

第三章 教育评价的类型与功能

一、依据目的来划分的评定分类

为什么要评定？对于评定来说，明确其目的是极端重要的。教育中的评定，其目的可以分为：(一) 旨在获得指导每个学生的资料；(二) 旨在检查和研究教案；(三) 旨在管理教育，等等。这里谈谈第一种评定目的。布卢姆(Bloom, B.S.)等把评定与教育工作联系起来，将评定区分为诊断性(配置性)评定、形成性评定和总结性评定。

(一) 诊断性(配置性)评定 这种评定是在各年度、各学期的开始和结束时实施的，它的目的，第一是了解学生是否具有新的教育目标(例如单元目标)所需要的基础知识与技能，亦即是否有准备；第二是旨在获得据以编班的资料，具体地说，1 了解学生的学力程度，2 探明学习上表现出来的缺陷及其原因，凡是为而实施的标准成就测验和准备(readiness)测验，均属此列。这种诊断性评定(diagnostic evaluation)的目的是，把最终获得的信息作为一个线索，以便进行适当的处理。不加处理，诊断是没有意义的。

(二) 形成性评定 传统的学习评定主要是学习结果的评定，形成性评定(formative evaluation)则是在教师的教和儿童的学的进程中，评定教学的成果，通过师生对于成果的了解，有助于教学工作本身的调节的一种评价。因此在形成性评定

中重要的一点是，教师和学生要明白下一步该教什么该学什么，重点何在。此外，教师要对自己的教案和教材进行检查，有时还得重新检验课程（参见图5—1）

具体地说，每当一个单元或一个章节这样较小的学习单位的教学告一段落的时候，要作几次测验（形成性评定测验）。一方面调节和修改整个教学工作，另一方面检查每个学生的学习进展状况。有时要诊断学习上的难点，并作出个别指导，帮助学生完全达到预期的学习目标。再有，形成性评定测验应视教学进度来实施，所以应由任课教师亲自完成。

（三）总结性评定 这是总结性地评定一学期或是一学年这种较长期的教学工作成果的手段（表5—2）。具体地说，就是根据学期末、学年末实施的学力测验与标准成就测验作出的评定。

总结性评定（Summative evaluation）的典型例子是成绩报告单。虽然，今日的成绩报告单，有许多是用了相对评定法和绝对评定法来表示学生的学力水准的实态的，但是它缺乏如下的本来应有所评定的最重要的侧面：为了使学习活动朝着更好的方向进展，学生自身应当怎么办以及家长应当怎样协助等。因此，尽管学生通过成绩报告单知道了自己的学力现状，但是不能期望他们把它同改进日后的学习活动和提高学习积极性结合起来。因此，人们希望就成绩报告单的应有状态展开探讨。

表 5-2 形成性评定与总结性评定的区别
(木村舍雄, 1975)

	形成性评定	总结性评定
目的	改进教授、学习与课程 (指出课题目标掌握与否的状态)	判定最终的学习成果 (分等、定资格)
职能	诊断教学过程, 提出改进措施	作出成绩评定
举行时间	教学进程中	教学结束时
评定对象	针对短期性的从属目标的形成性目标(局部的技能能力的评定)	长期性的综合目标(全面的综合性的技能能力的评定)
方法	在各单元中进行单元内的形成性评定测验	期终考试、学年考试、标准成就测验
测度	标准测度(是否达到教育目标的标准)	标准测度(反映全国同年级学生全境(population)情况的相对标准)
举行时间	分单元(小单元)每一至二周或一个月作一次	期末或学年末
评语	记述是否达到目标的要求、缺点、缺陷之所在	就综合性的技能和能力加以记述

大桥正夫编 钟启泉译:《教育心理学》上海教育出版社 1980年2月

二、自我评价

(一) 自我评价的一般概述

1. 什么叫自我评价？所谓自我评价乃是一个人对于自己生理、心理特征的判断，是自我意识的重要组成部分。随着个人自我意识的发展，其对个人本身的评价也经历着不同地阶段。

(1) 当处于生理的自我阶段，个体主要对自己身体、家庭和父母等对他的态度以及他对家庭成员的态度，自己所 物的判断，从而表现出自豪或自卑的自我感情。

(2) 当处于社会的自我阶段，自我评价主要表现在个体对自己在社会上名誉、地位、社会中其他人对自己的态度以及自己对周围人的态度等方面的评价和判断，也表现在对自己拥有的财产的感情的评价，从而表露出自尊或自卑的自我体验。

