷和量表,尤其是国际上多项学生学业成就测试中使用的问卷。其次,针对问卷中编制的试题实施了调研和访谈,较为广泛地征求了教育行政部门、校长、教师和教研员的意见,初步编写了问卷中的具体问题。同时,组织了心理测量学专家对调查问卷进行论证并提出具体的修改建议。最终依据项目分析和专家论证结果,对题目进行了修正、删减,形成了最后施测的问卷。

3、问卷试题样例

例3:你在家附近的超市买东西,在回家的路上发现收银员多找了5元钱。

(1) 你会把多余的钱退回超市吗?

不会 有时会 会

(2) 你认为你们班会有多少同学将多余的钱退回超市?

几乎没有人 小部分人 约一半人 大部分人 几乎所有人

本题旨在调查学生的品德行为中“利己的诚信和利他的诚信”。用纸笔问卷的方式来调查学生的品德行为在一定程度上受到了测量技术的限制。学生一般都有这样的信念,那就是每个人的行为都想迎合社会、家长、教师的意愿,知道什么样的行为是“好”的。因此,那些合乎社会规范或社会期望的行为很难反映一个人的内在特质。本题是利用投射作用原理而设计出的对个性心理进行测量鉴定的一种相对比较客观、准确、有效的测量方法。

(三) 体质监测

学生身心健康指数的调查是根据《国家学生体质健康标准》来开展测试。主要包括身体形态、生理机能、身体素质等3方面21项指标的内容。

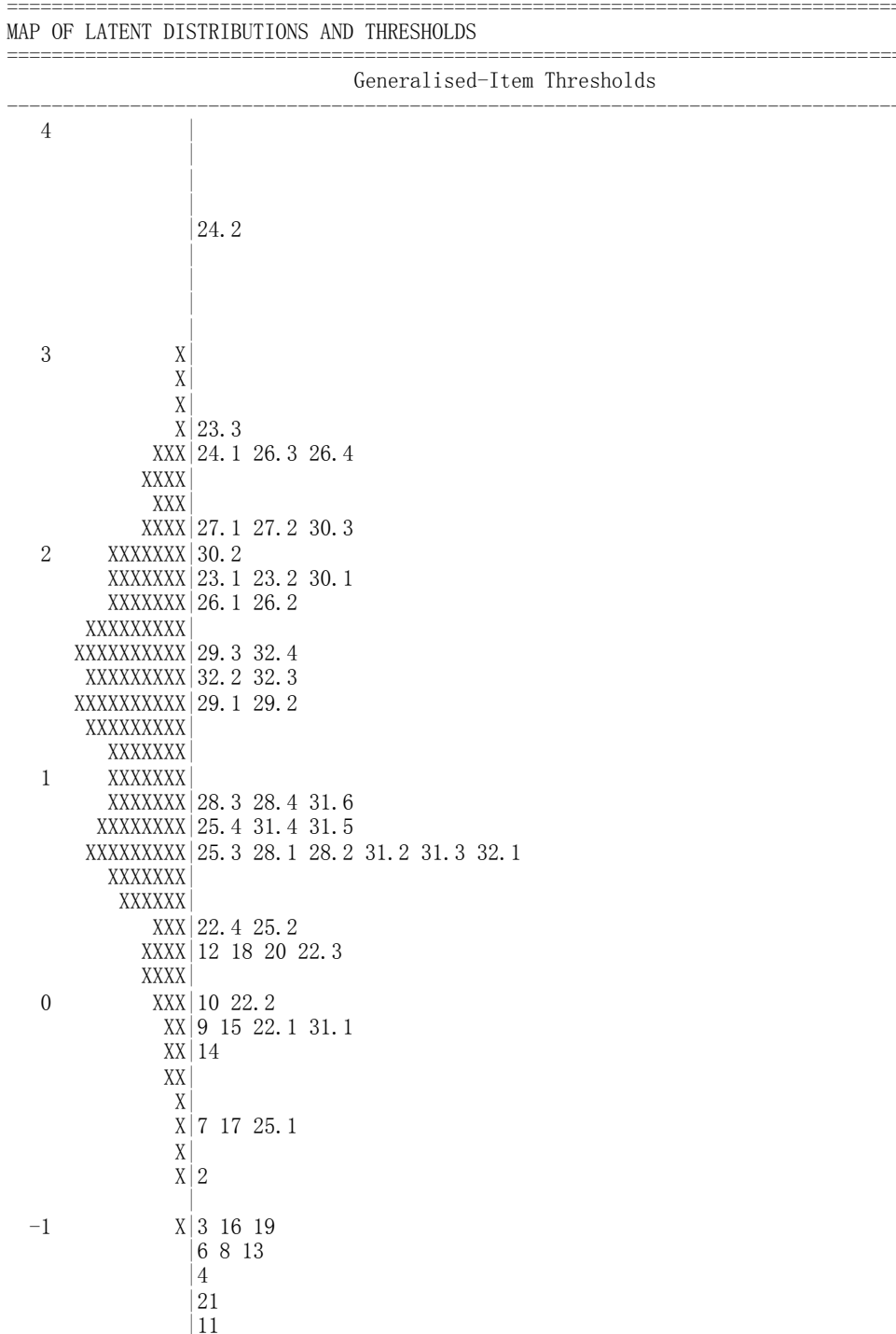
(四) 评价结果的处理

学生学业质量评价“绿色指标”项目广泛地借鉴了国际上大规模评价项目

如 PISA、TIMSS 和 NAEP 等数据处理的方法。

1、学业成绩的估计方法

学业成绩的处理主要是采用项目反应理论和经典测量理论对学科测试题进行分析,运用 RASCH 模型对试题的难度和学生的能力值进行估计。采用 ConQuest 软件对数据进行分析,分析的过程中,将试题的难度标定为均值为零,能力和难度标定在同一量尺上。得到考生能力分布和试题难度分布的对应图如下(图 1)。



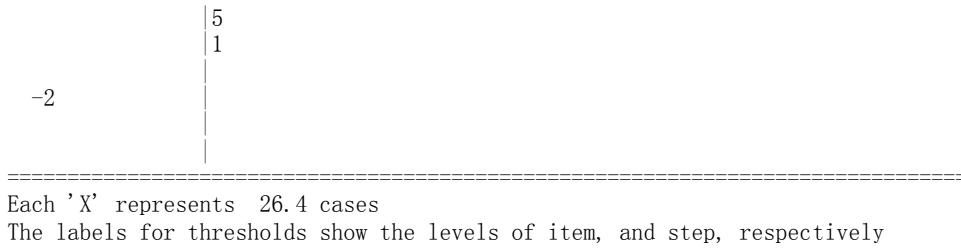


图 1 IRT 分析的试题难度与学生能力分布对应图

在项目反应理论中，学生的能力水平与项目的难度是处于一个共同的量尺之上，因此它们之间可以直接进行比较。图 1 是考生能力分布与项目难度分布图，图的第一列代表了量尺的刻度；第二列 X 代表学生的分布情况（注：每一个 X 代表了 26.4 名学生），它表明学生所处量尺的位置；第三列标明的是试卷中每一个题目的序号，其所处的位置表明估计得到的该题目难度的大小；对于解答题，题目后面所标数字表明该题目的第几个难度（如 24.2 表明序号为 24 的题目第二个域限值（难度值））。

