北京大学
中国教育财政科学研究所

简　　报

怀仁怀朴　唯真唯实

2014 年第 3 期（总第 91 期）　　**2014 年 6 月 24 日**

《幼儿学习环境评量表（修订版）》在我国使用的信、效度研究

黄晓婷　张眉　关可心*

摘要：近年来，随着我国学前教育总体规模的不断扩大，学前教育的质量问题也逐渐受到越来越多的关注。本文介绍了一套成熟的评价幼儿园整体质量的量表，即《幼儿学习环境评量表（修订版）》（ECERS-R），及其在我国西部某市 6 所幼儿园进行试测的情况。通过分析试测数据，本研究对 ECERS-R 在我国使用的信、效度做了初步的验证，探讨了这一量表在我国的适用性和使用价值。最后，本文详细分析了量表中各题项的得分情况，试图对我国幼儿园管理和教学工作的进一步改善提出相关建议。

关键词：学前教育　质量评价　ECERS-R　学习环境评量

* 作者简介：黄晓婷，北京大学中国教育财政科学研究所助理研究员；张眉、关可心，北京大学中国教育财政科学研究所科研助理。

一、研究背景

学前教育的意义和作用已经在世界范围内得到了广泛的认可[1]。我国自2010年实施“学前教育三年行动计划”以来，学前三年毛入学率、幼儿园数量、在园幼儿规模都实现了高速增长。随着政府对学前教育财政投入的持续增加，如何保证和提升保教质量，成为学前教育进一步发展的重要挑战。面对这一变化，我国多个省市制定了幼儿园评估标准，但这些地方性评估标准往往偏重幼儿园的硬件设施，忽略教学过程等软环境，或者是在可操作性、规范性及评价的时代性等方面不尽如人意[2][3][4]。过去的相关研究也存在类似的问题，少数现有的评价工具信、效度未经检验或不够高[5][6]。因此，我们需要科学有效的评价工具，来了解幼儿园质量的现状，特别是过去知之甚少的过程性质量[7]，从而促进学前教育质量的提升。

国际上对学前教育质量评价的研究已经有了一定的积累和经验[8][9]。其中，美国的Harms等研发的《幼儿学习环境评量表》（ECERS）最早于1980年发布，1998年又出版了修订版（以下简称“ECERS-R”）。该量表目前已经在美国、加拿大[10]、澳大利亚[11]、加勒比地区[12]、台湾[13]等多个国家和地区得到广泛应用，并有多项研究支持其在信、效度上的良好表现[14]。该评量表不仅关注幼儿园的物理环境，更着重强调了教学实施中活动安排、师生互动等多项对儿童发展有重要影响的因素。

这一量表近年来也逐渐引起了我国学者的注意[15][16][17]。例如，胡碧颖等[18]于2008年在北京市运用ECERS-R评量表，对7所幼儿园共计40个班级进行了试测。研究者们在评价实践中发现了量表中部分指标的不适用，存在一定与我国文化特点不符合的情况。但这些现有的研究较少深入探讨量表在我国使用时总体及子维度的信度和效度，且幼儿园类型比较单一，因此我们对ECERS-R在我国的适用性及其对我国幼儿园管理、教学工作的作用仍然缺乏细致详尽的了解。

本文详细介绍了ECERS-R量表及其在我国西部某市6所幼儿园进行试测的情况，并深入分析了该评价工具在试用中的信、效度，探讨其在我国的适用性和使用价值。研究还通过详细解读样本幼儿园在各个题项上的得分情况，对幼儿园在管理和教学中可能存在的问题提出改进建议。本研究旨在积累ECERS-R量表在中国应用的经验，同时也为我国学者研发更适合国情的学前教育质量评价工具

积累经验。

二、评价工具:《幼儿学习环境评量表(修订版)》(ECERS-R)

ECERS-R 由美国北卡罗来纳州立大学儿童发展中心的 Harms 教授研究团队经过几十年的积累研发而成。该量表是目前学前领域最具权威的评价工具之一,在美国和其他多个国家被广泛用来对幼儿园的总体质量,包括结构性质量和过程性质量,进行全面的评价。[19]

