



## “客观”记录是最大难点

高中生综合素质的档案，究竟应该记录多少内容，记录怎样的内容，是个众说纷纭的问题，仁者见仁，智者见智。但是，如果要做到《国务院关于深化考试招生制度改革的实施意见》所提出的“客观记录”，恐怕就超出了教育者讨论的范畴。长久以来，我们都心知肚明，教育者最不擅长的，恐怕就是“客观记录”了。在《实施意见》提出的各维度中，除了“学业水平”中的成绩，“身心健康”中的身高、体重、肺活量、体能测试等指标可以量化以外；如何客观反映学生在“思想品德”，“兴趣特长”方面的表现，就是难以攻破的堡垒了。可以想见，如果没有新的技术、新的思维方式的加入，学校最后出具的高中生综合素质档案，仍然只会是一本本加强版的“品德评语”，而这些不痛不痒的评语，对于招生学校来讲，提供的信息量殊为有限。

如果仔细分析“客观”记录中遇到的难题，大致可以分为以下三类：

**“所思”与“所为”的差别：**如果要实现客观记录，那对象必须是可观测的外显行为。但在有些领域中，具有代表性的行为可能非常少见。比如能够反映学生“思想品德”的行为，除了较为重大的违规违纪可以视为有悖道德，突出的好人好事可以被认为道德崇高之外，大多数人在成长经历中可能都没有相关的经历。如果单纯从外显行为进行评价，他们可能都处于中间的模糊地带。而对于这些学生来说，他们真实的思想道德水平，更多在于其在思想层面的认识。但是，这种发生在大脑中的“认识”，却是无法测得的——一方面是没有明显外在行为的表征，另一方面，如果用测试的方法去发现，智力正常的学生也很容易能够规避掉那些可能让自己被判定为“思想品德有问题”的选项。所以，当涉及到思想层面的评价时，客观记录是有其局限性的。

**“量”与“真实”的悖论：**在学生的成长过程中，有许多同质的行为会反复发生，比如阅读不同的书籍，去不同的地方旅游，不断在实验室中研究等。这样的数据，现在的信息技术完全有能力帮助我们分类和收集，学生阅读过多少书籍，参观过多少博物馆，加入过哪些社团等等。需要引起警惕的是，如果我们记录的，只是客观的“量”，那就很有可能引发一种误解：经历越多，综合素质也就越高——这对于习惯于追求高分、满分，讲求面面俱到的中国学校、家庭和学生来讲，几乎是一种定式思维。希望自己的综合素质评价结果更“漂亮”、“丰富”的学生，会大量“刷数据”，编织出一个并不真实的生活经历。如果这样，即使数据是“客观”的，其“真实性”也就大打折扣了。

**“过程”与“结果”的矛盾：**经历，真的是越多越好吗？教育者很难去判定，一个一年阅读过一百本课外书籍的学生，就一定比一年只读十本书籍的学生更博学；也很难判定一个自始至终只参加过一个社团的学生比三年中更换过十个社团的学生，在某个领域就一定更有钻研、水平更高。综合素质评价，让学生得以摆脱单一的结果导向式的考试评价方式，但对“过程”的重视，也同样会带来一系列的问题——“过程”比起“结果”而言，内容更加宽泛，指向更加多元。如果只是在档案中简单汇总与呈现，或许还可以得到普遍接受。但是当“过程”要被纳入评价时，其受到的争议必然不会小于对“结果”的评价。

“ 高中生综合素质评价的核心难点，在于技术储备。 ”

究竟在依靠什么知识生存，又在依靠什么知识创新？那些茫然的学生，可能从走下流水线的那一刻起，脑中装着的已经是过气的知识与理念……

教育，可能是社会上牵涉人群最广，最为庞大，效益又最难显现的一个体系。身处巨变的时代，在以上种种因素的作用下，原有的高考评价方式处于的窘境日益明显。所以，“高中生综合素质评价”的概念产生可以说是顺应社会需求的。

综合素质评价，其实并非一个新鲜的概念。甚至可以说是教育者长久以来的共同诉求。之所以其难以真正发挥作用，其关键在于对于方向的选择，也并非对于综合素质内容的界定，而是缺乏有效的技术手段来实现学生成长过程的“客观”记录。

## 高中生综合素质评价背后的技术储备

文 | 张韫

《国务院关于深化考试招生制度改革的实施意见》中指出：规范高中学生综合素质评价。综合素质评价主要反映学生德智体美全面发展情况，是学生毕业和升学的重要参考。建立规范的学生综合素质档案，客观记录学生成长过程中的突出表现，注重社会责任感、创新精神和实践能力，主要包括学生思想品德、学业水平、身心健康、兴趣特长、社会实践等内容。严格程序，强化监督，确保公开透明，保证内容真实准确。

