

科研动态

模最大的教育考试机构,而且是欧洲最大的考试评价机构。每年有来自150多个国家的800万学生参加剑桥评价主办的考试。

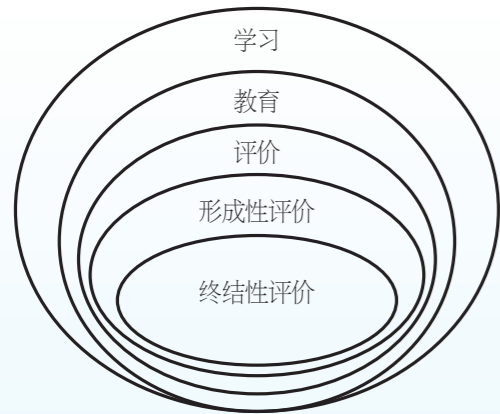
● 剑桥评价的发展理念

(1) 着眼学生未来发展——国际视野的人才观念

剑桥评价提出“剑桥人”的概念,指出剑桥人是主动的学习者,是见识多广的年轻人,是能深思熟虑、有独立思想、能解决问题的实践者。

(2) 全面、协调的学科评价——大教育背景下的评价理念

剑桥评价审视学习、教育、评价的关系,在大教育背景下形成全面、协调的评价理念,以人的发展为根本,提出了公民、信息交流技术和关键技能三项评价内容。



学习、教育和评价的关系

(3) 统一要求下的分层和选择设计——为了学习的评价

英国中等教育的评价将考试和评价区别开来,分为三种类型:质量评价、选拔评价和精英评价。采取进阶性与选择性结合,系统性与模块化结合的方式,提供专业化的评价。

(4) 恰当应用考试技术——理论、经验指导下的务实操作

● 剑桥评价机构的组成

考试部门	负责的考试项目
CIE	主要负责国际资格认证考试,考试项目有IGCSE、A-level、O-level、Pre-U、剑桥国际职业资格证书课程等。目前有150多个国家的国际学校为学生开设CIE组织的IGCSE、A-level等国际课程。
OCR	主要负责英国国内的资格证书考试(不包括剑桥大学的校内考试),承办GCSE、GCE、AS-level和GCE A-level等项目,是英国五家独立的学历与职业资格考试机构中最大的一家。
ESOL	主要负责英国本土以外的英语语言资格认证考试,在全球范围内提供一系列语言测评考试。包括雅思、剑桥商务英语、博思、剑桥少儿英语、剑桥通用英语五级以及英语教师资格证书等考试。

● 剑桥评价的高校选拔特色考试

(1) 开发目的

◇帮助高等教育机构对相同能力的学生进行区分

◇证明预测效度

◇测量学生成绩的一种附加方法

◇因为资质的多样性而采用的一种普通测量手段

◇允许世界范围内的学生有评等的机会

(2) 两种类型

◇学业测试(Specialist Tests),即评价学生在其申请的特定专业方面有无成功的潜力,如生物医学考试、第六学期数学纸笔考试、英语文学录取考试。

◇通用测试(Generic Tests),即评价学生学习任何一个专业都需要的思维技能,如思维技能考试(Thinking Skills Assessment,简称TSA)。

TSA是根据高校录取需要,不基于课程而侧重考查解决问题能力和批判性思维能力,用现代教育测量Rasch理论,一年中提供英国国内三个版本和国际两个版本的等值考试评价。剑桥大学于2001年开始使用TSA,实证结果证实TSA是大学学位成果的一个成功的预测变量。相关研究表明:根据效度研究中的相关系数解释指南,TSA的总分与学生在大学的专业学习有正相关性,与第一学年或第二学年的成绩相关性多处于“可能有用”和“非常有益”的范围。