此过程的分析可以得到具备等距量尺特性的 logit 分数，再通过等值技术，得到学生学业成绩的量尺分数。

2、学业水平的划分过程

采用国际上通行的 Angoff 和 Bookmark 方法进行学业水平等级划分，将学生的学业成绩划分为优秀、良好、合格和不合格的水平等级。学业水平划分的步骤主要包含以下几点：（1）选择标准设定的方法；（2）选择评委小组成员；（3）对评委小组成员进行标准划分培训；（4）实施第一轮标准划分；（5）对评委的评判结果进行统计分析；（6）反馈和讨论第一轮评判结果，并进行第二轮划分；（7）对评委的评判结果进行评价；（8）收集标准划分的效度证据和相关文档。

3、问卷结果的处理

问卷部分各个维度的调查试题大部分是李克特 5 点量表的形式，因此，对于正向的试题，按照“同意计 5 分，不同意计 1 分”的原则进行计分。反向的试题就按照“同意计 1 分，不同意计 5 分”的模式来进行统计。

例：你对以下问题的看法如何？

我很愿意参加集体活动

不同意 不太同意 不确定 比较同意 同意

如果学生选择“不同意”，则该生在这道试题上得 1 分，选择“比较同意”，则得 4 分。最后将学生在同一个维度的项目得分进行加总，就得到学生在这个维度上的总得分。

不是李克特 5 点计分模式的量表则采取其他的处理方法。

例：你母亲的职业是：(1)工人 (2)农民 (3)私营或个体经营者 (4)商业服务业人员 (5)政府工作人员 (6)军人 (7)教育、医务和科研人员 (8)企业管理人员 (9)进城务工人员 (10)其他职业 (11)无工作

此时需要将学生选择的答案与“国际标准职业分类 ISCO-88”职业编码中对应的职业进行匹配。通常情况下职业和社会经济地位 (SEI) 有一定的对应关系，根据文献资料中报告的处理方式，将学生父母的职业转换成相应的 SEI 分数，即社会经济地位的分数。

4、影响因素的分析

由于学业质量调查的数据结构具有嵌套性，即学生嵌套于学校，学校嵌套于区县这样两层的结构模型，因此在分析影响因素上广泛采用了多层线性模型 (Hierarchical Linear Model, HLM) 的统计分析方法。

多层线性模型分析的基本逻辑是以每个学生的学业成绩为学生层面的因变量，以学生层面的因素为预测变量进行分析，再在控制学生层面自变量作用的情况下，分析学校层面的变量对学生平均学业成绩的影响。模型表示如下：

第一层：学生层面

$$\text{学业成绩} = \beta_{0j} + \beta_{1j}(\text{变量 1}) + \beta_{2j}(\text{变量 2}) + r_{ij}$$

