

第一编

教育评估概述

第一章 教育评估的涵义及历史

第一节 教育评估的概念

一、什么是评估

对教育评估的涵义虽然在理论上有不同的见解，但是在我国的实际运用中，评估和评价通常是被人们当作同一概念在同一个意义上加以使用的。例如，在我国的教育行政管理活动中就可以看到二者是在本质意义相同的前提下并行使用的。在 1986 年 3 月 25 日国务院《关于第七个五年计划的报告》中提到：“要加强教育事业的管理，逐步建立系统的教育评价和监督制度。”而 1995 年 9 月 1 日正式颁布施行的《中华人民共和国教育法》规定：“国家实行教育督导制度和学校及其他教育机构教育评估制度。”又如，在高等教育领域多用评估，普通教育领域多用评价，实际上是同一个概念。在我国教育评估领域里就词源而言，评估与评价一样都译自英文的 *evaluation*。

由于英文 *evaluation* 一词的原义是引发出价值也就是作出价值判断的意思。因此，评估这一词汇的内涵就被解释为：依照一定的标准对客观事物进行观察并作出价值判断的过程。我们认为评估是人类社会生活中的一项基本活动，对于我们自己当前或过去的活动，人们无时无刻都在做出好与坏、成功与失败、长处与不足、有效与否的判断。但是，这些判断活动并非都是有意识的，并运用系统科学方法进行的。只有那些对我们认为是重要的活动所进行的有意识的评估，并运用科学系统的评估方法进行的评估，我们才称之为评估，才有积极意义。因此，严格地讲，评估是根据预定的标准，观察分析一项活动、一件事情做到什么程度，取得了什么结果，并分析解释实际效果与预定目标的差异，说明目标的合理程度和实现的水平等的过程。

二、教育评估的概念

那么什么是教育评估呢？

教育评估虽是现代教育研究的三大领域之一，但如何定义教育评估，到目前为止，尚没有一个被普遍认同的科学界定。

（一）国外学者的观点

国外学者在教育评估的发展过程中主要有以下几种主张。

1. 泰勒 (Tyler, R. W.) 在著名的“八年研究”(1933—1940)报告(《史密斯—泰勒报告》)中首次提出教育评价的概念。认为“评价过程在本质上是确定课程和教学大纲在实际上实现教育目标的程度的过程”。^①

2. 1963年克龙巴赫 (Cronbach, L. J.) 在题为《通过评价改进课程》的论文中对教育评价内涵的阐述是：“一个搜集和报告对课程研制有指导意义的信息过程。”^②

3. 1966年，斯塔弗尔比姆 (Stufflebeam, D. L.) 在对泰勒评价理论提出异议的前提下，主张“教育评价不应局限于评判决策者所确定的教育目标所达到预期效果的程度，而应该是收集有关教育方案实施全过程及其成果的资料，为决策提供信息的过程”。^③

4. 1975年比贝 (Beeby, C. E.) 把评价定义为“系统地收集信息和解释证据的过程在此基础上作出价值判断 目的在于行动”。^④

5. 1979年出版的日本辰见敏夫编的《教育评价小辞典》中，对教育评价的解释是：“依据教育目标和一定的价值观判断学生的学习成果、课程及教育活动状况的合目的性即实现目的的程度的过程。”

6. 1981年美国教育评价标准联合委员会对教育评价进行了综合性的界定，他们认为“教育评价是对教育目标和它的优缺点与价值判断的系统调查，为教育决策提供依据的过程”。^⑤

以上这些对教育评价内涵的界定是教育评价专家在教育评价实践和理论发展过程中提出来的见解，虽有不同之处，但也不乏相同之点。不同之处反映出他们是从教育评价的对象、目的、功用、手段等不

