

# 教师在终结性评价中的作用

这篇文章是 Nuffield Foundation 资助的“为了未来的评价系统”课题的研究成果。这个研究课题是由评价改革小组 2003 年建立的，旨在从对在校学生的终结性评价的研究和实践中寻找证据，从而获取使学生从终结性评估中收益的教学方法。怎样发挥教师在终结性评价中的作用是这个课题的重点。

这个课题是由剑桥大学教育学院的 Wynne Harlen 领导的，在文章的末尾列出了这个课题组的核心成员的名单，他们管理和参与了所有课题的活动和研究成果的出版。有关这个课题的详细信息和评价改革小组的其它活动及研究成果，请访问其网站 [www.assessment-reform-group.org](http://www.assessment-reform-group.org)。

## 综述

评价和考试对年轻一代的生活和工作有着重大的影响。学校所作出的决定，特别是国家统一考试的成绩极大地影响着学生的前途和就业机会。当这些考试和考评的结果被用来评判老师和学校时，也影响教师的教学手段。考虑到考试和评价的重要性，因此，终结性评价应该尽可能准确地，并以正确的方式去反映和影响学生在校的学习是非常必要的。

这篇文章着眼于怎样实现一个复杂的终结性评价系统，使其能够对学生全方位的能力提供可靠的证据和信息。目前，大量的研究结论得出，教育系统强烈地依赖于考试成绩有许多缺陷，特别是用这种方式来考察学生学习的可信程度，即正确性和可靠性值得怀疑。这篇文章提出恰当地运用教师判断能够减少终结性评价对教学所带来的负面影响。

这篇文章同时也注意到了在运用教师评价系统时应该考虑到的许多问题。特别强调了这几个方面：对教师的判断必须有强有力的和持续的质量保证及监控程序；提供衡量发展的标准去指导以达到特定的目标而进行的学习；教师可以得到精心设计的评价学生的技能和理解能力的考核任务，从而帮助他们在作出判断的时候考虑到全方位的学习目标；教师岗前和在岗培训应该提高他们对不同评价目的的理解和运用能力。终结性评价要与形成性评价相结合，使其透明，并且老师的判断要以事实为依据。所有的这些条件都具备了才得出值得信任的结论。

另外，为了避免通过一锤定音的的终结性评价去评估教师和学校所带来的负面影响，这里特别强调衡量学校优劣的标准不能单纯地依赖于从学生终结性评价

中得到的数据，考察学生成绩的标准应该来自方方面面所收集的证据而不是每个学生的考试结果。这篇文章对那些制定评价政策和实施这些政策的人都具有重要的指导意义。

## 前言

各种不同的评价方式被广泛地运用在教育中。为了改善和提高教学评价，最引人瞩目的是被称作教学过程中的评价（形成性评价）。这里主要讨论的是对教学效果的评价（终结性评价），一般为了总结学生在一定的阶段所学到的知识和掌握的技能以便汇报学生的学习成绩和进度。

自从 2002 年，英国的一些评价政策里已经显示要有选择性地使用外部的考试作为对学生的终结性评价。例如，英格兰的政策已经调整，倾向在终结性评价中强调以教师为中心和发挥他们作用。威尔士，苏格兰和北爱尔兰的改革，已从 2005 年夏季开始，赋予小学低段老师对学生进行评价的责任。

在威尔士，小学低段的考试（7 岁）已经在 2002 年就取消了，威尔士政府宣布从 2004 年夏天起废除对 11 岁和 14 岁年龄段学生的国家考试。在苏格兰，教师从电子库里抽取国家考试试题去帮助他们对学生的成绩作出判断，但是，政府不再统计考试成绩。北爱尔兰政府决定停止阶段性的考试和小学到中学的升学考试，而引入以教师评价为基础的学生个人档案。个人档案能让父母随时可了解他们孩子的学习进度和成绩，同时也为升中学和大学提供依据。

在 2005 年宣布的改革中，英格兰教育部长，Stephen Twigg 说他和他的同事将“给予教师完全的信任”。他强调在英国 4 7 0 0 所学校进行的实验表明，在小学低段（7 岁）尽管教师开始仍需用考试作为终结性评价过程中的一部分，但教师评价已经明显的得以“强化”。英格兰学历和课程委员会的执行主席在讲话中提出：