该量表包含 7 个子维度,即空间和设施、个人日常照料、语言一推理、活动、互动、作息结构及家长和教师。具体来说,空间和设施子维度主要评价幼儿园的物理环境,包括室内空间的面积和规划,日常学习或游戏用的设施、休息的空间和设施,户外活动的空间和设施等。个人日常照顾子维度着重评价教师在幼儿在园期间的卫生、健康和安全实务领域的工作。语言—推理子维度则对幼儿园在教学中对儿童语言发展方面的培养进行了评价。活动子维度主要评价幼儿园室内、外各种活动的计划性和多样性。互动子维度包括了在教学和活动中师幼间、幼儿间信息交换的质量。作息结构子维度针对幼儿园课程安排中的时间规则,评价教师面对不同个人需求的灵活性。最后,家长与教师子维度评价了家校互动、教师间互动合作,幼儿园对教师的个人需要、专业成长需要的供应,以及幼儿园对教师的督导与评量等多个方面。

这 7 个子维度分别包含 4-10 个题项,整个量表共包括 43 个题项。量表采用 7 点评分方式,1 分表明幼儿园在该题项评价的方面“未达到最低要求”(inadequate),3 分表明“仅满足最低要求”(minimal),5 分或以上为“表现良好”(good),7 分为“优秀”(excellent)。

量表的正确使用依赖于经过严格训练的评分员。ECERS-R 的作者们开发了一套标准化的评分员培训教材,主要包括量表解读、案例分析和评分练习三个部分。完成培训的评分员,需对每个参与评价的班级进行至少 4 小时的观察,随后完成评分。

由于本研究的评价实践进行时,还没有正式出版的简体中文版本,我们在台湾国立台东大学郭李宗文、陈淑芳翻译的版本(2006 年出版)的基础上,进行了中文的繁简体转换和习惯用语的修改。

三、研究方法

我国西部某市 6 所幼儿园（4 所城市园和 2 所农村园）的 45 个班级参与了此次评价实践。8 名评分员一起学习和探讨了与量表配套的标准化培训课程。在评价过程中，每个评分员一次只评价一个班级，每个班级观察 4 小时。对于不能直接观察到的信息，观察员需要安排一定的时间对幼儿园教师进行访谈。

本研究采用项目反应理论（IRT, Item Response Theory）中的多维度分部计分模型（Multidimensional Partial Credit Model，MPCM）[20]对评价数据进行分析，并验证量表的信度和效度。MPCM 适用于多维度量表中每题有多个分值、难度随分值的增加而增加的情况，常常被运用于各类包含多个子维度的测试和问卷（如托福、GRE、PISA 等）。专业软件 ConQuest[21]可以用来分析数据，除了每题的难度外，研究者可以同时得到量表总体和各子维度的信度系数。信度越高，表明评价结果的误差越小。

此外，本研究还请 8 位评分员对 2 所幼儿园的 16 个班级进行了交叉评分，即每个班级都由两位评分员分别进行了评分，用以计算评分员的内部一致性[22]。评分员一致性指的是采用多个评分员对特定题目进行评判的一致性。目的是评估题目既定特征的可信程度。评分员一致性良好，是保证评价结果有效的必要条件。

除信度外，效度是量表的另一重要性质。效度反映了量表在多大程度上实现了其测量评价的目的。效度研究不是单一的一个问题，而是包括内容效度（content validity，即量表的内容选择的适当性）、内部效度（internal validity,即问题设计的合理性）、外部效度（external validity，即与其他相关因素的关系的合理性）和影响效度（consequence validity，即使用该测量结果作出推断的科学性）等多个方面[23]。ECERS-R 是其作者们对学前教育质量理论和实践研究多年的产物，量表的内容效度已从多个角度进行了论证[19]。本研究针对量表的内部效度，对比了农村园与城市园的各部分的得分，看其差异是否符合一般认识。此外，本研究还通过教师问卷和教师对儿童发展多个方面的评价表，收集了相关教师背景和儿童发展情况的信息。这些信息和各个班级在 ECERS-R 中各部分得分的关系，成为量表外部效度的有力佐证。由于 ECERS-R 在本研究中未用来进行任何奖惩决定或其他判断，因此本文没有涉及影响效度的论证。