回望千年中华，不论兴衰沉浮，其对人才的选拔，始终为历朝历代所重视。从察举、征辟，到九品中正，科举；不同的选拔方式，无不表达着特定的时代与社会背景下，国家对于人才的审美与需求。由于国家选拔制度关系着社会阶层的流动可能，被千千万万寒门之家视作改变自身命运的仅有通道，也被名门世家看做光宗耀祖或捍卫门楣荣誉的象征，所以其变革就格外吸引万众瞩目。

任何一种选拔制度，都很难达到尽善尽美的效果。究竟是关注人的品行，还是强调人的才能；究竟是关注纸面的文采，还是关注实践的经验；这些问题始终在如潮的争议中来回摇摆、变迁。

时至当代，人才的选拔早已不仅仅是为“学而优则仕”这单一的价值服务，而是为各个社会部门，各行各业挑选潜在人才。而一旦人才的出路变得越多元，对选拔制度的考验也就越大——因为毕竟很难用一种统一的、单维的考察标准来为各种不同的专业、职业输送潜在人才。同时，随着社会文明程度的提升以及义务教育的普及，“知识改变命运”的思想就成为了大多数人的共识，于是在追逐“知识”的过程中，大量的家庭、学校、学生本人都有意无意将无法在考试中体现出来的“能力”都慢慢抛弃。其结果是，每年有数以千万计的“百无一用”的书生，在教育流水线上被塑造出来。更尴尬的是，在移动互联网时代，“知识”在以从所未有的速率进行着更新换代——今天，人们



## 可行性优先还是等待完美？

不管从哪个角度进行考量，不管有多少反对与犹疑，现代信息技术介入学生评价，几乎是毋庸置疑的发展趋势。而这个事实的背景在于，任何教育者都无法以人力来实现对学生成长经历的细致记录，而且从个人视角进行的记录也越来越难以取信于人。上海作为本次高考改革的试点区域，在其发布的高考改革方案中就已明确指出，需要“启用高中学生综合素质评价信息化平台，建立客观、真实、准确记录信息的监督机制”。

当然，我们必须看到，信息化方式实现科学的教育评价，仍然不可能是一蹴而就的。这是由多方面因素共同造成的。首先，相比于其他各个社会部门，教育领域的技术储备极其薄弱，从信息技术到评价结果之间，没有成熟的研究，不存在公认的科学计量模型。教育评价依旧是“学科考试+主观经验”。

而基数庞大的教育者中，信息技术素养与数据应用的水平参差不齐，对于新鲜事物的接受需要经历漫长的过程——更何况，这样新鲜事物将要直击教育中最核心的问题。

在这样的形势下，教育者对于信息化的诉求会呈现出两极分化的现象——部分人会怀有不切实际的乐观，认为信息化将解决教育评价中长期存在的不够客观与不够真实的弊端；另有部分人会怀有盲目的抵制情绪，认为信息化只会带来新的工作负担，或者只是花里胡哨的噱头，无法体现教育的内涵价值。

要推进技术的应用，就需要对技术能带来的改变与效果有一个适当的预期。如果跳出教育进行观望，不难发现移动互联网的普及，已经大大加速了各行各业的发展。过程性、实时化的简易记录与智能分析，都在大数据时代的移动终端上得以体现。

要记录一个学生在高中三年内经历过哪些实践活动，将时间花在了哪些感兴趣的活动中与课程上，都是完全可行的。唯一的问题在于，在这些数据有可能影响高校录取决定的巨大利益驱使下，学生、学校、家庭有没有可能为了获取一个看上去更“漂亮”的档案记录，而去炮制许多并不存在的经历？

这种可能性当然存在，但这种可能性不应成为反对信息化记录方式的理由——因为在这种情况下，作假的是人，而非技术本身。技术忠实地再现了相关人的行为——尽管这些行为背后也许“有目的”，是“刻意”的；而且技术能够很好地保存这些记录，进行智能地分类管理。

因此，如果我们理性地进行分析的话，可以发现，将今天的技术应用于教育评价，完全可以填补人类能力上的空白——实时地记录大规模学生的经历信息。但现有的技术目前还不能控制人性的缺陷——无法精准鉴别，这些信息是否完全是真实发生的，还是弄虚作假的。

剩下的就是教育者的抉择，究竟是可行性优先还是等待技术达到完美之后才敢应用？

不过，需要指出的是，与所有技术的发展路径相一致的是，惟有在使用中不断优化，才是技术发展的通路。而如果不使用技术，技术则永远无法发展——“完美”的状态永远不会到来，甚至不会离我们更近一些。