“与其它的竞争对手相比，我们国家已经有了实施严格的教师评价的基础，我们可以通过教师平时的，系统的收集证据去得到比目前的评价更可靠的结果”<sup>1</sup>。他预测对 11 岁和 14 岁学生外部的终结性考试将被恰如其分的教师评价所取代，但是这个过程至少要花 10 年。他希望有一天，当必须对学生进行考试的时候，英格兰的教师将可以从考试库中根据需要去选择相应的试题和考核任务。这篇文章表明了

一个观点，那就是在尽可能短的时间内推进教师在终结性评价中所肩负的重大责任有其充分的，以事实为根据的原因。

当然教师的终结性评价并不是灵丹妙药，也会存在一些问题。但是它在许多方面是优于以外部考试为标准的评价体系。这里提出的观点是以大量的研究证据为基础的，有的所引用的研究证据已经在这里给出了参考书目，有的引用了网上评价改革小组的著作和文章<sup>2</sup>。这篇文章首先探讨了运用教师评价作为终结性评价目的的优点和缺点以及为了确保它所提供的信息具有较高的可信性（可靠性和有效性的结合）所需要采取的行动；然后总结了以考试和测验为基础的终结性评价系统的一些问题和怎样才能确保终结性评价所提供的信息能够满足它所面对的不同的用途。这篇文章对于制定和实施教育政策的人具有重要的指导意义。虽然这里使用的是“中小学生”和“学校”这些词，但是这里的一些观点和意义与其他教育院校的学生也是相关的。

### 用教师的评价作为终结性评价: 优势和不足

通过对学生学习的评价去总结学生的成绩有着不同的目的，其多样性表现在从非正式的记录学习过程到一锤定音的证书(见表一)以及贯穿于从小学到成人教育的各个阶段。同时进行评价的方式也是多种多样的。这里主要讨论的是评价方式的采用必须基于所提供的信息符合其目的。

**表 1: 在国家评价体系中终结性评价的目的**

学生个体: 学校内部 - 为了保持记录和向其他教师, 家长和学生报告学习进度。 学校外部 - 包括证书, 选拔和适应法令规定的要求。
学生群体: 评估 - 主要是对教师, 学校和当地教育部门。收集哪些方面的数据通常由国家和当地的政策决定而不是学校。 督察 - 各个地区或整个系统每年学生的平均成绩的比较。这个程序的执行也是由学校外部决定的。

运用教师的评价作为外部终结性评价已经提倡了很长时间了。当我们想一想有效的终结性评价所应该具备的质量标准, 这种方式的价值就非常清晰了。与其它目的的评价相同, 终结性评价应该具备以下的一些质量要求:

- 1, 有效性: 评价必须覆盖所有的并且只能包括那些与评价目的相关的学生的成绩。
- 2, 可靠性: 评价应该精心设计让用户对结果的准确和目的的一致性有信心。
- 3, 冲击性: 评价不仅只是衡量表现情况而且还要对教学和学生的学习动机等起到促进作用。评价通常对课程的开设和教学有着强烈的冲击, 因此, 尽量减少任何的负面影响是非常重要的。
- 4, 可行性: 进行终结性评价所需要的资源 - 教师的时间, 专业知识, 费用以及学生的用于学习时间 - 应该与其所提供的价值相符<sup>3</sup>。

教师终结性评价怎样才能成功地满足这四个要求呢? 下面的表从正反两方面对教师终结性评价的优势与不足进行了比较。但是, 在我们展开这些争论之前, 非常重要的一点就是要对教师的终结性评价有一个明确的定义。从目前教师和专家的讨论来看, 它的定义仅限于教师对他们自己的学生的评价:

