

第 1 部分

理 论



考察现行评价 方法的偏差

尤其是在美国，由于定量指标及其教育功效的盛行，实质上出现了一股热潮，为了每一种可能的社会目的而设计各种测验形式……美国正处于变成一个“十足的测验型社会”的局面。我们可以将这种态势概述为：如果某种事物很重要，它就值得以某种方式加以测验；如果它不能被测验，那就可能没有价值。很少有观察者停下来考虑一下，在有些领域里这种测验的方法可能不恰当或效果不是最佳。

——霍华德·加德纳

《多元智能实践中的理论》

(*Multiple Intelligences : The Theory in Practice*)

你常常经历下列一种或几种情形吗？

- 一个学生能画出精致的图画来说明他理解了你所教的某些概念。然而你不能让他在正规测验中画一幅画，因为你没法编制一个评分标准。
- 在常规课堂中，你会碰到纪律问题。但是在音乐课上，这个学生

的行为是得当的。在你课堂上发生的不良行为，音乐教师并不认为是具有扰乱性的行为。

- 你花了过多的时间，为孩子们在地或州高效的、标准化的能力测验中获得成功作准备，而那些测验仅仅是测验了语言智能和数理逻辑智能。
- 一个非常可爱的学生拥有许多朋友，同时也是其他人的好朋友。在群体中，他相信每个人都能成功，但是由于某种原因，他自己却很少成功。
- 你知道，标准化测验不能确切说明你的学生能做什么和他们真正知道什么。然而，只要你的学生测验成绩不理想，你就会遇到麻烦。

所有这些情形在各个课堂中都是普遍存在的。我们知道，大多数学生比我们对他们的评价所显示的要更为聪明。每一位教育者都会碰到这样的学生，他们理解课堂讨论中的材料，他们拥有良好的关于信息的“实践知识”。他们知道该如何运用它，但是他们不能将其体现在一份笔答试卷上。格兰特·威吉斯（Grant Wiggins）针对这种情形提出了一个引起争议的问题：“在这种情况下，究竟是谁真的失败了？”他指出，在这类情形下，失败的是测验本身，而不是学生和教师。现行的评价工具使我们不能准确地评价学生的能力或他们对材料的理解力。那么，这些措施究竟有什么问题呢？

基于缺陷的评价

在我的第一本探讨评价的书中（Lazear, 1994），我提出，我们西方教育体制中的评价所面临的首要问题是，几乎所有评价方法都来自一个有缺陷的基础。这种方法最主要在两个领域盛行：能力测验（有时也叫做“诊断性评价”或“智力测验”）和学业评价。

诊断性评价及其局限性

基于缺陷的评价方法在测评能力时，常常是实施一组能力测验。设计

这种测验确切就是要查找学生的“缺陷”、他们的弱点、他们所做不到的。当然，在为我们自以为发现了的那些“学习缺陷”“更充分地提供服务”的幌子下，我们给学生贴上了标签。汤姆斯·阿姆斯特朗（Thomas Armstrong, 1987），著名的多元智能理论的作者和讲演者，他就这样的评价方法说了以下这段话：

很清楚，这些评价不能客观地测验一个孩子的能力……研究学习能力缺陷的专家在他们的工作中运用一种“一直测验直到发现问题”的方法。测验者对一个孩子实施评价，直到他们找到有嫌疑的学习缺陷，他们对这个孩子的测验才停下来，然后为他贴上标签。如果经过两三次测验，他们还没有发现缺陷，那他们就实施 15 或 20 次以上的其他测验，直到他们发现了缺陷或者耗费掉了他们的全部能量。这种与孩子打交道的方式助长了找差错的行为，并将发现优点和能力的机会减少到最低程度。（p. 30 - 31）

