引用格式： 孙旭东，任佳乐，刘霞，等. 儿童用品中吸入性化学应激源限量社会经济分析方法效度评估——以地垫甲酰胺为例[J]. 标准化学报，2026(11): 53-67.
SUN Xudong, REN Jiale, LIU Xia, et al. Validity Assessment of Socioeconomic Analysis Methods for Limits of Inhalation Chemical Stressors in Children's Products: A Case Study of Formamide in Floor Mats[J]. Journal of Standardization, 2026(11): 53-67.

儿童用品中吸入性化学应激源限量社会经济分析方法效度评估
——以地垫甲酰胺为例

孙旭东[1] 任佳乐[1] 刘 霞[2] 张 庆[2] 刘 晓[3*]
〔1.中国矿业大学（北京）；2.中国质量检验检测科学研究院；3.北方工业大学〕

摘 要：【目的】针对儿童用品中吸入性化学应激源限量社会经济分析方法构建了一套探索性的效度评估框架与方法，并以儿童地垫甲酰胺限量为例进行了初步验证。【方法】基于内容效度、结构效度、关联效度三维通用效度构建方法，搭建效度验证框架，通过内容效度指数、验证性因子分析、方差分析效应量求得相应的效度，并采用算术平均法合成综合效度。【结果】以儿童地垫中甲酰胺限量评价方法为评估对象，测评结果显示，该评价方法的内容效度指数为0.954，结构效度得分为0.873，关联效度效应量η^2为0.994。依据算术平均法，得到综合效度为94.03%（≥90%），等级为“优秀”。【结论】本研究构建的三维效度验证框架具有可操作性，为儿童用品中吸入性化学应激源限量社会经济分析方法的科学性评估提供了探索性工具，有助于推动相关决策向更规范、更科学的方向发展。

关键词： 效度验证；社会经济影响评价；化学品限量
DOI编码： 10.3969/j.issn.2097-857X.2026.11.006

Validity Assessment of Socioeconomic Analysis Methods for Limits of Inhalation Chemical Stressors in Children’s Products: A Case Study of Formamide in Floor Mats

SUN Xudong[1] REN Jiale[1] LIU Xia[2] ZHANG Qing[2] LIU Xiao[3*]
(1. China University of Mining and Technology, Beijing; 2. Chinese Academy of Quality and Inspection&Testing; 3. North China University of Technology)

Abstract: [Objective] An exploratory validity evaluation framework and method for the socioeconomic analysis of the limit of inhaled chemical stressors in children’s products was constructed, and a preliminary verification was conducted taking the limit of formamide in children’s floor mats as an example. [Methods] A validity verification framework was constructed based on three dimensions: content validity, construct validity, and criterion validity.

基金项目： 本文受国家重点研发计划项目“儿童用品中吸入性化学应激源限量研究”（项目编号：2022YFF0606205）资助。
作者简介： 孙旭东，博士，副教授，研究方向为能源经济与产业政策。
任佳乐，硕士研究生，研究方向为社会经济影响评价。
刘霞，博士，研究员，研究方向为产品质量安全。
张庆，博士，研究员，研究方向为消费品检测及风险评估技术。
刘晓，通信作者，博士，讲师，研究方向为决策评价。

Corresponding validity indices were calculated using the content validity index, confirmatory factor analysis, and effect size from analysis of variance, and the arithmetic mean method was applied for comprehensive validity synthesis. [Results] Using the limit evaluation of formamide in children's floor mats as the assessment case, the results showed that the content validity index of the evaluation method was 0.954, the construct validity score was 0.873, and the effect size η^2 for criterion validity was 0.994. Based on the arithmetic mean method, the comprehensive validity score is 94.03% (≥90%), which is rated as "excellent". [Conclusion] The three-dimensional validity verification framework constructed in this study is highly operational. It provides an exploratory tool for the scientific evaluation of socioeconomic analysis methods for limit values of inhalable chemical stressors in children's products, thereby promoting more standardized and scientific policy-making.

Keywords: validity verification; socioeconomic impact assessment; chemical limit values

0 引言

社会经济影响分析[1]（Socio-Economic Analysis，SEA）是一种广泛应用于化学品管理及政策制定的、用于评估社会经济综合影响的方法。我国GB/Z 24785—2009《化学品限制的社会经济分析指南》[2]为化学品限制的SEA提供了通用框架，消费品有害化学物质限量定值研究进一步明确，需在安全限量、科学限量约束下，综合经济、社会、环境影响确定综合限量[3]。儿童用品中化学物质限量的社会经济分析方法归属于综合决策研究范畴，是在公众健康保护、生态环境可持续性与产业发展活力之间达成平衡的关键政策工具[4]。如何评判所构建的化学物质限量研究方法与评估手段，即如何确保所构建的评价体系能够真正有效、无偏地测量“限量”所带来的“社会经济影响”，既是保障儿童用品中化学物质限量研究效度与质量亟须解决的关键问题，也是制定衡量研究方法科学性的技术规范时所关注的核心问题。评价方法若缺乏效度证据，其测量结果便无法被确信为对目标构念的有效反映，基于此得出的政策建议也将失去科学支撑。因此，构建一套方法效度评估框架，是提升本领域研究成果可信度与可比性的先决条件。

效度是衡量评价方法或评估方案质量的关键指标之一[5]，通过效度验证可保障评价活动与研究成果的高质量。当前效度的研究主要涵盖结果效度与方法效度2个维度。前者衡量具体研究结果的效度，包含实证研究得到的结论以及量表获得数据的有效性，后者衡量方法本身是否具有有效性。

结果效度分析可理解为基于评价结果所做出的判断和推断具备充分性、可靠性与合理性[6-7]。例如图书情报学实证研究效度检验[8]、学生创新能力的识别效度[9]、青少年关怀能力效度检验研究[10]和DeepSeek智能评分的效度研究[11]等通过统计分析方法检验特定实证研究的效度。此外双维创业韧性量表[12]、中国仲裁员胜任力量表与信效度检验[13]、青少年父母情感支持量表及效度信度的检验[14]和旅游目的地老化量表及信效度检验[15]等则针对不同问题开发量表并进行信效度检验，确保数据的有效性。结果效度分析通常难以对综合型或复杂性问题进行研究，因此更多研究建议围绕综合评价指标体系或综合评价方法展开。

方法效度分析可借鉴测量学中关于效度的定义，即测量工具能够准确测出研究方法所要测量特质的程度[16]。相关研究旨在构建针对具体问题的方法效果验证标准与评估规范。例如，采用等级相关系数法、肯德尔和谐系数法及兼容度优化模型，从结果相关度、次序一致性与方法兼容度3个维度，验证装甲兵数字化部队作战方案评估方法的相对效度[17]。在上述方法的基础上增加模糊聚类效度分析，评估装备保障机构能力评估中5种常用系统评估方法的有效性，以判断评