*教师终结性评价是教师通过一系列有计划的, 系统的方式去收集信息的过程, 并基于他们的专业判断去推断学生的学习状况和报告学生在特定时期的成绩。*

## 教师终结性评价的优势和不足

优势	不足
<b>有效性</b>	
<p>与统一考试相比，教师能够更全面地考查学生的成绩和学习情况。教师的评价所提供的信息既能反映学习过程也能反映学习的成效。</p>	<p>教师评价的有效性依赖于教学过程中所提供的学习活动和机会。</p> <p>繁琐的模式化的质量保证程序可能会限制教师的终结性评价，因此只能采用“安全的”和日常的一些方式。</p>
<b>可靠性</b>	
<p>通过适当的培训和规范教师的评价，其可靠程度是很准确的。</p> <p>在一些国家，继续教育和高等教育中有很多教师运用严格的终结性评价对学生进行评估的实例。</p>	<p>由于所用的标准各异，教师的评价经常被看作是，而且确实可能是不可靠的和有偏差的。</p> <p>作为对教师的判断的补充，一些统一的外部考试和考核任务可能始终是需要的。</p>
<b>影响力</b>	
<p>教师有充分的自由去设立适合他们自己的学生的学习目标；</p> <p>教师从学生正在进行的学习过程中收集的信息，既可以用作形成性评价去促进学生学习，也可以用作终结性评价的目的。</p> <p>规范化程序给教师提供了很有价值的专业培训内容。</p> <p>学生在分享自我评价的过程中能够形成一种有别于“以成绩为目标”的态度，更趋向于朝着“以学习为目标”的方向努力。</p>	<p>公众对教育系统的信心可能会降低，因为教师评价一直被看作是不如统一外部考试重要，特别是对 11 岁以上的学生。</p>
<b>可操作性</b>	
<p>由于减少了购买商业考试卷的数量，学校的经费变得更宽松。而当国家统一考试被用来去评估一个学校的时候，学校这方面的开销急剧增加。</p> <p>教师能够花更多的时间在教学上而不是去准备考试和改卷。</p> <p>通过用日常的作业而不是考试去评估进度，学生用于学习的时间每年至少增加了两周。</p>	<p>承担终结性评价的责任可能会增加学校和教师的工作量。</p> <p>教师可能会发现规范他们对学生学习的判断的过程是非常费时间的。</p> <p>如何解释和使用评价的标准需要进行培训。</p>

## 提高教师的终结性评价的可靠性

值得指出的是，研究中发现教师终结性评价中的不足，主要来自那些并没有尝试把教师的评价作为主要角色的项目。从其他把教师的评价作为终结性评价的目的的研究<sup>4</sup>和一些国家的经验来看，已经发现了影响评价的可靠性的条件和可以提高其可靠性的方法。

通过对英国，澳大利亚和美国（昆士兰州和加利福尼亚州）等实施教师终结性评价<sup>5</sup>的考察和共同探讨，我们得出了以下五个方面的经验。这些也适用于在工作岗位和职业学校怎样对学生获取国家认可的就业资格证书进行评价<sup>6</sup>。

1. 教师应该有明确的标准去描述不同方面的成绩，以及所达到的水平，比较理想的是，他们应该协助发展这些标准。同时，给解释所收集的信息提供一个基本原则，这些标准也应该讲清楚所要求的学习机会。这样会使教师在以日常课堂作业为基础对学生进行评价时感到容易一些。
2. 为了让教师按照程序去确保评价的可靠性，教师的专业培训是必须的。培训也应该以那些将要被采用的标准和那些经研究发现潜在的产生偏见的根源为核心。
3. 通过专业合作对教师的判断形成一个系统的模式对促进教学和评价都是有利的。模式影响评价的计划和实施，从而，影响教师对学习目标以及在实施这些目标的进度的标准的理解，它的作用远远大于质量保证的功能。
4. 提供一个精心设计的任务库以及评分标准能更好地帮助教师对自己学生的成绩做出判断。这些任务可以提供活动范例让学生能够去实现一些重要的目标，例如批判性的推理和将知识运用到新的环境中。作为评价的一部分，它们能够给学生提供有趣的和相关的学习经历。它们不能在评价过程中占统治地位，当然，也不应该被看作是游离于教师评价之外的单独的方式。它们也不是去为了肯定教师的判断。更确切地说，它们是所收集的信息的一部分，如果需要的话，能够用来帮助教师确保他们达到评价所想要达到的目标。
5. 当教师和终结性评价的使用机构已经变得依赖于外在的考试时，这就将需要更长的时间去增加教师运用动态评价的能力和建立对教师判断的信心。因