这种缺陷测验的逻辑外延是，矫正教育的方案是帮助学生们“克服”他们的“缺陷”。但是，实际上我们所开出的处方是什么呢？多数是他们不曾做到的！换句话说，我们只是矫正缺陷本身。我们苦心研究一个学生的缺陷，考虑如何能将各种信息灌输到他的头脑中。最近在由我召集的一个有关评价的研讨会上，我和与会者们长时间地讨论了矫正教育，尤其讨论了关于语言艺术和数学的矫正教育。一位与会者举手说，“我们学校所着手的矫正方式，在很大程度上，就像试图与一个完全耳聋的人交流时，开出的处方竟是‘让我们大点声讲话’！”

再读一读以“诊断性评价的测验”为开头的段落，并思考我们“天才”教育的方法。其中充满了许多与贴标签过程的另一种结果相同的问题。我不反对那些为有天赋、有才干者开展的教学方案，但是我对“我们为什么不将其提供给全体学生”存有重大疑问。似乎那些在常规课堂上表现不好的学生，恰恰是我们该努力寻找不同的教学方法去教的人。我们也该对我们所认为的“天赋”的含义进行一些慎重的反思。在各种语言和数理逻辑测验上得分优秀的那些学生就有资格参与到更多的天才教学

方案中？确实，一些学生在所谓的“学术”领域很优秀，并常常抱有一种势不可挡的学习进取心。但是为什么这些学生被认为比那些在所谓的“非学术”领域取得好成绩的学生要更有天赋呢？

学业评价和对失败的强调

当我们审视用那种基于缺陷的方法来评价学业进展的做法时，我们发现这与一个令人痛苦和沮丧的游戏相似。在那个游戏中，学生和教师是互相对抗的。一方面，教师的目标是设计一个测验来清晰地显示学生所不知道的内容。它经常会变成一个迂回的“学业了解”（gotcha）的游戏，其中充满了许多让人感到惊奇的东西、一些只有“真正了解”者才会理解的密码和意想不到的曲折和变化。另一方面，学生试图“琢磨”教师，推测考试内容。一旦他们真正处于参加考试的过程中，游戏就变成了一个再次猜测、揣摩教师心思的活动，学生们努力推断教师想要考察的内容和各种试题的正确答案。他们也经常花很多时间，通过隐瞒他们所不知道的东西以欺骗老师。很显然，这种描述有点儿讽刺意味。然而，在很多的场合，尤其是中学和大学中，的确是在运用这样的评价体制。想想当你将一份测试卷退还给学生们时，他们是怎么做的。你曾仔细分析过他们的第一反应吗？他们首先看的是分数，然后再寻找他们做错的地方。

加德纳（1993）以一名心理学家的视角来看待这个问题，他的评论对参与各种正式测验的教师和管理者同样适用：“正式测验的一个相当糟糕的方面就是对分数的运用。学生获得分数，看到他们的百分等级，得出了一个学业方面的结论，但这个结论并没有反映学生的全部优点和价值。我认为，心理学家花费了过多的时间给学生分等级，而没有足够的时间来帮助他们进步。评价的实施应首先有助于学生。评价者有责任向学生提供反馈，以帮助他们改进目前的状况。”（p. 178）

我们也能看到基于缺陷的评价方法在所有的所谓“标准化”测验形式中的力量，那是些错误地设想要提高水准的测验。我们希望所有的学生在这些测验中表现良好以便毕业。标准化测验存在许多问题，格兰特·威吉斯（Grant Wiggins, 1992）提出了其中的一个质疑，“有没有见到过一个标准的小孩？”当我们为了应答标准化的评价测试（例如，简答、对问

题、填空、多项选择)而瓦解整体性的知识体系的时候,我们讽刺性地简化了我们测验的标准。这些应答很少超出任何有关认知能力分类的最低思维水平。

标准化的学业评价方法也倾向于这样一个假设,“知识”和“学习者”是能够被客观化的某种单独实体。事实上,两者是紧密联系在一起的确,知识是一个客观体,但是,知识实际上是否被深入地了解和理解,取决于一个学习者的了解和理解情况。当我们意图评价这种了解和理解时,我们必须尊重学生和学习之间的这种深刻的、主观的关系。

