

The Starting Line: Developing a Structure for Teacher Ratings of Students' Skills at Kindergarten Entry

起跑线：进入学前班的学生技能教师评定量表的结构探讨

Jessica Goldstein & D. Betsy McCoach

University of Connecticut

摘要

为了评价幼儿和评价学习课程，我们需要发展适宜性、可靠的心理测验工具。在美国，在设计和使用能够涵盖多种幼儿发展技能的大规模评估工具方面的指导较少。在 2005 年和 2006 年，康涅狄克州通过了一项立法，要求在全州实施能够“测量儿童”进入学前班时入学准备水平的，发展适宜的评估工具。在这一立法的号召下，康涅狄克州教育部门设计了《学前班准入评定》。这一评定的设计以教师的观察为基础，提供了反映本州儿童在学前班（为 5-6 岁儿童提供的机构教育，美国设置在公立学校内）阶段初期幼儿表现出的技能的快照。本文运用一系列的探索性和证实性因素分析，调查了一个较大的城市学区中儿童进入学前班时教师对幼儿技能的评定。分析显示，学前班儿童入学准备的评估应该注重以下技能：表达性语言、接受性语言、对故事的反应、对书籍的熟悉程度、对字母的熟悉水平、早期书写、数数、形状和模式、测量、精细动作技能、矛盾解决、社会性参与、自我选择活动的参与和创造性技能。

引言

虽然在 2001 年的《不让一个孩子掉队法案》(NCLB)中已经明确规定一年一度的测评要求要从三年级开始，美国的教育者已经开始重新强调小学阶段教育的

重要性，因为它是未来学习阶段的基础。对年幼儿童教育发展的测查是任何一个综合测评体系的重要组成部分，但是它与年长儿童运用的测查方案不同。小斯科特，卡根和克利福德 (Scott-Little, Kagan, & Clifford, 2003) 提议，年幼儿童与年长儿童相比，运用更为片段式的方式进行学习，并且需要多种测评方式来获得对他们知识掌握的全面理解。对于年幼儿童来说，在一个时间点进行的单一测评不能正确地反映他们的发展。而且，全美幼儿教育协会 (NAEYC) 和美国教育部早期教育专家协会 (NAECS/SDE) 已经制定了早期测评的恰当运用范围。这些指导方针指出，早期测评的正确运用能够指导教育教学，鉴别可能需要集中干预的儿童，以及改善教育课程和发展干预手段 (NAEYC & NAECS/SDE, 2009)。从测查的角度来说，很显然，这些不同的主体需要不同的测评工具，对于这一群体的标准的测查手段至今还没有。

我们需要发展适宜的、可靠的心理测验工具来监控幼儿的发展，评价他们早期学习课程的有效性。但是在研究文献和实践中，几乎没有强调儿童的情感、认知和运动发展的大规模测评的设计和运用方面的指导。本文描述了在儿童进入学前班时教师依据州里的测查工具对学生技能实施测评的量表结构的实证调查。虽然本研究的结果对本研究工具的效度有重要意义，我们相信本测查的子量表得分能够被用来作为类似的用于评估学前班学生技能的工具的报告结构。

对学前班学生技能的理解

两个国家级数据库的建立以及对幼儿教学和评估日益增长的兴趣，引发了一些有关学生在进入学前班初期表现出来的技能的研究。美国教育部的国家教育统计中心 (NCES) 研发了一个名为儿童早期纵向研究-同龄群体 (ECLS-B) 数据库，聚焦于从出生到入学前班前这几年中孩子的健康、发展和教育。丹顿那根和麦克菲 (Denton Flanagan & McPhee, 2009) 发现，在进入学前班时，2001 年出生的孩子在阅读和数学知识技能的得分上会因为民族、家庭类型、贫困状况、主要的家庭语言和他们进入学前班前一年中最初的早期保育和教育环境等因素的不同而不同。尤其值得注意的是，白人和亚洲儿童比黑人、西班牙人、美洲印第安人和阿拉斯加当地儿童的阅读和数学测评得分更高。而且，家庭中有双亲、收入在贫困线和贫困线以上家庭的儿童、以及英语为主要家庭语言的儿童比同伴的阅读和数学得分也更高。作者也发现，在入学前班前一年参加过正规早期保育和教育机构