此，给予足够的时间去尝试和评价创新的实践活动以及清晰的执行程序 and 建立内在的安全机制去确保评价的可靠性都是非常重要的。

## 以考试作为终结性评价存在的问题

以教师评价作为终结性评价并不是没有缺陷的，它有着与其它的终结性评价共同存在的问题，特别是当其结果被用来作为外部以择优选拔为目的的手段时。但是，这些问题应该提出来与那些以考试成绩为准的评价所导致的问题进行比较。大量的证据表明，以下一些原因造成了以考试为标准的教育系统的问题：

- 它不能够提供一个反映飞速发展的社会和技术的变革所需要的全方位的教育成果的信息，因此，它不可能去鼓励发展这些方面的技能。主要包括高层次的思考技能，适应不断改变的环境的能力，对怎样去学习的理解以及在工作和学习中相互合作的能力和独立工作及学习的能力。
- 它限制了已被证明能够提高学习成绩和缩小优生及差生之间的差距的形成性评价（或为了促进学习的评价）的发展。
- 它所提供的数据并不象人们期待的那么可靠。例如，根据对英格兰学校系统的重要阶段考试（Key Stage Tests）的结果估算，至少有三分之一的学生在第二阶段的考试（KS2）和百分之四十的学生在第三阶段（KS3）<sup>7</sup>的考试是被给予了错误的结论。
- 考试结果缺乏可靠性意味着将影响一些学生，并对他们在校内、校间<sup>8</sup>和校外发展作出不公正和不正确的决定。
- 目前没有确凿的证据支持考试可以提高学习成绩的主张。
- 它降低了一些学生的学习动力。
- 考试造成的紧张氛围，阻碍了一些学生发挥他们的最佳水平。
- 它鼓励以肤浅的和表层的学习为主的教学方式，而不是深层次的，对本质概念的理解。

研究也表明，一些考试和考核结果被用来应付各种目的，而它们所提供的价值是极为有限的。把每个学生的考试结果用于各种目的，从设定目标到学校排名，这种做法实在是过于简单化。为了一个目的所收集的信息并不一定能满足其它的

目的的要求，用于收集某些类型的学习成果的方式也不能适应所有的目的的需要。通过让教师在评价个体学生<sup>9</sup>中扮演重要的角色和用多样化的方式去评估和考察教师及学校，可以减少终结性评价的负面影响。在讨论这个问题之前，很重要的一点是要说明目前采用的终结性评价的方式不能充分衡量现代教育的一些关键性成果，它对学生，课程，教学以及通过评价去促进学习都有负面的影响。

## 评价什么？

许多终结性评价，例如，英格兰学生在 11 岁和 14 岁参加的国家考试，取决于学生在规定的时间内所进行的笔试成绩。就像已经提到的，这种形式限制了可以进行评价的学习成果的范围，把许多高层次的认知和交流技能，独立学习和相互合作的学习能力都给排除在外了。考试成绩所具有的“一锤定音”的特性鼓励以应付考试为目的教学和为了考试而进行的过多的练习。研究证明<sup>10</sup>，其导致的结果是学生尽管并不具备考试中所涉及的能力，但他们学会了如何去通过考试。一项英国教育和技能部资助的研究项目得出了这样的结论：当训练 11 岁的孩子去通过国家考试，很可能会提高学生的考试成绩，但并不能确定对学生今后的学习带来多少帮助<sup>11</sup>。利用考试对学习成果的评价的狭窄性是和教育与技能部在“关注孩子”的文件中所提出的全方位的学习目标的观点是相违背的<sup>12</sup>。

## 什么应该评价？

至关重要的一点是评价要覆盖那些在一个迅速缩小的世界和改变的社会里，年轻人在工作和生活方面所必须要学会的东西。任何学科都有两个重要的目标：

- 学习去理解
- 和
- 理解学习

第一个目标是就发展“大思路”而言 – 概念能被应用到不同的环境，能够使学生通过思考本质联系去判断不同情况下的各种现象。单纯地去死记事实和一套固定的程序是不能帮助年轻人将所学的知识运用到各种环境中去。

第二个目标是与发展学习意识相关的。“学生不可能在学校里学到他们将来在社会生活中所需要的一切”<sup>13</sup>已经是被广泛认可的了。因此，学校必须给学生提供



