新世纪统计与经济优秀著作文库

本研究获得国家社科基金项目(编号06BTJ010)的资助

经济评价的有效性及其检验方法研究

董麓 著

中国统计出版社
China Statistics Press

新世纪统计与经济优秀著作

本研究获得国家社科基金项目(编号06BTJ010)的资助

经济评价的有效性及其检验方法研究

董麓 著

中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

经济评价的有效性及其检验方法研究 / 董麓著. —北京:中国统计出版社,2014.2

ISBN 978-7-5037-7052-4

Ⅰ.①经… Ⅱ.①董… Ⅲ.①经济评价-研究 Ⅳ.①F014.9

中国版本图书馆 CIP 数据核字(2014)第 023079 号

经济评价的有效性及其检验方法研究

作　　者/董　麓
责任编辑/徐　颖
封面设计/黄　晨
出版发行/中国统计出版社
通信地址/北京市丰台区西三环南路甲 6 号　邮政编码/100073
电　　话/邮购(010)63376909　书店(010)68783171
网　　址/http://csp.stats.gov.cn
印　　刷/河北天普润印刷厂
经　　销/新华书店
开　　本/710×1000mm　1/16
字　　数/195 千字
印　　张/11.5
版　　别/2014 年 3 月第 1 版
版　　次/2014 年 3 月第 1 次印刷
定　　价/29.00 元

总前言

自改革开放以来，我国的统计科学研究进入了一个新时期，并逐渐与国际接轨。随着我国加入 WTO，我国经济进一步融入国际经济大环境，我国的统计科学研究更注意学习和掌握国际先进的统计理论、方法和技术。为了推动我国统计科学研究发展，中国自然科学基金和中国社会科学基金每年都有专门资金资助和鼓励统计科学研究项目，国家统计局每年也有专门资金资助统计科学研究，因此，每年都有很多统计专家学者在为统计科学的发展进行着辛勤的研究工作，他们的这些科研成果也以不同的形式服务统计事业，服务中国的社会主义建设。为了总结和整理这些珍贵的科研成果，我们决定编辑出版“新世纪统计与经济优秀著作文库”丛书，把这些优秀的统计科研成果，以丛书的形式，奉献给广大读者。

“新世纪统计与经济优秀著作文库”主要编辑出版由国家科学基金和国家社会科学基金资助的优秀统计科研项目成果，我们认为，对于我国统计事业来说，这是一件好事：首先，这是我国第一次把优秀统计科研成果以丛书的形式系统地整理和出版，这不仅能使这些优秀成果不致散落在各科研单位和使用单位，而且能使它们在时间的流逝中沉淀下来，给统计后人留下宝贵财富。其次，这套丛书将成为展示我国当代统计科学研究成果的重要舞台，反映我国统计科研人员的整体水平，显示我国统计科研事业的实力和发展方向。第三，通过这套丛书，我们将可以客观地评介我国统计科研水平，并找出我们与国际统计科研水平的差距，以使我国的统计科研尽快赶上世界先

进水平。第四，这套书的出版，不仅会使我们进一步认识那些活跃在统计科研战线上的老同志，还将使我们认识统计科研事业的后起之秀，而这些统计科研新人的出现和成长，是我们的统计事业得以继续发展壮大，永葆生命活力的保证。

希望这套丛书的出版能对统计科学研究事业和统计事业的发展有所贡献。

“新世纪统计与经济优秀著作文库”编辑部

2006 年 6 月

前言

近年来，经济评价应用研究的文献数量呈现快速增长趋势，其社会影响力也与日俱增。一些著名的评价研究，如国家竞争力评价、区域竞争力评价、城市竞争力评价、公司治理评价、农村社会发展评价、环境评价等被长期、定期开展。评价结果受到政府、企业、学术界和社会大众的普遍关注，具有十分广泛的社会影响。然而，从方法论角度来看，评价方法也是一种测量方法，评价结果是一种测量结果。当人们看到一项评价结果时有理由询问，所进行的测量是科学进行的吗？评价标准是客观的吗？评价结果及其产生的应用对所有被评价对象是公平对待的吗？评价是否准确地评价了所要研究的内容？实际上，这些都是每项测量研究必须回答的问题。