的儿童比那些没有去过早教机构的儿童在阅读和数学评估中得分更高。在精细动作技能方面也有类似发现；在精细动作技能评估中得分较高的儿童倾向于是男孩、白人或者亚洲人、家庭中有双亲、家庭收入在贫困线或以上、入学前班前一年有过正规的早期保育和教育的经历。

较早的一个类似研究，1998-1999年的儿童早期纵向调查-学前班班级(ECLS-K)，追踪了一个有全国代表性的22,000名学前班的儿童样本，从1998年秋季追踪到五年级。韦斯特，丹顿和吉敏楼-侯斯肯(West, Denton, and Germino-Hausken, 2000)报告了儿童进入学前班时的技能。在早期读写能力方面，66%的幼儿熟练认识字母，29%的孩子能够熟练理解词汇的起始音，17%的幼儿熟练理解词汇的末尾音。在数学方面，几乎所有的学前班孩子能够熟练区分数字和形状，58%的孩子熟练理解相对大小，20%的幼儿熟练理解顺序。关于社会技能方面是教师的评定，大约75%首次进入学前班的孩子能够接受同伴的建议，能够发展友谊。对于样本学生来说，教师的报告认为，71%首次进入学前班的孩子能够经常或者很经常坚持任务，75%的孩子很愿意学习，66%的孩子能够大部分时间集中注意力。拉斯和韦斯特(Rathbun & West, 2004)运用ECLS-K的数据描述了儿童从进入学前班开始到三年级期间阅读和数学方面的进步。

除了描述学生的技能，ECLS-K还提供了教师对学前班儿童准备状态的看法方面的资料。林，劳伦斯和歌瑞尔(Lin, Lawrence & Gorell, 2003)开展过这样一项研究。对于准备状态的定义，学前班老师倾向于更强调学校教育的社会性方面的需求，学业技能的发展方面其次。特别是，准备状态的定义主要集中于孩子的社会性行为，比如“说出需要和想法”，“不扰乱课堂”，“能理解指导”和“知道轮流和分享”。较少提到的社会技能是“能安静地坐和集中注意”，“完成任务”，“具有问题解决技能”和“对他人敏感”。在他们的研究中，教师一般较少地把学业技能包含其中，比如“能数到20或更多”，“知道大部分字母”“能说出颜色和形状”，以及“运用铅笔和毛刷”。

除了ECLS-K的设计以及随后的成果发表，其他早期教育专家也尝试在早期发展和学前班准备状态研究领域内使用一种共同语言。这些努力主要围绕着发展的多维视角，包括：（1）身体和运动发展；（2）社会和情感发展；（3）学习方式（如创造性、主动精神、学习态度、任务技巧等）；（4）语言；和（5）认知和综合性知识(Kagan, Moore, & Bredekamp, 1995; Love, 2001; Meisels, 1999)。这

一多维结构解释了家庭和早期教育机构对儿童发展的作用，并强调了孩子学习和成为群体中的一员的发展方向。学业知识仅仅被认为是广泛多样的技能体系中的一个方面。

研究表明，学前班教师支持这一观点。研究发现，公立学校学前班教师认为，学业准备最基本的前三个素质是：儿童身体健康、睡眠和营养良好；能够用语言交流需要和想法；以及对探索新的活动有热情和好奇心(Heaviside & Farris, 1993)。十年后进一步的研究证实，教师认为学前班的成功在于孩子健康、有社会感受性、能够与人交流和理解指令(Lin, Lawrence, & Gorrell, 2003; Wesley & Buysse, 2003)。其他研究显示，父母和幼儿教师比学前班教师更强调学业能力和基本的知识准备，如字母表中的字母(Harradine & Clifford, 1996; Hains, Fowler, Schwartz, Kottwitz, & Rosenkoetter, 1989; West, Germino-Hausken, & Collins, 1993)。