(3) 当处于心理的自我阶段，自我评价表现为自己的智慧、才干、自己的道德水平等方面的评价与判断，从而产生出自我优越感等的自我体验。

由于自我意识的不同发展阶段，自我评价的主要内容也随之发生变化，由此，在不同阶段具有不同的自我追求，在自我意识发展的第一阶段，是属于生理的自我阶段，在其自我评价的基础上，自己主要追求其身体外表，物质欲望的满足，获得家庭人员的关心与爱护等；在社会的自我阶段，在其自我评价的基础上，则自己处处追求他人的注意与重视，追求他人对自己的情感，追求名誉地位，追求金钱，这阶段上竞争心特别重；自我意识发展到了心理的自我阶段，在其自我评价的基础上，主要追求自己在

政治上、宗教上、道德上能上进，并且努力发展自己的智慧。

2 自我评价的功能。“人贵有自知之明”，正确的自我评价对于个人的心理生活及其行为表现有重大关系，对于协调社会生活中的人际关系亦是不可缺少的一个主观因素。早在二十世纪初期心理学家柯里就指出：“在人们的心理生活中，自尊或自卑的自我评价意识有很大作用。人们经常会把自己看作是有价值的、令人喜欢的、优越的、能干的人。如果一个人看不到自己的价值，只看到自己的不足，什么都不如别人，处处低人一等，就会丧失信心，产生厌恶自己并否定自己的自卑感，这样的人就会缺乏朝气，缺乏积极性，但柯里又指出：“如果一个人只看到自己比别人好，别人都比不上自己，这样就会产生盲目乐观情绪，自我欣赏，自以为是。因此就不能处理好人际关系，调动主客观双方的积极性，而且还会遇到社会挫折，产生苦闷。”

一般讲，个体对自己生理、心理等方面的评价不可能做到各方面都恰如其份。人们认识客观世界总是从不全面到全面，从不正确到正确，何况正确认识并估计自己是个更为复杂的过程，因为除了认知因素外，还会受到其动机、需要、愿望等其他心理因素的影响，因此往往容易过高或过低地估计自己。

问题在于如果自己的估计与社会中其他人对他的客观评价距离过于悬殊，就会使他与周围人们之间的关系失去平衡，产生矛盾，常此以往，将会形成稳定的心理特征——自满或自卑，并形成个人对他们的某种态度，不利于个人心理上的健康成长。

3 自我评价的形成途径，自我评价的意识并不是天生的，

它是在社会生活中逐渐形成起来的。

(1) 根据社会上其他人对自己的态度估计自己，个人评价自己是以别人对自己的评价为参照点。例如，有人能经常完成任务，把工作做得格外出色，常常受到社会上其他人的赞扬，尤其是受到了他所尊敬的老师、父母、领导的表扬，从而容易获得自信心，相反，如果有人对于布置的任务总是完不成，常常遇到困难而失败，因此受到人们对他的不满，批评，这样常此下去，就会使他缺乏自信心，看不到自己的力量，所以柯里指出，别人对自己的态度是自我评价的“一面镜子”，一个人处在一定的社会关系中，通过与他人相处，从他们对自己的评价中，看到了自己的形象，为自我评价提供基础。他人评价这面镜子并不是指某个人的某一次评价，而主要是指对自己有影响的关系较为密切的周围的人，从一系列的评价中概括出来的某些经常的稳固的评价，这才是自我评价的基础。

(2) 通过与社会上与自己地位、条件相类似的人的比较来估计自己，马克思指出：“人类起初是以别人来反映自己，名叫彼得的人把自己当作人，只是由于他把名叫保罗的人看作和自己相同。”

一个人自我评价高或低，不是孤立地进行的，他总是把自己与他相类似的人加以比较而作出的。例如，一个人用十五秒钟时间跑了100米，自己判断为快还是慢，是通过比较后得出的，他不可能与奥林匹克短跑运动员的记录来比较，也不会和幼儿园小孩子的速度来比较，他总是找一些与他相当的人加以比较。个