最后，我们还对 ECERS-R 的题项内容进行了逐条分析。参与研究的幼儿园

大多数得分为“未达到最低要求”或“仅满足最低要求”的题项，可能揭示了该市幼儿园在这些方面存在普遍性问题，或是这些题项不适合该市的实际情况。而样本幼儿园大多数得分为“表现良好”或“优秀”的题项，则可能彰显了该市幼儿园的特点和长处，或可能是这些题项对该市幼儿园缺乏区分度。本文根据这些题项的具体内容，逐一进行分析和推论。

四、研究结果

1. 量表的信、效度

ECERS-R 的 43 个题项中，有一个评价幼儿园对特殊儿童需要的供应的题项。由于本研究的样本中，仅有一个班级有特殊儿童，该题缺失数据太多，没有纳入下面的分析。表 1 显示了 45 个班级在余下 42 个题项的总平均分和各个子维度的平均分。结果显示，整个量表的总平均分为 4.08，7 个子维度的平均分从 3.56 到 4.79 不等。因此，整体来说，样本幼儿园目前的质量已满足最低要求，但尚未达到良好水平；在互动领域的水平较高，但在活动安排的多样性和灵活性方面相对不足。

表 1 ECERS-R 总平均分及各子维度平均分

	ECERS-R 整体	7 个子维度平均得分						
		空间和设施	个人日常照顾	语言推理	活动	互动	作息结构	家长与教师
平均分	4.08	3.71	3.90	3.96	3.56	4.79	4.10	4.51
标准差	1.21	1.42	1.00	1.66	1.14	1.34	1.76	1.27

ConQuest 运行的多维度分部计分模型分析结果显示，ECERS-R 的总体信度系数（同α系数）高达 0.95。表 2 显示了各子维度的信度。除“个人日常照顾”外，其他子维度的信度系数都较高。

表 2 ECERS-R 整体及各子维度信度

	整体信度	空间和设施	个人日常照顾	语言推理	活动	互动	作息结构	家长与教师
信度系数	0.95	0.89	0.54	0.86	0.85	0.79	0.77	0.74

通过分析交叉评分的数据，我们发现在此次评价中，评分员的内部一致性偏低，为 0.53。在此类通过实地观察对行为表现等进行评价的研究中，评分员的内部一致性往往出现偏低的现象。在未来使用 ECERS-R 的评价实践中，我们还应该通过延长培训时间、增加评分练习等方式，提高评分员对评价标准的把握。

此外，本研究通过多个证据验证了 ECERS-R 的内部效度和外部效度。表 3 总结了城市园与农村园在各个子维度的得分，从中我们不难发现城市幼儿园在各个领域的得分都高于农村园。其中差异最大的两个维度是“作息结构”和“空间和设施”。农村园硬件较城市园落后是不争的事实。同时，由于农村园师幼比往往更低，导致教师在时间分配上无法照顾到个体需求，这一情况与评价结果显示的两类幼儿园在“作息结构”上的差异相呼应。

表 3 城市园与农村园的得分对比

	空间和设施	个人日常照顾	语言推理	活动	互动	作息结构	家长与教师
城市幼儿园	4.26	4.29	4.33	3.77	5.29	4.70	4.49
农村幼儿园	1.86	2.76	2.02	2.48	3.45	1.83	2.70
得分差异	2.40	1.53	2.31	1.29	1.84	2.87	1.79

最后，研究发现，ECERS-R 量表中园所对教师个人需求提供和专业成长机会的得分与教师问卷中教师对工作环境的满意度（1-非常满意，5-非常不满意）呈显著正向影响作用（B=-.39, t=-4.35; B=-.33, t=-.24）。ECERS-R 量表中的作息结构得分与儿童健康和体能发展有显著的影响（B=.27，t=3.90）；园所的活动安排得分对儿童的社会心理（B=.52,t=3.68）、语言阅读（B=.31,t=3.77）及数学科学（B=.17,t=2.90）等多个方面都有显著正向影响；园所的“语言-推理”得分与儿童的语言阅读能力也理所当然的呈现了显著正向影响（B=.15,t=2.85）。这些结果有力佐证了 ECERS-R 量表在中国使用时同样具有外部效度。

