

教育对象之间的一致性程度。

因此，建构效度不仅是要关注理论或构念系统（有时可能并不需要某个构念系统，如依据一项结果是否达到教育政策的要求来给予评价），更要关注评价体系是否真实反映了评价客体的心理状态、行为取向、能力水平各方面的变化。同时一个具有良好建构效度的教育评价工具还可以准确地将一组被试和另一组被试区分开来。比如，一个评价工具要检测学生面向社会收集信息的能力，在检验这一评价结果时，如果所设计的评价项目能够将收集信息能力强与能力弱的学生分别开来，那么这个评价工具就有了建构效度。反之，如果评价的体系、维度不能反映学生在收集信息方面真实存在的能力差别，这个评价工具就没有建构效度。

（三）内容效度，即教育评价需要考察的内容或行为样本的适当程度

教育评价在实际操作时，需要对教育评价的总目标进行分解，总目标又具体分解为具体的教育子目标，而子目标又规定了哪些内容需要被考察以及相应的权重大小如何。因此，内容效度是对于要考察的内容或者行为样本的适当程度。在实际的评价过程中，内容效度是通过对评价目标以及具体的可操作的内容维度目标来进行检验的：一是反映总目标的维度内容的充分代表性，即维度目标是否被覆盖了所要考察的全域范围中的总目标要求；二是要评价的项目应是已界定好内容范围的代表性样本；三是要对每个项目所要评价的具体行为（如知识与技能）与上一级评价目标应保持一致，同时各子目标要相互独立，不能重叠，要力图反映总目标的要求。考察内容效度可采用复本法、再测法或者经验判断法。

（四）交叉效度，即各种教育评价方法相互检验和验证的一致性程度

实际评价中，多元性评价标准给评价带来了一定的困难。有时候，评价者在操作层面上将不同的标准加以融合是出于某一实用的动机，但往往忽略了“程序或技术的方法”与“逻辑证明的方法”之间的区别，对自己使用的方法背后的方法论缺乏自觉和反省。^[7]结果，专家群体则很难对他们的研究进行质量评价。由于标准不同，这时在评估教育评价的效度时需要充分考查各种评价方法和标准的交叉效度。三角互证法可以用来检验不同的证据来源或不同的资料收集方法，其目的是为了在不同的标准

之间进行相互证实。因此，可将此方法用作评估检验交叉效度的一种方法。

在教育评价中一般有两种基本评价方法：量的评价方法（亦称量化评价方法）和质的评价方法（亦称质性评价方法）。由于这两种方法采用不同的研究范式，因此，在研究方法、路径和步骤等方面具有很大的差异，评价效度的类型也有本质差异。如表1所示归纳了两种不同研究范式下的代表性方法以及据此建立的评价效度类型。

表1 两种评价方法的比较及效度评价类型

方法类别	代表性评价方法举例	评价效度类型
量的评价方法	数学方法：数理统计方法，如多元统计（包括聚类分析、判断分析、相关系数检验法、因子分析、Bayes方法等）、回归分析、相关系数检验法；模糊数学法：模糊综合判断、AHP（层次分析法）、模糊距离模型等；灰色系统理论如灰色综合评价法。	关注统计结果效度、实证效度、结构效度等
质的评价方法	基于主观评价法：同行评议法、专家评议法、德尔非法、调查研究法、案例分析法、思辨分析；基于情境和内容的评价法：民族志、现象学、访谈法、内容分析、叙事研究、实地研究、传记法、会话分析、实地考察、档案文件研究、哲学等。	描述效度、解释效度、理论效度、推广效度（内部、外部）、评价效度 ^[8]

有必要指出的是，当两种方法整合在一起使用，这时如何考察评价方法的效度？其总的原则是在讨论量的评价结果时，则应考虑按数学概率的规则。将研究的结果推广到从中抽样的总体当中的适当性。在讨论质的评价结果时，则要根据质的研究目的、研究方法和策略以及研究步骤采用不同的效度评价标准及效度类型。如表1所示。无论是量的评价方法还是质的评价方法都需要从多方面寻找“效标证据”，同时考虑效度的测量指标的易获取性和可能性等因素。

（五）评价主体效度，即各种评价主体对评价客体评价结果的正确性和公平性

现代教育评价实践中，往往存在多个评价主体和多个利益相关者。因此，不同评价主体的评价目的不同，评价观念、评价标准、评价方法、评价工具的使用都可能带来很大的差异甚至相反，从而使评价失去效度。从本质上说，评价效度是针对于目的而言，不同评价主体由于评价目的、价值观、立