第二层：学校层面

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{变量 3}) + \gamma_{02}(\text{变量 4}) + \gamma_{03}(\text{变量 5}) + \dots + \mu_{0j}$$

四、学生质量绿色指标评价的样本情况

1. 样本区

2012年，全市17个区县均抽样参加了“绿色指标”测试。在抽样设计时，采用的是两阶段不等概率抽样方案，即第一阶段抽取学校，第二阶段在抽中的学校中抽取学生。在区县中抽取学校时按照学校性质及学制对学校进行分层，再根据与学校学生数成比例的PPS(probability proportional to size)抽样方法抽取学校，第二阶段按照有序系统抽样方法抽取学生。从而有效保证了各区县样本学生对该区县总体学生的代表性。此外，除全市统一抽样外，有9个区（黄浦、静安、徐汇、长宁、普陀、闸北、虹口、杨浦、奉贤）要求测试学校全覆盖，有3个区（松江、金山、青浦）要求不同程度增加样本校数量。

2. 样本校

全市有569所小学、428所初中（九年一贯制学校按小学、初中分别计数）参加了测试，分别占到全市学校总数的62%和68%。2012年参与测试的学校数量分别比2011年增加103所和90所；其中以招收进城务工随迁人员子女为主的民办小学66所，比2011年增加21所。所有“新优质学校”（43所）均参加了抽样测试。

3. 样本学生

2012年，实有四年级学生41672人、九年级学生50195人参加测试，分别占到全市该年级学生总人数的31%和55%。2012年参加测试的学生数分别比2011年增加了41%和47%；测试学生总数91687人，比2011年增加44%。

五、学业质量绿色指标综合评价的主要结果

1. 上海学生在达到课程标准的基本要求方面,表现出很高的学业水平。

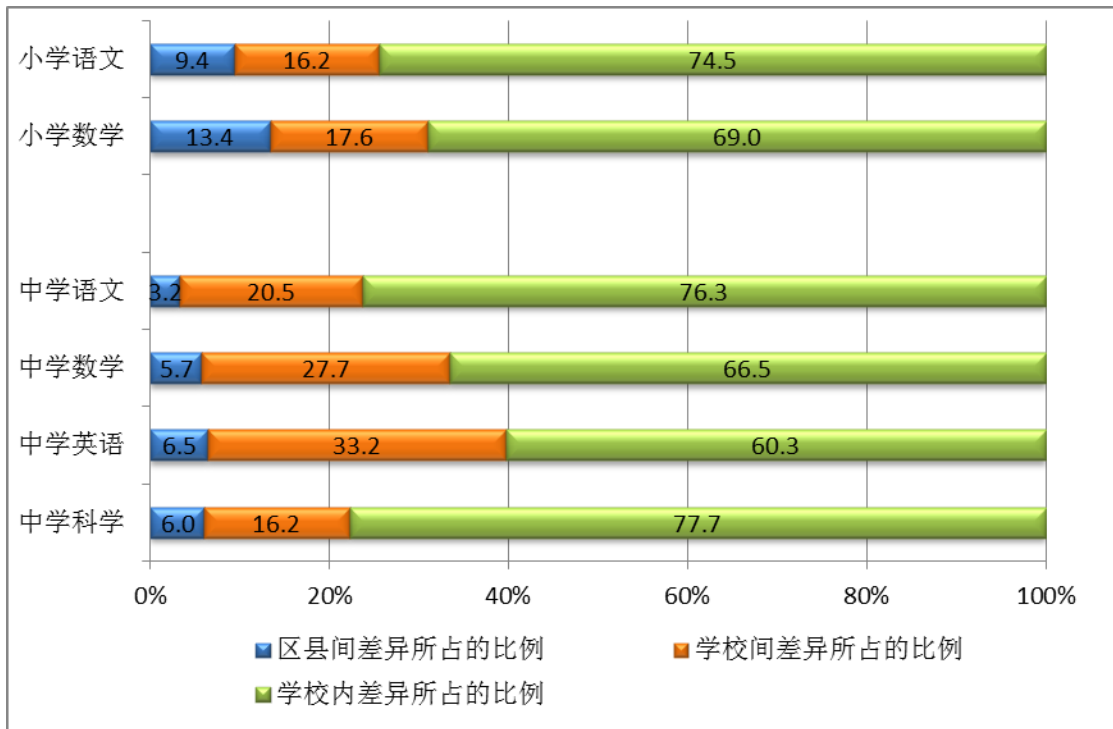
四年级有99%的学生达到了语文和数学课程标准的基本要求。九年级分别有94%、96%、94%和91%的学生达到了语文、数学、英语和科学课程标准的基本要求。

2. 学生学业成绩的均衡度处于比较高的水平上

将学生学业成绩的差异分成区县间、学校间和学校内（即学生个体）三个层

面，并通过分析区县层面因素、学校层面因素和学校内部（学生个体）层面因素对学生成绩差异造成的百分比（解释率），来判断上海市学生学业质量的均衡度。

总体而言，2012年由区县层面的因素造成上海市义务教育阶段学生学业成绩的差异，小学阶段高于初中阶段。在所有测试的学科中，中学英语和中学数学由学校层面的因素造成学习成绩的差异最大，分别为33.2%和27.7%。具体数据如下图。