参看瞿葆奎主编：《教育学文集·教育评价》人民教育出版社，1989年，第263页。

同上，第82页。

王致和主编：《高等学校教育评估》，北京师范大学出版社，1995年，第18页。

同上，第19页。

钱在森等主编：《普通教育评价原理与方法》辽宁大学出版社，1992年版，第5页。

同角度去揭示教育评价的基本特征。各种认识都在一定意义上揭示了教育评价的涵义，为我们理解教育评价提供了依据和启示。相同之点就是都认为教育评价是一种判断（包括描述），是一个判断确定的教育目标在实际上实现程度的过程，或者说是对教育活动的目的性的判断过程，对教育活动效果的价值判断过程。在以上几种观点中，1981年美国教育评价标准联合会给教育评价下的定义是需要给予重视的，“因为它是综合了各家之长的一种较为全面的界定，也是结合了教育评价的形成和发展，反映教育评价的实质和功能的界定，它侧重于评价的信息作用，同时兼顾了教育目标与价值判断”。^①

（二）我国学者的观点

我国教育评估的实践和理论研究始自 80 年代，虽然起步较晚，但发展迅速。其中对教育评估的理解及其内涵的界定是我国教育评估理论研究者与实践工作者共同努力探讨的一个问题。在探讨中他们本着“面向世界”、“洋为中用”的原则，学习和借鉴上述国外的观点，把握其内核，并本着“实事求是”、“立足本国”的精神，从我国的教育实际出发对教育评价的含义作出了自己的阐释。一般人们把教育评估定义为：“是根据一定的目的和标准，采用科学的态度和方法，对教育工作中的活动、人员、管理和条件的状态与绩效，进行质和量的价值判断。”^②

在理解这个定义的内容时要注意：（1）这里的“一定目的”一般是指教育总目标、各级各类学校的目标、各种教育活动及各科教学的目标、教育管理和学校管理的目标、教育改革的目标等。（2）定义中的“一定的标准”一般是指某一次评价所依据的评价指标体系，它是从质和量两个方面规定的评价内容和标准。（3）定义中的“采取科学的态度和方法”是指实施教育评价要采取实事求是的态度，选用规范的观察、访问、问卷、测量统计、评比评定等评价工具和手段。（4）定义中的“教育工作中的活动、人员、管理和条件的状态与绩效”主要是指评价对象的范围，它包括了教育的全部领域，或者说与教育活动相

钱在森等主编：《普通教育评价原理与方法》辽宁大学出版社 1992 年版 第 5 页。

② 王汉澜主编：《教育评价学》河南大学出版社，1998 年 第 15 页。

关的各个领域。

在理解这个定义所揭示的教育评估的特点时要注意:(1)教育评估是一个组织化了的的过程。就是说,它是一个有目的、有计划的自觉活动,是有规范性程序或步骤的活动,即它是由确定目标、搜集资料、分析整理资料、形成判断、指导行动等环节所构成的连续活动。(2)教育评估的根本是用价值观念对教育教学、管理活动中的人和事的状态进行描述进而作出判断评定。(3)教育评估为科学的评估标准、客观而可信度高的相关资料,以及科学的测量、评估工具与技术所支持。

三、教育评估与相关概念的辨析

(一) 教育评估与教育测量

教育评估与教育测量是既有联系又有区别的两个概念。但我们在日常工作中经常能看到将评估和测量混同的现象,因此要加以分析辨别,以期科学地使用这两个概念。教育测量主要是定量地描述所要实现的教育目标或目标实现的程度,或者说它是对测量标准的参与回应,是以多或少的数量表示的。而教育评估是对教育活动、教育系统工作情况进行分析判断,它是对评估标准的参与回应,是定性的表述,如合理、不合理、适当、不适当、好、不好、优、劣等。应该说二者是不同的,但它们又并非对立或毫无联系。教育评价的产生与发展,与教育测量的产生发展是分不开的。教育测量是教育评价的基础,现代教育评价强调它包括了教育测量。一般说,教育评价在与教育测量相区别使用时,多为对儿童、学生进行综合评价,对整个教育活动、教育系统工作进行评价。

(二) 教育评估与教育评比

教育评比是教育行政管理经常使用的一种传统的管理方法。80年代以来,我国教育评估的理论研究与实践倍受关注,在这一进程中也出现了把教育评估与教育评比混用的情况。虽然评比与评价都把比较作为一个基本的方法,但从比较的目的、标准、手段、功用等方面看还是有着诸多不同的。教育评比的目的重在确定比较对象,即教育工作或活动中的人和事的孰优孰劣、优劣的程度及其顺序名次,以利实施选拔与奖惩。教育评比也依据一定的条例和准则、标准,但其多为管理者依据自身的经验和上级权力要求或一定时期的工作任务的