美国早期学习标准详细说明了进入学前班以前儿童的学习和发展的期望，国家层面的标准也可以看作是州层面对这个阶段儿童技能期望的一个概念基础。这些标准文件对于理解学前班准备状态非常重要，因为它们代表了从早期学习到正式学校教育之间的桥梁。小斯科特，卡根和弗瑞乐 (Scott-Little, Kagan, & Frelow, 2006) 对 2005 年一月份以前所有由州一级组织制定的 46 份早期学习标准文件的内容分析，发现州一级早期学习标准更关注语言和认知技能而不是其他领域。作者认为，这一重点可能因为是对儿童早期学习标准与 K-12 (学前班-12 年级) 标准一体化的努力，同时也因为将有更多学业知识内容放入早期阶段的学习中的倾向，因为它与未来几年的成绩有关。作者也测查了每一个领域所覆盖内容的深度和广度。在身体健康与运动发展领域，发现运动技能 (大动作、精细动作、口部的、感知觉的) 和机能表现以及自理技能是标准项目中的主体，而不是身体健全或者全面的健康。在社会情感领域，与同伴的社会交往技能是标准项目中经常反映出的指标类型。其他的指标包括，情感表达、自我概念和理解他人的情感。几乎没有州把关于发展与同伴和成人的关系技能和孩子的自我效能放入标准中。在学习品质领域主要有以下四个指标：(1) 反思和解释方式；(2) 对新任务和挑战的好奇心；(3) 创造和想象能力；和 (4) 主动性、任务坚持性和注意力。在语言和交往领域，有 16 个不同的指标都是口头语言或者早期读写技能。在认知和综合知识领域，几乎 80% 的认知标准项目都是关于物理世界的知识或者数理逻辑知识。

早期教育专家告诉我们，看儿童的视角必须是多维的，但是对于政策制定者和实践工作者来说，有关幼儿的知识技能测量方法方面的操作性指导很有限。来自 ECLS-B 和 ECLS-K 的数据提供了对儿童能力的看法，但是各个领域的评估和技能清单至今还没有，所以也不能使用。迄今为止，25 个州已经开始对进入学前班的准备状况进行评估，而且还有四个州已经对发展进行评估；大多是教师个人运用学生技能检查表(Stedron & Berger, 2010)来考察。但是，至今还没有相关的大规模的研究。进入学前班初始阶段学生技能的信息对于如何努力建立对进入学前班儿童的准备状况的期望以及孩子的学习准备是非常有必要的。本研究通过运用一个州的学生样本，致力于探究如何确立学生进入学前班时技能领域的结构。具体来说，我们运用探索性和实证性因素分析来考察教师评估量表的结构，努力将学生的技能进行分类。我们的目标是，描述一个学生样本在学前班初始阶段的技能。我们希望，这一努力将对当地的有关学前班准备性概念的研究和政策计划有启发作用。

研究方法

这一部分包括对工具和数据收集技术的总体介绍，研究对象和数据统计分析方法。

研究工具

在 2005 年和 2006 年，康涅狄克州通过立法，要求对进入学前班的儿童的入学准备水平实施全州范围内的评估。为了响应这一立法，康涅狄克州教育部门制定了学前班儿童入学准入技能清单。这一技能清单的设计以教师的观察为基础，旨在提供全州学生在进入学前班的初始阶段所表现出来的技能的缩影。该清单的指标由康涅狄克州学前课程框架和州立课程关于语言艺术和数学的标准发展而来，并且以康涅狄克州的教育标准为基础。幼儿学校（3-5 岁儿童教育机构）和学前班的教师群体代表了农村和城市地区、特殊教育和英语初学者不同的儿童样本，考察了技能清单中的指标，他们给教育部门提出了这种以使用测查为主的指标适宜性的建议。清单中所选的指标是全体成员输入的结果。