人评价自己的其他方面如道德品质、能力等都是如此。社会心理学家费斯汀格指出，一个人对于自己的价值“是通过与他人的能力和条件的比较而实现的。”费斯汀格把这称之为“社会比较过程。”

(3) 通过个人对自己心理活动的分析来实现自我评价。

前面讲过，个人的自我评价常常是根据他人对自己的评价与态度，从而估计自己，在这同时，自己也对自身的心理活动及其行为表现加以主观的分析，因此，自我评价不等于是他人评价的反映。华东师大心理学系学生会调查了大、中、小学生三百多人，要求他们既作关于道德品质的自我评价，又评价小组中其他人，结果表明，比较多的学生的自我评价符合或基本符合他人评价，有少数学生的自我评价与他人评价有较大距离，个别学生的距离很大。例如，有的学生评价自己的品质（诚实，大公无私等）都是“最好”，而小组同伴都评他“最差”、“很差”。也有个别学生把自己评价的很低，而同伴都给以公正而客观的评价。

这一研究表明，自我评价不是完全以他人评价为根据，乃是通过自我分析独立地完成的。为什么有的人自我评价远离客观评价而作心理分析，这既有自我认识的原因，又有其本能动机的原因。一种是由于自我认识不清而自我欣赏或自感不如他人，另一种是由于如实反映自我的真实评价怕他人取笑，有人想作如实的自我评价，而且评价确实很低，但又怕人因此轻视自己，有人实际上自我评价应该高些，却怕人指责为自高自大等，因此作出了不符

合实际的自我评价。

4. 自我评价的主观性。个人自我评价往往取决于他本人的社会角色、社会地位与价值观。有的人在客观上取得了很大成绩，周围人们也对他作出了高度的评价，但他自己却表现出缺乏自信而苦恼；也有人被他人看来並沒有值得他夸耀的地方，甚至表现很差，但他却怀有很大的自信心与自尊心。这实质上是自己对自己的要求不同所致，如果对自己要求很低，那么只要有一点成功就会变得自满自足。

由此可见，人们的自尊心和幸福感决定于自身内在所具有的心理结构，也就是说，个人对自身的评价最终还是操在自己的手里，生活中常讲“身在福中不知福”，在他人看来，那个人很幸福，生活不发愁，工作也有保证，该满意了，但他自己却感到物质生活条件还不如他条件好的人，工作条件也比不上人家那么轻松，因此常常自怨自艾，自叹命苦，如果换了一个人就会感到十分自在。生活也常常听到人们说：“以苦为乐，为苦为荣”，有人生活虽然艰苦些，工作也繁重，但他本人比比过去就感到很知足，能自得其乐，但这种条件放在另一个人身上就会感到苦不堪言。可见一个人的自我评价既是客观的，也是主观的，自我评价，取决于自己的心理结构、自我理想、自我要求等主观因素，从而决定了自己对自己的感情与态度，决定了自己对自己的判断与估计，因此产生了自尊、自信、自大、自卑等自我意识。

自我评价的意识十分多样，可以从三个方面加以分类。

(1) 以自己理想的形象作为标本的自我评价意识，如自我满

足感，自我不知足感。例如，一个人一心要追求成为一个医术高超的医生，当达到这一目标，就会产生自我满足感。如果由于种种条件的限制，达不到这一目标，就会产生自我不满足感。

(2) 以周围其他人作为标准的自我评价意识，如自卑感、自我优越感。例如：一个人千方百计要超过他周围的某个人，当达到要求时，就产生一种优越感，否则就产生一种自卑感。前者是自以为比他人好，后者则认为处处不如人家、低人一等。

(3) 以自我评价的意识为基础的情绪方面的表现，如自信心、自我胆怯等，看到了自己的力量就产生了自信心，否则就会胆怯。

(二) 关于自我评价的研究

早在两千多年以前，古希腊哲学家苏格拉底对于自己的人格就发生了研究的兴趣，他号召大家要“认识自己”。关于自我评价问题，国外心理学界有较多的研究，本世纪二三十年代，美国哥伦比亚大学教授霍林沃斯(Hollingworth)对于自我评价与他人评价进行了比较，探讨两者之间有何差别。他采用排队法进行研究，选择了25个被试。他们相互之间都比較熟悉，霍林沃斯提出九种品质(文雅、幽默、聪明、交际、清洁、美丽、自大、势利、粗鲁)，要求每一个被试根据这些品质把这25个被试(包括自己在内)依次排队，程度最高的排在第一，程度次高者排在第二，依次下去，一直到程度最低的一个，列为第二十五。一种品质评完后，再评价第二种品质，一共评价九种品质。