需要而定，带有较强的主观随意性。教育评比在方法上，虽也常用积分法，但其评比项目的部分主要是依经验而定，且积分方法多为简单的单项分值相加。而教育评估的目的不仅是对教育工作、活动及其人员和对象的优劣的比较，更本质的是对评估对象的价值判断。其功用也不仅是选拔与奖惩，更为关注的是促进评估对象的改进与发展。其评估标准是以科学方法从量和质两个方面对评估对象的规定。评估方法更不是简单的横向比较，而是多种科学评估手段技术的综合运用。因此，教育评比远不能代替教育评估。

第二节 教育评估的目的与功能

一、教育评估的目的

进行评估活动 首先 就是要确定评估目的。换一句话说 就是在每一次评估活动启动之际，操作者都必须先回答这样一个问题，即该次评估活动要达到什么目的。不管是哪一个领域的评估，如教育评估，不管是哪一类评估如政府评估，包括教育督导评估或民间评估，从实施程序上讲，概无例外。

目的是对所进行的活动未来状况及结果的预期。通常人们首先从哲学意义上理解，它是人类活动的根本特征之一。马克思主义根据意识与存在的关系，第一次对人类活动的目的性作出了科学的说明。马克思指出 人的活动是有目的的 即在活动结果产生出来之前，先在头脑里以观念的形式构造出来。对此我们可以理解为，人们进行任何活动的目的都是由人们所规定的。当然，这并不是说活动目的是人在头脑中主观臆造出来的，马克思深刻指明这是客观见之于主观的东西，是主客观的统一。这正是人类活动与动物活动的根本区别，动物的活动只是一种本能的生命活动。目的，对于评估活动及具体到教育督导评估活动来说，同样是第一重要的。其次，随着管理实践与理论的发展，目的也成为了管理学的一个基本范畴。任何管理活动都有目的，其目的至少有两层意义：一是管理活动的终极目的是实现管理客体的使命和需求；二是管理职能和手段的选择、发挥和使用的目的，而后者是为前者服务的。因此一般管理理论认为，管理活动的功能（或称职能）及手段与管理活动的目的有着密切的关系，在一定意

义上，确定目的就是选择功能。

作为管理活动过程中不可或缺环节的评估，具体到教育评估活动目的确定，通常也是从评估功能的选择入手的。

（一）从评估功能的内容上对评估目的选择

就理论探讨而言 教育评估的功能是多方面的 如检查鉴定功能，反馈指导功能 激励功能 预防功能 补救功能 成果推广功能等等。这反映了教育评估活动的发展及教育评估内容的扩大和作用的变化。但就其功能的内容而言，最根本的是监督鉴定与指导改进这两个方面，其他功能都可以看作是这两个功能的延长、作用、结果或方式、手段。比如，反馈功能就是观察、监督检查鉴定功能的延长，即检查鉴定结果的处理应用。预防、补救等功能可视为指导活动的作用与结果。沟通功能实际是检查和指导功能发挥的方式、手段。因此在评估中 功能的选择主要是考虑这两个功能，以哪一个为根本或主导功能的问题 而不是二者择一的问题。通常我们在谈到评估目的时 常问 是鉴定、评比还是指导、改进、发展 其实并不是二者择其一 也不可能绝对地做二者选一，因为两个功能是有着一定联系的。特别是指导、改进、发展功能的实现是需要检查、鉴定为前提、为条件的。严格地讲，只能是以哪一个功能为主，或者为最终目的的选择确定问题。

（二）从评估功能作用的领域选择上确定其目的

教育评价理论研究成果显示，就教育整体而言，教育评估是为了端正教育思想 提高教育质量。但是 对教育不同领域活动的评估（或者说教育评估的对象领域不同）其要求不同 具体的目的也不同 尽管它们的最终目的也是为教育质量的提升服务。到目前为止教育评价研究把教育诸领域的评估的具体目的的概括地分为以下四个方面：为指导教育、教学活动服务 为学生学习服务 为教育调查、研究活动服务；为学校管理活动服务。以下稍做具体阐述。