国际上的一些权威测量机构、研究组织和专家学者很早就开始重视这些问题。早在20世纪30年代，这一领域的问题已经形成了一个明确的概念——有效性(Validity)，并长期被加以研究。20世纪50年代，美国心理学会(American Psychological Association)将有效性检验列入了《心理学测量标准》。1999年，由美国教育研究协会(American Educational Research Association)、国家教育测量委员会(National Council on Measurement in Education)和美国心理学会三个权威机构共同发布的《教育学和心理学测量标准》中，将有效性作为一个独立概念明确规定在测量标准中。

从有效性研究的内容看，它应该属于评价研究的一个组成部分，它是以理论逻辑和经验证据为基础，对评价系统设计和评价结果得分，及其应用后果合理性和正确性的检验。为了证明评价的有效性，研究者应该收集相关的证据，提供评价有效性检验的结果。否则，一项评价结果的科学性就难免受到质疑。正如权威所言：如果一项评价研究不能提供其有效性的证据，那么它的评价结果就可能是一堆没用的数字。

然而，考察目前在经济领域开展的评价研究可以发现，对有效性问题并未给予足够的重视。在经济评价理论研究方面，也缺少对有效性问题的系统性研究成果，不能对应用研究提供有效性检验方法论方面的指导。这一现状显

然与当前经济评价应用研究的蓬勃开展是不相称的，因此本研究是一个很好的补充，对于开展经济评价应用研究具有很好的实用参考价值，对于完善经济评价理论体系具有一定的学术价值。

本书主要包括三个方面的内容：一是吸收相关学科的研究成果，结合经济研究特点，建立了一个经济评价有效性检验的理论框架；二是针对有效性检验的各方面内容，建立一个统计检验框架；三是针对经济评价研究的不同环节和目标，提供一套完整的统计检验方法。本书内容的结构安排是：首先理清有效性的基本概念、性质与任务。其次，充分考虑经济评价在研究对象、任务、目标等方面的自身特点，阐述经济评价有效性的内涵。并从模型设计和结果应用两个角度，建立一个经济评价有效性研究的理论框架，在该框架范围内提出有效性检验的统计分析框架和步骤。最后，对每一类检验设计一套或多套对应的统计检验方法，并通过具体案例说明方法的使用。

本书在方法与技术方面主要有两个特点：一是吸收、融合了多个学科的理论和技术，并予以创新应用。在有效性理论方面，吸收、借鉴了教育学和心理学评价的研究成果；在经济评价有效性理论方面，结合了经济研究的特点予以创新；在有效性检验方面融合了多种统计技术，包括相关分析、回归分析、多元统计分析、定性资料统计分析、结构方程分析、Meta 分析等。引入了一些较新的统计方法，如跨样本验证性因子分析、相关特质相关方法验证性因子分析、高阶结构方程分析、Meta 回归分析等。二是列举了大量案例。几乎每种检验方法都配有具体案例，这就使得从事实际研究的工作者可以方便地参考和使用书中提供的方法。所有案例都使用 SAS 软件进行了计算，充分证明了完成这些复杂的有效性检验是完全可能的。

本书的主要观点体现在经济评价有效性研究理论框架的设计和统计检验方法的构建方面。本书提出的理论框架包括：经济评价有效性的概念、性质、研究内容与检验思路。经济评价的有效性是指：以经济学理论逻辑和实际经验证据为基础，对经济评价系统设计和评价结果，及其应用后果合理性和正确性的推断。经济评价有效性的研究内容主要包括以下几个方面：①评价的内容有效性；②评价的构想有效性；③评价结果的外部有效性；④评价结果的社会有效性。

内容有效性反映了评价研究内容与其依赖的经济理论的吻合程度。设计内容覆盖的理论内容范围与实际理论范围吻合的越严密，说明所研究内容越具有代表性，即评价模型设计的内容有效性越高。构想有效性用于反映评价模型构想的测量方法和测量体系结构，与经济理论描述的要素间关系和经济系统结构关系是否一致。两者的一致性程度越高，则模型的构想有效性越高。