终生学习所需要的技能，理解力和求知欲。由于评价什么强烈地影响着教什么和怎样教，所以我们必须认真地看待评价什么的问题。如果不能覆盖所需要去评价的学习成果，那么就需要改变评价方式。

## 对学生，教师和教学的冲击

就像前面已经强调的，目前的考试制度对学生的学习动机和学习经历都有相当重大的影响。研究<sup>14</sup>发现：

- 考试的成绩似乎已经变得比学生到底学到了什么更重要；
- 考试能降低成绩不佳的学生的自尊，更难使其相信他们可以在其他的方面获得成功。
- 在考试练习时连续的失败会挫伤一些学生的士气，从而增加优秀生和差生之间的距离；
- 考试焦虑症对女生的影响比对男生的影响更大；
- 教学方式可能局限于那些考试所必须的内容（例如，忽视实践课）。

这些负面影响能直接作用于学生准备来自外部的“一锤定音”的考试，或间接地给教师提供一个运用于所有年级的拙劣的评价模式。

## 为了促进学习的评价

通过评价去促进学习(缩小高分和低分之间的距离)通常被忽视了。更糟糕的可能是对形成性评价的歪曲理解，评价改革小组将形成性评价过程定义为

*寻找和解释那些被学生和教师用来决定学生目前的学习进度，需要达到的目标和实现这些目标的最佳途径的信息<sup>15</sup>。*

很多学校在实施形成性评价中发现,在实际教学中，其所要求的教学理念的改变并没有发生。这种情况可能出现，例如，教师受到了超出他们控制的外部的考试限制。其结果是教师不可能像形成性评价所要求的那样，为了发展学生终生学习的能力，在指导学生的学习过程中让他们扮演主要角色。以考试为主的课堂评价的特性（如表2）。

表 2 国家考试是怎样驱逐评价为了促进学习

PACE<sup>16</sup>课题对英格兰小学的一组学生进行了为期 8 年的跟踪研究，这项研究开始于英国决定对 7 岁学生进行全国统一考试之前。其研究发现自从引入了外部考试以后，教师自己的课堂评价变得更像终结性评价。学生感到在引入考试之前教师的评价帮助他们的学习，但是随后他们注意到老师更加关注考试成绩而不是学习过程。学生自己开始采取终结性评价的标准去评论他们自己的作业。

## 满足终结性评价的目的的需要

### 形成性评估和终结性评估协调统一

考试对形成性评估的负面影响，是强化教师评价的主要原因之一，因此，把二者结合起来是非常重要的。在课堂教学中，教师能够通过观察，提问，倾听学生的交谈和批改作业去收集学生的信息。在形成性评价中这些信息可以立即用来促进学生的学习或可以储存起来用于去计划将来的学习机会。通过这种方式收集到的信息通常是不确定的，也可能是相互矛盾的，因为众所周知，学生所能做的是受特定环境的影响的。这给终结性评价造成了问题，但对达到形成性评价的目的却是有用的。在为这两个目的服务时，在信息和信息的解释之间需要区别对待。

对于形成性评价而言，信息的解释是相对于一个学生朝着一个特定阶段的学习目标迈进的进度。然后按照什么已经达到了和遇到了什么问题去决定下一步计划。说明的是需要做什么去帮助今后的学习，而不是什么水平和一个学生已经达到的级别。为了满足终结性评估的目的，需要运用一些对所有的学生具有相同含义的常规的标准和总的成绩，如水平和级别。这就是说，如果已经收集的，用作形成性评价的信息被用作终结性评价，必须要满足能够明确报告水平和级别的标准。在终结性评价过程，为了及时掌握反映学生全方位的成绩的最准确的信息，必须动态地考虑不同时间的变化。

反映学生学习成绩的信息既可以用来促进学习也可以用来作为汇报的目的，提供终结性评价并不是对形成性评价的一个总结，而是按照更广泛汇报成绩的标准而进行的重新评估。固定的记录级别和分数不是形成性评估，而是一系列的‘小终

终结性’评估。当然，对终结性评估必须有必要的质量保证。终结性判断所占的分量越重，越需要严格的质量保证，最好是包括学校间和学校内的判断模式。

采取措施增加以外部需求为目的的教师的终结性评价的可靠性将不可避免地提高他们以内在的需求为目的的评价能力。然而，为了避免对形成性评价的负面冲击，非常重要的一点是，内部的终结性评估不能过于频繁，不能超过所真正需求的对学习进度和成绩的报告。经常用水平和等级去评价学生意味着他们收到的反馈是评判为主的，鼓励他们把自己与他人进行比较。在这种状况下，教师和学生不会把注意力放到形成性评价上。