然后实验者加以统计，把各人在每种品质排列中自己所占的地位和其余24人排他的地位(平均数)进行比较，发现有很大

差异。例如，有一被试自以为他的“文雅”程度应该排在前几名的，可是把其余24人对他的评价平均起来，他的名次却排在二十名以后；另一被试对于“清洁品质”的评价，自己排的位置要比别人评价的平均的位置提前五名，“聪明”、“美丽”提前六名，“幽默”、“文雅”“势利”、“自大”、“粗鲁”等的自我评价要比他们评价的平均数分别要退后五名到六名。

这一实验结果表明，“优良品质的自我评价常常比别人的估计来得高，不良品质的自我评价却比別人低”，也就是说，符合社会要求的良好品质会自我夸大，不符合社会要求的不良品质会自我忽略。

2. 沙平(Sarbin)于1954年采用选择法来研究自我评价，他向被试提供一系列描写个性品质的形容词，要求被试从中选择符合其自我形象的词汇作为对自己的评价。结果发现，正常人中有70%的被试都提出了有利于自己的形容词，如“有夏心的”“爱社交的”、“自制力强的”、“积极的”等等；而神经病患者中有77%的被试提出了不利于自己的形容词，如“不安的”、“情绪容易冲动的”、“自我中心的”“神经质的”、“烦恼的”等。后来日本心理学家加藤隆胜于1964年用同样方法作类似研究，以高中学生与大学生为对象，让他们作自我评价，正常人也获得同样结果。许多学生对于一些优良品质如“诚实”、“忍耐”、“开朗”等都作了自我肯定，对于一些不良品质如“孤独”、“懒惰”等大都作了自我否定。

3. 日本的大平于“1957年调查了自小学一年级到初中

三年级学生对于自己行为动机的评价，发现低年级小学生对其动机多半作了不自觉的权威性评价，随着年龄的增长，逐渐上升为自主的独立的判断，研究还强调指出，从小学五年级开始，能够以社会标准来判断其行为的动机。

4. 美国的凯吉塔于1962年通过研究指出：自我评价高的人，容易接受他人对他的过高评价；而自我评价低的人，则容易接受他人对自己的过低评价。

5. 日本的田于1975年研究15~24岁于高中生与大学生对于自信心、自尊心、自我肯定三项个性品质的自我评价，研究结果发现，上述三项个性品质的自我评价存在着男女生性别差异。

6. 关于自我评价问题，苏联心理学界也很重视，取得不少研究成果。他们用“理想的自我”与“现实的自我”加以比较，以确定其自我评价的程度，方法是向被试提供一系列反映个性品质的词汇表，如谦虚、骄傲、粗鲁、关怀、认真、原则性等。先要求被试把这些词汇加以排队，把自己认为最有价值的词汇排在第一，其次的排在第二，第三等，这一顺序代表了该被试的“理想的自我”，然后，要求被试根据自己本身实际情况，把上述的一套词汇排列成一个系列，这一系列反映了该被试“现实的自我”，当这两个方面自我比较之后，如果差距小的就意味着自我评价高，差距越大则说明自我评价越低。

研究者进一步要求查明个人的“期望评价”，以及“现实评价”、“个人对团体评价”，所谓期望评价是个体想象自己被团

体中他人所作出的公正评价。所谓现实评价是个人在团体中他人实际上给予的评价，研究表明：“期望评价”与“自我评价”是不一致的，可能是接近自我评价，也可能是不太接近自我评价，一般讲，自我评价高的人他的期望评价是低的，个人给团体评价高，则团体给他的现实评价也会相应提高，自我评价与现实评价不符合的人，他的期望评价也是低的。