估方法的相容性与评估结果的合理性[18]。从内容效度、结构效度与效标关联效度3个角度评判高校德育评估的效度，其中内容效度侧重内容代表性与目标覆盖范围；结构效度关注对假设能力的测量程度；效标关联效度强调与外部效标的一致性[19]。内容效度是评估方法效度的基础维度，旨在检验方法所纳入内容是否充分、适切地覆盖研究目标，常通过内容代表性、覆盖度、专家评价、条目效度指数等方式验证。在大规模语言测试领域，内容效度可依据权威理论框架，从题材、体裁、难度、考核技能等维度，对照考试大纲开展系统性覆盖度分析，判断测试内容与目标构念的匹配程度[20]；在量表开发研究中，内容效度检验常采用德尔菲专家咨询法，通过多轮专家评议，结合条目水平内容效度指数、量表水平内容效度指数及矫正Kappa值量化内容适切性与清晰度[21]；在职业考试评估中，还可通过知识类型、认知层级分布的量化分析，检验考核内容与考试目的的契合度[22]。

当前，方法效度研究主要集中于教育、心理、管理及军事等领域，而针对儿童用品中吸入性化学应激源限量的社会经济分析方法效度评估研究目前仍处于空白状态。鉴于此，本研究结合该交叉领域特点，沿用上述方法效度评估思路，从内容效度、结构效度和关联效度维度对评价方法本身进行系统检验，具体验证以下3个方面：（1）测量内容的全面性；（2）内在结构的合理性与区分度；（3）对政策变量（限量）变化的敏感性。进一步，以儿童地垫中甲酰胺限量评价方法为评估对象，展示方法效度评估技术的完整应用流程，验证其可操作性。本研究不仅为特定物质的评估提供方法效度证据，更期望推动“化学品风险—社会经济影响评价”向标准化、科学化的方向发展。

1 儿童用品中吸入性化学应激源限量社会经济分析方法简介

本研究依托国家重点研发计划课题“儿童用品中吸入性化学应激源限量研究”开展。该课题以研发共性关键技术为目标，建立适用于儿童用品吸入性化学应激源限量的社会经济分析方法，具体包含3部分内容：一是构建儿童用品中吸入性化学应激源限量社会经济影响评价指标体系；二是实现各指标的货币化度量；三是完成限量社会经济影响的福利化度量。对化学品限量社会经济影响进行社会经济分析[23]，其本质是解析“限量”这一政策干预如何通过复杂路径传导并作用于不同的利益相关方。

1.1 儿童用品中吸入性化学应激源限量社会经济影响评价指标体系及货币化方法

前期工作已构建并发表了消费品中化学物质限量社会经济影响评价的指标体系[1]。本文为实现评价指标体系在儿童用品中吸入性化学应激源限量社会经济分析的落地应用，采用的限量影响评价指标体系涵盖健康安全、生态环保、产业发展3个维度，包含8个三级指标，其中，$V_1 \sim V_5$的计算方法同前期研究一致。

为将3个维度的异质性影响纳入同一量化框架，采用货币化度量的方式，将不同性质的影响转换为统一的货币单位，便于后续的福利化分析。记$f(V_i)$为指标V_i的货币化公式，各指标的具体计算公式如表1所示。

健康安全维度的货币化度量涉及非不良反应风险和器质性病变风险（V_2）。$f(V_1)$与$f(V_2)$均基于伤残调整寿命年框架，将健康风险暴露导致的疾病负担折算为经济损失。

生态环境维度的货币化度量涉及全产业全过程链废气排放量（V_3）、废水排放量（V_4）和固体废物排放量（V_5）。其中，$f(V_3)$和$f(V_4)$采用污染当量法与排污税机制，以环境税费形式反映货币化成本；通过$f(V_5)$区分一般工业固废与危险固废，分别计税并加总。

产业发展维度下的三级指标包含企业生产成本（V_6）、企业投资改造成本（V_7）和消费者支付意愿（V_8）。企业投资改造成本（V_7）包含工艺、固定资产、研发等可变成本。消费者支付意愿（V_8）在原有指标计算的基础上增加了产品销量变动的影响。

表1 儿童用品中吸入性化学应激源限量社会经济影响评价指标体系的货币化公式

一级指标	二级指标	三级指标	货币化公式
健康安全	健康风险（U_1）	不良反应风险（V_1）	$f(V_1)=\dfrac{X_1\times X_2\times CR\times X_3\times CN\times RAI\times CPP+CN\times CR\times X_3\times CPP\times CDD+CN\times CR\times X_3\times AR\times SAW}{10\ 000}$ X_1（V_1）：非致癌物造成人体健康风险的危害商值 X_2：非致癌风险下的伤残调整寿命年，年 X_3：0~14岁儿童患病概率，% CR：就诊条件概率，% CN：儿童总人数，人 RAI：居民年平均收入，元 CPP：儿童用品使用比例，% CDD：该化学物质导致不适的平均诊断检查费用，元 AR：儿童产生不良反应家长误工天数，天 SAW：社会平均日工资，元
		器质性病变风险（V_2）	$f(V_2)=\dfrac{X_4\times X_5\times DP\times X_3\times CN\times RAI\times CPP+CN\times DP\times X_3\times CPP\times CDT+CN\times DP\times X_3\times OL\times SAW}{10\ 000}$ X_4（V_2）：致癌的终身概率，% X_5：致癌风险下的伤残调整寿命年，年 CDT：该化学物质导致疾病的平均治疗费用，元 DP：确诊概率，% OL：儿童产生器质性病变家长误工天数，天
生态环保	绿色环保（U_2）	全产业链全过程废气排放量（V_3）	$f(V_3)=\sum_{m=6}^{9}P_m\times\dfrac{X_m}{W_m}$ m：废气中污染物种类，6~9分别代表二氧化硫、氮氧化物、颗粒物及挥发性有机物 X_m：m污染物排放总量，万吨 W_m：m污染物的污染当量值，kg P_m：m污染物单位计税单位征收税额，元
		全产业链全过程废水排放量（V_4）	$f(V_4)=\sum_{n=10}^{13}P_n\times\dfrac{X_n}{W_n}$ n：废水中污染物种类，10～13分别代表化学需氧量、氨氮、总氮及总磷 X_n：n污染物排放总量，万吨 W_n：n污染物的污染当量值，kg P_n：n污染物单位计税单位征收税额，元