## 为了评估的评价

评估存在于各个层次：个人，学校，当地教育部门以及整个国家的教育系统。评估在每个层次的冲击取决于其考虑的信息的类型，用来判断其效果的标准和评判之后采取的行动。如果评估的信息是为了其它的终结性评价的目的而收集的，那它是不可能为评估的目的服务的。通过学生达到某个水平的百分比来对学校进行判断的不足如下：

- 其结果不可能反映一个学校所努力去实现的全方位的教育成果；
- 忽视了对“边缘”学生的注意；
- 鼓励以通过考试和测验为主的狭窄的需要。

评估最好以学生全面的成绩和学习行动的信息为基础，根据学校的环境和条件去作出判断。因此，学校应该提供有关课程，教学方式，以及相应的学生的背景和过去的学习情况等方面的信息。在学校自我评估中有一些很好的例子<sup>17</sup>。

## 为了系统的监控的评价

就教育系统而言，“监控”所指的是学生的学业成绩，教育的提供和教学等方面的变化。它所关注的通常是“标准”是否提高了，降低了或者是保持在原有水平。虽然它所用的信息包括学生的成绩，但是它的目的是为了帮助制定和完善相应的政策和实践，而不是对每个学生进行评判。学校一级的监控最好是通过自我评估来进行，同时收集其他能够对学生的评价数据进行解释的信息。尽管整个

系统每年的变化不可能有太大的意义，但是一定时期的发展趋势可以提供非常有用的信息。

系统监控的价值依赖于所收集的各个方面的信息。英格兰的方法是收集每个学生在全国统一考试的成绩。为了确保督查的有效性，需要收集从观察学生的实际操作技能和各种评价的结论等全方位的信息。

收集已有成绩的方式具有其成本优势，但必须取决于是否能够提供有用的和相关的信息，如用中学最后阶段的考试去判断整个国家的教育发展情况。同样的，建立和运作覆盖各个方面的教育成果的普查的花费必须取决于能否给政策决策者和实践者提供详细的有用的信息。把监控从个体学生的表现中分离出来，可以避免集中收集每个学生的分数。这将消除对“一锤定音”的考试的“需要”，能够确保评价不再被考试所限制，更重要的是，教什么不再被考什么所限制。在一些国家<sup>18</sup>进行的持续的，定期的对部分学生的抽样调查证实了这种作法的价值。

## 结论:

这篇文章解释了为什么以考试为主的评价系统所提供的学生的成绩，其信息的有效性和可靠性是不足的。究其原因，部分是由于所评价的知识和能力，部分是由评价的方法和怎样去用这些结果。其后果是限制了课程和教学方式以及削弱为了促进学习的评价的实施。同时它也导致了终结性评价太过频繁。为了减少这些负面影响，终结性评价应该设计成为为特殊的目的提供信息，并且只能当学习过程中需要总结和评估才进行。其它的时间教师应该集中于形成性评价。

从那些已经用了和增加使用恰当的教师终结性评价的教育系统所提供的信息来看，这种方式并没有什么负面影响。但是，必须承认，这种方式需要采取一系列步骤去确保教师终结性评价所提供的学生成绩的信息是可靠的，这是一个巨大的挑战。关键的几点如下：

- 需要对教师的判断有一个强有力的和持续的质量保证和控制程序，以确保教师终结性评价能够对学生的学习提供有效的和可靠的评估。
- 教师的岗前和在职培训应该发展教师对不同的评价目的的理解和技能，强调教师评估中存在的潜在的偏见，从而帮助教师减少评估对学生造成的负面影响。

- 必须投入精力和资源去制定达到一个特定目标的学习进度，并能够运用到各种活动中去的发展标准。
- 教师应该得到设计得比较完善的方案去评价学生的技能和理解，这能帮助他们在考虑全方位的学习目标的基础上作出判断。
- 程序应该是非常透明的，教师的判断要以事实为依据。
- 终结性评价必须融入形成性评价的程序中，应该通过合理的设计去减少教师和学生的负担。