区县、学校、学生三个层面因素造成学业成绩差异比例

从上图的数据可以看出，上海市学生学业成绩的均衡度处于比较高的水平上。

3. 学生学习自信心比较强

根据学生自我报告的学习自信心情况进行分析，四年级学习自信心较强的学生所占的人数比例为70%，九年级为41%。在四年级，学习自信心较强的女生所占人数比例为73%，男生为67%，女生比男生更加自信。在九年级，男女生之间则没有表现出这种差异。无论是四年级还是九年级，学生的学习自信心越强，学业成绩越高。

4. 学生内部学习动机较强

内部学习动机较强的学生所占比例小学四年级为 69%，初中九年级为 49%。对小学四年级，测试结果表现出学生的内部学习动机越强学业成绩越高的趋势，九年级则没有表现出这样的趋势。

5. 学生感到学习压力较大

无论是小学四年级还是初中九年级，学生们感到学习压力较大，而且不同群体学生之间几乎没有差异。连续几年以来，感到学习压力较大的学生所占的比例没有显著变化。以 9 年级为例，从 2008 年到 2012 年连续五年来，感到学习压力较大的学生所占的人数比例分别为 31%、33%、25%、24%和 28%。

6. 学生对学校的认同度很高

对学校的认同度比较的学生所占的人数比例小学四年级为 82%，初中九年级为 76%，学生对学校的认同度处于比较高的水平上。连续几年来学生对学校认同度呈现出增加的趋势。学生对学校的认同度与学业成绩的相关分析表明学生对学校的认同度越高，学业成绩越好。

四年级学生对学校的认同度，公办学校学生与普通民办学校学生基本相同，但都高于随迁子女民办学校学生。上海户籍学生要高于非上海户籍学生。与男生相比，女生对学校的认同度更高。

九年级不同类型学校、不同群体学生对学校的认同度基本相同。

7. 学业负担仍然偏重

调查数据表明，本市学生的学业负担仍然偏重。中学九年级学生的学业负担要高于小学四年级。减负增效的力度需要进一步加大。

(1) 睡眠时间

小学四年级平均每天睡眠时间在 9 小时以上学生所占的人数比例，从 2008 到 2012 年连续五年分别为 45%、47%、47%、44%和 42%。初中九年级平均每天睡眠时间在 8 小时以上学生所占的人数比例，连续五年分别为 14%、15%、11%、13%和 16%。

(2) 做作业时间

参照有关文件和相关研究成果，小学生每天做作业时间在 1 小时以下，中学生每天做作业时间在 2 小时以下为适宜。

四年级学生在做作业时间较适宜的比例上，公办学校学生所占人数比例为39%，民办学校为27%，公办学校高出民办学校12%。

九年级学生在做作业时间较适宜的比例上，公办学校学生所占的人数比例为40%，民办学校为22%，公办学校高出民办学校18%。

8. 学生对师生关系的评价较高

四年级对师生关系评价较好的学生所占人数比例为71%，九年级为44%。无论是四年级学生还是九年级学生，学生对教师的师生关系评价越高，学业成绩越好。

9. 学生对教师教学方式的评价有待提高

对教师教学方式的评价较高的学生所占人数比例，小学四年级为59%，初中九年级学生为45%。无论是四年级学生还是九年级学生，学生对教师教学方式的评价越高，学业成绩越好。

10. 教师对课程领导力评价尚可

在对校长课程领导力的评价上，对于小学教师，2012年对校长课程领导力评价较高的教师所占人数比例为72%，2011年为67%。对于初中教师，对校长课程领导力评价较高的教师所占人数比例2012年为52%，2011年为45%。校长课程领导力得到一定程度的提升。

11. 学生社会经济背景对学业成绩的解释率比较低

数据分析的结果显示，学生家庭背景因素对学生学业成绩差异的解释率都处于比较低的水平上。这表明学校教育给来自不同家庭背景的学生同等受教育权和教育过程的均衡化程度比较高。