1. 指导教育、教学活动的目的。

这是基于教师等对于学生施以教育教学的指导者立场而利用评估的目的，也是为选择、决策更加有效的教育教学计划，选择决策适合每一个学生的教育教学方法而利用评估的目的。具体一点讲，以下的诸种评估都是以此为目的：(1) 为了获得编制指导学生学习的长期

计划的信息依据而实施的评估，有关学生的智力、性格、兴趣等的标准化检查都是以指导学生的教育教学活动为目的的评估。(2)为选择各科教学的教法编制单元教学计划而实施的事前评估。(3)为防止学生学习跟不上教师的教学，在教学过程中实施的学生对所教的每一个基础、基本的内容的理解、掌握的状况的评估检查。(4)在单元教学结束时，进行的以了解学生学习问题及给予辅导为目的的事后评估，学期末及学年末为反省教学计划的恰当与否，实施得如何，今后怎样改善等为目的的终结性评估。

2. 为学生学习服务的目的。

这是将学生等学习者自身作为评估的当事者，以自己评估和学生相互评估的形式让学生进行的评估，是以学生对学习的自我调整为目的的评估。学习心理学指出，将学生的学习结果交给学生评估，这对他们来说是最好的学习。在这个意义上，这一评估目的，较之过去，现在普遍受到了极大的重视。作为具体的手段，除了传统的成绩通知单的发给以外，还有日常教学中学生对教师提问的回答、学生学习作业的查阅情况的不断反馈等。当然，对学生的自己评价采取放任自流的态度和方式是不行的，因为学生不喜欢也难于自己进行评价，所以教师给予经常的指导和帮助是十分必要的。

3. 为管理服务的目的。

就学校管理而言，班级的编制、学习小组的划分，成绩的记录与通知，高中与大学的入学选拔录取，企业里员工的录用决策，各种资格的认定，奖励与惩罚的依据获取等，都是利用评估的管理目的。这种评估目的，在评估活动的历史上曾经是最重要的。严格说这是以发挥评估的鉴定、评比的功能为根本的评价。今日评估目的选择的重心已移向对管理行为、管理活动的改善及发展的促进。但前者在学校管理实践中、在社会中依然是非常重要的。此外，在实践中这一评估目的的实现尚存有许多问题，例如高中及大学的入学选拔、各种资格的考试、学生的综合评估表 过去的成绩单 筹。因为这一评价目的较之其他的评估目的有着浓厚的个性特征，即它与个人的人权、社会的公正有着更加密切的关联。因此 在实施这一目的的评估时 必须尊重、重视它的客观性和可信度。

4. 为调查、研究服务的目的。

为了适应社会发展的需求，研究开发教育课程，研究开发教法、教材和教具等常利用教育评价。教育评价的这一目的在教育评价的发展史上无疑是比较新的，或者说是近年来才被重视的评价目的。特别是在美国，它有着日益被国民所关心的趋势，但在另一些国家尚未引起足够的重视。其原因一方面是由于教师、国民对教育课程改革的关心度一般较低，另一方面是国民对教育结果的知道权要求较低。在评估实践中，以调查研究为目的的评估对象范围在不断扩大。例如，一个国家、一个地区（包括省、市、县、乡甚至村）为了解其实施的教育计划取得了哪些成效，还有哪些不足、缺陷而进行的评估。小到一个班级、一所学校所进行的学力标准测验，学期末、学年末所实施的综合评估等都同时具有着调查、研究的目的成分。

总之 确定评估目的是评估程序的第一步 它的重要性无须赘言。简单地讲，它决定着所进行的评估采用什么样的评估标准，也决定着评估工具、手段的选择 影响着评估的组织实施的一些具体做法。

二、教育评估的功能

教育评估的功能是指教育评估活动本身所具有的能引起评估对象变化的功用和能力。它通过教育评估活动与结果作用于评估对象而体现出来。其功能的内容取决于评估活动的结构及运行机制。