(3) 主要特点

◇重视关键能力的考查

◇注重专业发展潜力的考查

◇重视大学入学考试的效度研究

◇重视入学考试的共性与个性的关系

● 剑桥评价对教育考试评价改革的启示

◇考试评价改革的核心目标是提高教育质量,促进学生全面发展

把改革考试评价制度作为教育改革的重要环节和突破口,从对课程学业的评价向人力资源的评价过渡,使每个人得到全面的、个性化的、充分的发展。

◇评价多元化是教育考试评价改革的主题
根据不同类型、不同层次教育形式的差异性需求,建立多渠道、多层次的教育考试评价模式。

◇确保评价的有效性

◇探索多样化、个性化、信息化的考试评价方式
研究考试形式和考试载体的改革,充分利用考试数据信息资源,向社会、学生、学校提供各具特色、内容丰富的评价报告,使考试具有更多的评价内涵。

——有关资料由办公室朱文琪提供

科研动态

编者按:《国务院关于深化考试招生制度改革的实施意见》中指出,到2020年基本建立中国特色现代教育考试招生制度,形成分类考试、综合评价、多元录取的考试招生模式,健全促进公平、科学选才、监督有力的体制机制,构建衔接沟通各级各类教育、认可多种学习成果的终身学习“立交桥”,明确了我国教育考试评价改革的顶层设计和实现路径。本期就国内外关于教育评价的研究成果进行简要介绍,以期引起大家对教育评价的关注和探讨。

教育评价的理论与实践

一、教育评价的概念

对于教育评价的概念界定,学术界存在着多种观点,教育评价的理论体系、指导思想、评价模式和方法技术等,都是在教育评价实践的进程中不断丰富和发展起来的。

美国“教育评价之父”泰勒在其著名的“八年研究”报告中,首次提出并正式使用“教育评价”这一概念,在20世纪初将教育评价解释为:“评价是一个验证课程与教学效果达到教育目标的实际程度的过程”。

20世纪60年代,克龙巴赫把教育评价定义为:“为制订教学计划和教育决策去收集和使用信息”。这一观点强调通过评价促进发展,拓宽了评价功能、范围。斯塔弗尔比姆进一步提出“评价是一种划定、获取和提供叙述性和判断性信息的过程。这些信息涉及研究对象的目标、设计、实施和影响的价值及优缺点,以便指导如何决策,满足教学效能核定的需要,并增加对研究对象的了解”,他强调“评价最重要的意图不是为了证明,而是为了改进”。

1981年由美国教育学会、心理学会、心理测量学会三大学会共同组成的教育评价标准联合委员会(Joint Committee on Standards for Educational Evaluation)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”。

我国《教育大辞典》对教育评价的定义是:“通过系统收集信息,对教育目标及实现目标的教

育活动进行分析和价值判断的过程。”我国学者陈玉琨认为“教育评价是对教育活动满足社会与个体需要的程度做出判断的活动,是对教育活动现实的或潜在的价值做出判断,以期达到教育价值增值的过程”。这一界定中的评价不仅关注当前的现实,还着眼于未来的可能性,并以教育价值增值为目的,赋予了评价发展性的功能。

基于国内外评价学者对评价概念的阐释和对资料的分析整理,可以看出,他们对教育评价的认识有些共同特点:1.承认评价是一个过程;2.价值判断是评价的本质特征;3.以一定的教育价值观为依据;4.采用一切可行的科学手段。

二、教育测量与教育评价

教育测量是根据测量学的原理,运用测量学的有关方法,对教育现象及其属性进行量化研究的过程。教育测量是教育评价的基础,是获得事实的有效途径。测量的结果,只有通过评价的解释才能揭示其实际意义。

教育测量	教育评价
◇通过非人为的操作,以严密的、正确的、数量的方式表达学习效果	◇与价值相关,关注学生在现实变化的教育目标中成长及发展过程
◇注重测量工具的客观性、可信度与统计标准	◇认为工具的适宜性是决定因素
◇主要涉及学科的达成状况、知识水平以及儿童的独立行动	◇主要涉及儿童的经验、学习的动因以及所有的行为