12. 学生体质健康达标率较高

小学四年级学生的体质健康达标率为92%，初中九年级学生的体质健康达标率为86%，全部达到了教育部达标率高于85%的要求。

13. 学生对道德认知及相应行为表现有比较正确的认识

调查问卷的结果表明，大部分学生在热爱祖国、自尊自爱和尊重他人、利己的诚信和利他的诚信、爱护公共环境和维护公共秩序、责任、同学间的关怀与关

心弱势群体、公平对待他人等方面表现出积极的倾向。

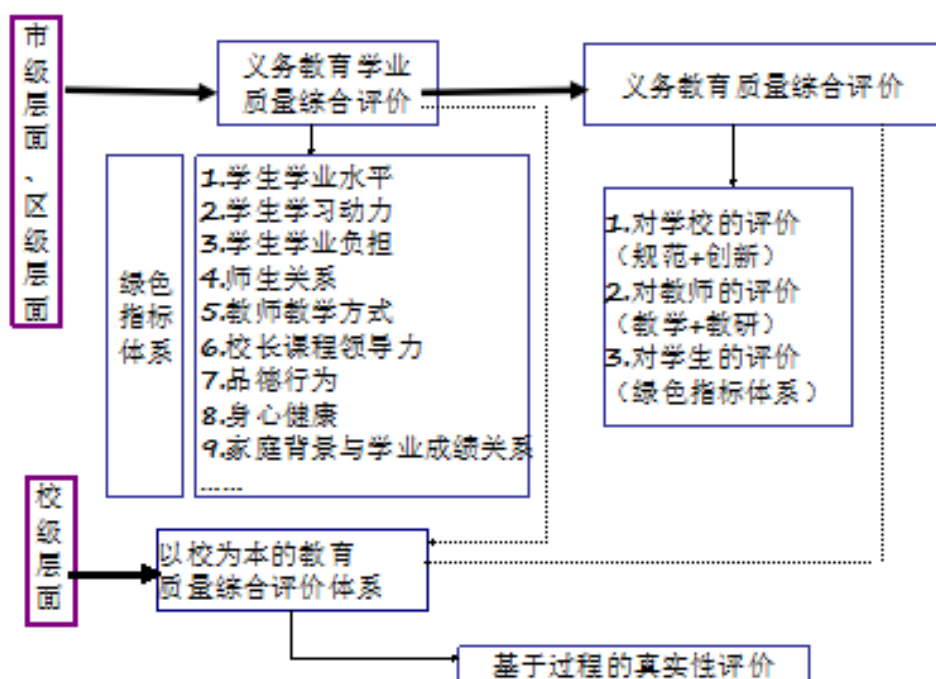
六、评价改革的初步成效与建议

（一）评价改革工作主要成效

随着实践的推进，“绿色指标”评价改革实施两年来，全市评价改革的思路 and 方向渐趋清晰，人们对绿色指标的认同逐渐加深，科学评价的强大正能量逐步凸显。

1. 形成教育质量评价改革基本框架

“绿色指标”发布之前，上海参加了两个测评项目，一是参与创建“建立中小学生学习质量分析反馈与指导系统”项目，并连续 6 年参加测试；二是参加 PISA 测试。两个项目评价结果都显示：上海学生学业成绩非常优秀，但学业负担较重，学习压力较大。因此，在关注学业成绩的同时，必须关注取得学业成绩的过程、方式及为此付出的代价，必须建立促进学生健康快乐成长的教育质量评价指标，引导区县、学校和社会突破以学业成绩和升学率为主的单一质量追求倾向，更加关注以学生全面而有个性发展为指向的全面质量。2011 年 3 月，上海市委、市政府召开基础教育工作会议，提出“在教育价值取向上，从过度追求现实功利转向更加追求教育对人的发展的价值；在教育质量评价上，从过度注重学科知识成绩转向全面发展的评价”。根据这一要求，我们研制了教育质量综合评价改革方案：



上海义务教育质量综合评价的基本框架

2. 构建有上海特点的绿色指标

学业质量是教育质量的重要组成部分，但不等同于学业成绩；它既是一个构成性概念，指学生学业表现、品德行为、身心健康等诸多领域的发展水平；又是一个关联性概念，包括影响学业表现的各相关因素；还是一个发展性概念，体现发展的均衡性、连续性和个体差异性。评价的目的和主要功能，不是为了排名，而是为了全面了解教育教学状况，发现和诊断教学问题，改进教学行为，持续提高教育质量。学习结果与取得学习结果的路径综合起来，教育教学改进架构到实证分析上来，被认为是教育评价领域重要的理论创新和制度创新。

3. 健全绿色指标评价实践的操作机制

上海市教委出台了《上海市教育委员会关于〈上海市中小学生学业质量绿色指标（试行）〉的实施意见》（沪教委基〔2011〕86号），保障了评价改革有序推进。在专业能力上，组建了部市联合的项目专家指导委员会，部市签订了上海基础教育质量综合评价改革项目合作备忘录，建立了市区两级专业化的评价队伍。在组织实施上，建立了由教育行政、教研、督导、考试等部门组成的工作小组，负责政策跟踪、学业测试、数据处理、质量分析、宣传引导等任务，保障了评价

改革有序运行。在专项经费上，市政府每年投入 800 万，连续五年用于项目的实践与研究。在操作程序上，构建了从测试框架制定到评价工具开发、科学抽样、封闭施测、结果处理、报告撰写、指导反馈、系统反思等一系列环节。目前正在谋划建设上海市基础教育质量数据库管理系统，已基本形成评价报告自动生成工具。

2011 年 10 月 27 日，本市进行第一次绿色指标测试，2012 年 7 月 1 日，发布 2011 年度“绿色指标”评价结果；2012 年 7 月 2 日，以“绿色指标”为主题举办本市中小学（幼儿园）校（园）长暑期培训；2012 年 8 月 28 日，向区县、学校集中反馈 2011 年度“绿色指标”评价结果，引导区县正确解读、分析和应用评价结果。2012 年 10 月 18 日，举行第二次绿色指标测试；2013 年 6 月 26 日，向区县、学校集中反馈了 2012 年度评价结果。