在教育教学活动、学校管理过程中，教育评估有着不同的类型。不同类型的教育评估所具有的主要功能是不同的，而教育教学活动进程、学校管理进程的不同阶段所需的教育评估的类型也是不同的。系统理论告诉我们，教育活动是一个由教育目标、教育教学活动开展、教育评估所构成的系统过程。学校管理也是一个由教育决策、决策执行、反馈控制活动所组成的系统。不管在哪一个系统中，作为具有反馈控制功能的教育评估，在整个活动中都占有不可或缺的地位。它是教育教学过程中一个基本环节，也是学校管理过程的一个基本步骤。从整个教育活动、学校管理活动来讲 如果没有教育评估工作，只有教育目标、教育教学活动，只有决策及决策执行活动 实践中必然会使教育教学活动缺乏现实基础和科学依据。缺乏检查指导，必然会使学校管理的决策缺乏科学论证，执行缺乏检查监督，最终影响教

育目标的实现，降低学校管理的效率和效益。

具体说到教育评估对教育教学活动和学校管理活动的功用，可以通过教育教学活动、学校管理过程中所需的不同类型的教育评估加以分析阐明。

一般讲，在有目的、有意识、有计划的活动过程中需要四种类型的评估。第一，活动开始之前对活动计划的可行性论证，即评估活动计划是否切实可行。第二，在活动开展进程中对活动是否有效推进的评估，也是为修正调整活动自身运行服务的评估活动。第三，在活动进行到一个段落时为了解、把握其活动成效而开展的评估。第四，站在活动以外的立场为活动的改善提供意见的评估。教育教学活动、学校管理活动也需要以上四种评估。在教育评估的发展历史中，迄今比较重视的是第三类评估，即为了把握教育成果的评估，人们称它为终结性评估。而最近，其他几种评估也受到了更多的关注。相对终结性评估，第一类的活动前的评估，人们称之为事前评估或诊断性评估；第二类活动中的评估被称之为形成性评估。第四类由外部进行的评估被称为外在评估或他人评估。第四类评估，首先，对被评估的教育教学、学校管理活动有着不同的功能 其次 也包含着某些共同功能。这是因为不同类型的教育评估 其目的、结构、运行机制有所不同 而作为教育评估活动又有着共同质的规定。如果在归纳的基础上大致区分一下，可概括地分为：

（一）导向功能

教育评估的导向功能是指教育评估本身所具有的，引导评估对象朝预定目标前进的功用和能力。它内含于教育评估的评估目标、评估指标这些构成因素中 通过教育评估的结果、信息的反馈、利用 指导评估对象行为的运行机制实现。

（二）鉴定功能

鉴定是指对人和事的优缺点，优劣程度的鉴别和评定。教育评估的鉴定功能就是指教育评估对教育评估对象的目标性、合格与否、优劣程度、水平高低的判断确定的功用和能力。它主要通过对教育评估对象相关资料的收集、整理、分析、判断的运作机制得以实现 通过检查、比较、判断等评估工具的利用而获得。

(三) 激励功能

激励是一种引起需要、激发动机、指导行为以有效实现目标的心理过程。教育评估的激励功能是指，教育评估活动所具有的使评估对象产生或形成逼近及实现预期目标而不断进取的内在动力的功用和能力。它通过向评估对象反馈评估的信息这一评估的运行机制加以实现。

(四) 改进功能

改进有改变旧有状态，使其有所前进的意思。教育评估的改进功能就是指教育评估活动内含的使评估对象反省自身状态、克服不足、促进发展、实现目标的功用和能力。它主要是通过教育评估的信息反馈、评估结果对评估对象行为的指导等运行机制得以显现。

第三节 教育评估的产生与发展

教育评估作为科学的概念是在 20 世纪 30 年代正式被提出来的，它发端于教育测量运动。

一、教育评估的源头

一般说来，人们对自己或他人的能力、适应性等的评估行为，是人们的社会生活中不可或缺的，也就是说是必要的。众所周知，连猴子都知道判断猴群中其他个体的能力强弱，进而了解自己，把自己放在一个适当的位置上。而人类呢，可以说，人类的个体在朝向他人的一切目光里，都包含着对对方的判断及与自己进行比较的方面。

