

国际展望:

数学教育评价研究

GUOJI ZHANWANG

SHUXUE JIAOYU

PINGJIA YANJIU

ICMI 研究丛书之二

唐瑞芬 李士琦 等编译

上海教育出版社

国际展望：

数学教育评价研究

GUOJI ZHANWANG

SHUXUE JIAOYU

PINGJIA YANJIU

ICMI 研究丛书之二

唐瑞芬 李士琦 等编译

上海教育出版社

国际展望:数学教育评价研究

ICMI 研究丛书之二

唐瑞芬 李士琦 等编译

上海教育出版社出版发行

(上海永福路 123 号)

(邮政编码:200031)

各地 新华书店经销 上海市印刷十二厂印刷

开本 787×1092 1/16 印张 18 插页 4 字数 377,000

1996 年 12 月第 1 版 1996 年 12 月第 1 次印刷

印数 1—2150 本

ISBN 7-5320-4804-7/G · 4774 定价:(软精)20.50 元

序 言

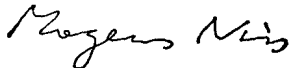
非常乐意为 ICFI 研究丛中译本第二卷撰写序言, 本书选编了 1990—1993 年期间, 国际数学教育委员会用英文出版的四本最新的研究专著的内容, 这四本专著是:《数学与认知》,《数学大众化》,《数学教育评价研究》以及《数学教育评价的现状》。

中国数学教育工作者的干劲、热情与效率给我留下了深刻的印象, 他们积极了解世界各地数学教育理论与实践的最新发展信息, 并努力传播给遍布全国的同行们。特别要感谢与赞扬上海华东师范大学以唐瑞芬教授、张奠宙教授为首的一支编译队伍, 虽然我不懂中文, 但我可以断定, 他们正以巨大的努力, 力图使世界潮流的动向与趋势, 通过多种渠道渗入中国的数学教育界。

然而, 在这样做的过程中, 我们的上海朋友与同行们也在数学教育的国际交往中形成了一种值得注意的不对称现象: 作为中国的数学教育工作者, 你们确是更好地认识了世界各地数学教育的现状与进展, 可是, 生活在那些地区的我们, 却没能更好地了解中国数学教育的优良传统与当代实践。事实上, 我们经常从中国的不同地区获得令人激动的印象; 因此, 我能肯定, 中国数学教育的传统、经验与思维方式, 也值得世界其他地区的数学教育工作者更好地学习。所以, 请允许我在结束这个序言的时候, 提出一个愿望, 希望当代中国的数学教育专家学者们, 能通力合作, 创作一本综合专著给全世界的读者, 介绍中国数学教育的过去、现在和未来, 阐述她的本质特征。这样就可以改变我前面提到的不对称现象。虽然还没有得到正式同意, 但我相信国际数学教育委员会将会欢迎这一创举, 并尽可能促使其实现。

最后, 衷心期望本书内容(至少某一些), 能激励我那些相识与不相识的中国同行们, 使他们感兴趣, 甚至对他们有用。

Mogens Niss



数学与数学教育教授

国际数学教育委员会秘书长

1995.3.28.

目 录

第一篇 数学教育评价研究.....	(1)
1. 数学教育评价及其影响: 导言.....	(3)
2. 数学评价的历史: 链与箭.....	(23)
3. 评价、课程与社会之间的关系.....	(36)
4. 数学评价的国际展望——一个模糊分类.....	(44)
5. 从认识论角度看评价存在的问题: 什么是数学知识?.....	(57)
6. 怎样才能学会知识: 数学学习的模式和理论.....	(64)
7. 评价意味着什么? 数学知识评价的情况.....	(76)
8. 数学教育评价的一个重要观点——如何体现学生 是主体?	(97)
9. 改进评价工作的要求.....	(108)
10. 学生数学知识评价理论的设想.....	(117)
第二篇 数学教育评价的现状.....	(125)
1. 美国数学评价中的争议问题.....	(127)
2. 数学教育改革中的评价: QUASAR项目中的评价设计.....	(139)
3. 英国数学教育评价: 理论与实践的发展.....	(147)
4. 学校数学设计: 英国首创的中学评价方法.....	(158)
5. 教/学过程与评价工作: 数学教育中两个交织在一 起的侧面.....	(168)
6. 问题的类型以及挪威学生是如何解决的.....	(174)
7. 丹麦高中数学的评价.....	(183)
第三篇 数学与认知: 数学教育心理学国际研究综述.....	(189)
1. 引言.....	(191)
2. 数学教育的认识论和心理学.....	(200)
3. 语言与数学.....	(213)
4. 高等数学思维.....	(225)
5. 数学教育心理学研究展望.....	(240)
第四篇 数学大众化.....	(251)
1. 数学大众化研究概述.....	(253)
2. 数学与游戏.....	(272)
编译后记.....	(283)