格兰特·威吉斯(1992)通过提出“看待评价时不能回避其道德维度——一个我们忽略了的角度”(p.8)来阐述了这种关系。他对这些类型的测验的不道德之处评价如下:

当我们惟一的目标是测量时,就不可避免地将孩子作为一个测验的对象来对待了……如果我们将测验理解预定为这样一个过程,真实评价的教育意义的实现就处于危险之中。在这个过程中我们坚持自己提出的问题、规定答题的时限、强行限定参考资料和问题的优先答案,并且极为重视打分和报告。当我们在一次测验中将学习者的“知识”从学习者的品质中分离出来时,我们就不再感到有责任仔细了解这项评价(或甚至是追求一个答案的意义),不再感到有责任给学生提供机会,让他们通过测试取得进步,不再感到有责任为他们创造机会来展示他们能够做的。(p.4)

语言数理谬误

评价中的另外一个主要问题,似乎涉及到西方所有的正式教育,这个问题威胁着评价的信度和整个教育事业自身的建构。我把这个问题暂称为“语言数理谬误”。简言之,这种谬误认为,对一个人的学习进行真实的、最终的、主要的测试,就是测试他们以书面的、逻辑的形式展现知识的能力和进行精确数学运算的能力。来看一下在目前的教育实践中,关于这种

谬误的几个有趣的统计例子。

一、我最近收到了全国教育目标报告（1997）的摘要（如果你不熟悉这些目标，请在附录中查找）。第一眼看上去，它们都是合理的，每个人都可能会认为这些目标是绝妙的。报告包括一个“美国记分卡”，它显示了1997年美国学生对这些目标达成情况。报告首先致力于讨论应当怎样做才能保证美国在数学和科学领域的世界领先地位。但请注意报告中对“我们如何能将数学和科学成就提高到世界级水平”这个问题的回答是带有偏见的。

- 与世界最高的水平作比较，在数学和科学领域制定强硬的标准。
- 将这些标准运用于其他教育系统，包括课程、教学、教科书、评价和学校政策。
- 提高教师在数学和科学领域的专业知识水平和教学技巧水平，将国家的教师政策与体现国家标准的教学目标更紧密地结合起来。

二、州或地方层面上标准化的能力测验，在美国和加拿大不断得到应用。在此必须要提到这些测验中存在的一些问题，其中主要的几个问题是这样的：什么是真正应该被测验的？应当如何进行测验？如果要为学生在社会中成为积极而负责任的人做好准备，这些测试又能起到什么作用？这些测验主要是针对语言和数学能力的测试，我们听得最多的就是“3R”（读、写、算）。这“3R”一直是很重要的，但是仅仅靠它们无法保证一个人在我们的后现代的、科技主导的、多元文化的社会中取得成功！我们为什么不就“当今社会需要什么基本能力”和“21世纪需要什么”而进行一次全国性的讨论呢？

原则上，我并不拒绝接受能力测验，但是这些问题有很大疑问，测验如何实施？分数如何解释和运用？关于熟练程度的范围可能狭窄和偏倚，那测验的价值和主旨如何保证？

在《多元智能评价方法》一书中，我与艾米利·格瑞德（Emily Grady, 1992）就档案袋评价问题作了共同的研究。这里请让我重复一下其中的一段评论：

看上去，测验和培养成功的学生之间是没有什么关系的。在工业

化国家中，美国学生被测试出处于最缺乏学术性的行列。根据许多教育家和心理学家所言，甚至是教育测验服务中心领导的观点，标准化测验作为一种提高学生成绩的方法是不符合要求的（pass muster）。事实上，它们可能恰恰违背它们本该要达到的目的。（p. 6）