科研动态

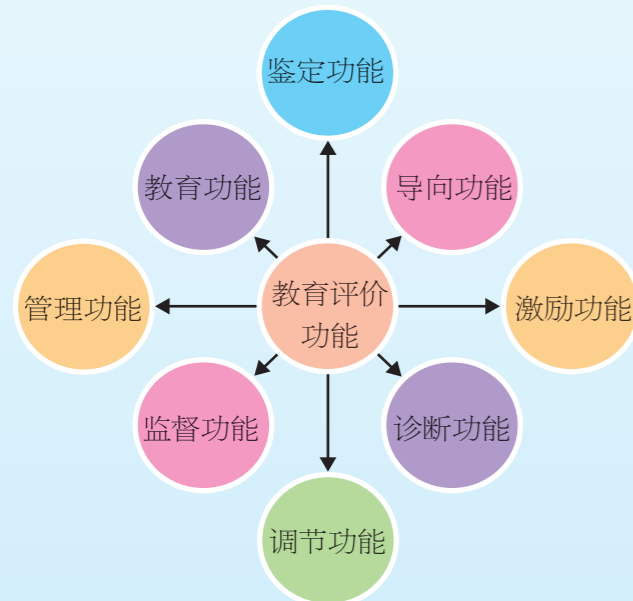
三、教育评价类型分类

教育评价项目多,且内容丰富,相互之间呈现出各种各样的关系,它们或因某一标准而成为同类,或因另一标准而分属不同类型。



四、教育评价功能

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



五、教育评价模式

教育评价模式是教育评价基本理论与方法的总体概括,它包括评价的大体范围、基本程序、主要内容和一般方法。



六、国际教育评价发展的主要趋势

- 1.由强调减少教育投入成本,转变为重视提高当前投入的教育产出。
- 2.地方性评价活动逐渐增多,不再局限于大规模的全国性评价。
- 3.评价不再局限在量化方法上,更加注意引入质性方法,强调结果解释的人性化和民主性,充分体现以个体需要为导向的心理建构思想。
- 4.重视发展性教育评价理念和被评者的参与。
- 5.重视元评价。
- 6.从重结果转向过程评价和结果评价相统一,从重鉴定功能转到重改进功能的发挥。
- 7.许多国家涌现出多种评价的学术性组织,为广大研究者提供了交流的平台,在体制上促进教育评价研究向纵深发展。

七、国外教育评价体系的模式借鉴

(一) NAEP——美国国家教育进展评价(National Assessment of Educational Progress)

●概况

NAEP,又称为国家教育报告卡(The Nation's Report Card),是美国联邦政府拨款资助,对美国中小学生学习成就进行连续和长期评价的项目,是世界上类似教育评价项目历史最长、技术最为领先的项目之一,被教育测量和评价业内称为“皇冠上的明珠”。NAEP的设计目标是向美国公众报告美国教育政策的成就和学生的基本文化素质。

●实施对象

美国4年级、8年级、12年级的学生。

●类型

国家主评价(Main NAEP Assessment),考查美

国学生在主要学科知道什么和能做什么。

长期趋势评价(Long-Term Trend Assessment),通过比较十年乃至几十年内各个学科成就水平的变化趋势,来衡量在这段时间内教育质量的变化。

州评价(NAEP-State)

城市级评价(NAEP Trial Urban District Assessment)

●评价报告

(1) 报告对象

NAEP评价并不为单个的学生和学校报告分数,评价所报告的是以性别、种族为特征的群体的结果。NAEP针对不同公众的理解水平和需求,设计了不同类型的报告卡,如下表所示:

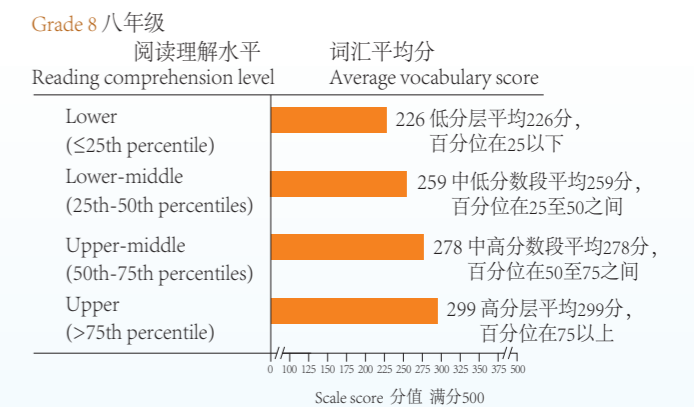
类型	目标公众	目的/内容
标准报告卡(NAEP Report Card)	政策制定者	向所有参加测验的人及不同人口群体呈现结果
焦点报告卡(Highlights Reports)	家长、学校董事会成员、普通公众	用非专业化的语言回答常见的问题
教学报告卡(Instructional Reports)	教育者、学校管理者、学科专家	包括NAEP评价中的许多教育和教学资料
州报告卡(State Reports)	政策制定者、教育部门官员、公立学校的首席官员	向每个州的所有参加测验的人及不同人口群体呈现结果
跨州数据汇编(Cross-State Data Compendia)	研究者、州测验指导者	含在州报告卡中,呈现州际结果,作为其他类型报告的参考文件
趋势报告卡(Trend Reports)	无特定对象	描述长期趋势评价学生的成就变化
焦点报告(Focused Reports)	教育者、政策制定者、心理测量专家、感兴趣的公民	开发富有教育含义的深度问题
概括的数据表格(Summary Data Tables)	无特定对象	基于从学生、教师、学校调查问卷中概括表格格式的数据
技术报告(Technical Reports)	教育研究者、心理测量专家以及其他技术公众	评价的细节,包括样本设计、试题开发、数据收集及分析

(2) 报告形式

NAEP的成绩报告形式是与表现标准紧密联系的,报告形式主要有成就水平(achievement level)和量尺分数(scale scores)。学业成就水平主要描述学生在知识和思维能力方面的表现程度,是告知学生在应该知道和应该做的事情上达到的标准。分为三个级别:达标(basic)、熟练(proficient)、优秀(advanced)。

科研动态

量尺分数是从参与评价的学生群体的总体成就水平导出,根据学生答对的题数,用特定的数学公式换算而成。NAEP典型的平均量尺分数分为0-500(阅读、数学、历史和地理)和0-300(科学、写作和公民课)。这些平均量表分数是用来表述学生在量表上所处位置的描述性指标。结合解释的目的,量尺分数针对特定的学生群体,提供与学业成就相关的信息。



●特点

(1) 连贯的测评设计。NCLB法案规定NAEP全国评价和州评价至少每两年开展一次,评价结果依据学区、州和国家的需要作为一个整体使用。历次评价结果的收集形成了一个根据时间维度的连贯的测评设计。

(2) 多维的数据收集。NAEP从建立伊始就关注并收集学生、学校、学区的背景数据,除对不同年级学生的不同科目成绩开展测试外,还配套丰富的调查问卷信息,目前已经建立了非常强大的基准数据库,对基础教育学业质量的评价有重要参考价值。

(3) 多元的评价服务。NAEP体系的多元评价服务主要体现为服务内容和对象的多元化方面。首先,NAEP在学生阅读、数学、科学、写作等学科领域提供评价结果反馈,同时还可以根据已经建立的多维数据库实现对学生长期趋势的评价和预测,这对基础教育学业质量评价有重要作用;其次,NAEP的评价结果可以为国家教育科学决策提供依据,为各州、学区和学校的改进提供参考,为学生和家长提供评价反馈。

(4) 有效的公共服务。NAEP的基准数据库向教育研究机构、研究人员和社会公众开放,借助社会研究力量实现了对数据的挖掘和深入解读,为教育改革和发展提供了有效的公共服务。

(二) 剑桥评价

剑桥评价(Cambridge Assessment)不仅是英国规