外部有效性对应那些试图将评价结果推广到样本以外更广范围的情况。例如使用一个上市公司样本得出的实证结果，去说明上市公司整体水平，这就涉及评价结果的外推有效性或外部有效性问题。由于多数经济评价使用的样本很难做到随机性和客观性，因此评价结果是否具有外部有效性必须进行严格的检验。社会有效性是指评价结果的社会价值，是经济评价有效性检验中最为重要的内容，也是最困难的内容。从理论上来说，任何一项评价研究都应该是客观的，无论对被评价者，还是对评价结果的使用者来说都是公平的。然而在实际研究中，评价系统的构建大都受研究者自身价值观的影响。有些时候，研究者的价值观取向与社会主流价值观取向并不一致。也有一些研究，由于系统设计不正确而扭曲了研究的价值取向。由于经济评价具有广泛的社会影响，而且常常与政府的经济政策和改革措施相联系，因此评价结果与社会主流价值标准的一致性，应该作为评价结果有效性的一个重要考量标准。

本书将经济评价有效性的性质归纳为五个方面：第一，经济评价的有效性是评价研究本身的一种属性，而不是问卷、评价指标体系或其他测量工具的属性。第二，经济评价有效性研究更加侧重于评价结果的社会性和政策性。所谓社会性是指，要注重评价结果与社会主流价值观的吻合程度及其可能引起后果。所谓政策性是指，在根据评价结果制定政策时要注意评价结果的适用范围和边界。第三，对经济评价的有效性应该从多个方面进行验证，因为有效性的各方面都不是可有可无的，任何单一方面的有效性都不能说明评价结果的科学性。第四，有效性是一个程度概念，不适合使用有效或无效这样的判定结论来进行评价。第五，经济评价有效性的检验有时需要一个持续过程。特别是那些为政策服务的评价研究，应该随着研究主体、时间、所处环境等条件的变化，对其有效性进行重新评估。

本书共分为三个大部分、总共六章。第一章构成第一部分，主要对有效性研究的发展历史进行考察、回顾，对已有的研究成果和观点进行评述，总结了现有研究成果关于有效性的概念、意义、性质和任务。这些内容绝大部分来自于经济学和经济评价之外的领域，对其进行考察的目的，是展现当前有效性问题研究的全貌，介绍这一领域最新的研究思想和技术成果。

第二章构成第二部分。在这一章中，我们从经济学角度，结合经济研究的对象、任务和特点，尝试为经济评价研究提供一个更为贴切、更为简洁的有效性检验理论框架，这也是当前其他领域有效性研究的一个发展趋势。本章提出了经济评价有效性的概念、意义、性质、研究内容和检验思路。在这一章的末尾，我们还讨论了另一个重要内容——为经济评价有效性检验制定一个基本的统计研究框架和工作程序。

本书的第三部分包括第三、四、五、六章，这一部分主要围绕各类有效性检验构造对应的统计检验方法。在第三章中，我们提出了一个建议，即在建立经济评价模型时除了要进行内容有效性检验外，同时还要重视构想有效性的探索分析。这就把通常建模之后才进行的构想有效性检验工作，提前到建模之前和之后同时进行。相信这种做法有助于提高研究的总体有效性，并可以提高工作效率。在内容有效性检验方面，重点讨论了专家评价结合列联表分析的方法；在构想探索方面，讨论了探索性因子分析和聚类分析方法的应用。在第四章中，我们讨论了使用验证性因子分析模型进行构想有效性检验的统计方法；针对经济评价的特点，讨论了关于多特质、多方法综合评价构想检验的相关分析方法和结构方程分析方法；针对复杂评价构想设计，讨论了高阶构想结构模型的统计检验方法。在第五章中，我们讨论了关于各种效标关联分析的统计方法。包括实用参数方法、非参数方法、均值检验和回归分析方法，定性指标分析的统计检验方法等。还使用具体案例介绍了实际应用中，通过数据转换屏蔽失真数据的方法。在第六章中，我们特别强调评价在政策方面应用的有效性问题。讨论了关于外部有效性检验的统计方法，包括针对不同研究对象样本的协方差矩阵检验；跨样本结构方程模型分析方法；合并独立研究的 Meta 分析方法。为了给政策研究者提供更多的有效性检验方法，我们还讨论了对独立研究进行异质性分析的 Meta 回归分析方法。