续表1

一级指标	二级指标	三级指标	货币化公式
生态环保	绿色环保（U_2）	全产业链全过程固废排放量（V_5）	$f(V_5)=X_{14}\times P_3+X_{15}\times P_4$ P_3：一般工业固体废物排污税征收标准，元 P_4：工业危险固体废物排污税征收标准，元 X_{14}：全产业链全过程一般工业固体废物排放量，万吨 X_{15}：全产业链全过程工业危险固体废物排放量，万吨
产业发展	产业承受度与韧性（U_3）	企业生产成本（V_6）	$f(V_6)=V_6=[MC\times(1+X_{16})]\times PUP\times(1+X_{20})\times ASV\times(1+X_{21})$ X_{16}：原材料成本增加比例，% X_{20}：价格变动比例，% X_{21}：销量变动比例，% MC：产品中原材料成本，元 PUP：当前市场的产品单价，元 ASV：产品年销量，万件
		企业投资改造成本（V_7）	$f(V_7)=V_7=[PC\times(1+X_{17})]\times FAC\times(1+X_{18})+PRU\times(1+X_{19})\times PUP\times(1+X_{20})\times ASV\times(1+X_{21})$ X_{17}：工艺成本增加比例，% X_{18}：固定资产成本增加比例，% X_{19}：专利及研发成本变动比例，% PC：产品中工艺成本，元 FAC：产品中固定资产成本，元 PRU：产品中专利及研发成本，元
	市场环境（U_4）	消费者支付意愿（V_8）	$f(V_8)=V_8=PUP\times(1+X_{22})\times ASV\times(1+X_{21})$ X_{22}：消费者意愿支付价格涨幅均值，%

产业发展维度下的指标本身即为货币化单位指标，故其货币化度量公式等于对应指标的计算公式。

1.2 儿童用品中吸入性化学应激源限量社会经济影响的福利化

本研究的“社会福利”是指当儿童用品中吸入性化学应激源限量治理后，在健康安全、生态环保与产业发展3个维度所获得的综合性福祉。福利值用于量化限量标准变动所引发的各维度、各指标福祉变化，即为“社会福利”的变化。针对某一评价指标V_i，其福利值$F(V_i(L))$可定义为当限量由当前基准限值调整至目标限值L时，该指标货币化值的变动差额，计算公式如（1）所示，根据指标类型分情况确定：

$$F(V_i(L))=\begin{cases} f(V_i(L_{base}))-f(V_i(L)), \text{成本型指标} \\ f(V_i(L))-f(V_i(L_{base})), \text{收益型指标} \end{cases} \quad (1)$$

式中：$F(V_i(L))$为目标限值L下评价指标V_i的货币化值，$f(V_i(L_{base}))$为基准（当前）限值下评价指标的货币化值。若$F(V_i(L))>0$，表明该限量变动提升了社会福利；若$F(V_i(L))<0$，则表明该限量变动降低了社会福利。

为消除量纲影响，构建相对指标——福利指数。针对某一评价指标V_i，其福利指数$M(V_i(L))$可定义为当限量由当前基准限值调整至目标限值

L时，该指标福利值$F(V_i(L))$与基准情景下所有指标货币化值绝对值和$f_T(V_{base})$的比值，计算公式如（2）所示，根据指标类型分情况确定：

$$M(V_i(L)) = \frac{F(V_i(L))}{f_T(V_{base})} = \begin{cases} \dfrac{f(V_i(L_{base})) - f(V_i(L))}{f_T(V_{base})}, & \text{成本型指标} \\ \dfrac{f(V_i(L)) - f(V_i(L_{base}))}{f_T(V_{base})}, & \text{收益型指标} \end{cases} \quad (2)$$

式中：$M(V_i(L))$为由当前限值变为限值L下评价V_i指标产生的福利指数；$f_T(V_{base})$为基准（当前）限值下所有指标货币化值绝对值和。综合福利指数$M_T(L)$为健康安全、生态环保与产业发展维度福利指数的代数和，如公式（3）所示，反映不同限量标准下整体社会福利的相对变化。当该指数达到最大时，即视为实现了社会福利的最大化。

$$M_T(L) = \sum_{i=1}^{7} M(V_i(L)) \quad (3)$$

如表1所示，健康安全维度基于疾病负担理论、生态环保维度基于环境治理成本法，两者属于成本型指标；产业发展维度基于成本收益分析，属收益型指标。

2　方法效度的评估技术体系构建

2.1 框架图

本研究构建了一个用于评价儿童用品中吸入性化学应激源限量社会经济分析方法的效度评估技术框架，如图1所示。本研究效度评估的对象是“完整的评价方法流程”。该流程涵盖了从“指标数据采集”到“货币化度量”再到“福利化合成”的全过程。效度验证不仅要检验指标是否选择准确，还要检验货币化换算与福利合成逻辑是否科学。

该框架从内容效度、结构效度、关联效度3个维度[19]对分析方法指标体系进行系统性测评。内容效度验证指标是否全面覆盖评价目标的所有重要方面，通过结合专家逻辑判断与量化分析来确保内容的代表性。结构效度验证指标体系的维度划分是否合理、清晰，通过专家对维度独立性的打分矩阵来检验其内在逻辑结构。关联效度验证指标体系对不同限量水平变化的敏感性和区分能力，通过不同限量水平的数据并运用方差分析等统计方法进行检验。即“评价指标全不全”“结构划分清不清晰”和“能不能体现出变化差异”3个层面的测评共同构成了一个完整的验证闭环，为指标体系的科学性、合理性和实用性提供了系统的量化证据与优化依据。

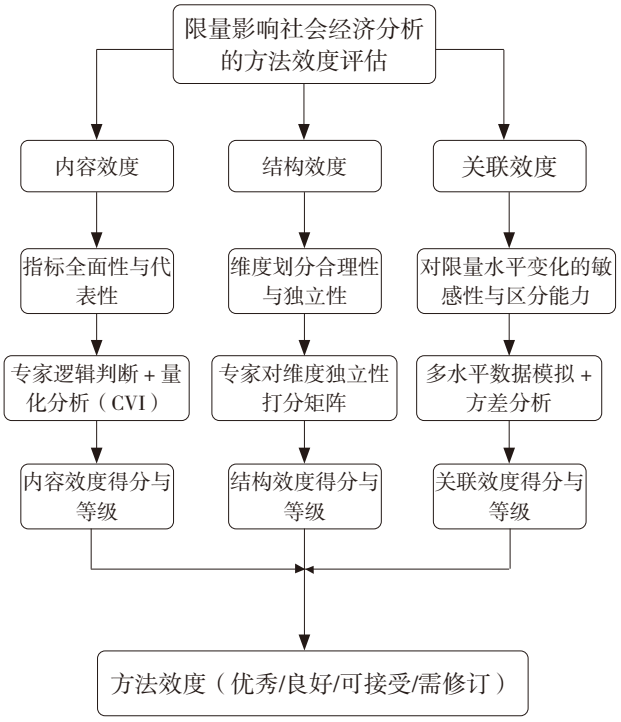


图1 方法效度评估技术框架图

2.2 内容效度评估方法

内容效度是评估指标体系的内容是否具有代表性，即是否测评到了应测评的全部内容。在儿童用品中吸入性化学应激源限量社会经济分析中，内容效度意味着所构建三级指标必须能够全面、无偏地反映化学物质在健康安全、生态环保、产业发展3个维度上引发的所有关键影响因素。