可以想见的是，在技术臻于完美之前，我们可能还没有办法完全信任技术给我们提供的学生成长数据。值得庆幸的是，对于这些信息赋予怎样的权重，决定权在教育者手中。而权重不断调整的漫长过程，也就是技术不断优化，人们对于技术的理解逐渐深入的契机。同时，有了实践的积累，我们就可以探索在大量数据的积累基础上找到甄别虚假数据的方法。而且，值得期望的是，在不久的将来，随着移动设备的进一步细分，数据采集的准确性与真实性的保障也将更有可能得到解决。

如果不是经由这样一条探索的道路，那么，实现科学与公平的评价，将永远只是纸上谈兵。优先使用可行性强的数据采集，逐步将一部分过程性数据纳入评价，在实践的过程中谋求评价数据模型与采集方式的完善，一定是有价值的尝试。

## 工具是第一诉求

教育之所以在信息技术方面的积累十分薄弱，与其业务性质的繁杂有着密不可分的联系。学生的一切行为都与教育有着千丝万缕的联系——教育不一定要发生在课堂，也不一定发生在校园，可以在世界的各个角落；教育可以借由语言的交流产生，可以借由书本和信息设备得以传递，也可能是完全无表征的在脑海中的遐想。

这也造成了，“思想道德”“学业水平”“身心健康”“兴趣特长”“社会实践”这些听上去很符合教育人审美的词汇，其实并非是严格界定场景和行为的特定表现。教育人长久以来一直秉承着“评价全面发展”的理念，但在实际操作上却捉襟见肘，甚至连蓝图也很难描绘。直白地说，学生提供怎样的证据才能够自己的思想是道德的，学业水平是高的，身心是健康的，某些领域确有特长，参加社会实践是积极的？这些问题，或多或少存在争议。



而争议之所以有大有小，关键的区别在于是否有普适的工具来反映问题。比如，学业水平如何，考试院能够提供统一的考试，而考试作为一种客观、标准的工具，具有公信力与公平性，尽管在内容上颇有非议，但在形式上仍然是受到青睐的。而另外几个领域的评价就相对困难了，以“兴趣特长”为例，如果说“特长”的鉴定，可以由不同协会、组织、赛事提供的等级证书与获奖证明予以佐证，那么，“兴趣”的厘定就显得非常困难了。

要解决这个问题，可能就需要从工具化的思维来进行考虑。比如，什么工具才能采集到学生的“兴趣”呢？过去，我们是在表格中让学生进行填写，但是，这样获得的数据并没有太大的参考价值。一方面，学生自然而然会存在粉饰与包装自己的倾向；另一方面，即使学生确实有某方面的兴趣，有时也很难进行描述。所以，被动的填报不能有助于获得有质量的数据。

但是，如果将“兴趣”转化为“为某种活动所投入的时间”，就成为了一种具备可行性的解决方案了。因为，时间的记录完全可以使用“信息化工具+标签”的方式进行。就好比一个学生，如果他打开电脑后，70%的时间是用来浏览各种新闻网站，查找学习资料，进行网络阅读，30%的时间用来玩游戏；相比30%的时间用来学习，70%的时间用来游戏；那么，前者声称自己爱好“阅读”显然是更合理的，而后者与“游戏”的联系则会更紧密一些。

如果用这样的思想去设计高中生电子记录档案平台，那么其首要的因素就是需要一系列工具——这些工具在不同场景下采集着学生的行为，这些行为发生的时间，持续的时间长短，行为发生的场景，互动的对象等等；这些信息都是客观的。而通过大量数据的累积，研究者可以形成一定的算法模型，判断某学生在该行为上所处的水平，进而产生客观的量化描述。

教育之所以比其他社会部门显得更难进行评价，原因在于教育牵涉的场景繁多，发生的方式也有万千种变化，但是教育者一直不愿意用统一的数据标准对相关行为进行梳理并进行采集。

如果对照医疗领域我们就不难发现，为什么一个病患的病历档案比较容易客观呈现？而且不同的病历档案，不同地区的医生都能够进行阅读和参考？正是由于医疗领域存在非常系统、客观、科学的诊断标准，而现今的诊疗设备能够提供真实的数据。根据症状不同，病人可能会接受不同诊疗设备的检查，每一种诊疗设备都会提供其应用场景需要提供的数据。

这样的思维方式，教育同样有理由借鉴，如果我们在不同的教育过程中，都有相同的教育数据采集工具，将我们需要的“诊断”数据都用相同的标准予以保存；那么这将极大地便利数据间的比较——这种源于实际发生数据的比较，要比制定考察标准后的比较要来得更为准确；准确的评价模型也会随之产生。

教育评价要真正走向信息化，那么在数据采集工具上的技术储备是最为优先需要进行的。