在人类团体以及社会性的组织中进行能力与适应性的评估活动有着非常久远的历史。例如，人类文化学的研究成果表明，许多民族都有举行成人仪式的习惯。在那个仪式上，达到一定年龄的青年由年长者主持并指令他们进行各种各样的训练，通过观察确认他们是否具有了作为部落一员的能力。可以说这就是原始部落在集团中进行的一般性的能力测定。

在国家这样大规模的人类社会组织被建立起来之时，对人的能力及适应性的评估活动在人才的选择、录用及调配上就具有了重要的意义。例如，6 世纪 公元 606 年在我国的隋朝开始实行了科举制 即通过分科考试来选用人才 录用官吏 采取了贴经、墨义、策问、

诗赋等方法去测试考生。这种制度建立在打破门阀与实力主义的原理之上，在当时是非常进步的，作为一种制度是具有划时代的意义和合理性的。宋代以后，其内容虽有所变化，但这一制度一直运作到 20 世纪初的清末，给各国的公务员制度，特别是采用考试的方法选拔录用公务员以极大的影响。

教育评估可以说是随学校教育的出现而产生的。也就是说，自实施系统教育的机构——学校产生以来，就有了对学生入学、升级、学习情况、毕业的认定，就有了对教师教学质量的考核与评定。因为这是学校管理运作上不可缺少的要素。也就是说，不管以什么样的形式，测定也好，评估也好，总是要对学生的学力以及教育结果进行检查确认的。在我国，《学记》中就有记载：“比年入学 中年考校，一年视离经辨志 三年视敬业乐群 五年视博习亲师 七年视论学取友 谓之小成。九年知类通达 强立而不反 谓之大成。”又如，日本创立于 7 世纪律令制下的大学，就规定有旬试和年终考试两种考试。19 世纪上半叶在西方国家的学校里，广泛采用了口试、面试的方法检测学生的知识、能力和教育结果，其起源最早可以追溯到古希腊苏格拉底的产婆术，即问答法。最初教育评估的最大特征是它的主观性，就是说这些评估缺乏客观标准。具体表现在两个方面：（1）考试方法单一 试题少，不足以反映学生所得知识与能力的全貌；（2）评分上受主考人、阅卷人主观标准的左右 口试还受到考场气氛、学生个性的影响。随着学校教育的发展，这类评估的弊端日益被关注。为了确保评估的客观性，矫正弊端，出现了新的评估方法和评估体系，这就是教育测量运动。

二、教育测量运动

19 世纪末到 20 世纪 30 年代，欧美一些学者在当时的实验心理学、统计学与智力测验发展成果的影响和推动下，为排除以往评估试题编制的片面性和评分方法上的主观经验性，对学校中的考试进行了积极的改革，逐渐形成成为教育测量运动。教育测量运动的发展历程大致可分为以下几个阶段：

（一）初始阶段

改革首先是否定口试实施笔试。笔试先在大学里使用，最早是英国的剑桥大学于 1702 年以笔试替代了口试。而首先在公立初等中等

教育阶段真正使用笔试的当为美国。19 世纪，美国的初等学校普及，中等学校学生人数日益增多，对学生以至毕业生一一进行口试已不可能。于是，1845 年在美国著名教育家贺拉斯·曼(Horace Mann) 的倡导下，美国的波士顿文法学校首先引入了笔试。

贺拉斯·曼在阐述笔试较之口试具有客观性也更为妥当时指出四点理由：(1)可以获得有关每一个学生学力的更多的信息资料。(2)可以留下关于学生的学力情况记录。(3)可以用同一的试题测试众多的学生。(4)可以减少教师对某些学生的偏爱、偏袒及不公正评估。

应该说，从口试转向笔试，确实使评估向客观化迈出了一大步，使评估或测验手段更加灵活。但是，笔试也就是书面考试，也存在着不容忽视的不足，即其试题内容仅为能够用文字在纸上表现出来的知识性教育成果，而无视对学生情意等方面发展的检测。此外，成绩评定即评分的客观标准问题尚未解决。