本研究采用广泛接受的内容效度指数

（Content Validity Index， CVI）进行量化评估[21]。根据社会经济分析理论框架，依据国家相关化学品安全管理法规、技术标准，编制“指标—目标双向细目表”，明确每个三级指标旨在测量的具体影响维度。邀请多位化学品安全、环境健康、产业经济、公共政策等相关领域的专家，审阅指标体系及“双向细目表”，依据四点量表对每个指标与评估目标的相关性进行逻辑判断：“1=完全相关”“2=密切相关”“3=有点相关”“4=完全不相关”。前两者（1和2）被视为“强关联项”，后两者（3和4）被视为“弱关联项”。

将多位专家两两配对，形成若干专家组合。针对每个专家组合（*A*，*B*），汇总对所有指标的判断结果，形成如表2所示的判断矩阵。

表2 专家判断的分类指标

专家A\专家B	强关联项	弱关联项
强关联项	*a*	*b*
弱关联项	*c*	*d*

表2中，*a*代表2位专家均判定为“强关联项”的指标数，*d*代表均判定为“弱关联项”的指标数，*b*和*c*代表判断不一致的指标数。该专家组合的内容效度指数（CVI）采用一致性百分比（Item-level CVI, I-CVI）方法进行计算，见式（4）：

$$CVI=\frac{a}{a+b+c+d} \tag{4}$$

该系数范围为[0，1]，数值越大，表明2位专家对指标关联性判断的一致性越高，即内容效度越好。计算所有专家组合CVI的平均值，作为本次评估指标体系内容效度的最终量化估计值，并乘以100%作为其百分制得分。

2.3 结构效度评估方法

结构效度关注评价工具的内在逻辑结构是否与理论假设相符，是检验指标维度划分是否合理、各维度间是否具有足够区分度的核心环节。

界定“消费品化学物质限量社会经济影响”构念，并对“健康安全影响”“产业发展影响”“生态环保影响”3个一级指标进行精确定义，明确其理论边界与核心测量内容。完成指标体系评价后，将所得的综合评价结果与一套效度公认较高、结构已得到验证的同类过往评估体系的评估结果进行比对。根据积差相关或斯皮尔曼等级相关计算两套体系评估结果之间的相关系数。如果相关系数显著且较高，则表明新构建的指标体系在结构上具有较好的效度，其所测量的结构与成熟体系相近。

为进一步验证维度结构的客观性，本研究将构建的“健康安全—生态环保—产业发展”三维结构与现行权威标准进行了比对。依据我国GB/Z 24785—2009《化学品限制的社会经济分析指南》[2]及欧盟REACH法规社会经济分析指南（ECHA, 2015）[24]，化学品管控的评估核心均涵盖人类健康影响、环境影响及经济成本效益三大支柱。本研究的维度划分与上述国际标准及国家标准的核心架构完全吻合，从政策逻辑层面证实了结构效度的合理性。

当缺乏合适的成熟效标体系时，可采用专家对维度独立性的直接判断进行量化分析。邀请*m*位领域专家，对每对一级指标（共3对）及重点二级指标之间的概念独立性进行判断。采用5级评分法评价维度间区分程度：“5=完全独立”“4=基本独立”“3=中度相关”“2=高度相关”“1=几乎同一”。

1）计算每一对维度（如维度A与B）的专家评分均值如式（5）所示：

$$M_{AB}=\frac{1}{m}\sum_{k=1}^{m}S_{AB}^{k} \tag{5}$$

式中：S_{AB}^{k}表示第*k*位专家对维度*A*与维度*B*间独立性的评分。

2）计算所有维度对评分的整体平均值，作为维度独立性的整体度量，计算公式如（6）所示：

$$\overline{M}=\frac{1}{p}\sum_{j=1}^{p}M_j \tag{6}$$

式中：*p*为参与评价的维度对总数（一级维度

3对，或包含重点二级维度对），M_j为第j对维度的评分均值。

3）将平均得分转换为百分制，为结构效度得分S_s，计算公式如（7）所示：

$$S_S=\frac{\overline{M}}{5}\times100\% \quad (7)$$

得分越高，表明专家认为维度间的独立性越强，指标体系的内部结构区分度越好，结构效度越高。

2.4 关联效度评估方法

关联效度核心是检验所构建的评价指标体系能否对“化学物质限量水平”这一核心自变量的变化做出敏感、一致且符合理论预期的响应。目标是量化验证在给定的化学物质限量范畴内，指标体系的综合评价值能否有效区分、排序不同限量水平所导致的社会经济影响，并检验其变化趋势是否符合预设的理论逻辑。

明确所研究的限量范围作为效标范畴。在该范畴内，依据政策现实与研究目的，选取具有代表性的若干限量水平点。针对上述每一限量水平，通过文献参数化、暴露—效应模型推演、生产工艺模型模拟等方法，生成一套完整的、对应于该限量水平的三级指标的估计值，形成“限量水平—指标数据集”。应用预设的指标权重体系，对每一限量水平对应的指标数据集进行合成，计算出该水平下的社会经济影响综合评价值。

为量化限量水平对综合评价值的解释力度，本研究采用方差分析及其效应量指标。

1）以限量水平L为自变量，综合评价值Y为因变量，构建二次多项式回归模型，如式（8）所示，以捕捉可能存在的非线性（如“倒U型”）关系：

$$Y_i=\beta_0+\beta_1L_i+\beta_2L_i^2+\varepsilon_i \quad (8)$$

式中：β_0为截距；β_1、β_2为回归系数；ε_i为随机误差项。通过模型的方差分析，得到F统计量及其对应的P值。若P值小于显著性水平0.05，则拒绝“所有回归系数均为零”的原假设，表明限量水平作为一个整体对综合评价值有显著的解释作用。

2）计算效应量，用以量化该二次回归模型所解释的因变量变异占总变异的比例，公式如（9）所示：

$$\eta^2=\frac{SS_{reg}}{SS_{tot}} \quad (9)$$

η^2值介于0到1之间，与样本量无关，能更稳健地反映关系的实际强度。η^2值越大，表明评价指标体系对限量水平变化的区分能力越强，关联效度越高。本研究以η^2作为关联效度的核心量化指标。

3）将η^2转换为百分制，作为关联效度的量化得分S_r如公式（10）所示：

$$S_r=\eta^2\times100\% \quad (10)$$

2.5 综合效度合成

本研究建立如表3所示的四级评级标准，分为优秀、良好、可接受、需修订。依据内容效度系数CVI、结构效度得分S_s和关联效度得分S_r，确定各维度所对应的效度等级L_c，L_s，L_r。

表3 分级等级评级标准

评级	标准	含义
优秀	得分≥90%	表现卓越，几乎不存在方法学缺陷
良好	85%≤得分<90%	达到良好水平，满足科学研究与应用的基本要求
可接受	70%≤得分<85%	存在一定瑕疵，但结果仍有参考价值，方法需进一步优化
需修订	得分<70%	存在严重缺陷，评价结果可信度低，方法必须进行重大修订