课程框架和标准的内容被放入清单中，它们代表了学生在学前班初期需要表现出的最重要的技能。这些技能和行为分成 6 个领域，每个领域有 3-5 个特殊的

指标——语言技能、读写技能、计算技能、身体/运动技能、创造和审美技能和自我及社会技能。举例来说，语言领域包括以下指标：参与对话，复述故事中的信息，能理解简单包含了 2 个步骤的语言指令，能运用至少 5 个词语的句子说话，能表达自己的情感和需要，以及专心地倾听。本研究将对指标之间的关系进行分析。工具在 2007 年秋季开始使用。

在工具实施过程中，要求每个教师能够按照领域将学生在班级中的表现水平分为三种；例如，每位老师针对六个领域中的每一个领域给每个学生评定一个等级。要求教师以学生表现的技能水平的一致性和所需要的指导帮助为基础进行评定。评定量表有三个水平：

水平 1：能初步表现出制定领域中的技能，但需要较多的指导帮助。

水平 2：特定领域中的技能表现前后不太一致，需要一些指导帮助。

水平 3：能始终表现出特定领域中的技能，几乎不需要指导帮助。

学生在某一个领域内一系列技能方面如表现出前后不一致的情况下，关于如何给学生评定等级，本研究没有给老师提供指导。

在 2009 年秋季，市区的一个较大学区的行政领导向该州政府申请要求完成学前班准入技能清单的指标和领域水平两个层面的数据（州政府仅要求领域水平方面的数据）。这些领导者认为，这一清单所提供的数据可能比它的原始形式的工具所提供的数据更为有用。本研究就运用了这一清单的数据。总体来说，这些教师给班上每个学生在六个领域的 32 个指标进行了等级评定。

研究对象

来自 24 个不同小学的 95 个学前班教师给 1670 个学生进行了等级评定。分配给每位老师的学生数量从 1 到 34 个不等，平均为 18 个。其中 5 位教师教半日班，8 位老师仅负责提供 10 个以下学生的数据。个人背景信息数据由学区行政部门提供。数据显示，49% 的学生是女生，27% 的学生英语不熟练，9% 的学生需要特殊教育，97% 的学生享受免费的或低价午餐。大部分学生是西班牙人（60%）。在其他学生中，32% 是黑人，6% 是白人，2% 是亚洲人，0.1% 是美洲印第安人。

数据分析

学前班准入技能清单的设计是用来测查儿童进入学前班时的准备状态。学前班准备状态被看作是隐性的、不易观察的和可变的。在本数据库中，进入学前班技能的准备状态是通过一个3分的等级量表来测定，它考察学生所表现的一系列可观察技能和知识的一致性和独立水平。本研究运用指标得分，通过探索性和实证因素分析程序来确定学前班准备状态的界定和结构，来自2009年数据库的数据被随机分成两个样本用于分析。第一个子样本用来做探索性因素分析，来自探索性因素分析的数据结构再运用第二个子样本进行证明。虽然本研究的结果对本研究工具的效度有意义，我们相信本测查的子量表得分能够被用来作为类似的用于评估学前班学生技能的工具的报告结构。

首先，针对第一个样本，本研究运用直接斜交转轴的主轴因子法(PAF)进行探索性因素分析。在主轴因子法中，因子以指标中的共变为基础进行界定。相似地，直接斜交转轴法假定，因子之间有相互关系，这在初步的相关分析中很明显。本研究运用一些标准来定义数据库中一定数量的因子，包括凯撒格特曼规则(Kaiser-Guttman rule) (Kaiser, 1991), the Scree Plot (Thompson, 2004), 和平行分析(Hayton, Allen, & Scarpello, 2004; Fabrigar, Wegener, MacCallum, & Strahan, 1999)。