4. 评价的导向改进功能逐步显现

在“绿色指标”引领下，各区县都在积极构建区域教育质量评价与保障体系，许多区县安排了评价改革研究的专项经费，建立了区县的教育质量评价机构，配备了专业人员开始与市级层面的评价改革相协调。9 个区聚焦重难点问题开展评价改革子项目研究，有的重点研究若干指标，有的着重探索校园生活品质，取得了积极的效果。还有一大批学校开展“绿色指标”评价的校本化实践与研究。在全市绿色指标政策的引导下，黄浦、闸北、闵行、松江、青浦等区县出台区域贯彻落实“绿色指标”、推进减负增效的政策文件。

绿色指标评价在改进激励方面的功能也得到了较好的发挥，主要表现在：教育管理上，通过评价结果的分析，全面认识区域和学校学生学业状况，发掘教育教学的有效经验和不足之处，探求分数背后的影响因素，引导学校进一步理解绿色指标的内涵，变革教学方式，优化教学行为。教学研究上，通过解读评价报告，深度分析评价信息，给学校和教师提出具体细致、有针对性的教学建议。学校教学上，将评价信息，运用到下一阶段的教学改进中，逐步建立“标准→教学→评价→改进”的良性循环，逐步形成以校为本的教育质量保障体系。评价结果显示：一些原先不引人注目的学校，学业质量比较好，学生负担比较轻，展现了过硬的育人质量；有的择校热门学校，通过分析评价结果，也揭示了潜在的不足和隐患，如学生的校园生活质量、学生对学校的认同感，明确了改进和提高的空间。绿色

指标,既在一定程度上改变了家长、社会对学校“好差”口口相传的主观评价状态,又帮助每一所学校寻找持续发展的方向,从而不断推进育人目标的有效落实。

5. 评价改革初步营造了良性社会效应

“绿色指标”及其评价结果发布以来,受到社会广泛的关注,引发了教育内外更深度的热议,《人民日报》、《中国教育报》、中国教育电视台等媒体进行了相关报道,绿色指标也成为人民群众关注的热点之一。

2012年,国家教育体制改革领导小组印发简报《上海构建“绿色指标”评价体系,引导学生全面健康发展》(2012年第31期);中宣部理论局出版的《2012理论观点面对面:辩证看,务实办》中,绿色指标成为回应素质教育的重要命题。2013年,在“绿色指标”评价基础上,国家层面出台了《教育部关于推进中小学教育质量综合评价改革的意见》(教基二〔2013〕2号),在全国范围内推广“绿色评价”;教育部对国家教育体制改革试点项目推进情况进行考查,绿色指标成为素质教育领域4个国家级示范项目之一。

教育视界

今天怎样评价教育(上)

考分高就是好学生?升学率高就是好学校?在这个似乎“公认”的评价标准下,“减负”“通晓”“升学”随之,应试教育逼得素质教育节节败退,学生高分低能,学业负担过重,创新人才匮乏……
有识之士疾呼:“好学生”、“好学校”的标准应该改一改了,教育呼唤“绿色标准”!

什么是好学校?什么是好学生?

教育呼唤“绿色标准”

本报记者 张 杰

标准不变,改革只能沦为“花架子”

前不久,记者走进一堂中学语文优质课,课堂气氛热烈,学生兴趣盎然,短短的45分钟让人意犹未尽。这样和煦又高效的课堂,让我看到了中学语文教育可喜的变化,一位观摩课听课者感慨道:然而,更多的是一位教师并不认同,一位听课者直率地说:“考试分数直接决定着学生前途,这种方式如果不能提高学生考分,只能是‘花架子’。”

为什么会这样?原因并不复杂,旧有的评价方式并没有随着新课改的推进而调整。衡量好教师的标准,仍是能否培养出高分学生;衡量好学校的标准,仍是能否在考试中“所向披靡”;衡量好学校的标准,仍是升学率,“北清率”。在这样的标准下,任何雄心勃勃的改革,诸如减负和素质教育,都难逃尴尬。用上海市静安区教育学院附属学校校长张人利的话,就是“素质教育是素质教育,扎扎实实的才是应试教育”。

这也就不难解释,为什么初一、初二课外活动丰富多彩,到了初三就戛然而止了;为什么减负换来的却是家长的“集体不安”,纷纷给孩子报辅导班补课,甚至联名要求多布置作业;为什么明明知道在经历有害于孩子成长,家长还要逼着孩子去学奥数、“奥数”、“奥数”……
“评价对教育改革的意义重大。”上海市教委副主任尹后庆说,“长期以来,我们评价基础教育质量的‘法宝’,方法陈旧,对评价过程的跟踪和评价结果的运用不科学。”
如今,纸笔考试依然是我国义务教育阶段学生学业或成绩评价的主要方式,被教育专家称为“高利害测试”。尹后庆认为,“高利害测试或导致教师不得不带领学生进行超负荷的练习,学生在一个真正有趣的课堂这一系列问题之外,以测验为中心的练习活动赶走了。”他同时认为,高利害测试是把“双刃剑”,教师也是“受害者”,为了自身利益,必须让学生高分,于是,不得不把课堂变成“刷题”、“刷题”。“如果教师做的就是让学生去刷题地练习,那么,一些学生选择去大量放弃学习也就不足为奇了。”尹后庆说。

小学里成群结队的“小蜜蜂”,口袋里不完定的辅导书,稚嫩的肩膀上沉重的大书包,基础扎实但缺乏创新火花的孩子们……评价标准不变,指挥棒不变,一切努力只能徒劳无功。教育评价已经到了非改不可的地步!