教师们暗中期望着根据这些测验教学。他们非常明白，他们的业绩和学校的业绩最终都将由学生分数的好坏来判断。

不断地对学生进行测验，提高测验的难度，以此作为改进学校教育的一种方法，仅根据经验就可以知道这个策略是不可行的。然而，近些年来，美国政府已经在部分地区推行全国性的测验。这就类似于，一个准备减肥的人，被告知要使用更好、更精确的秤，并进行更严格、更频繁的称量！

三、我们渴望那种容易打分、高效省时的测验程序，但在认识上却存在错误。如果不是由于认识不充分，那一个简单而关键的原因就是，我们一般认为语言和数理逻辑方面的材料比由其他智能而得来的材料更“客观”。我们也错误地认为每个人都能理解那些材料。例如，如果一项作业是描绘一幅壁画，那给这项作业打分就会非常困难。在我所召集的多元智能研讨会上，我从教师们那里得到的最多的评论之一就是，“我确信我有其他测验学生的方法，它会更公正，对学生的学习更有益，但是我就是没有时间来编制、实施这类评价！我们每天要看到那么多学生，这样的评价是不具操作性的！”相反，我们选择了各种语言—逻辑的简答题和记忆性的测验。威吉斯（1997）针对这类测验说了如下的话：“我们认为测验很值得，因为我们乐意将‘评价’缩为‘测验’，对把测验看做是与学习相分离的事情感到满意——这样便于你在教学一旦结束之后，就能知道学生们掌握得怎么样（通常是为了其他人而非学生的利益）……不可避免地，我们逐渐依赖于能被可靠使用的作业，最简单、有效而琐碎的作业。节省时间和人力成为测验设计考虑的一个极为重要的优点，并使得每个测验设计者都依赖于简单的、使用快捷的和可重复利用（因而‘安全可靠’）的项目。”（p. 3）

换句话说，我们追求最小公分母，即那些容易理解、容易打分、容易

执行的，并容易与其他语言一逻辑结果相比较的评价。当某种场合下需要评价的效率时，很可能会违背教育的本来目标，即对学习加以深入理解和消化，以便为未来的有效生活作准备。而且，它与当前关于如何确保大脑最佳运作和如何学习的脑科学研究结果完全不一致！格兰特·威吉斯（1997）就这一点说，在一次教师会议上，不仅仅是专业的测验编制者，还有教师们也表示愿意使用地方性的机械评分测验，根据标准答案对学生的回答快速打分。他们认为人类的判断力在合理的评价中是不必过度宣扬的……那为什么乐队指挥、辩论赛的教练、科学公正的法官、影视剧导演和排球教练，没有屈服于同样如此轻率的或宿命论的权宜之计呢？因为他们懂得“测验”或成绩只是一个过程，而不是你随后要做的事。（p. 3）

在各种研讨会上，常常会出现对高效率有异议的推论：“那么，要获得并维持一份好工作，应该做什么？当你的老板想要一份报告时，她当然不会是想让你唱首歌！你不能依赖于其他智能来谋生！”即使这是一个相当单纯的并有点儿天真的异议，然而，对于这样的异议，我可以指出，实际上有意义、有成效的职业中有一些是与其他智能领域紧密相关的。虽然，我不相信任何智能和职业之间有既成事实性的联系，但一个智能强项确实可能对一个人从事任何可能的职业产生一种指示。请阅读今天市场上过剩的关于高层管理方面的书，你将很快注意到这些出版物所共同宣传的几个关于多元智能的主题。你会发现雇主想要的是那些拥有有效团队意识的人。他们希望雇员有创新思维、能解决问题，也能将他们的创造性想法与其他人的想法结合，提出更好的、更具创造性的想法。雇主想要那些善于自我反思、可以自我评价并改进工作的雇员。换句话说（用我自己的话说），雇主想要那些能运用所有的智能而不仅仅限于语言和数理逻辑智能的人。