本书的主要贡献是：①系统阐述了经济评价有效性的概念、性质、研究内容、研究目的和意义，提出了一个经济评价有效性研究的理论框架，推进了该领域的理论研究。②提出了一个经济评价有效性检验的统计分析框架，并设计了一套完整的统计检验系统，将定性资料分析、结构方程、Meta 回归分析等多种统计方法引入到经济评价有效性检验研究，形成了完整的统计检验系统，丰富了该领域的方法论研究。③提出了为政策服务的外部有效性检验观点，拓展了有效性研究的领域。④强调经济评价的社会价值有效性观点，有利于促进经济评价研究科学、健康的发展。

本书为开展经济评价应用研究提供了有效性检验的理论和方法支撑。这有利于促进人们从更严谨、更科学、更规范地角度去看待和从事这一领域的研究。也为评价研究的社会监督提供了有效工具。有理由相信，这项研究成果会很快应用于实践，并取得良好的社会影响和效益。

鉴于笔者水平有限，文中难免存在不当和错误之处，恳请各位专家学者和广大读者批评指正。

作者

2013 年 5 月

目 录

第一章

有效性的概念、研究内容与意义

一、有效性的概念和研究意义

综合评价方法从本质上来说是一种测量技术。它是通过收集被研究对象多方面的行为和表现，经过技术分析，对研究对象某个方面或者多个方面的特质进行测量和评价。评价结果也称评价得分，通常是数值指标或有序指标，在特殊情况下也可以是名义指标。[①]同使用其他测量方法一样，使用综合评价方法得到的测量结果也存在误差。如果假设第 i 个研究对象待测量特质的真实得分为 T_i，综合评价的结果为 X_i，测量误差为 E_i，则测量结果和特质真实分数之间的关系可以用下面的公式来表示：

$$X_i = T_i + E_i \tag{1.1}$$

从上面的公式可以看出，评价结果包含了真分数和测量误差两个部分。显然，社会公众希望看到的是一个具有较小测量误差的评价结果。测量误差主要来源于以下三个方面：

1. 随机误差

随机误差是测量过程中偶然发生的，由不可控制因素引起的测量误差。

① 有关指标类型的讨论请参阅张尧庭，定性资料的统计分析，广西师范大学出版社，1991。

例如连续两次测验一个学生 50 米跑成绩，即使在完全相同的外部条件下，两次测验的结果可能仍然不同。再如一个班级考试，即使使用同一张试卷，班级学生两次测试结果仍然可能不一致。这些误差是由客观的、随机因素引起的，在研究中不可能被完全消除。

2. 系统误差

系统误差是由于使用的测量方法不当，或由于其他因素、特别是主观因素干扰导致的测量误差。例如，使用外语试卷去考察学生的文学水平，此时导致学生成绩差异的原因不仅由学生文学水平决定，而且还由学生的外语水平决定，使用的测量方法明显偏离了评价目标。再如，使用调查问卷收集评价数据时，由于种种原因，答卷者没有进行真实回答，从而导致使用错误数据进行评价造成系统误差。在进行评价模型设计时，由于研究者对理论和研究目标把握不准确，导致评价结果存在系统误差，是评价研究中最常犯的错误。由于人类认知的局限性，系统误差在实际经济评价研究中也不可能完全消除。