另外，为了避免用“一锤定音”的终结性评价去评估教师和学校所带来的负面影响：

- 学校系统的评定不能单纯地依赖，或主要来自于终结性评价的数据。这些数据应该放在反映学校的其它成就的指数的环境中去报告和解释。
- 监测学生所达到的标准应该从各种信息中得来而不是每个学生的考试成绩。教师评价在一个教育系统的作用是以一定数量的学生为样本去收集各个方面的信息。

## 意义

这些结论对那些在国家和地区一级制定方针，政策的，对那些在学校和大学里实施评价政策的，对顾问，督学和对师范教育工作者都有重要的指导意义。由于实施是以政策为依据，所以这种改变要求那些在实践中具体操作的人真正地理解这些政策，并采取相应的行动，特别是包含着信任和责任。因此，下面所归纳的对政策决策者和实施者的意义是没有等级区别的。

### 对国家和地区一级的政策决策者的意义：

- 认识到目前在国家和学校各个层次上以考试为基础的终结性评价政策所投入的经费和时间并没有得到相应的价值的信息。
- 通过教师判断的模式去报告学生的成绩，从而取代现有的国家考试，把节省下来的一些时间和经费投入到提高教学的质量中去。
- 回顾和检查教师评价在考核 16 至 18 岁的学生时的作用。

- 提倡在教育系统进行为什么需要改革和怎样进行改革的公开讨论。
- 对任何新的以教师判断为基础的终结性评估系统应该给予至少两年的实验和评估期，以及同样的时间段去宣传和培训教师。
- 评估学校和鼓励自我评估要以各个方面的指标为依据，而不只是学生的成绩，降低“一锤定音”的外部考试所带来的负面影响。
- 建立一个以国家监测为目的的抽样学生成绩系统，因此，在降低全面考试的负担的同时增强信息的广度和相关性。

## 对学校管理层的意义

- 建立一个任何时候都以评价去支持和促进学习的学校评价政策，并只在必要时为了检查和报告进度才需要终结性评价。
- 对所有终结性评价有质量保证的安排，包括任何教师给予的考试，使学校能够基于可靠的信息对学生的进度作出决定。
- 确保家长理解评价怎样帮助学生学习 and 怎样的标准是被用来报告一年中各个阶段的进度。
- 抵制考试的硬指标所带来的压力，鼓励用能够说明学生学习情况的各种类型的信息。
- 通过模式化给教师评价提供质量保证和时间保障。

## 对教师的意义

- 确保评价总是用来去帮助学习，当需要一个终结性评估报告时，最好要对照相应的标准和对证据的可靠性进行判断。
- 把学生包括在自我评价的程序中，帮助他们理解以报告为目的，去评价他们的学习情况的标准，以及终结性判断是怎样作出的。
- 参与模式化的终结性判断和其他质量保证程序。
- 只在最合适的时候用考试，而不是例行公事。

## 对督学和顾问的意义

- 检查学校的政策和实践以确保评估是用于教学的过程中而不是被终结性的考试和任务所代替。
- 鼓励运用能说明学生各个方面的成绩的信息。
- 确保将评价纳入教师的继续教育中，以满足那些需要这方面知识的教师的需要。
- 回顾和检查评价是否完全合适和其它质量保证程序，以及哪些对教学是有利的。
- 帮助学校发展以各种评价方式为基础的自我评估行动计划，而不只是停留在学生的成绩上。

## 对岗前和在职专业发展课程提供者的意义

- 确保课程的设计中提供充足的时间去：
  - 讨论评价的不同目的和使用已有的评估数据。
  - 让参训者去辨别，举例和评估不同的收集学生成就的信息的方式；
  - 分享设计与特殊的学习目标相关的评价标准的经验；
  - 思考评价中存在偏见事例和其它问题的根源，以及如何把它们最小化。