四、大学入学考试几乎仅仅是测验了语言和数理逻辑智能，即使最初的设想是要覆盖所有的学科领域。大学典型地依赖于这些测验分数和其他数学、科学和英语等级，来决定是否接受学生入学。家长们经常对我说，“所有这些多元智能的材料都是有趣的，孩子们真的都很喜欢它，可它有助于测验吗？既然不允许他们在 SAT 考试中扮演角色或画一幅壁画，那么，为什么还要在这方面浪费时间呢？”教师们常常在多元智能研讨会上

发表类似的评论，“是啊，我们应该给孩子们运用所有这些智能的策略和工具。我确信它有助于孩子们的学习。可问题是，这些美妙的多元智能内容与测验无关，因此我们不能在上面花太多时间！”

在《多元智能：实践中的理论》一书中，加德纳（1993）讨论了一个现实问题。许多在正式的标准化书面考试中表现很好的学生，“常常对这类测验‘得心应手’，甚至在他们对学科内容的理解显然只是一般的情况下，在这种测验中也能取得很好的成绩”（p. 16 - 169）。因为他们已经理解了测验是如何构成的，哪种回答是符合要求的，我们错误地认为他们理解了学习材料，其实我们真正评价的是他们应付这类测验的能力！

当然，最大危害是，当分数被公布于地方报纸时，会被假设为给予了公众一个有关“全国教育状况”的准确指示，并且似乎没人对这个假设有所质疑。在《多元智能的多元评价》一书中，吉姆·布兰卡（Jim Belanca）、卡罗林·查普曼（Carolyn Chapman）和贝丝·斯沃特兹（Beth Swartz）（1994）指出了发生在我们学校中的、隐藏在多数测验背后的、可能最真实的基本原理：

标准化测验，除了有某些预言价值，除了能够取悦那些机会主义的政治家和那些拥有良好生源的学校管理者，几乎一无是处。这些测验很少被用于确定受制度影响的学习过程的缺点。这些测验常常被用于惩罚一个没有让全班或学校“超过标准”的教师或校长，或是强迫教师们严格忠诚于一项没有说服力的标准。

这些测验是揭示语言—数理谬误束缚我们教育制度的绝好例子。我们知道，培养和发展所有的智能对孩子是有好处的。它有助于他们开展学习活动，它提高了他们的自尊，使得他们成为一个完整的人，它有助于他们尊重组成人类大家庭的奇妙的差异。不过，既然它不被体现在测验中，那我们就别为它费心劳神了！

简言之，我提出的这些问题，使我们看到了整个西方教育体系中，大多数评价问题的核心：如果一个学生不能写、读、谈论和计算某项内容，那他就没有理解它。同样，相反的假设就是这样的情况：如果他们能在测验中做出所谓的“正确”答案，他们就理解了。这个假设是错误和幼稚的，它对许多每天生活在我们课堂的学生来说，无疑是一种巨大的伤害。

在本章里，我仅仅讨论了坏消息。然而，尽管我这么评述，但我感到这是个做一名教育家的大好时机！人类从没有像现在这样了解教和学的过程。而且，关于我们职业的艺术级水平的研究比从前更广泛、更深刻。我们从没有像现在这样了解到更多的关于人脑的知识，了解如何运用与人类官能的真实活动相一致的方法来教学。在下面的章节里，我转向讨论好消息——哲学理论和课堂实践，用以解决许多的评价问题。

2

探讨采用有效
评价方法的可能性

有关人脑如何学习的“最可信的知识”撞开教育之门已经有 20 多年了，然而我们进入大门的步伐却非常缓慢。为什么？原因是众多的，但其中的主要原因是……现有大脑研究的结论与传统的教育实践相抵触，与我们本来认为它应是怎么样的图景相抵触，以致我们既忽视和误解了这些知识，又想当然地没有正确运用这些知识。

——苏珊·卡瓦连克 Susan Kovalik 和克伦·奥尔森 Karen Olsen)