效度是一个多维构念，任一维度的严重缺陷都将削弱整个评价方法的可信度。为了避免单一维度缺陷对整体评估的过度惩罚，并更全面地反映方法学的综合性能，本研究采用算术平均法对各维度效度得分进行合成。

该方法假设3个维度的效度在整体评估中具有同等的重要性（权重相等），计算各维度百分制

得分的平均值，得到综合评价方法的综合效度得分C，计算公式如（11）所示：

$$C=\frac{1}{3}\ (CVI+S_s+S_r) \qquad (11)$$

依据表3中的四级评级标准（优秀≥90%、良好85%~<90%、可接受70%~<85%、需修订<70%），确定评价方法的整体效度等级。

3 儿童用品中吸入性化学应激源限量社会经济分析方法效度评估

3.1 方法评估应用对象选择

本文将第2章提出的方法应用在地垫的甲酰胺限量研究中，甲酰胺是地垫生产过程中常用的发泡剂残留物，具有潜在的神经毒性和生殖毒性。不同国家和地区对儿童地垫中甲酰胺限量标准存在差异。本研究选取5个代表性限量水平L：293.03 mg/kg（最宽松）、159.31 mg/kg、80.73 mg/kg、51.26 mg/kg、34.97 mg/kg（最严格），覆盖了由宽松至严格的不同限量设置。受实际监测数据可得性所限，本研究采用基于公开文献参数与生产工艺模型推演生成的示意性模拟数据，来表征上述各限量水平下的指标取值。其中随限量调整发生变化的22个决策变量详见表4。

模型中涉及的消费者支付意愿（V_8）与企业改造成本（V_7）等关键参数，来源于课题组于2025年期间开展的专项调研。调研采用分层随机抽样法，

表4 不同限量下决策变量X_1–X_{22}取值

变量	限量/（mg·kg^{-1}）				
	34.97	51.26	80.73	159.31	293.03
X_1	5.91	8.67	13.65	26.94	49.55
X_2	0.37	0.37	0.36	0.64	0.63
X_3	0.35%	0.37%	0.41%	0.45%	0.50%
X_4	8.29×10^{-5}	1.21×10^{-4}	1.91×10^{-4}	3.78×10^{-4}	4.74×10^{-4}
X_5	0.39	0.58	0.76	0.75	0.83
X_6	0.000 07	0.000 10	0.000 16	0.000 31	0.000 56
X_7	0.000 81	0.001 18	0.001 86	0.003 68	0.006 77
X_8	0.000 56	0.000 83	0.001 30	0.002 57	0.004 73
X_9	0.000 23	0.000 33	0.000 52	0.001 03	0.001 90
X_{10}	0.000 13	0.000 20	0.000 31	0.000 61	0.001 12
X_{11}	0.000 01	0.000 01	0.000 02	0.000 03	0.000 06
X_{12}	0.000 04	0.000 06	0.000 10	0.000 19	0.000 35
X_{13}	0.000 000 59	0.000 000 87	0.000 001 37	0.000 002 71	0.000 004 98
X_{14}	2.259 7	3.312 4	5.216 7	10.294 5	18.935 4
X_{15}	0.167 0	0.244 8	0.385 6	0.760 8	1.399 5
X_{16}	4.7%	4.7%	3.5%	2%	1.4%
X_{17}	4.7%	4.7%	3.8%	1.9%	1.4%
X_{18}	4.5%	4.5%	3.6%	1.8%	1.4%
X_{19}	4.7%	4.7%	3.4%	1.9%	1.4%
X_{20}	5%	5%	4.4%	2.6%	2.1%
X_{21}	4.8%	4.8%	3.7%	2.4%	1.7%
X_{22}	20%	18%	15%	10%	5%

覆盖华东、华南地区共计30家地垫生产企业（含原材料供应商、制造商、品牌商）及1 000名0~14岁儿童监护人，通过线上线下结合的方式发放问卷，回收有效企业问卷25份，消费者问卷850份。

除上述决策变量之外，模型中所涉及的其余参数如表5，其取值来源包括《中华人民共和国国民经济和社会发展统计公报》《中国统计年鉴》《中华人民共和国环境保护税法》的相关规定与税率标准以及相关调研问卷。

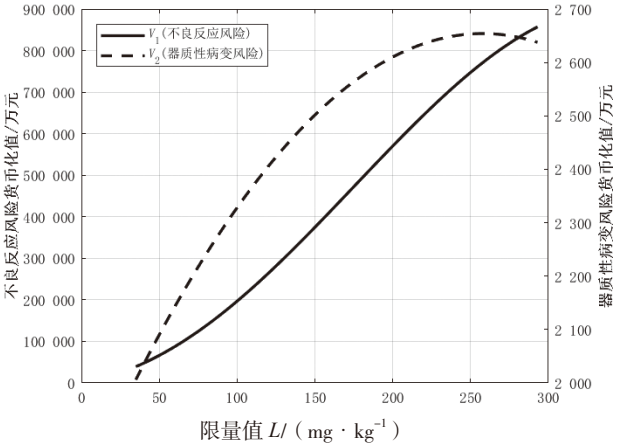
表5 儿童地垫甲酰胺中其余参数取值

参数	取值
CR	0.62
CN	222 400 000
RAI	39218
CPP	6%
CDD	207.9
AR	2.5
SAW	109
CDT	2 892.4
DP	0.34
OL	6
W_m（$m\in[6,9]$）	0.95
P_m（$m\in[6,9]$）	1.2
W_n（$n\in[10,13]$）	1；0.8；0.8；0.25
P_n（$n\in[10,13]$）	1.4
P_3	25
P_4	1 000

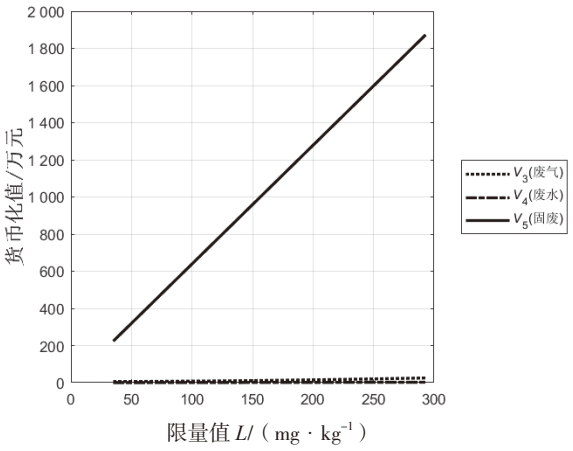
将上述所确定的各个决策变量及参数的具体取值，依次代入$f(V_1)$ ~$f(V_8)$ 8个核心指标各自的货币化计算公式中，通过系统计算汇总，便可以求出这8个核心指标所对应的货币化数值，最终计算结果分维度展示，如图2所示。