然后，我们运用Mplus(Muthén & Muthén, 2007)进行多重实证性因素分析，来检测因子结构对第二个数据样本的适宜性。起初，对学生的评定按照不分等级的方式进行分析，并按照连续处理的方式进行。我们承认，这些数据更适合分类作为有顺序的类别。但是，我们以数据的连续性为基础，进行分析来测查修正指数。修正指数提供了教师看待清单指标方式的初步理解，因为它们反映了指标之间的协方差。把数据作为顺序数据看待的话，修正指数在Mplus中并不可见。

单一水平的实证性因素分析之后，我们在多水平背景下考察了数据的顺序性。我们相信，这一多水平技术是更为合适的，因为它考虑到了教师评定的聚类特点。多水平实证性因素分析(MCFA)精确地建立了教师内和教师间两种水平的因素结构模型。我们有几种理由可以期望，数据应该表现出非独立性。首先，尽管有州政府的培训和专业发展的进步，每一位教师可能对工具有自己独特的解读。一位教师可能对“最低限度的指导帮助”或者“表现一致”的理解比另外一位老师更为精确。其次，教师对等级量表的解读可能建立在学生与所在班级学生群体进行比较的基础上。而其他教师可能以全州所有学前班学生的背景来解读量

表。最后，多水平框架解释了可能存在于同一班级中学生之间的本来联系。在我们的实证性因素分析中，我们运用几个常见的拟合指数来评估模型的适宜性，包括近似误差均方根 (RMSEA)，塔克 - 刘易斯指数(TLI)和比较拟合指数(CFI)。我们也测查了标准化回归权重（模式矩阵），多元相关平方和标准化残差。

在多水平的实证性因素分析中，方差-协方差矩阵被分为两个矩阵——一个用来获得教师内方差和协方差，另外一个用来求取教师间的方差和协方差。学校间的方差占总方差的比例是组内相关(ICC)，它随着教师间异质性和教师内同质性的共同增长而增长。组内相关系数的范围是从 0 到 1。较高的组内相关系数说明，项目的方差大部分来源于教师之间。同一个教师评定的学生之间存在一定的同质性，不同的教师对学生的评定存在一定的异质性。换句话说，知道一个学生的老师是谁的话能够帮助预测学生的评估成绩。如果组内相关系数为 0，则同一个班级的学生比其他班级的学生并没有更多的相似性。如果组内相关是 1，则同一个班级的学生则几乎是另一个学生的复制品。如果教师给所有学生评定同一个等级，那么组内相关就是 1。

可预测的参数数量受到教师数量($n = 84$)的限制。正是由于这个原因，有必要对多水平实证性因素分析模型加一些限制。首先，我们限制所有的指标组间变量和组内变量平等。另外，我们限制组间的误差变量为 0。限制组间的误差变量为 0 意味着，组平均数中所有的变化均能够用公共因子平均数中的差异来解释。赫克斯(Hox,2002)陈述到，当水平 2 的样本较小并且组间方差接近于 0 时，样本在多水平实证因素分析中，将组间的残余方差固定为 0 是必要的。相反，允许组间残差的存在意味着，一些组内水平变量对于每一个测查的变量来说都是特定的 (Kamata, Bauer, & Miyazaki, 2008)。这些是较为普通的限制。

结果

每一个项目的平均数和标准差见表 1。总体来说，身体和运动领域以及创造性领域中的指标平均分较高。读写和数学领域中的指标平均分较低。学生样本随机平分，每一个样本均做分析。

表 1
项目分类，平均数和标准差($n = 1659$)