第 一 篇

数学教育评价研究

1. 数学教育评价及其影响: 导言

Mogens Niss

丹麦 罗斯基得大学

1. 本书乃 ICMI 专题研究成果

ICMI 专题研究会议“数学教育评价及其影响”的成果有二:一是本书“数学教育评价研究”,另一是“数学教育评价的现状”。均属于由 ICMI 的主席与秘书长任主编的 ICMI 研究系列丛书,但侧重面不同,所以可以各自独立成章。本书致力于分析评价的主要方面,另一书则讨论评价的实际执行情况,两者相辅相成。

为使读者了解本书的背景,下面简单介绍一下 ICMI 专题研究的性质与范围。

自 1986 年起,国际数学教育委员会(简称 ICMI)开始出版数学教育研究系列丛书,已经出版的有(全部由剑桥大学出版社出版):“九十年代的中小学数学”(Howson & Wilson, 1986),“计算机与信息科学对数学与数学教育的影响”(Church house 等, 1986),“作为服务性学科的数学”(Howson 等, 1988),“数学大众化”(Howson 等, 1990),“数学与认知:数学教育心理学国际研究综述”(Nesher & Kilpatrick, 1990)。

根据主题的不同,有的侧重研究,有的侧重行动(或者两者兼有)。目的都想对该主题的现代发展水平作一介绍与分析,或是描述当前的研究与发现,或是讨论关键性的、具有实质意义的争议问题,以及数学教育工作者持有的不同看法。

ICMI 专题研究通常按以下步骤进行(认知研究例外):ICMI 执行委员会任命一个相当小的国际“程序委员会”,它的第一个任务是草拟一个“讨论文件”,提出研究的主题、目的与范围,以及需要论述的题目与争端,该文件发表在国际杂志(包括 ICMI 的正式刊物《数学教育》)与通讯上,数学教育工作者会作出反应,并申请参加专题研究会议。

专题研究会议参加人数有限,约 50—100 人,其观念、阅历与特长均有差异,与会者以“讨论文件”为指导,提出论文、报告,进行辩论与小组活动,形成有关主题的研究论坛。

最后,由 ICMI 的主席与秘书长为主编,根据专题研究会议所做工作与书面材料,出版相应的专题研究文集;因为它是以国际学术界为广泛基础编写的独立出版物,因而它就不同于普通的会议论文集。1989 年 5 月,ICMI 执行委员会任命以下国际程序委员会:

Claudi Alsina, 当地组织者, 加泰罗尼亚工业大学, 西班牙, 巴塞罗那;
Desmond Broomes, 西印度群岛大学, 巴巴多斯, 布里奇顿(译注: 拉丁美洲);
Hugh Burkhardt, 诺丁汉大学数学教育中心, 英国;
Mogens Niss, 程序委员会主席, 罗斯基得大学, 丹麦;
Thomas A. Romberg, 威斯康星大学麦迪逊分校, 数学教育研究中心, 美国;
David Robitaille, 不列颠哥伦比亚大学, 加拿大, 温哥华;
Júlianna Szendrei, 国立教育学院, 匈牙利, 布达佩斯。

讨论文件公开发表于《数学教育》以及若干杂志与通讯上。

专题研究会议于 1991 年 4 月 11—16 日在西班牙卡隆赫的 Cap Roig 召开, 有来自欧洲, 北美洲, 拉丁美洲, 亚洲, 大洋洲, 非洲与中东的 25 个国家的 80 名代表参加。

关于术语的一点注记

对一些专门名词想作点说明。我们将要论述的领域经常使用诸如评价、评估、测验、考试一类词汇。可是, 在不同教育制度背景下, 在不同的语言中, 这些词会有很不相同的内涵; 甚至同一个词对不同民族而言就具有截然不同的意义。这里, 我们想对评价(assessment)与评估(evaluation)两者作一区分, 其他问题以后再说。