绿色指标,考分不再是一选项

作为教育改革的先行者,上海市推出了“绿色指标”的概念,如果将高分低能称作“黑色GDP”的话,那么,促进学生全面发展和健康成长的指标就是“绿色GDP”。

“绿色指标”与传统的以分数为单一内容的评价不同,是对学生自主学习、幸福指数、身心健康、学业负担、教学方式的综合评价标准,使评价真正成为学生健康成长的助推器。”上海市教委负责人介绍。

翻开《上海市中小学生学习质量绿色指标基本结构》,“绿色”之绿扑面而来——绿色指标分为学业表现、身心健康和品行三个方面,共计10个“指数”,仅学业表现,就分为学生学业水平指数、学习动力指数和学业负担指数。

继续细化,每个指数包含若干具体内容。“学生学业水平指数”分为学业成绩达标度、高层次思维能力指数和学业成绩均衡度,不仅对语文、数学等学科的成绩做要求,还有对学业成绩均衡度的考量。“学习动力指数”占据重要地位,包括学习自信心、内部学习动机、学习压力和对学校认同度,其中,学习自信心是“对个人学习能力的自我评价,尝试解决困难问题的意愿、对取得优异成绩和完成学习目标的预期等”;内部学习动机是“对学习本身的兴趣,对学习目的和意义的认识等”;学习压力指数“指个人评价标准,指学习过程中产生的心理负担和焦虑,作业量的多少及难度、考试次数及学校公布成绩、考试前的感受等”;“学业负担指数”,则将课内时间、作业负担、补课时间纳入进去。
“根据绿色指标,如果学生负担过重,师生关系差,那么即使升学率再高,也不是好学校。”上海市教委基础教育处处长倪闽江说。

值得注意的是,绿色指标对学业质量的考量,突破了以往单一纸笔测试模式,增加了学生、教师和家长的问题调查,学生健康指数

和自评,并收集相关教育数据。学业测试结果不再是终结性评价,不与毕业、升学挂钩,旨在将测试或教学内部的一项活动,发现问题加以改进。“绿色指标的更大价值,在于从传统教育评价体系的‘结果证明’模式,转向‘过程改进’突破”,有专家指出。

2012年7月,上海市首次发布中小学生学习质量绿色指标,结果显示,在达到课程标准的总体要求上,上海学生表现出较高的学业水平,学习自信心和内部学习动机,但学业负担普遍较大,以睡眠时间为例,四年级学生睡眠9小时的不超过4%,九年级学生睡眠8小时的不超过13%。

“上海迈出了教育发展的关键步伐,标志着上海教育新起点,对学业的评价不再是仅仅以分数为唯一标准。”教育部基础教育课程教材发展中心主任助理刘教授表示。

标准再造,身心健康也是教育质量

近日,复旦大学研究生黄洋被投毒案热议如潮,联想起儿时的马加爵案,人们不禁要问:我们的学生怎么了?一位大校长痛心疾首地说:“无论如何,教育培养的是正常的人,培养正常、合格公民比什么人都重要、更迫切。”
是的,这个命题程度,远远大于考试成绩,我们需要的是身心健康的孩子。

让广大家长期待的是,上海绿色指标中,体现了关注学生健康成长的核心价值观,师生关系指数、学生品行行为指数、学生身心健康指数等,成为评价“好学校”、“好学生”的重要标准。师生关系指数考察教师是否尊重学生,是否公正、平等地对待学生,是否信任学生等;学生品行行为指数,包括学生的理想信念、公民素质、健全人格;学生身心健康指数,分为体质健康和心理健康两个方面。
“学生身心健康的理念是我们衡量教育质量的必要指标之一。”上海教育行政学院院长王卫华和学生个体情况。”尹后庆说。

绿色指标的考核让一些原本忽视的问题凸现出来。上海市中小学生学习质量绿色指标评价结果显示,一些学校学生成绩相对偏低,主要是由于师生关系紧张,教师教学方式不当等因素造成的。“这需要提高教师能

力,引导教师主动走近学生、关心学生。”绿色指标评价结果给出了这样的“诊断书”。

“这对校长、教师的导向作用很明显,让他们知道什么是教育品质,同时也提醒家长,光成绩好是有风险的,身心健康也是教学品质。”倪闽江说。

新的评价体系作用逐渐显现,在2012年上海市实验性、示范性高中“提前录取”面试中,很多学校将学生的可持续发展能力、学习意愿及身心健康状况纳入考核范围。

上海市中小学生学习质量绿色指标

- 学生学业水平指数
- 学习动力指数
- 学业负担指数
- 师生关系指数
- 教师教学方式指数
- 校长课程领导力指数
- 学生社会经济背景对学业成绩的影响指数
- 学生品行行为指数
- 学生身心健康指数
- 跨年进步指数