指标内容	<i>M</i>	<i>SD</i>
语言 1: 参与对话	2.10	0.77
语言 2: 复述所听故事中的内容	1.81	0.74
语言 3: 理解包含简单的两个步骤的语言指令	2.12	0.74
语言 4: 用至少 5 个词语的句子说话	2.11	0.79
语言 5: 交流情感和需要	2.07	0.74
语言 6: 专心倾听别人说话	2.02	0.75
读写 1: 会拿书并从前到后翻页	2.32	0.73
读写 2: 理解文字能表达意义	2.07	0.77
读写 3: 独立地探索书籍	2.19	0.74
读写 4: 识别书面字母, 特别是自己名字中和熟悉词语中的字母	2.00	0.78
读写 5: 能将字母和发音联系起来	1.82	0.76
读写 6: 鉴别一些词汇的起始音	1.86	0.77
读写 7: 表现出初步的书写	1.73	0.72
数学 1: 能数到 10	2.36	0.76
数学 2: 数数时表现出一一对应	2.15	0.78
数学 3: 能用各种常见物品测量物体	1.73	0.70
数学 4: 认识简单形状	2.15	0.76
数学 5: 识别模式	1.95	0.74
数学 6: 能按照大小、形状、用途或者其他特点给物体分类	1.95	0.73
数学 7: 理解事件顺序	1.74	0.70
社会 1: 参加自我选择的活动	2.35	0.67
社会 2: 与同伴互动共同游戏	2.23	0.69
社会 3: 运用语言表达自己的情感或运用语言来识别冲突	2.11	0.73
社会 4: 寻求同伴或成人的帮助来解决冲突	2.14	0.72
社会 5: 遵守班级常规	2.20	0.71
身体 1: 跑、跳和平衡	2.51	0.61
身体 2: 踢球或抛球, 爬楼梯、跳舞	2.49	0.63
身体 3: 用书写工具写字或画画	2.30	0.73
身体 4: 完成任务, 如走迷宫、跳绳或用剪刀剪开	2.28	0.73

创造性 1: 通过画、涂、造型或建造来表征经验	2.19	0.70
创造性 2: 参与装扮性游戏	2.29	0.63
创造性 3: 愿意参加音乐活动	2.37	0.67

探索性因素分析(EFA)

我们用直接斜交转轴的主轴因子法(PAF)对指标的等级数据进行探索性因素分析。在这些数据中，三个因子是建立在凯撒格特曼规则（Kaiser-Guttman rule）（特征值大于 1）的基础上，两个因子由陡坡图来说明。在平行分析中，出现类似的随机“干扰”数据并与从这些数据引出的因子进行了比较。我们期望在这些干扰数据中并没有出现因子，研究数据中的因子特征值应该比随机数据中特征值的百分比数据大。以特征值平均数和百分比数据为基础，分析说明了工具中出现了三个因子。我们选择继续进行三个因子的解决方案，因为我们的主要目标是测查教师是如何运用指标来定义学生进入学前班时的准备状况的。在对三个因子进行旋转和抽取后，我们获得了取样適切性量数(KMO = .97)，并被 KMO 指导方针评定为“很好” (Pett, Lackey, & Sullivan, 2003)。

大体上，新的因素包含了最初的量表。以项目相关为基础，读写、语言和数学领域结合起来成为“学业准备”因素。身体和运动以及创造性和美学领域内的指标结合起来成为第二个因素，就是我们所说的“活动准备性”。自我和社会量表的指标与语言领域内与和别人打交道的两个指标一起结合起来成为第三个因素。我们称这个因素为“社会准备性”。这三个因素和项目内容见表 2。

表 2 来自 EFA 的模式矩阵

指标内容	Factor		
	1	2	3
读写 5: 能将字母和发音联系起来	0.94		
读写 6: 鉴别一些词汇的起始音	0.91		
数学 7: 理解事件顺序	0.85		
读写 4: 识别书面字母，特别是自己名字中和熟悉词语中的字母	0.85		
数学 6: 能按照大小、形状、用途或者其他特点给物体分类	0.83		
数学 5: 识别模式	0.8		