将不同指标的货币化值按照评价维度分别代入对应的福利指数计算公式（2）中求得不同限量值下3个维度的福利指数，如图3所示。

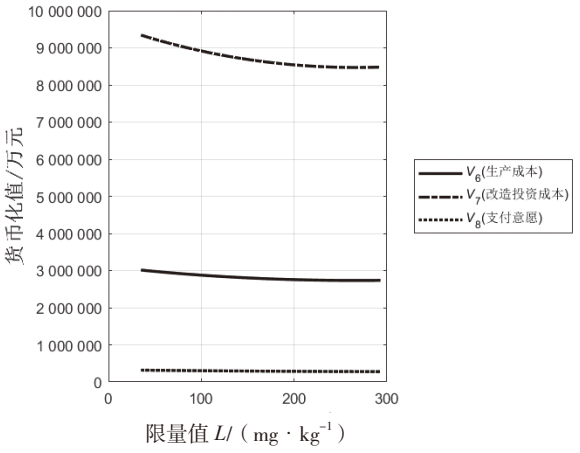
将3个维度下的福利指数曲线合并为综合福利指数，求得5个代表限量水平下的综合福利指数，见表6。



（a）健康安全维度货币化值曲线



（b）生态环保维度货币化值曲线

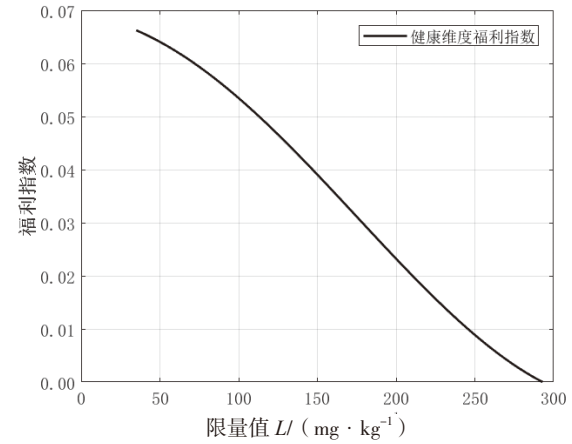


（c）产业发展维度货币化值曲线

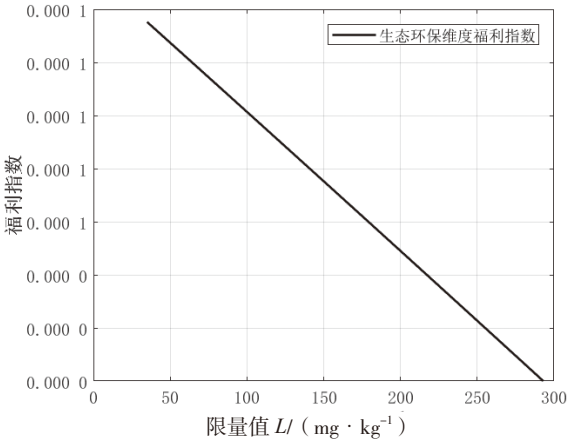
图2 不同限值下指标的货币化值

表6 代表限量水平下的综合福利指数

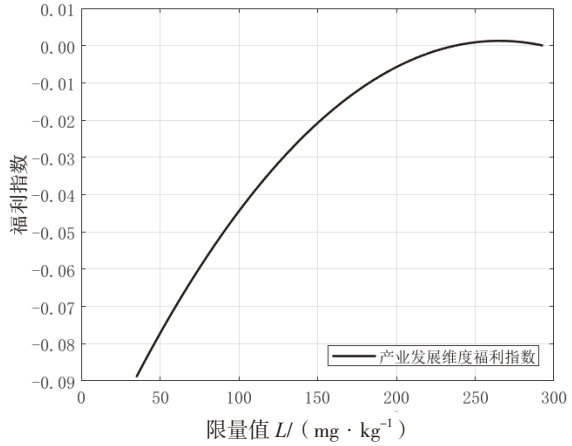
限量	34.97	51.26	80.73	159.31	293.03
福利指数	−0.022 6	−0.012 3	0.002 2	0.018 9	0



（a）健康安全维度福利指数曲线



（b）生态环保维度福利指数曲线



（c）产业发展维度福利指数曲线

图3 不同限值下3个维度的福利指数曲线

综上，限量水平变化会通过多种路径产生影响。在健康安全方面，限量越严格，甲酰胺暴露风险降低，健康风险得分下降；在生态环保方面，限量变化可能影响生产工艺，进而影响排放量；在产业发展方面，限量越严格，企业改造成本可能增加，但消费者支付意愿可能提升。这种多维度、多路径的影响机制，使得构建科学的评价指标体系尤为重要。

为确保效度评估的专业性与权威性，本研究制定了严格的专家遴选标准：入选专家需具备副高级及以上职称，或拥有硕士学位且从事环境健康、产业经济、化学品风险管理等相关领域工作不少于8年，并有参与政策咨询或标准制定的经验。邀请到的20位专家，涵盖正高级职称8人（40.0%）、副高级职称9人（45.0%）及资深中级职称3人（15.0%）；所属机构类型包括高校/科研院所（60.0%）、政府监管部门/标准化机构（20.0%）、行业检测/认证机构（15.0%）及企业研发部门（5.0%）。咨询问卷的回收率为100%，确认不存在任何可能影响本研究客观性与公正性的商业或利益冲突关系。

3.2 方法内容效度评估

在内容效度方面，专家依据“指标—目标双向细目表”及四点量表，对13个三级指标的必要性（与“甲酰胺限量影响”目标的相关性）进行了独立判断，结果见表7。判断结果显示，绝大多数指标被专家认定为“强关联项”（即打分为“1=完全相关”或“2=密切相关”）。

共有190组专家组合，其内容效度指数的直方图如图4所示，最小值为0.846，最大值为1.000，取其平均值0.954代表整体方法的内容效度系数。因此，本研究的内容效度为0.954，则内容效度得分为95.4%，评级为“优秀”，表明从量化角度看，整个指标体系具有优秀的内容效度。

3.3 方法结构效度评估

在结构效度方面，表8中专家对一级指标独立性的打分结果显示，健康安全—生态环保维度配对平均得分为4.55分，健康安全—产业发展

表7 三级指标内容效度专家判断统计

三级指标	判断为“强关联”的专家数量	占比/%
非不良反应风险	20	100
不良反应风险	20	100
诊断检查费用	20	100
疾病治疗费用	19	95
全产业链废气排放量	19	95
全产业链废水排放量	19	95
全产业链固废排放量	19	95
全产业链碳排放量	20	100
企业成本	19	95
产业链结构风险	20	100
政府管理费用	19	95
消费者推荐意愿	20	100
消费者支付意愿	20	100