我们该怎样超越有偏见的评价实践呢？考虑到我们已经构建起的美国公共教育的方式，这个问题的答案就很复杂，并有点儿缠绕不清了。但我们必须关注显而易见的政策性答案，尤其是考虑到校务委员会在州和联邦一级的教育政策制定过程中严重的政治属性。然而，这不是我在这里所关注的，尽管我对我认为要发生的事情作了一些观察。

这个问题的答案，一部分是要教育那些普通的纳税人，使其关注当前的教育研究，关注如果我们想使今天的学生为他们 21 世纪的生活作准备（顺便提及的是，他们将在那个世纪里度过他们自己绝大多数时间），为什么今天的学校教育不能仅仅是以“如果我是个孩子，我应怎样”的方

法。我们必须承担一项重要任务——研究理查德·斯廷吉斯 (Richard Stiggins) 称之为“评价盲”的问题 (1991; 也见 Lazear, 1994, 第 7 章), 即我们必须把公众培训成为对他们所获得的各种评价数据有所挑剔的消费者。

答案的另一部分包括设计各种方法, 激发学生深入的、内在的学习需求, 帮助他们理解教学的过程。不研究好这个问题, 真实的教育是不可能发生的, 因为我们将使学校继续充斥着那些仅仅靠“玩学校提供的游戏”便能毕业的学生。玩游戏既不服务于社会的需要 (培养未来公民), 也不服务于学生的需要 (使他们对将来过有效的生活而作准备)。

本书中, 我致力于在这些方面作出回答。转变我们现行的有偏见、不公平的评价实践最有效的方法是, 实施多元智能模式。在我的《学习的七种路径》(1993) 一书中, 我研究了帮助学生认知、欣赏和理解他们的多元智能的主题。威廉·格兰瑟 (William Glasser) (《课堂控制理论》), 阿尔菲·科恩 (Alfie Kohn) (《奖励的惩罚》) 阿特·考斯塔 (Art Costa)、大卫·柏金斯 (David Perkins), 还有其他许多人的研究也是很重要的, 可以帮助我们发展积极的、负责的终身学习者。

在以后的论述中, 每次我使用“评价”这个词, 都是指“教学评价”, 因为它们确实是一枚硬币的两面。我们今天所面对的许多真实困难都来自于教学和评价的分离。如果我们准备以可信的方式评价学生, 真正服务于促进学生深入学习的最终目标, 教学与评价必将再次结合。

接下来, 基于目前对认知科学、多元智能和神经系统科学的研究, 我描绘了变革的四个领域。我不仅要讨论第 1 章中的有关评价的基本假设的研究, 而且更为重要的是, 我根据多元智能理论和研究, 建议实施新的评价方法, 并提供了实用可行的启示——这些方法将带我们超越现有的谬误、超越“最新教育趋势或流行思想”、超越无休止地“快速修正”的要求。所提出的这些变革领域与第 1 章里所讨论的问题并不是一一对应的, 只是这些问题自身是交叉的、复杂的, 因此所提议的解决办法要涵盖所有的问题, 并且相互关联。不过, 一般说来, 可以按照一种我称之为多元模式的视点, 来研究基于缺陷的诊断性评价和学业评价。“超越标准化测验”、“超越语言—数理谬误”、“超越非情境性的学习”部分涉及到了语

言—数理谬误和超越它的各种方式。

超越基于缺陷的能力和学术测验：评价的一种多元视角

假如评价时要寻找的不是学生所做不到的，而是每个孩子的真实天赋和卓越才华，那会怎样呢？如果我们的矫正方法着眼于学生的长处和能力，然后利用那些长处去帮助他们发展弱的能力，会怎样呢？托马斯·阿姆斯特朗（1987）曾说，“我宁可 not 利用‘学习缺陷’的标签，因为人们倾向于过于强烈地认同这个术语。这个标签伴有某种程度的宿命论——这个观点认为，这是一个人的学习生涯中难以抹掉的某个部分，而不是某种能被转变的事情。相反，我喜欢思考这些学习困难或学习机会之间微弱的联系，因为它们向个人提出挑战，以使达到新的高度”（p. 30—31）。阿姆斯特朗甚至提议，应该重新解释“学习困难症”（LD）的标签，并重新理解“学习差异”的意思。我尤其是喜欢这个提议，因为它恰好与当今的人脑研究结果相一致——研究表明每个人的头脑都是独一无二的，这个星球上没有两个头脑是相同的。在我们现有的评价实践中，写作和口头表达是最高层次的！所有学生学习相同的教材，被按照同样的方法教育，以同样的方式测验，这就暗示着所有的大脑是基本上相同的。