读写 7: 表现出初步的书写	0.78		
数学 2: 数数时表现出一一对应	0.77		
数学 4: 认识简单形状	0.75		
读写 2: 理解文字表达意义	0.74		
数学 3: 能用各种常见物品测量物体	0.74		
语言 2: 复述所听故事中的内容	0.72		0.17
	0.67		0.
数学 1: 能数到 10		18	
语言 4: 用至少 5 个词语的句子说话	0.56		0.32
读写 1: 会拿书并从前到后翻页	0.55	0.21	0.16
语言: 参与对话	0.54		0.31
语言 3: 理解简单的包含两个步骤的语言指令	0.54		0.32
读写 3: 独立地探索书籍	0.53		0.27
身体 2: 踢球或抛球, 爬楼梯、跳舞			0.89
身体 1: 跑、跳和平衡			0.87
身体 4: 完成任务, 如走迷宫、跳绳或用剪刀剪开	0.2		0.61
身体 3: 用书写工具写字或画画	0.2		0.58
创造性 3: 愿意参加音乐活动			0.51 0.26
创造性 2: 参与假扮性游戏			0.47 0.35
创造性 1: 通过画、涂、造型或建造来表征经验	0.23	0.4	0.3
社会 1: 参加自我选择的的活动		0.35	0.49
社会 2: 与同伴互动共同游戏		0.21	0.71
社会 4: 寻求同伴或成人的帮助来解决冲突			0.85
社会 3: 运用语言表达自己的情感或运用语言来识别冲突			0.83
社会 5: 遵守班级常规			0.68
语言 5: 交流情感和需要		0.39	0.5
语言 6: 专心倾听别人说话		0.42	0.49

单一水平的实证性因素分析

实证性因素分析的主要目的是证实探索性因素分析所表明的三个因素的数据结构, 而预测的结果给有关各个指标之间的关系和教师可能导致的差异提供了有用的指导。我们用最大近似估计值来预测模型。正如上面所述, 数据样本被随机分开用来做探索性因素分析和实证性因素分析。第二个样本用来做该实证性因素分析。

首先，我们测验了以表 1 中详述的因素结构为基础的假设模型。在探索性因素分析中，语言 5 和语言 6 分别属于两个因素。为了保持子量表完整地包含仅仅一个因素，语言 5 和语言 6 被指定仅仅属于社会准备性因素。结果显示了模型和数据之间的不合， $\chi^2(461, N = 797) = 5195.25, p < .001, \chi^2/df = 11.27$, Tucker-Lewis 指数(TLI) = .805, 比较拟合指数 (CFI) = .819, 均方根近似误差(RMSEA) = .114 CI (.111, .116)。卡方统计值对于大样本量来说是显著的，TLI 值高于.95，CFI 值高于.95，并且 RMSEA 值低于 .06 or 置信区间包含 .05，这都说明模型合适度是可接受的 (Browne & Cudeck, 1993; Hu & Bentler, 1999)。

我们测查了修正指数来帮助我们更好地理解工具的结构和作用。有几种理由可以相信修正指数能够反映教师对工具的运用。一个指标的残差与另外一个指标的残差相关时，就有了相关误差。在这种背景下，老师可能以以往知识或者对基本能力的假设为基础而不是以所述技能的评估为基础，在多个指标上评定相同的等级时，相关误差就产生了。同理，当指标实际上只强调同一种内容时，也就可能产生相关误差。模型不拟合可能是由一个项目包括了多种因素而不仅仅是一种因素而导致的，三个指标被包含在两种因素中。相关误差和交叉包含从工具发展的角度来看是有益的，因为它们帮助我们鉴别多余的语言指标或者参与的老师对指标的多余处理，表 3 显示了修正指数基础上每一个领域内指标的分组。该表也包括了“潜在的子领域”一栏。一般来说，修正指数暗示了具有相同内容的指标的组合。强调相关误差和交叉包含改善了模型的拟合程度， $\chi^2(440, N = 797) = 1948.81, p < .001, \chi^2/df = 4.43$, Tucker-Lewis 指数 (TLI) = .935, 比较拟合指数 (CFI) = .942, 均方根近似误差(RMSEA) = .066 CI (.063, .069)。子领域的含义和运用将在讨论中进行强调。