他们的研究还指出，自我评价的特点和其自身活动的特点（如独立性、目的性等）之间存在着一定的联系。

7. 自我评价的研究，目前我国开展较少。华东师大心理学系师生于1983年调查了小学、中学、大学等五个年龄组（8、11、14、17、20岁，相当于小学二年级、五年级、初中三年级、高中二年级、大学二年级）三百多名学生自我评价特点，准备两套道德品质评价表，一套是正面的积极品质，共八项，如“诚实”、“做事认真”、“帮助人”、“礼貌”、“大公无私”、“勤劳”、“清洁卫生”、“遵守纪律”；另一套包括正面的积极品质与反面的消极品质各四项。前者如“严格要求自己”、“团结友爱”、“学习努力”、“文明礼貌”，后者如“损人利己”、“欺骗行为”、“办事拖拉”、“不讲卫生”。两套品质把被试一分为二，分别让学生既评价自己又评价他人，评价采用等级法，整理材料时，将等级转换为分数，然后把自我评价的分数与小组对他的评价的平均分数加以比较。

研究结果表明以下几点：

(1) 从总体看，优良道德品质的自我评价有高于小组对他的评价的倾向，全体被试有64%的人次是自我评价比小组对他的

评价高，这一结果与霍林沃斯的研究结果是一致的。

(2) 自我评价高于小组评价的倾向表现出年龄差异，尽管各年龄组都有一半以上被试的自我评价高于小组评价，但其中小学二年级的学生尤为严重，83%人次有这种倾向，从小学五年级开始，这一倾向不很明显。

(3) 自我评价高于小组评价的倾向还表现了男女生性别差异，有意义的是这种差异随着年龄的推移，又发生的逆转，即小学五年级女生这一倾向显著大于男生，而到了高中与大学阶段则相反，男生显著高于女生，这一逆转可能反映着男女学生社会化的过程（参看下图）

(4) 对于不良品质的自我否定强于对优良品质的自我肯定。这不管是小学生，中学生或大学生都普遍产生同样的倾向。而且这种倾向还存在普年龄之间的显著性差异，小学二年级学生的这一倾向比起其它几个年龄组学生来说，在强度上要差一点，这说明他们的自我防御功能比不上其他各组。

(5) 学生道德品质主客观评价的吻合存在着明显的个别差异，前面说过，学生优良品质的自我评价一般都高于小组对他的评价，有一部分学生的自我评价则低于客观评价，但这主客观之间的差距有大有小，很不平衡，显示了个别差异。根据材料的分析，约有60%以上学生的自我评价与客观对他的评价比较接近，其中少数学生基本上是吻合的，尚有三分之一的学生主观评价与客观评价的距离较为悬殊，个别学生的主观评价距离甚大。进一步调查表明，主客观评价距离接近的学生，在班级中多数是小干部，

享有一定的威信，人际关系也比较和谐，一般都是公认的好学生，而主客观评价差距过大的极少数学生，一般都是属于班级中的后进生，表现较差，同伴对他有一些共同的想法，而这些学生本人对于自己道德行为方面的一些缺点或弱点却看不到，或者虽然意识到却要故意装饰自己，甚至全面肯定自己。有一被试自我评价八项道德品质是属于“很好”，而小组评价他是“差”。

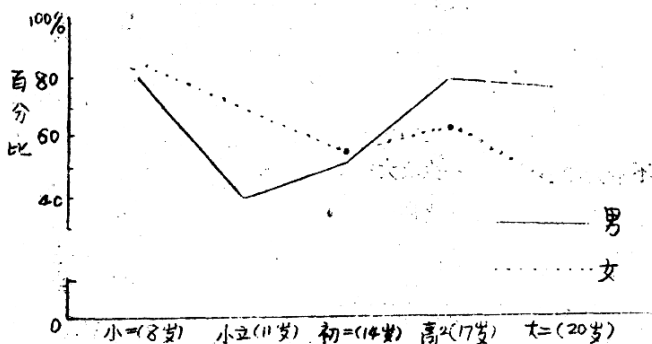


图 自我评价高于小组评价倾向的男女发展趋势

研究指出，学校教师要重视对这一类学生的教育与引导，经常对他们作出正确评价，以培养他们正确评价自己的能力。在教育中要贯彻一致性原则，因为成长中的学生，尤其是青少年，他们对于周围人们给他的评价是非常敏感的，他们还没有形成关于自己稳固的形象，他们非常强烈地感受与体验着周围人们对自己的态度，因为如果各方面给他们的评价不一致，就会使他们感到无所适从，不能正确认识自己的品质，不利于自我教育和自我提