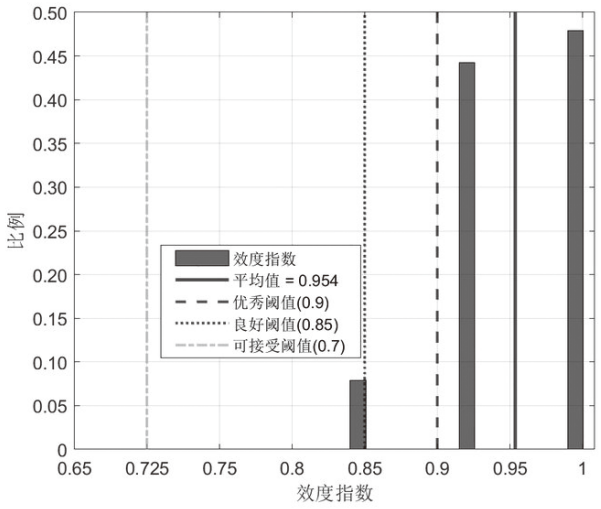


图4 内容效度的直方图

维度配对平均得分为4.20分，生态环保—产业发展维度配对平均得分为4.35分。整体结构效度为4.366，则得分为（4.366/5）×100% = 87.3%，评级为“良好”。结果表明，3个一级维度划分清晰、独立性良好。健康安全—产业发展维度得分相对较低，说明这2个维度之间存在一定关联性，但仍在可接受范围内。

表8 专家匹配得分统计

维度对	5分	4分	3分	2分	1分	平均分
健康安全—生态环保	11	9	0	0	0	4.55
健康安全—产业发展	7	10	3	0	0	4.20
生态环保—产业发展	8	11	1	0	0	4.35
综合						4.366

3.4 方法关联效度评估

为检验指标体系的关联效度，本研究以甲酰胺的5个限量水平为自变量，以前述公开文献参数推演的示意性的福利指数为因变量，进行二次函数拟合，结果如图5所示。根据图5和二次函数回归结果表9可知，F值为179.93，R^2为0.988 9，P值为0.005 5。P值小于0.05，表明二次回归模型在统计上高度显著，方差分析的效应量 $\eta^2 = 0.994$，表明限量水平解释了综合评价值变异的绝大部分，效应极强。则关联效度得分为99.4%，评级为“优秀”。

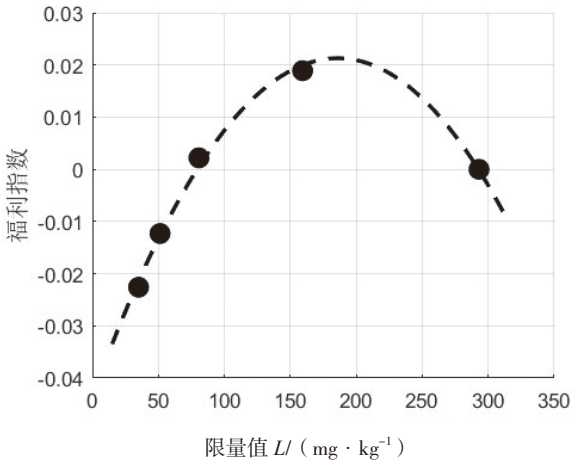


图5 福利指数的二次函数拟合

表9 二次函数回归结果

变量	系数	标准误	t值	P值
常数	−0.043 491	0.002 278 3	−19.09	0.002 733
L	0.000 696 8	0.000 039 1	17.81	0.003 137
L^2	−0.000 001 9	0.000 000 12	−16.16	0.003 809

该结果反映了“限量趋严对健康安全和产业发展的影响存在权衡”的理论模型。具体表现为：当限

量从293.03 mg/kg加严至159.31 mg/kg时，因健康风险显著降低，综合得分上升（福利改善）；但当限量进一步加严至80.73 mg/kg和51.26 mg/kg时，企业为满足更严标准而产生的改造成本、技术升级费用等负面效应开始凸显，并可能超过健康风险的边际改善收益，导致综合得分转而下降。这种先升后降的非线性趋势更符合复杂社会经济系统的实际情况。

3.5 综合效度评价

根据表10中展示的三维效度得分情况，采用算术平均法计算综合效度得分：

$$C=\frac{1}{3}(95.4\%+87.3\%+99.4\%)=94.03\%$$

依据分级标准，综合效度得分94.03%≥90%，因此本案例中针对甲酰胺限量影响所构建的评价方法的综合效度最终判定为“优秀”。这一结论表明，该评价方法整体上科学、可靠，能够有效用于甲酰胺不同限量政策的比较评估。

表10 效度评估结果

效度	量化得分	百分制得分/%	等级判定
内容效度	0.954	95.4	优秀
结构效度	4.366（5分制）	87.3	良好
关联效度	0.994	99.4	优秀
综合效度	0.940 3	94.03	优秀

从三维效度分值可以看到，各项效度维度表现存在差异。

1）结构效度（87.3%）相对偏低，是制约综合得分进一步提升的关键因素。这反映了评价对象固有的经济学属性。在“健康安全”与“产业发展”两大维度之间，存在着客观且难以剥离的内生性关联与成本—收益权衡。甲酰胺限量的加严虽然在理论上降低了儿童群体的健康暴露风险，但在现有生产技术条件下，往往直接引致原材料升级、工艺改造及检测认证等合规成本的跃升。这种此消彼长的动态博弈关系，导致2个维度在统计学上难以实现完全的线性无关。专家在评分时敏锐地捕捉到了这种现实经济逻辑，从而在维度独立性判断上给出了相对保守的评分。本研究通过专家评分及与国际国内权威指南的逻辑对照，验证了三维结构的合理性。然而，专家评分本质上仍属主观判断，可能受专家对概念熟悉程度及个人认知偏好的影响。因此，在构建跨学科的综合社会经济影响评价体系时，指标体系优化应更注重维度间交互效应的建模与解释。未来研究应在收集真实评价案例数据的基础上，运用客观统计手段，进一步验证维度结构的稳定性和独立性，以提升结构效度的证据强度。

2）关联效度（η^2=0.994）表现卓越，验证了该方法作为政策评估工具的核心价值。0.994的效应量表明，限量标准这一政策变量能够解释综合评价值近99.4%的变异，证实该框架具备极高的灵敏度与政策响应能力。二次回归模型不仅精准捕捉到了限量变动引发的社会经济涟漪效应，还拟合出了“先升后降”的倒U型曲线，这与理论预期中“过严标准导致边际收益递减甚至负增长”的规律高度吻合。然而，这一拟合结果在很大程度上得益于数据生成机制的理想化设定。本研究采用的5个限量水平下的指标数据，是基于公开文献参数与生产工艺模型推演生成的示意性数据。在真实的复杂市场环境中，消费者偏好的随机漂移、供应链上下游的博弈扰动以及宏观经济周期的波动等噪声因素，极有可能稀释限量标准对综合评价值的解释力。因此，当前的高关联效度更多代表的是一种“理论上的可达性”，其在实际政策场景中的稳健性尚需真实世界数据的进一步校验。

3）内容效度（95.4%）的优异表现为整个评价体系奠定了坚实的基石。这表明指标体系的设计全面覆盖了健康安全、生态环保与产业发展的关键要素，得到了领域专家的广泛认可。尽管本研究采用了标准的CVI指数对专家判断进行了量化，但内容效度本质上仍属于主观性评价。专家的学术背景、对行业现状的理解深度以及个人偏好，仍可能在一定程度上影响对指标“相关性”的判断。未来研究可结合认知访谈等方法，进一步校准指标措辞与测量意图的匹配度，从而降低主观偏差。