我们经常交替使用评价与评估两词, 按讨论文件提出的意见作以下约定: 数学教育的评价是指对学生个人或是小组的数学能力、表现与成绩(三者均以广义理解)作出评定, 而学生则包括从幼儿园一直到博士生; 因此评价涉及学生层次的数学教学。数学教育的评估则是对全部或部分教育或教学体系的评定, 当然也包括数学教学。评估可以涉及体系, 诸如课程、计划、教师、师资培训以及学校或学区教育体系的特定部门等; 所以, 评估涉及体系层次的数学教育。

测验、考试作为评定学生成绩的方法, 是特殊的评价形式, 应该归入评价的范畴。与此相对的是, 测验、考试作为一个教育(子)体系运行方式的一部分, 或是测验、考试结果作为这一体系质量的标志, 就像在国际比较研究的情况下, 考试、测验就属于评估的范围。这种二重性反映了评价与评估关系的特点: 评价项目——特别是评价结果, 也包括评价方式——可以包括在体系的评定中, 所以它们组成评估活动的一部分。反过来, 却不成立; 例如对教师的评估经常包含许多方面, 其中不一定涉及对学生的评价。所以, 评价与评估之间的关系是不对称的。

本书的研究偏重于上述意义下的评价而不是评估。但由于刚才提到的二重性, 评估问题不会一点也不考虑; 当然只注意与学生评价有关的方面。

为什么研究评价?

近年来, 国际数学教育界越来越关心评价。理由很多, 主要的一条似乎是: 过去几十

年内,数学教育在观念与目标,理论与实践各方面都有了长足进展,唯有评价的发展相形见绌。

数学课程领域提出了新要求。首先,在内容方面,应用与建模,与其他学科共同感兴趣的问题的协作,数学哲学与数学史,着重于问题的创造力,计算机与信息科学辅助下的探索与实验,已经包括在不少国家的教学计划与课程之中。其次,我们已经看到了学生活动与活动形式的范围有显著扩展。对纯数学与应用数学的进一步研究,课题规划,科学探究与思考,课外活动,实验,小组活动等等,已不再是数学教学中的乌托邦。所以,对数学与数学教育已经形成了更广的概念。

可是,在评价的意义、概念、理论、实践、方式、步骤等方面都没有相应地发展。因而,在数学教育现状与现行的评价实践之间的不相称与紧张情况日益具体化。数学教育的观念与目标从未真正与数学教育工作者使用的评价方式相一致,只是以前的中等数学教育只提供给少数青少年,也许不相称的问题不太严重,或者至少在数学教育工作者看来不太严重而已。无论如何,数学与数学教育概念的扩展肯定会在当代的数学教学与传统的评价实践中拉开差距。

这一差距将评价提到了数学教育工作者的日程上,应该说这是关心真理的一个崭新现象。数学教学在过去的三、四十年内强调课程改革——各种不同类型的改革,有时甚至真的是相互矛盾的类型,——是最重要的任务。同时,数学教育又是一个学术领域,注意力也集中在数学学习的条件与过程,特别是关于数学概念的形成与获得。可就是忽视了评价,将评价看作数学教育中一个不重要的因素,并且,从许多方面来看,对数学教育都是一个“外来的”因素。评价之所以引起数学教育工作者的注意,一定程度上是由于对它的作用与功能的担心。传统的评价方式,特别是考试与测验“来自外部”,许多情况下,都成为阻碍课程改革或改革速度缓慢的因素之一。

目前,课程改革已经在许多地方实行,情况有所改变。数学教育评价的作用、功能与影响不应再被忽视;相反,出于多方面的理由,应使之成为研究、考察的对象(参见国际数学教育委员会,1990):

- 对当代评价方式的作用,功能与影响不明确,也缺乏理解。
- 现行的评价方式与实践注意点不同,目标分歧,且无意中带来了令人失望的副作用。特别是很难设计这样的评价方式,使它同时满足:(a)允许我们以有效且可靠的方法,对数学知识、见解、能力与技巧的理解与掌握,从根本上进行评定;(b)对个别学生在获得数学见识与能力的过程中,进行调整与改进方面提供真正的帮助;(c)帮助个别教师调整并改进其教学、指导、管理与评议;(d)帮助课程设计者与当局、教科书作者与在职教师培训人员,形成适当的数学教学结构。
- 在设计与使用有效且协调的评价方式时,如果没有严重的内部、外部问题的影响,所遇到的困难性质似乎是生来固有的,而且也是普遍的,所以值得从国际研讨的角度