2. 对题项内容的质性分析

为进一步深入探讨 ECERS-R 的适用性和使用价值，本研究统计了三分之二以上幼儿园被评为“未达到最低要求”或“仅满足最低要求”的题项。共有 11 题出现了这种情况。其中有一些反映了我国目前大多数幼儿园仍无法满足的质量要求，如罕有满足幼儿个体化需求的隐秘空间、柔软性设施；缺少供幼儿游戏的

多样化设备、沙/水玩具；此外，样本园普遍存在对纪律的过度强调、过多以团体为单位组织的活动的问题，也导致了个别题项整体得分偏低。此外，统计结果反映出对教师个人需要供应不足等，也是样本幼儿园亟待解决的问题。

另外有少数题项是因为与我国国情不符合，导致了样本园整体得分偏低。如允许家长接送时进入教室与我国大部分幼儿园的安全规定不符合。又如“自由游戏”中，样本园所在地教育管理部门对自由游戏的时间有明确的规定，因此实际情况与题项要求不符。还有“鼓励接纳差异”一题出现了由社会文化差异导致的不适用性，该题要求在教材教具上具有明显可见的种族及多元文化上的多样性，与美国多种族文化背景不同，我国大部分地区的教材教具上没有明显可见的种族及文化上的多样性。

除低分题项外，我们还统计了 95%以上样本园被评为“表现良好”或“优秀”的题项。这 10 个题项主要包括基础设施设备的提供和修缮、安全监护、教师的责任划分、对教师的督导评价和教师在职培训几个方面。由于 6 所样本园，特别是 4 所城市园均属于该市优质幼儿园，因此这些题项是否对我国幼儿园缺乏区分度还需要进一步实验和探讨。

综上所述，ECERS-R 虽然有少量题项不符合我国幼儿园的实际情况，但主要的评价内容能帮助研究者和实践者们发现目前学前教育发展中的长处和问题，在我国也极具使用价值。

五、结论和探讨

本研究使用 ECERS-R 对 6 所幼儿园 45 个班级进行了评价。分析结果表明，ECERS-R 在这次试用中展现了良好的信度和效度。深入的题项内容分析发现样本园在教师的督导与评量、安全实务、专业成长的机会、幼儿之间的互动、教师的互动与合作这些方面得分较高；而在隐秘空间、沙/水设施、休息和安抚设施以及对教师个人需要的供应方面得分较低。大部分得分高的项目是较容易通过硬件投入就能实现的，而更多体现教育理念的“软性”质量指标则得分偏低。

ECERS-R 评价标准中贯穿的价值判断是当今学前教育领域所倡导的以儿童为中心、尊重儿童的个性成长的理念。而我国幼儿园在过去的实践过程中更重视课堂目标的实现和管理的便利性。如何以儿童发展为中心，安排和实施幼儿园的

活动，对很多幼儿园来说仍然需要逐步摸索。ECERS-R 在评价幼儿园质量的同时，也为幼儿园如何改进工作提供了参考和线索。

在此次评价实践中，我们为在中国使用 ECERS-R 积累了一定经验。评分员一致性不够高是本研究最大的局限所在。在未来的评分员培训中，我们建议延长培训时间，增加中国本土的评分实例。此外，还需要在培训过程中明确量表中涉及的概念，量化一些较为模糊的描述（如“足够的设备”等），来统一评判标准，提高评分员的内部一致性。

最后，未来研究还可以调整、修改或删除不适用题目，增加样本量和幼儿园类型，以提高研究结论和政策建议的可推广性。

参考文献

[1] OECD. (2012).Starting Strong III: A Quality Toolbox for Early Childhood Education and Care. 2012, Paris: OECD.

[2]朱家雄. 从幼儿园评估的角度谈理论与实践的脱节[J]. 幼儿教育，2009(5).

[3]康建琴, 刘焱. 制定幼儿园评估标准需要澄清的几个问题[J]. 学前教育研究，2011(1).