结合内容效度、结构效度和关联效度三者来看，本研究的评价方法在“指标设计”上表现优异，能实现“政策响应”，但在“维度解耦”上受制于客观经济规律。综合而言，该评价方法已具备一定的科学水准，但在将其转化为实际政策工具时，仍需警惕因结构耦合带来的解释复杂性。

4　结 论

本研究旨在解决化学品限量社会经济影响评价领域缺乏效度验证工具的问题，主要工作与结论如下：1）构建了一个系统、可量化的三维效度验证框架。将“方法效度”操作化为内容效度、结构效度与关联效度3个可检验的维度，为同类研究的规范化构建与质量评估提供了初步验证的分析工具和明确的技术路径。2）提出了以各维度效度得分的算术平均合成综合效度的方法，弱化了单一维度缺陷对整体评价的过度惩罚，能够更客观、全面地反映评价方法的整体性能。3）以儿童用品中甲酰胺限量评价为方法效度评估对象实证了框架的实用性。结果表明，该框架能够系统指导从数据模拟、多维效度计算到综合等级判定的全流程评估。针对该评估对象，最终测得其综合效度得分为94.03%（≥90%），等级为“优秀”，这为其用于甲酰胺限量决策支持的可信度提供了直接证据，同时也验证了本效度验证方法自身的可行性与价值。

尽管本框架在理论构建与初步应用中表现稳健，但仍存在以下局限：1）数据理想化。当前指标数据基于文献参数推演，缺乏全链路实测数据的支撑，且关联效度分析仅设5个样本点，统计功效受限。2）主观依赖性。内容与结构效度的量化高度依赖专家判断，虽经严格遴选，仍难以完全剔除认知偏差。3）案例单一性。仅验证了“地垫–甲酰胺”单一组合，尚未在多品类、多化学物场景下检验其通用性。

基于此，未来研究应着力于2个维度的拓展：第一，多场景鲁棒性测试。将框架迁移至其他高风险化学应激源及不同材质儿童用品的限量评估中，验证其跨场景适应性。第二，现实数据融合。基于行业协会统计、电商消费大数据及环境监测实测值，构建数据驱动的动态效度评估模型，降低模拟假设带来的偏差。

综上所述，本研究构建的效度验证框架，为化学品限量社会经济影响评价领域提供了从方法构建、效度评估到结果报告的科学范式和实用工具，有助于推动相关研究与实践的规范化、标准化进程，从而为基于证据的化学品风险管理决策提供更坚实的支撑。

参考文献

[1] Pearce D W, Atkinson G, Mourato S. Cost–Benefit Analysis and the Environment: Recent Developments[M]. OECD, Paris, 2006.

[2] GB/Z 24785—2009 化学品限制的社会经济分析指南[S].

[3] 周荣喜，杨跃翔，孙榛，等. 消费品中有害化学物质限量定值标准研究[J]. 标准科学，2020（12）：115–120.

[4] 胡雯娴，严江涛，刘庚慧，等. 消费品中化学物质限量影响的评价体系构建研究[J]. 标准科学，2025（2）：117–122.

[5] McMillan J H. Student perceptions of assessment[M]// Brown G T L, Harris L R. Handbook of human and social conditions in assessment. New York：Routledge，2016：221–243.

[6] 赵斌，张梦娇，周闻哲. 莫道桑榆晚：基于个体优势的老年员工工作重塑概念维度与效应检验[J]. 管理评论，2026，38（2）：197–211.

[7] 郝旭光，张嘉祺，雷卓群，等. 平台型领导:多维度结构、测量与创新行为影响验证［J］. 管理世界，2021，37（1）：186–199

[8] 彭妍，金胜勇，李小北. 逻辑与规范：图书情报学实证研究效度及其检验[J]. 情报理论与实践，2026，49（3）：77–83.

[9] 李彦，蒋佳.“拔尖计划”对学生创新能力的识别效度

与公平性再审视[J]. 中国高教研究, 2025(11): 57–65.
[10] 蒲可喻, 贝冬丽, 李虹漫, 等. 中文版青少年关怀能力问卷在中学生中的信效度检验[J]. 中国临床心理学杂志, 2026, 34(1): 96–101, 112.
[11] 陈曦, 胡中锋. 基于DeepSeek的智能评分: 效度、信度与可行性研究[J]. 高教探索, 2025(3): 62–67.
[12] 刘晶晶, 吴周玥, 王侨, 等. 双维创业韧性: 理论构建、量表开发及效度检验[J]. 科技进步与对策, 2026, 43(4): 24–33.
[13] 姜金栋, 魏萍. 中国仲裁员胜任力量表的编制与信效度检验[J]. 社会科学研究, 2025(6): 156–165.
[14] 唐雅君, 黄守桢, 许世梅, 等. 青少年父母情感支持量表的编制及效度和信度检验[J]. 中国心理卫生杂志, 2026, 40(3): 249–256.
[15] 吕兴洋, 杨雨晴, 艾进, 等. 旅游目的地老化: 概念辨析、量表开发与检验[J]. 旅游学刊, 2026, 41(1): 160–174.
[16] Pieschl S. Metacognitive calibration: an extended conceptualization and potential applications[J]. Metacognition and Learning, 2009, 4(1): 3–31.
[17] 马亚龙, 张伟, 朱敏洁, 等. 评估方法的效度分析[J]. 系统工程与电子技术, 2007(6): 907 –908, 925.
[18] 孙明, 刘星伯, 裴宏. 系统评估效度分析的定量化方法[J]. 装甲兵工程学院学报, 2008(3): 86–90.
[19] 夏婷, 赵祖地. 高校德育评估效度的鉴定方法研究[J]. 辽宁行政学院学报, 2014, 16(12): 100– 102.
[20] 徐鹰, 邓雅玲. 大规模语言测试翻译任务内容效度研究: 以CET为例[J]. 山东外语教学, 2023, 44(3): 26–37.
[21] 陈佳佳, 蔡毅媛, 杨薇, 等. 机构实施循证实践准备度评估量表的汉化及内容效度验证[J]. 协和医学杂志, 2025, 16(3): 765–776.
[22] 曾守锤, 徐馨怡, 张欢欢. 社会工作师职业水平考试的内容效度研究[J]. 社会工作与管理, 2024, 24(3): 12–21, 68.
[23] Wu H, Zhang Y, Zhi Y, et al. A comprehensive review on the application of socio–economic analyses in chemical management: challenges and opportunities[J]. Environmental Science & Policy, 2024, 154: 103694.
[24] ECHA.Guidance on the Preparation of an Annex XV Dossier for Submitted for Authorisation or Restriction–Socio –Economic Analysis [R].Helsinki: European Chemicals Agency, 2015.