加以论述。

2. 评价的目的

数学教育评价的目的大多数不限于数学,也适用于其他学科。所以将要谈到的数学成绩评价的许多方面对多数学科同样成立。

只要有学数学的学生,就要用这种或那种形式来评定,以确定他们的成绩与受益情况。任何人(或是专业教师,或是亲戚朋友)想教某人一点数学,也总要评定一下教学的结果——也许是无意识的,也许是心照不宣的,也许是明显的,也许是专门出题。

只要学校(中小学、大学等)教数学,主管学校或是承担学校经费的人就要求对教学结果作某种形式的汇报——正式的或非正式的,口头的或书面的——当然个别学生也经常希望知道。这种汇报活动要求有一个基础,根据它对教学成果进行评定与交流。这就促使人们去创造确定的、标准化的评价方式。

只要数学能力与获得职位、提升、特权、执照、升学或是掌握某种任务与职责相联系,学生数学成绩的汇报也就成了关心学生数学教育的校外人士的需要。

只要存在不同的经济、政治、社会与文化兴趣,不同的社会价值与思想意识,对于获得职位和机会,对于在此背景下评价的作用与功能,就会存在论战与斗争。

由此推出:数学教育评价涉及形形色色的目的(这些目的不一定相容),因而也就必然引起各种论战。

评定学生某门学科特别是数学成绩的目的,似乎可以分成三个不同而又相互依赖的主要范畴:提供信息,建立决策或行动的基础,最后是形成社会现实。

提供信息

数学教育评价的基本目的是给一些选民提供有关数学教育成果的信息。其最终目标——可能非常遥远——是通过信息的供求,帮助选民作出某些决策,并采取相继的行动,即便决策也许是什么也不做。

所寻求与传递的信息是一个多变量函数,变量有:信息内容;收集、记录、处理与传递信息的方法;收集信息的人或系统;想接受信息的人及其需要。本节论述将信息提供给某人的目的,强调信息作为接受者的函数。

学生数学表现(成绩与能力)的评价可以为不同类型的接受者提供信息。

给学生个人的信息

- 在课程进行中,或在课程结束时,按照课程的内在要求,提供学生有关的表现的绝对信息。包括学生在内容掌握中的强项与弱点,活动的质与量,工作习惯等等。目的可能是对学生表现给出一个静态的/积累的画面(他目前掌握了什么?),或是给出一个

动态的/区别的画面。(学生现在掌握的与以前相比有何变化?)

- 提供学生有关的表现的相对信息,即相对于类似情况下的其他学生,比如同组,同班,同校,同一学区,同一国家,同一区域,或是全世界。所寻求的画面也可能是静态的或是动态的。

将这类信息以绝对的或是相对的措词提供给学生的最终目标是帮助他们进一步改善处境或取得发展,使他们能掌握必要的手段,以调整并控制自己的工作、学习及其结果。带有这一目的的评价常称为形成性评价。

给教师的信息

- 以绝对的或相对的措词,提供有关学生个人的情况与发展。可能出于不同的目的:教师能给学生提出建议;可帮助教师评价教学及其对特殊学生的效果,以便调整、发展或是作根本的改变,使之更符合学生的需要;作出影响学生行为的决策和行动;提供依据,使教师可向家长,学校,当局高一级的教育机构,以及个别学生的雇主进行汇报。
- 提供有关学生集体(也许是全体学生,也许只是某一范围,某一班级,或某一年龄层次的学生)的教学效果,以便判断是否应该改变或是朝某一方向发展。

给体系的信息

(这里的体系包括家长,学校,毕业生的接受单位(高一级教育机构、雇主),行政管理当局)

- 关于个别学生的表现,以便提供基础,作出选择与安排的决策或采取行动,例如提升,升学,开业执照,对该学生的雇用(如果这类信息包含对学生表现的简要综合的概括,则经常称这种评价为总结性评价);
- 关于学生成绩的分布,作为对教师、机构、学区、计划、课程作评估的输入资料(如前所述,这并非本书重点)。

采取决策与行动

任何社会都要采取大量决策与行动,对个人进行筛选,使个人获得各种机会、职位、工作与特权。这里,“社会”这个词是广义的,包括个体(作为私人,作为公民,或是公私部门的决策者),也包括大范围、有结构、有组织,且有文化传统的政治-经济体系。因而上面所说的体系只是社会的一个真子集。决定选择的基础与方法并非简单的中性事物,而是深深地植根于作为社会特征的一系列物质条件、观念、价值与传统之中。