3. 抽样误差

当使用的数据资料不是来自研究对象全体，而是来自其中一部分时，评价得分称为样本统计结果。由于样本仅是研究对象的一个子集，因此不可能包含总体的全部信息，当使用的样本不同时，测量的结果也不相同。因此，如果用样本结果推断总体情况必然存在误差，由抽样导致的误差称为抽样误差。抽样误差是一种可以控制大小的误差，增加样本量通常可以减少抽样误差，当样本与总体重合时抽样误差即被消除。

随机误差和抽样误差不依赖于具体的研究学科，无论在自然科学、还是社会科学中都具有共同特点，因此形成了独立的研究领域。抽样误差只有在通过样本推断总体情况时才存在，只与抽样方法和使用的样本有关，而与经济理论和研究内容无关。因此如果使用的是总体模型，则模型中仅包括随机误差和系统误差。例如假设模型(1.1)为总体模型，则误差项 E_i 只包含两种类型的误差，即随机误差和系统误差。随机误差研究属于可靠性研究，衡量系统可靠性的指标称为信度。有关这一领域的研究至少已有二百多年的历史，目前已经形成了专门的研究领域。随机误差和抽样误差影响经济评价的有效性，是影响有效性的必要条件，但都不是充分条件。例如一个具有较高信度的测量结果，仅能保证在测量技术方面具有可靠性，但不能保证真正测量了想要测量的内容。

对系统误差的研究，需要结合具体学科理论和研究使用的技术方法展开。在评价研究中，虽然大多数有效性度量指标涵盖了整个误差范围，包括了系统误差、随机误差、抽样误差的影响，但有效性研究主要针对的是系统误差分析。

在本书中，我们不单独讨论某一种具体误差的影响，而把它们包含在一些具体的有效性测量中。测量有效性的指标通常称为效度。

今天，关于有效性的概念、有效性的研究意义，人们仍然存在很多不同的观点。早期的学者并不区分有效性是评价研究本身的的属性，还是评价结果的属性，或是使用的调查问卷、测量工具的属性。直到20世纪50年代，人们才认识到有效性的真正研究意义并非针对某份调查问卷，或者是某个使用的测量工具。“有效性不是一个问卷的属性，询问一个问卷是否有效是幼稚的”。(Cronbach 和 Meeh，1955)因为一份问卷对有些对象可能适宜，但对另外一些对象却不适宜。例如大学英语六级考试对测量大学生的英语水平很适宜，但对中小学生却很不适宜。脱离评价研究本身，我们就无法判断一份问卷的有效性。因此，有效性研究不是针对一份评价问卷，也不是针对使用的测量工具或评价指标体系，而是针对评价研究本身及其结果。正如 APA(American Psychological Association，1974)在其公布的测量标准中所阐述的：有效性永远是指证据支持从评价结果所得出推论的程度。

对有效性概念的争论经过了一个漫长的过程，直到20世纪80年代中期才形成了具有主流观点的有效性概念。目前主流观点的有效性概念是：“以经验证据和理论逻辑为基础，对基于评价得分和评价方法所得出的结论，以及由此引发的行为的充分性和适当性的综合性度量。”(Messick，1989)

在20世纪70年代以前，有效性的内涵仅局限在传统的测量误差概念范畴。直到20世纪70年代以后，才进一步扩展到重视评价结果的社会影响。Messick 是这一思想的倡导者和积极推动者。他认为：“一项评价的有效性依赖于该评价的社会影响，这一点必须得到公认。……在分析(评价的)有效性时，社会价值和社会影响不能被忽视。”(Messick，1980)

今天，人们已经普遍接受了对评价的社会有效性进行检验的观点。现代有效性的研究意义，已经远远不再局限于评价的误差问题，而是扩展到评价构想和评价结果的合理性，以及评价结果社会影响和社会后果的综合性考量。为了能够准确地理解和把握有效性的概念和意义，下面有必要先对有效性研究的历史发展过程作一个较为全面的考察和回顾。