在《多元智能评价方法》（1994）一书中，我拓展了加德纳关于为每个学生建立一个智能档案夹的概念。通过运用各种观察措施，我们的确能收集到关于个体学生独一无二的智能构成的信息。那么我们就帮助他们运用他们的强项致力于发展那些不是强项的智能。

当评价学生学业的过程开始时，如果不是用所设计的测验来指出学生的失败和错误，而是我们给学生提供机会展示或证明他们所知道的，那会怎么样呢？如果我们的评价是对他们学习的一种真诚的赞美，那又会怎么样呢？

在得克萨斯州，我曾多次举办过多元智能的专业发展研讨会，英语文学教师威瑟（Wieser）女士叙述了下列体会：

那是关于莎士比亚的一个重要单元的结尾。整个单元给了学生们许多机会运用多元智能来学习材料。威瑟女士不断教给学生许多认识方式。

到了教研室所要求的单元测验的时候，威瑟女士给全班复习了教材，确信使所有学生都理解了他们应该掌握的知识。接着，她对全班说：“你们每个人都要以你所能的最佳方法向我证明，你已经真的理解了这篇教材。你们可以做任何事，但是你们必须以某种方式向我展示你们已经理解了。”

正如我确信你们已经猜到的那样，某些学生就立即着手写一篇短文。然而，有一个年轻人说，“威瑟老师，我的生活不能依赖于写篇短文，我能不能画一幅画？”

威瑟女士说，“可以，但是你记住，通过画你必须能够证明给我，你真的理解了教材。”

威瑟女士告诉我，这种体验改变了她的一生。那个年轻人带来了他的画，并把它铺开在地上。他们就这幅画展开了一个 4—5 分钟的对话，年轻人解释了画中的各部分的含义，威瑟女士问了他许多问题，证明他的确理解了。

她告诉我，在那个对话结束的时候，她明白了，这比让他写一篇短文更能知道他对莎士比亚的真实理解。

霍华德·加德纳（1993）的引语确实为多元模式的评价方法创造了条件：

关于个人强项的各种文献资料和关于多种认知风格的研究发现，都有重大的教育意义。首先，早期就辨别一个人的强项和弱项是非常重要的，这样可以使它们能成为教育计划的一部分。关注个体间的差异也能引发这样的问题：每个人是否该选修同样的课程？在统一课程下，是否要以相同的方式呈现给所有的学生？

只有当正式的测验被用来凸显而不是遮蔽不同认知特征的时候，测验才能实现对不同认识特征的褒奖。……设计一些用于“把关”的措施，允许学生展示他们的强项和最佳表现（就像大学的使命）是特别重要的。然而直到现在，就此我们几乎没有做过多少努力，测验被更加频繁地用于指出学生的弱项而不是指明其强项。（p. 170）

在《多元智能评价方法》（1994）一书中，我介绍了多元模式评价的概念，以此确保所有学生都能以公正的方法证明自己对各种学科领域和学校课程概念的理解程度。加德纳有关理解力的理论和过程的研究，引导他提出了这样的建议，如果我们关注对学生理解力的整个评价过程（针对仅仅再现记忆过的或先前学习过的教材的能力），我们就必须以至少三种不同的方式评价他们。如果他们能准确地以至少三种方式“再现他们的理解力”，那他们极有可能对被评价内容掌握得很熟练了。换句话说，如果他们准确写出了你所评价的材料，你就也得要求他们通过绘画、雕刻、油漆或身体性的方式如戏剧扮演、跳舞和体育运动，栩栩如生地再现他们的理解力。