有时作出的选择决定完全是非正式的,基于个人的接触或了解,一时的印象,纯粹的巧合,或是在统治集团中有影响的权力与地位。有时更正式的就建立在评价过程与评价结果的基础上。大多数社会广泛使用正式与非正式两种选择方式,但是不在同一背景

下。为什么社会要用正式的以评价为基础的选择方式?要圆满回答这一问题需要分成几个组成部分。

有些职业确实需要特定的能力。一些明显的例子如数学教学与数学研究,以及在特殊背景下的数学应用与数学建模。在这种情况下,即使不是实际必需,至少在原则上需要引用最小能力的概念。对此概念要得出一致的定義恐怕十分困难,也很难对它作具体描述,因为它会随着时间与空间而变化。然而,有一点也许都会同意,即雇用毫无数学知识的人去担任刚才提到的三类工作是不恰当的。所以,对某些职业设想存在一个最小能力还是合理的。

这就导致执照(或证书)的概念:任何社会对某些职业,例如医生、数学教师或保险统计员的正式承认,都要求有最小能力的证明。由于现代社会劳动分工的复杂性,这种能力由教育机构提供,而不是由雇用这类专业人员的机构或公司提供。所以这就需要配备某种形式的“通过”标记或是一个开业执照,而决定候选人是否应当通过就不可避免地要应用某种形式的评价过程。

这种(通过,不通过)的二元决策不仅用以决定谁可以取得开业执照。在大多数社会中,这也是决定谁能进入下一轮机会的关键因素,例如学校中的升级,选拔到某个小组或机构,在等级制度中的提升等等。

除了决定应该允许谁参加选择以外,各社会部门还经常要在达到最小能力要求的候选人中进行挑选,这就引出了候选人的排序概念,或者说是分等概念。通常的序关系(可能是线性的,这时经常称之为排名次,但也不一定是线性的)建立在个人表现的基础上,或至少以其表现为基本成分。这又要求设计某种形式的评价过程。候选人排序的主要目的是提供一个灵活机动的选择方法,特别在各种机会(例如工作与薪水,学校中的职位,拨款与奖金,免费的设备,以及俱乐部、社团或委员会的成员等)不足的背景之下。

正常情况下,对判断通过/不通过的评价过程,社会必需有相当大的信心。特别是,社会必须相信具有“通过”资格的人确实满足最小能力的条件,否则就没有意义。达不到标准的人被授予了执照,而实际上达标的人却被否定,在这两者之间,社会更担心的是前者。任何社会对正确衡量通过不通过的规定极感兴趣,换句话说,需要设计达到这一目的的有效且可靠的评价过程。

当涉及按表现排序时,情况似乎有点不同。当然,如果存在表现的有效排序,那将是很值的重要的社会手段,可用来在一组有能力的候选人中挑选最好的。然而,即使对流行使用的排序计划存有或多或少的怀疑,特别是对其有效性有怀疑,社会还是经常使用它们而不太顾虑其欠缺。这可能有四点理由:

(1) 如果我们没有完全满意的评价方式,我们必须满足于尽可能找一个比较好的,这总比没有好。这里“没有”是指不明显的、非正式的评价方式,而并非不存在评价,因为对表现的判断总是在进行。

(2) 不考虑各种保留意见,我们还是偏爱通用的明显而正式的评价方式,因为相对于

不明显、非正式的评价方式而言,它们至少有一个优点,那就是它们是看得见的,因而也就易于分析,辩论,或者改变。这样,就有助于消除评价可能产生的任意性与混乱,受某些不成文规则的支配,甚至重用亲戚。

(3) 如果评价规则模糊,那些规则的解释者就会有特权。这就会造成对社会部分成员——这部分人往往由其经济地位,能力,级别以及家族关系决定——的偏爱。于是就会造成一个“机会不平等的社会”。所以,主张平等的社会经常使用正式的评价方式来挑选各种类型的职位与工作。

(4) 即使那些根据正式的评价过程采取决策或行动的人对所用步骤的有效性或全面质量没有信心,社会也仍然要使用这些步骤,使决策或行动正规化以经得住批评。至于是否真正应用评价方式得出该决策或行动,或只是用以掩盖在完全不同的基础上作出的决策,那并不重要。重要的是使用客观的评价过程,就使有关的决策与行动合法化。