二、有效性研究发展综述

(一)早期的有效性研究

根据 Lissitz 和 Samuelsen(2007)的考察，有效性概念在文献中的正式定

义大约出现在20世纪30年代。但实际上，早在19世纪末期就已经出现了与有效性研究相关的思想萌芽。1890年，Cattell在他的名著《心理测验与测量》中提出：测量需要有常规标准以便于进行比较。根据Peterson(1926)在其书中的记载，1893年，Jastrow在美国芝加哥举办的哥伦比亚展览会上所做的实验，已经使用了将测量结果与常规标准之间进行比较的方法。在这一时期的心理学和教育评价研究中，出现了将被评价对象进行个体之间的关联性比较(Sharp，1898～1899)，[①]将评价结果与教师评价之间进行关联性比较(Bolton，1891～1892；Gilbert，1894)，将评价结果与学生学业表现之间进行关联性比较(Wissler，1901)，或将评价结果和被评价对象个人多方面表现的测量结果进行关联性比较(Spearman，1904)等研究方法的使用。Binet(1908)还针对个体的心理测验，建立了一个总体判别标准。[②] 研究者使用这些相关分析或判别分析方法的目的很明确，就是试图寻找到一种证据，以此来证明自己评价研究结果的正确性。这一时期存在的一种朴素认识是，只要评价结果与某个已经存在的、具有标准意义的结果相关，那么该评价结果就是有效的。这种思想可以看作是效标关联有效性检验的早期雏形。

大约在1915年，出现了所谓效标关联有效性(criterion-related validity)的概念(Lissitz和Samuelsen，2007)。Thorndike(1921)阐述了当时人们对这类有效性的认识："(因为)在实践中验证观点需要一些衡量标准，标准被假设反映或近似反映了变量的真实价值……。任何能为标准提供准确估计的测量就被认为是有效的。"通过使用测量以外的价值标准来判别测量的合理性，这说明人们当时已经认识到单纯依靠提高测量工具的精确度，即提高测量的可靠性并不能保证测量结果的正确性。因为提高测量精度，并不能告诉人们是否测量了想要测量的东西。如果不知道测量结果和预测意味着什么，也就无法对其进行应用(Gulliksen，1950)。

20世纪30年代以前，有效性研究仍然是作为可靠性研究的一部分，被包含在可靠性研究中。在20世纪30年代至40年代，有效性从可靠性研究范畴中分离出来，开始成为一个独立的概念。这一时期出现了很多关于有效性问题的专门讨论和研究成果(Thurstone，1931；Guilford，1936；Rapaport，Gill和Schafer，1945；Googenough，1949；Thorndike，1949；Weber，1949)。1931年，L. L. Thurstone在他编写的《测量的可靠性和有效性》一书中，明确区分了可靠性和有效性这两个在评价研究中具有重要意义的概念。J. P. Guilford

① 参阅J. Peterson，Early conception and tests of intelligence. Yonkers，N. Y.：World Book Co.，1926.

② 转引ANASTASI，A. Psychological testing (5th ed.). New York：Macmillan，1982.

(1936)在《心理测量的方法》中对有效性的思想进行了系统总结。这些研究对后来有效性理论的发展和研究产生了很大影响。

为了寻找能够证明有效性的各种证据,早期的研究者们从不同的角度发展了不同的方法。只要一种证据能够证明评价结果确实测度了人们想要测量的信息,它就被定义为一种有效性。于是大量的有效性概念在这一时期相继出现,比较著名的定义有:效标关联有效性(criterion-related validity)、内容有效性(content validity)、实证有效性(empirical validity)。此外还有 Guilford (1946)提出的因子有效性(factorial validity),Mosier 提出的表面有效性(face validity),Gulliksen(1950)提出的内在有效性(Intrinsic validity),Cronbach (1949)从实证有效性中分离出的逻辑有效性(logical validity)等。

虽然早期有效性的概念众多,但基本局限于将有效性仅仅作为对测量技术方面的评价标准。这使得有效性成为一种狭义的概念,其目的仅仅就是为了指导研究者如何去做好技术方面的测量工作而已。20 世纪 50 年代前后,这种狭义思想开始遭受到很多批评,其中 Rulon(1946),Cureton(1951)等人的观点最具有代表性。显然,狭义的有效性检验没有将有效性与评价研究所依赖的学术理论背景相联系,也没有和评价结果的应用,及其引起的社会结果相联系。然而如果不能说明评价结果的理论价值和社会价值,就很难说一个评价是真正有效的。正如 Messick(1989)所言:"检验有效性的关键在于(其理论的)解释力、关联度和应用性。"