在使评估走向客观化的早期努力中，还有一个研究不能忘记，那就是致力于评估标准的客观化、明确化探讨的英国的学者费舍(Fisher, G.)，他被称为是致力考试客观化的先驱者。为了尽可能地对自己学校的学生的试题答案和作品进行客观评分，他收集了大量的学生成绩样本，区分优劣，作为评定学生成绩的客观标准，1864 年编制成《量表集》。它表示依据一定的价值程度进行评分的标准的尺度表或量表。虽然费舍的这一尝试性研究在当时并未引起教育界的反映，但在 50 年后桑代克(Thorndike, E. L.) 筹开拓教育测量运动时被作为依据，给予了极大的重视。也有人视这一研究成果为绝对(到达度)评估的一个努力。此外还可举一例是美国的莱斯(Rice, J. M.) 的拼字测验。其测验目的是要辨别学生学力的个人差异，了解学校教育的成果，具体是希望通过对学校学生的答案或作品尽可能地作出客观的评分，提供学校教育改革的依据。1897 年莱斯发表了他对 20 所学校 16000 名学生的拼字测验结果。莱斯的测验表明：8 年中每天用 45 分钟同每天用 15 分钟进行拼字练习的学生的测验成绩并无多大的区别。这一结论引起教育界的强烈反响，使人们对测验的方法、测验的客观性产生了普遍的关注。莱斯的研究是出于适应时代

进步的要求，进一步丰富教育课程的考虑，所以后人高度评价了他的拼字测验，称其为客观测量的先驱、教育测量的创始人。严格地讲，至此的研究依然是教育测量的初始阶段，或称之为胎动时期。对教育现象进行客观测量活动日益活跃起来是进入 20 世纪后的事情。

（二）形成时期

为什么 20 世纪教育测量能引起人们的普遍热情，其原因除了教育实践的需求以外，还有两点：一是 19 世纪物理学、医学、天文学等自然科学的研究，由于使用了数量测定法而获得了快速的发展，这样的发展给人文科学以极大的刺激和影响；二是受 19 世纪后半叶高尔顿 (Galton, F.)、卡特尔 (Cattell, J. M.)、比纳 (Binet, A.) 等的个人差异心理现象的研究和统计理论、方法发展的影响与推动。

国外一般把美国的桑代克作为教育测量运动的开拓者，称其为测量运动之父。他于 1904 年发表的《心理与社会测量导论》(An Introduction to the Theory of Mental and Social Measurements) 一书被作为这一时期的标志，并被高度评价为测量学上划时代性的著述。在该书中桑代克提出了一个著名的论断：“凡是存在的东西都有数量，凡是有数量的东西都可以测量。”这一论断推动了他与他的弟子们共同投入对教育成果的测量研究。此后，陆续编制出测量标准、标准测验。斯顿 (1908 年)、高斯 (1909 年) 相继编制出了算术的标准测验。1909 年，桑代克也发表了他的书法量表，此后又编制出了拼音量表、作文量表、图画量表。到 1928 年在美国就有约 1300 多种标准化测验问世。

另外，应该提到的是作为标准化测验发展的强有力支持的一个研究，即对以往的由教师编制、实施、评分的论文题式考试信度的实证性研究。该研究结果表明，以往的考试及评分的可信度很低，必须进行客观性测验的研究。客观测验最初只是利用标准量表，后又用教师自编的测验，其开拓者普遍认为是美国的麦柯尔 (McCall, W. A.)。他在 1920 年前后发明的 T. B. C. F 制可以说使教育测量在科学化的道路上迈进了一步。

以上的测验是以学生的学力测验为主的，而同时期作为教育测量运动的另一个内容是智力测量的出现。1905 年，法国的比纳和西

蒙 (Simon, T.) 提出了他们的第一个智力量表即《比纳—西蒙量表》,1908 年、1911 年又两次修订了这一量表。其中 1908 年的量表中引入了智力年龄的概念。他们的智力量表奠定了此后智力测量的基础。在此之后,美国斯坦福大学的心理学教授推孟 (Terman, L. M.) 对《比纳—西蒙量表》进行了五年之久的研究修改于 1916 年发表了他修订后的新量表。该量表不仅引用了德国人斯腾提出的智商概念,并提出了计算智商的公式,从而使心理测验走向成熟。嗣后,在欧美日本等国兴起了利用教育测量进行教育研究之风。