无论最初编制和实施这种多元模式评价方法有多么困难，但这种做法能将我们带进更有力度、更准确、更公平的评价实践中。就如我在第一章结尾所说的，仅仅因为一个学生能够阅读、写作或计算，并不能说明他真的理解了。只是单单表明他用词良好，或者他也可能在想出和表达正确答案方面运气不错。反之亦然。仅仅因为他不能用语言表达出他的理解力，并不意味着他真的缺乏理解力。提出这种理解力是多元智能理论革命性意义之所在。

超越标准化测验

在目前的各种测验模式中，智能被看做是不可变化的。我们生来就带有某种“先天能力”，它以一种相对常规的、可预言的方式得以发展。所以，同一类型的测验措施能够用于各个年龄段的所有学生，并可以指出谁掌握了、谁还没有掌握教材。

越来越多的高等教育界的知名人士认识到，大学入学考试的分数不一定能准确地说明一个学生在大学里的表现会如何。任何标准化测验，几乎都着眼于从一个人教育历程的记忆库中提取出正确答案或对没有情境的问题加以精确计算。任何一所最受欢迎的大学的入学考试，都没有涉及产生正确答案或实际问题解决策略的认知过程。然而，这些恰恰是大学生成功应具有的能力和技巧。加德纳（1993）就人类知识的本质作出了下述令

人鼓舞的评论：“完成某项任务要求具备的知识全部储存在一个人的大脑中，这种推断是不正确的……说一个人‘有’或‘没有’必需的知识，就太过于简单了……正式测验总是聚焦于个别时刻、储存在个别人大脑中的知识，这样可能会歪曲、夸大或明显低估个体在更广阔的社会情境中能够作出的贡献”。（p. 172 - 173）

正确答案是暂时的，学习过程和整个人类成长和发展过程却是终身的。加德纳继续建议道，我们需要学会以一种很不同的方式来看待人类的认知，“人类认知能力是一种显露的能力，可能明显地是由三种不同的要素交叉组成：拥有能力、知识和目标的‘个人’；使能力得以激活的‘知识领域’的结构；一系列制度和规则——一个被环绕的‘领地’——当一个特殊的表现被接受时，当它不能迎合规范时，可以据之加以判断。这样的认识是非常有意义的。”（p. 173）加德纳所提到的三种要素，都需要被呈现在任何可行的测验情境中，尤其是那些决定谁将进入大学的测验情境。现有的测验，诸如 SAT 和美国大学测试（ACT），仅仅涉及这些要素中的一个——知识领域；甚至这个要素也常常被呈现在人类认知的全部范围中相当低的水平上（按照布卢姆的分类学）。

闭上你的眼睛，尽可能生动地想像下列场景，这些场景是根据戴夫（Dave）和简·尤利瑞（Jan Ulrey）（1992）的研究工作设计的，这两个人是全国有名的顾问，专门研究如何推行发展性的、适当的评价方法。在每一个节点之后，停顿几分钟，允许你进行想像和反思，如果当前的学校按照这种方式运作的话，会有怎样的不同。

你已经神秘地进入到一种教育情境中，在这里人们认识到，通过积极地参与学习的过程，无论是哪个年级水平，孩子们都学得最好。

- ▶ 这些人进行的教学实践研究了学生需要的全部范围，包括智力的、社会的、精神的、身体的、创造性的和审美的。
- ▶ 教师们能强烈地意识到他们的学生有各种学习风格和智能，要运用这个知识来开发课程和教学策略。
- ▶ 学校团体对所有的孩子进行评价，重点在于发展他们的强项。
- ▶ 孩子们分享了八种智能中激发他们成长和发展的各种学习经验和活动，这些学习经验与活动受到平等重视。