高。

时蓉华：《社会心理学》 武汉师范学院

1984年

三、到达目标与到达度评价

△ 问题的提出

(一) 背景 五级相对评价

众所周知，日本中小学长期以来考试频繁，竞争激烈，升学考试是各级学校实际上的最终目标，学校的一切工作都围着它转。层层筛选的“考试地狱”使得本来应该是黄金时代的欢乐愉快的学校生活变得紧张不堪，毫无情趣。这不仅在中小学里造成了种种问题，甚至还影响到幼儿教育。

造成这种现象的因素很多，既有社会的也有教育内部的。在教育内部的诸因素中，日本战后在中小学里实行的“五级相对评价”是影响重大的一项。

日本战后在美国的直接参与下进行了彻底的教育改革。

1948年新制定的学校教育法施行规则第十五条规定，学校必须给每个学生建立一本“学籍簿”（后改名为“指导要录”），记录学生在校期间的成绩、操行、健康等情况，既为教学记录，又作对外证明之用。学习成绩的评定方法参照美国的办法，在1948年规定“一般程度者为0，较优者为+2，较劣者为-2。一般情况下，0当占多数，+2及-2当为少数。”这种“-2、-1、0、+1、+2”（后为方便起见，1955年

起改记为1~5)式的评分方法,要求按照正态分布规律把全班或全校学生的成绩分成两头小中间大的五级,故称为五级相对评价。

虽然日本文部省从未正式规定过五级之间的比例,但在实践中人们一般都根据正态分布曲线给出的近似值,按1:2:4:2:1的比例打分。这样的成绩,不仅登记在学籍簿(指导要录)上面,也登记在成绩报告单上,以及升学时送往高一级学校的“内申书”上。

然而,实践证明五级相对评价不利于学校教育。一方面,硬性制造正态分布往往不符合班级实际情况;另一方面,它给中小学带来了激烈的竞争和筛选。因为五级相对评价的成绩并不表示该生掌握知识技能的实际水平,只表示他在班级里或学校里属于第几名。由此,师生们在学校里首先关心的不是知识技能掌握得如何,而是考试的名次如何。考试规模越大,名次越重要,统考(在日本称作统一学力调查)分数经标准化之后,就成为指导学生去向的依据。一次次考试就是一次次竞争、一次次预选(如果入学考试算最后的筛选的话),无形之间,教育的目的变得不是为了受教育者的发展,而是为了使他们通过考试了。人们把这种教育称为“考试主义教育”。

考试主义教育中的五级相对评价的又一弊病就是抹煞师生的积极性。因为由于种种原因学生的成绩可能长期固定在某一级上,这样他们就看不到自己的发展和进步,成绩差的就必然会心灰意冷,不求上进。同时,教师也无意于努力提高全体学生的水平,

因为相对评价总是在否定这种努力的结果,即无论师生如何共同努力,用成绩来反映成果的话,终归是两头小中间大。而放任自流的结果,必然是学生之间的差距越拉越大。

应该指出,随着问题的暴露,日本文部省也曾对成绩评定方法作过一些调整。例如,1955年修订的关于指导要录的规定中,要求不划定五级之间的比例,允许一定程度上根据班级实际情况打分。但是,对正态分布的形状仍有规定,要求“一般来说,3分者占大多数,5分及1分者为极少数”。1961年修订的条文进一步删去了这种形状上的要求,增加了对照目标的规定,成为“参照教学大纲所规定的该学科的学科目标及学年目标,给班级或年级中的一般程度者打3,特别优于3者打5,特别劣于3者打1,介于3.5之间、3、1之间者,各打4、2”。

但是,这样的修订只是放宽了根据正态分布曲线机械地分级打分的要求,而对相对评价的原则却仍然坚持不放。既然如此,五级相对评价仍然在促进着中小学里的竞争和筛选,由它造成的问题也随着竞争的激化而日趋严重。

(二) 到达目标概念的提出

问题怎么才能解决呢?1966年日本国立教育研究所的研究人员板仓圣宣在日本教育心理学会的杂志上发表了题为《正确的学力评价》的论文,试图从探讨教育目标与教育评价的关系入手,解决中小学的教育评价究竟应该是什么的问题。

我们知道,目标与评价是互为表里的关系。制订目标的目的之一,就是为了日后对照着进行评价;有什么样的目标,就有什