从以上教育测量运动的形成中,可以看出教育测量是以心理学、测量学、统计学为指导原理的,其实质是改变过去传统考试的主观性即信度低的问题,使考试客观化。从中也使人们认识到,不基于客观的、可信度高的教育情报所编制的教育计划和进行的指导是十分危险的。正是在这个意义上我们可以说,教育测量运动是现代教育评估的原点 教育评估必须是客观的、有信度的。

(三) 发展时期

进入第一次大战后的 30 年代,教育测量运动的发展出现了转机,那就是心理学、教育学的新发展。心理学和教育学的新研究成果表明 人的教育不是单一的知识传授 而是智力、社会、情绪、身体的统一的、全面的教育。用这一观念审视教育测量可以看到,教育测量固然能使学生学习知识的考查客观化、数量化、精确化,但不能测量学生学业的全部 如学生的学习态度、兴趣、鉴赏力、品德、性格、情操等方面,它只能侧重于可进行数量化知识方面的测量,应该说这正是教育测量的重大缺陷。这一缺陷往往导致学生死记硬背教材,不利于学生个性的发展。反省这一缺陷产生的根源,当时专业领域普遍认为是忽视了教育目标,而教育活动是有自身的明确目标的,教育的价值也正体现于此。不管是教育测量也好,测验也好,如果不与教育目标紧密结合是毫无意义的。基于这样的观念,教育测量转向了教育评估。这一转变过程在美国大约发生在 1930 年 此时也正是美国新教育运动兴起的时期。

三、现代教育评估的产生——教育评价运动

为避免阅读其他书籍中有关内容时产生混乱,这里使用多数人

翻译介绍泰勒等学者的研究成果时使用的评价一词。

现代教育评估诞生于教育测量的发展，而推动教育测量走向新发展的是教育评价运动。为教育评价运动奠定基础的是美国进步教育协会(Progressive Education Association)进行的有关课程与评价的“八年研究”。这一研究是从1933年至1941年进行的。在教育应该从重知转向全人教育的理念引导下，首先对课程内容进行了实验研究。与此同时，为了检查、比较两种教育形式的效果，以由美国大学著名教授泰勒等人为成员的评价委员会组成了。该委员会一方面批评教育测量存在的问题，另一方面积极研究考察教育效果的方法。通过研究，泰勒等不仅提出了“教育评价”这一新概念，并依据此概念设计了一套评价方法。泰勒认为评价就是测定目标已达到的程度。基于此概念而设计的教育评价方法，就是根据预定的教育目标，通过系统地有目的地收集资料来判断实际教育活动是否达到了预定的教育目标或判断达到目标的程度。这种评价方式被后人视为“行为目标评价”的雏形。对至此的发展人们概括地说，引导30年代的教育测量走向教育评价的理念无疑是教育学的理论。这在教育评价的发展史上具有十分重要的意义，它表明评价不是为评价而评价，必须是为了实现教育目标。美国进步主义教育协会为评价“八年研究”成果组织成立了评价委员会。1942年该委员会提出了报告书，报告中概括了“八年研究”的基本指导思想，具体是：(1)所谓教育，是使人的行为样式发生变化、改善的过程；(2)种种的行为样式的变化就是教育的目标；(3)依据教育目标在实际中达成的程度去评价教育过程；(4)由于人的行为是复杂的，所以评价必须对各个相关方面进行评价，另外，评价不仅是专项的还必须进行综合性的评价；(5)作为评价方法，只用纸和笔的测验已经不够了，必须利用包括行动观察等在内的多种方法。应该说以上这些思想也是在教育评价运动中发展起来的。

四、教育评估的发展

如果说以泰勒等人的“八年研究”为内容的教育评价运动是现代教育评估的产生源，教育评价运动时期为现代教育评估的产生时期

参见[日]《现代教育评价论》，金子书房出版，1984年版。