在提升或挑选某个学生参加高层次的数学活动或进入数学专业时,学生的数学表现作为评价的决定性目标,这不足为奇。可是,值得注意的是,许多国家在提拔、筛选学生研究与数学毫不相关的内容时,也唯独以数学表现的评价作为一个基本条件。正如 Jean-Pierre Kahane 在一次谈话中曾提到,“也许,问题是数学中的评价还没有评价中的数学那么多”。

为了非数学的目的使用数学评价,似乎依赖于两个相容的假定。我所以用“假定”这个词,因为这里不讨论它在理论或经验上是否合理的问题,也不问是否将其看作是观点、信念、事实或是科学结论。

第一个假定来源于传统的能力心理学,认为数学能力与一般认知能力以及任何智力活动的技能密切相关,且十分重要。数学能力与一般的智能究竟是因是果并不重要,重要的是前者可作为后者的一种标志。第二个假定更具有教育性。它强调某人如果在困难的数学学习中取得成功,就证明他具有一定水平的才能,信心,自我控制力以及不屈不挠的精神,从而增加其他学习(不论是否数学)成功的可能性。至于成功是否能归结为一种特殊的数学天才(如果这存在)或是善于从事艰难的工作,这也并不重要。重要的是在困难的训练中取得成功。

虽然本书范围限于评价而不是评估,但应该指出,大多数社会,评价的重要目的之一,总是调节、控制教育体系,以及学校、教师及其教学,特别是各种课程。于是,许多地方将学生评价结果作为涉及教师生涯的提升、雇用以及解雇、调整薪水的决策因素。进一步还可提供背景情况以修改或改革课程,或是保持不变。

社会现实的形成

当数学教育评价根据以上目的进行时,自然会对学生、教师、家长、学校的社会现实产生强烈的影响。评价的实际方式与步骤会影响许多人的生活条件。但是,许多后果都是

附带效应而并非评价的直接目的。

然而,有理由相信,社会也想使评价具有一些职能,如通过对学生、教师与各种机构的控制,形成更宽广的社会现实;创造对艰巨工作作出承诺的“大棒与胡萝卜”气氛;竞争与奋斗;接受或至少采纳分而治之的现实;服从于权力、权威以及思想意识。这类意图强调到什么程度,以及所追求的敏感程度,均随社会而变,但是任何社会在某种程度上,都掌握着这点且必须掌握它,以便使公民们在为教育、工作、职位以及物质财富,一句话,为成功而进行的竞争中,能承受困难、紧张甚至遭受挫折。

总 结

以上考虑的问题可以概括成下面的目的矩阵(图 1),它展示了评价目的(各行)与那些目的所针对的对象组(各列)之间的关系。矩阵的各项表示目的与相应的对象组如何相关。括号中的事项更适合于评估时所必须注意而不是评价所关心的。箭头应读作“供给”或“为之提供基础”。

接受者/ 对 象/ 代理人 目的 关于	学生	教师	体系
分别给学生、教师、体系提供的信息,有关:	他/她自己的成绩	他/她的学生的一成绩 * 他/她自己的教学	→ 个别学生的成绩 * (全体学生,教师,学校,课程)
分别由学生、教师、体系作出的决策与行动,涉及:	控制他/她自己的学习活动	↓ 控制他/她的教学 * 对个别学生的活动与安排进行控制并提出建议	↓ (对教师、学校与课程的监督与控制) * 筛选,选拔并安置学生
分别形成学生、教师、体系的社会现实,关于:	↓ * 精力与时间投入的优先次序 * 工作习惯与态度 * 对竞争与社会规程的态度	↓ 教学内容,教材与教法的选择 * 课堂环境 * 精力与时间投入的优先次序 * 工作习惯与受雇者的态度 * 对竞争与社会规程的态度	↓ 体系的维持与自我保护 * 体系的有控制发展

图 1 目的矩阵

3. 数学教育的评价方式

我们将使用评价方式来表示所实施的或者可以被实施的评价模型中所有的组成部分。每一给定的教育体系中的任何一种评价方式,或多或少反映了(虽然不一定完全同构)该体系内部与评价相联系的目的,正如前节所述以及目的矩阵中所示。

换句话说,评价方式可以看作是一个“向量”,它有一组分量,其中包括:

- 评价的对象。即评价谁,对哪类学生进行评价? 个别学生,学生小组,班级,同年龄组,还是全校学生等等。
- 评价的目标。即评价什么,哪类数学内容(包括方法,内部与外部关系)和相应的哪类学生能力组成评价的目标?
- 评价的项目。即评价哪类输出信息,如何根据目标提出评价测试的要求? 评价对象应该展现出哪些特定项目以供评价?
- 评价的时机。即什么时候进行评价,什么情况下收集评价的信息,是在一门课、一个学期、一个方案或全部研究之前、中间、或是之后?
- 评价的过程与情况。在评价的时候,发生了什么以及要求谁去做什么?
- 评价的判定与记载。为了判断评价的项目,突出什么,准则与步骤是什么,以及记录什么作为评价过程的结果?
- 评价结果的报告。评价过程及其结果向谁报告,报告些什么,报告应采取何种形式及设计?

以下各节将分别阐明每一个分量。

评价的对象: 评价谁?

根据我们对评价的定义,回答很简单:学生个人是数学教育的最终评价对象。这不只是一种形式定义,也是各种评价目的的必然结果。所以,大多数情况下,学生个人不仅是最终评价对象,也是进行评价的直接对象。

可是,近年来,全世界很多地方在数学教育中引进了小组活动。这就意味着现在的直接评价对象有时会是一组学生。但这并非表示小组成了最终评价对象。活动小组经常只是简单地当作一个小范围而又更自在的环境——“一种包装”,学生个人的评价可在小组活动的基础上进行。可以间接地从小组活动的整体评价中得出学生个人的评价,或是直接地从其对小组活动的贡献来评价学生个人。以活动小组作为一个集合,不分成个人层次而成为最终评价对象,这种例子很少遇到。

数学教育以学生、班级、学校、同龄组等较大集合作为评价对象,经常是因为涉及这些集合之间某种形式的竞赛,或是事实上的兴趣已从学生评价转向学校、课程、教师或体系的评估。

评价的目标：评价什么？

这在很大程度上随地区、教学水平与课程而变化。从国际研讨的角度来看，评价目标中占主导地位的内容如下：

- a) 数学事实，即定义、定理、公式，某些特殊的证明，历史、传记的资料（但遇到的机会非常罕见）。
- b) 获得数学结果的标准方法与技能（例如，质或量的结论，问题的解，结果的展示）。
- c) 标准的应用，即典型的非数学情境，通过明确的数学手段完满处理。

评价目标正在日益增加但目前还为数有限，通常还包括：

- d) 启发式的证明方法，即非常规背景下形成数学结果的方法。
- e) 非典型的也许是复杂的、开放性的问题解决。
- f) 开放的、复杂的、不明确的、有时甚至是真实的，属于其他学科或实践领域的非数学情境，由个人以任何数学方法加以处理。

我们很少遇见以

- g) 探索与假设的形成。

作为评价目标。

关于学生能力的评价，往往集中于与以上提到的数学本质相匹配的各种类型的能力。

最经常评价的是学生的以下能力：

- 事实的理解，特别注意对事实的正确记忆；迅速、精确、条理清楚地表达这些事实的能力；以及在不同背景下适当地选择以及综合这些事实的能力；
- 获得数学结果的标准方法与技能的掌握，特别注意学生能设想、综合且灵活使用一系列的方法与技能，并且可以顺利地应付复杂的背景；
- 典型情境中对数学的标准应用的表现，特别注意学生能掌握的情境的范围与复杂性，以及对情境处理的独创性与深度。

自然，不常见的内容目标也有相应的学生能力可评价，例如学生创造启发式的证明方法与步骤的能力；解决开放的复杂问题的能力；为开放的非数学情境建立数学模型的能力；以及探索情境并形成假设的能力。

将这一类内容与能力项目选为数学教育评价目标，其哲学背景是什么？

至少，普遍认为以上提到的所有项目是和数学与数学活动真正有关的实质性表示。项目 a) — g) 以及对应的能力确实体现了数学及其应用的特征，无论以过程或是结果的方式来描述，也不必要求其详尽无遗。表中前几项，a) — c) 可以认为是强调数学的简单常规特性，而其余的就是更为复杂与更高要求的，无论如何，这说明前者是具有数学知识、见解与能力的必要而不一定充分的标志。

如果项目 a) — g) 全是重要的，那么为什么数学教育中流行的评价方式经常只限于前几条，即 a)，b) 与 c) 呢？对此问题的一个答案也许是：