(二)构想有效性概念的建立

经过 20 世纪 30 年代至 40 年代的探索和争论,人们对有效性问题的认识大大深化,有效性的概念进一步清晰。到了 20 世纪 50 年代初,有效性的概念被集中到少数几种类型,主要包括效标关联有效性、内容有效性、表面有效性和构想有效性,下面是对这些有效性概念的解释。

1. 效标关联有效性(criterion-related validity)

效标关联有效性是最早出现的有效性概念,它是指评价结果是否符合某个独立于评价之外的价值标准。所谓"效标"指的就是在该事物上已被人们普遍接受的独立价值标准。效标可以是一个已经存在的、或者是一个先前获得的、并被证明是有效的同类评价的客观结果。效标关联有效性通常使用评价结果和效标之间的相关系数来进行度量,也可以使用相关系数的平方来进行度量,称之为效标关联效度。

按选择的效标时间来划分,效标关联有效性又分为:预测有效性(predictive validity)和当前有效性(concurrent validity)。预测有效性是指以

一个未来出现的标准为效标，考察评价结果与未来标准的关联程度。预测有效性效标的确定受研究所在学科的理论影响，具有明确的理论意义。一般说来，评价结果和效标之间应该存在明显的因果关系或相依关系。例如，以未来企业的经营业绩为效标，对企业竞争力的评价结果进行检验。显然这里的效标具有明确的理论含义，即随着竞争力的提高，企业的经济效益也将提高。当前有效性也称为同时有效性。它是指以一个当前存在的效标为标准，考察评价结果与现实标准的关联程度。与预测有效性一样，当前有效性的效标也具有明确的理论意义。例如，以当前企业的经营业绩为效标，对企业竞争力的评价结果进行检验。这里的理论含义是，竞争力强的企业应该比竞争力弱的企业有更高的经济效益。

2. 内容有效性(content validity)

根据 Lissitz 和 Karen Samuelsen(2007)的考察，内容有效性的概念大约出现在 20 世纪 40 年代。Rulon(1946)是最早在实践中使用这类研究的学者之一。所谓内容有效性是指：评价的内容是否很好的覆盖了所要研究的内容范围(通常称其为内容域)，调查问卷中提出的问题，或评价指标体系是否与所研究的内容相关，并且能够代表所研究的内容域。例如对学生某门课程的学习情况进行评价，其内容有效性检验，就是要验证评价内容是否很好的覆盖了学生所学课程的全部知识点，在测试使用的题目在全部可选题目中是否具有代表性。

通常规定，在检验内容有效性之前，需要根据研究的理论背景确定一个评价范围的内容域。然后将评价研究的内容和理论内容域进行比较，进而根据两者之间得吻合程度来确定内容有效性。在现实研究中，由于理论内容域往往比较复杂，研究者基于不同的研究目标和研究角度，对内容域的理解也不完全一致，因此很难建立评价内容和理论内容域之间比较的客观标准。所以对内容有效性的检验大多依赖专家的主观判断，专家的水平和专家数目都对内容有效性的检验结果有很大影响。内容有效性的度量结果称为内容效度。

3. 表面有效性(face validity)

表面有效性是指从被评价者或使用者角度来看待评价测量的目的。有些学者认为，表面有效性不是真正具有评价意义的有效性(Nevo，1985；Messick，1989)。Messick(1989)在对表面有效性的讨论中写道：从技术角度来看，表面有效性并不是真正意义的有效性。但是对于被评价者来说，他可以通过表面有效性形成对评价研究目标的看法。如果被评价者认为评价结果对其不利，他可能在测量中采取不合作态度。反之，如果被评价者认为某类评价结果对其有利，他可能会主动引导测量转向对其有利的结果。