[4]刘霞. “托幼机构环境评价量表(修订版)”述评[J]. 教育导刊，2004(2、3).

[5]郭良菁. 上海市幼儿园教师专业发展自我评价体系研制简介[J]. 学前教育研究, 2007(07-08).

[6]虞永平. 幼儿园教学活动的评价[J]. 早期教育：教师版，2005(3).

[7]黄晓婷, 宋映泉. 学前教育的质量与表现性评价——以幼儿园过程性质量评价为例[J]. 北京大学教育评论，2013(1).

[8] Hamre, B. K., Mashburn, A. J., Pianta, R. C., Lacasle-Crouch, J., & La Paro, K. M. (2006). Classroom Assessment Scoring System Technical Appendix.2006.

[9] Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. Early Childhood Research Quarterly, 19, 375-397.

[10] Perlmana, M. Zellman,G., Leb,V.(2004).Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R)，Early Childhood Research Quarterly Volume 19, Issue 3, 3rd Quarter 2004, P.398–412.

[11] Ishimine,Karin (2010). Quality of Australian childcare and children's social skills. International journal of early years education，ISSN 0966-9760.

[12] Lambert, Michael Canute (2008). "Are the indicators for the Language and Reasoning Subscale of the Early Childhood Environment Rating Scale-Revised psychometrically appropriate for Caribbean classrooms?". International journal of early years education (0966-9760), 16 (1), p.41.

[13]傅馨仪. 运用 ECERS-R 提升幼儿园品质之研究[D]. 台湾国立政治大学硕士论文，2010.

[14]李克建, 胡碧颖. 国际视野中的托幼机构教育质量评价——兼论我国托幼机构教育质量评价观的重构[J]. 比较教育研究，2012(7).

[15]郑晓边. 学前教育机构环境质量评价--Harms Clifford 环境评价介绍[J]. 教育研究与实验，1988(4).

[16]马洁然. 美国儿童早期教育环境评估量表介评[J]. 早期教育，2011(7-8).

[17]陈香艳. 托幼机构环境评价量表（修订版）简评[J]. 基础教育研究，2012(9).
[18]胡碧颖, 朱宗顺. 美国《幼儿学习环境评量表》及其在中国的初步应用[J]. 幼儿教育，2009(33).
[19] Harms,T.,Clifford,R.M. ,Cryer,D. (2004) .Early Childhood Environment Rating Scale.(Revised Edition) (ECERS-R).Teachers' College Press.
[20] Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
[21] Wu, M., Adams, R. J., & Wilson (1998). ConQuest. Hawthorn, Australia: ACER Press.
[22] Fleiss,J. L., B. Levin, et al. (1981). "The measurement of interrater agreement." Statistical methods for rates and proportions 2: 212-236.
[23] Wilson, M. (2005). Constructing Measures: An Item Response Modeling Approach. Mahwah, New Jersey: Lawrence Erlbaum Associates.

上期回顾

2014 年第 2 期（总第 90 期）

学前教育公共财政投入、事业发展趋势及公平性挑战

摘要：自 2010 年中国政府开始加大学前教育财政投入以来，学前教育事业发生了显著变化，同时也面临着严重的公平性挑战。本文首先概述了 2010 年以来中央和地方政府对学前教育的财政投入，其次展示了学前教育在学前三年毛入园率、幼儿园规模、在园儿童规模、公民办供给及结构质量指标等方面的变化趋势，并剖析了学前教育面临的公平性挑战。最后，本文提出了改进学前教育不公平现状的相关政策建议。

本《简报》由北京大学中国教育财政科学研究所主办；旨在反映本所最新的学术科研活动；相关内容仅体现作者本人观点，并不必然代表本所的立场。

简报内容仅供参考，如需转载须事先征得本研究所同意。

本期印发：1800 份

下载网址：http://ciefr.pku.edu.cn

主办单位：北京大学中国教育财政科学研究所

电子信箱：newspaper@ciefr.pku.edu.cn

责任编辑：毕建宏

电　　话：010-6275-9700

地　　址：北京市海淀区颐和园路 5 号
北京大学教育学院楼 413 室（100871）

传　　真：010-6275-6183