“面向未来的评价系统”项目核心小组成员：

**Mr David Bartlett** Co-ordinator for Assessment, Birmingham City Council Children and Young Person Services

**Professor Paul Black** King's College, University of London

**Professor Richard Daugherty** University of Wales , Aberystwyth

**Dr Kathryn Ecclestone** University of Exeter

**Ms Janet English** Head Teacher, Malvern Way Infant and Nursery Schools, Croxley Green, Hertfordshire

**Professor John Gardner** Queen's University, Belfast

**Professor Wynne Harlen** University of Bristol , ASF project director

**Ms Carolyn Hutchinson** Head of Assessment Branch, Scottish Executive Education Department

**Professor Mary James** Institute of Education, University of London

**Mr Martin Montgomery** Assessment Development Manager, Council for the Curriculum, Examinations and Assessment, Northern Ireland

**Dr Paul Newton** Principal Assessment Researcher, Regulation and Standards Division, Qualifications and Curriculum Authority

**Dr Catrin Roberts** Assistant Director, Nuffield Foundation

**Mr Jon Ryder** Teacher, Lord Williams's School, Thame, Oxfordshire

**Professor Judy Sebba** University of Sussex

**Dr Gordon Stobart** Institute of Education, University of London

**Ms Anne Whipp** Curriculum and Teacher Assessment Manager, ACCAC  
(the Qualifications, Curriculum and Assessment Authority for Wales ).



参考文献和注释:

---

- <sup>1</sup> Ken Boston , speech at the launch of the Annual Review of QCA, March 2005. ([http://www.qca.org.uk/251\\_12898.html](http://www.qca.org.uk/251_12898.html) )
- <sup>2</sup> Particularly in Gardner , J (ed) *Assessment and Learning* (2006) London : Sage.
- <sup>3</sup> See Appendix on costs of assessment in Working Paper 3 on the ARG website at [www.assessment-reform-group.org](http://www.assessment-reform-group.org)
- <sup>4</sup> Summarised in Working Paper 2 (ASF, 2004) on the ARG website at [www.assessment-reform-group.org](http://www.assessment-reform-group.org).
- <sup>5</sup> See report of seminars 2 and 3 and summary of the main points in Working Paper 1 on ARG website at [www.assessment-reform-group.org](http://www.assessment-reform-group.org).
- <sup>6</sup> See report of seminar 5 and Working Paper 4 on the ARG website at [www.assessment-reform-group.org](http://www.assessment-reform-group.org).
- <sup>7</sup> Wiliam,D. (2001) Reliability, validity, and all that jazz. *Education* 3–13, **29**(3) 17–21.
- <sup>8</sup> See Chapter 7 in J. Gardner (Ed) (note 1) and J. Gardner & P. Cowan (2005) *Assessment in Education* 12 (2) 145–164.
- <sup>9</sup> See ASF Working Paper 2 available on the ARG website at [www.assessment-reform-group.org](http://www.assessment-reform-group.org).
- <sup>10</sup> See the summary in the ARG pamphlet *Testing, Motivation and Learning* on the ARG website at [www.assessment-reform-group.org](http://www.assessment-reform-group.org).
- <sup>11</sup> <http://www.dfes.gov.uk/research/programmeofresearch/projectinformation.cfm?projectId=14390&type=5&resultspage=21>.
- <sup>12</sup> DfES 2004 see [www.everychildmatters.gov.uk](http://www.everychildmatters.gov.uk)
- <sup>13</sup> OECD, 1999, *Measuring Student Knowledge and Skills*. OECD Programme for International Student Assessment ( PISA ). Paris : OECD, p9.
- <sup>14</sup> See *Testing, Motivation and Learning* (2002), available on the ARG website.
- <sup>15</sup> ARG (2002) *Assessment for Learning: 10 Principles*, available on the ARG website.
- <sup>16</sup> Pollard, A.,Triggs, P., Broadfoot, P, McNess, E and Osborne, M. (2000) *What Pupils Say: Changing Policy and Practice in Primary Education* London : Continuum.

---

<sup>17</sup> For example: How Good is Our School in Scotland, in England the emphasis on schools self-evaluation in *A New Relationship with Schools* (DfES and Ofsted, 2004) and in Wales *Guidance on the Inspection of Primary and Nursery Schools* (September 2004) and *Guidance on the Inspection of Secondary Schools* (September 2004) (Estyn).

<sup>18</sup> For instance, the National Assessment of Educational Progress (NAEP) in the United States, the National Education Monitoring Project (NEMP) in New Zealand and the Scottish Survey of Achievement (SSA).