力,引导教师主动走近学生、关心学生。”绿色指标评价结果给出了这样的“诊断书”。

舆论处三

公众满意度能否由虚变实

教育是通向文明的通道,到今天为止,仍然没有任何一种手段能够替代教育,有被提高个人素质,完善个人修养,但生命之途这样一个过程经济环境下,教育似乎已经成为一种长期的投资,为公众提供多元的选择。于是,这样的情形日益多见:高分考生放弃北大、清华抛去的橄榄枝,选择香港高教院校;既出国留日且品行优秀生源外流,内地高校纷纷“喊‘渴’”,如此种种,正是公众对于优质教育资源作出的主动选择。

义务教育均衡发展如此,近年来,转身的中小学生在从北上广深,涌向中小城市,即便是北京这样优质教育资源相对丰富的地区,也在学生从中心城区、郊区等相对薄弱地区涌入建设。在城区等优质资源更为集中地,就读趋势,即便在同一区域,学生和家长的“择校”的方式选择和优质学校的现象也十分常见。

这是公众对于教育“用脚投票”,是公众对于教育资源和教育水平满意度的体现。我们常提及“教育满意度”,但究竟应该如何评价教育对于教育的满意度?公众对于教育资源的实际选择才最能真实的体现。

几天前,教育部发布实施全国县域义务教育均衡发展督导评估办法,这意味着,从今年起,全国2000多个县义务教育均衡的进程将记录在案并接受“国检”,通过督导和评估的方式,倒逼“县县均衡发展”。缩小优质学校和普通学校的差距,使得这一方面的同时,如何让“公众满意度”变为科学、准确、可量化的数据,成为教育主管部门未来考虑的重点。毕竟,如果只是在走访与问卷调查等形式,容易出现问题结果不够全面,不够客观的可能。

这让人们不禁想起“教育”制度,这一由美国经济学家阿尔弗雷德·弗雷德提出的理念,是指政府通过直接投入公立学校教育经费而形成的单位成本制,以后,以美国国定的评估法,即教育基的形成直接交给家庭和学生,学生教育基自由选择政府所认可的

人民日报报道上海“绿色指标”

（二）深化评价改革的设想和建议

基于实践的反思，我们深刻认识到“绿色指标”还需要在实践中进一步完善。教育部对上海评价改革有着更高的期望，社会对“绿色指标”的深入推进有更高的期待，这都需要不断努力，让教育评价真正成为教育转型的“指挥棒”。

1. 加强评价专业能力建设

向全国的辐射需要更强的专业能力，因此需要在现有的工作基础上，进一步整合更广泛的力量，如高等师范院校、教育科研机构、招生考试服务机构等。加强评价的理论建设、制度建设和队伍建设。特别是要提高市、区、校评价专业人员的专业能力。成立上海市基础教育质量评价专业工作组，为评价改革提供更有力的专业规划和推进指导。根据教育质量评价改革发展的要求和上海课改的实际，集合专业优势，开发高质量的绿色指标评价工具。

2. 完善基础教育质量评价体系

根据教育部相关要求，吸收最新的研究成果，优化“绿色指标”，特别是加强对品德行为指数、心理健康指数的研究，努力解决这两个长期困扰教育评价的难点问题。在网络阅卷系统和报告生成工具的基础上，建设功能强大的上海市基础教育质量数据库管理系统，形成基于信息化的评价运行平台，打造上海基础教育质量评价改革高地。

深入指导各试点区县、试点学校探索教育质量评价和保障体系建设，巩固市、区县、学校联动推进评价改革和提高教育质量的局面，形成各具特色的常态化教育质量评价机制，特别是要深化以校为本、基于过程的真实性评价体系建设，让教育实效建构在动态评价信息之上。

3. 优化评价结果的反馈应用机制

通过专题研究、专项培训、针对性指导，加大对“绿色指标”反映的教育教学规律的总结和推广，增强区县教育部门和学校正确解读、有效应用“绿色指标”评价报告的能力，切实改进相关工作。

围绕运用“绿色指标”评价结果改进教学管理、教学研究和学校教学，举办系列交流、展示活动，充分发挥科学“指挥棒”推动基础教育转型的作用。

4. 增强广大教师的学生评价能力

教育教学的真正变革，很大程度上依赖于教师专业素质的提高。针对当前教师教学能力较强、评价能力偏弱的情况，要将教师评价能力建设作为新时期教师专业发展的一个重点，着重加强教师对学业质量的分析应用、对学生动态表现的综合评价能力，以评价促进学生的发展。

5. 加深教育质量评价领域的国际交流

加强“绿色指标”与经合组织、美国、欧盟、日本等国家和香港、台湾等地区教育质量评价改革的比较研究，注重运用世界先进技术和方法完善“绿色指标”，使上海基础教育质量评价走向世界前沿。