表 3
以多余指标为基础的修正指数

原始 领域	潜在子领域	原始指标
语言	表达性语言	语言 4: 用至少 5 个词语的句子说话 语言 5: 交流情感和需要 语言 1: 参与对话
	接受性语言	语言 6: 专心倾听别人说话 语言 3: 理解简单的包含两个步骤的语言指令

	复述故事	语言 2: 复述所听故事中的内容
读写	对书籍的熟悉程度	读写 1: 会拿书并从前到后翻页 读写 2: 理解文字能表达意义 读写 3: 独立地探索书籍
	对字母的熟悉程度	读写 5: 能将字母和发音联系起来 读写 4: 识别书面字母, 特别是自己名字中和熟悉的词语中的字母 读写 6: 鉴别一些词汇的起始音
	早期书写	读写 7: 表现出初步的书写
计算	数数	数学 1: 能数到 10 数学 2: 数数时表现出一一对应
	形状/模式	数学 4: 认识简单形状 数学 5: 识别模式 数学 6: 能按照大小、形状、用途或者其他特点给物体分类 数学 7: 理解事件顺序
身体和运动	测量	数学 3: 能用各种常见物品测量物体
	精细动作技能	身体 3: 用书写工具写字或画画 身体 4: 完成任务, 如走迷宫、跳绳或用剪刀剪开
	粗大动作技能	身体 1: 跑、跳和平衡 身体 2: 踢球或抛球, 爬楼梯、跳舞
自我和社会	冲突解决	社会 3: 运用语言表达自己的情感或运用语言来识别冲突 社会 4: 寻求同伴或成人的帮助来解决冲突
	参与	社会 2: 与同伴互动共同游戏 社会 5: 遵守班级常规
创造性和审美	自我选择的活动	社会 1: 参加自我选择的活动
	创造性和审美	创造性 1: 通过画、涂、造型或建造来表征经验 创造性 2: 参与装扮性游戏 创造性 3: 愿意参加音乐活动

多水平实证性因素分析(MCFA)

每个因素都被看作是独立的, 因为可估计的参数数量已经受到教师数量的限制。许多并不是所有在单一水平的分析中必需的修正指数, 在多水平背景下需要用来完成模型适宜性考察。另外, 我们使得组间水平的残差误差变量为 0, 并且以修正指数为基础释放一些项目的差异 (如表 3 所示)。由于对模型的限制, 每

一个因素都表现出了可接受的模型适应性的程度。表 4 中包含了每一个模型的结果，包括相关误差和残余方差的处理。

我们对每一个指标都计算了组内相关(ICC_s)。正如前面所提到的，较高的 ICC_s 说明项目变量的很大部分来源于教师间的变量。同一个老师对学生的评定存在一定程度的同质性，不同的教师对学生的评定存在一定程度的异质性。举例来说，如果一个老师比其他老师更倾向于评定较高的等级，或者如果教师对一些项目的意义解释存在差异，这样的情况能够导致较高的 ICC_s。表 5 按照 ICC 列举了指标。指标的 ICC_s 较低，则对不同的教师样本更容易进行一致的解释（即差异性小）。这些项目几乎与教师因素无关。

表 4
各个模型中每个因素的结果

指标	三种模型		
	学业准备性	社会准备性	活动准备性
	语言 1-语言 4 读写 1-读写 7 计算 1- 计算 7	语言 5-语言 6 社会 1- 社会 5	身体 1-身体 4 创造性 1- 创造性 4
χ^2 (df)	304.145 (31)	39.216 (7)	46.129 (5)
RMSEA	0.105	0.076	0.102
CFI	0.973	0.886	0.892
TLI	0.994	0.951	0.934
相关误差			
聚内	读写 6 和读写 5 读写 3 和读写 1 语言 4 和 语言 1 计算 2 和 计算 1 计算 7 和计算 3	社会 4 和 社会 3	身体 1 和身体 2 创造性 3 和 创造性 2 身体 3 和身体 4
聚间	读写 6 和 读写 5 读写 3 和 读写 1 语言 4 和语言 1 计算 2 和计算 1 计算 7 和计算 3	社会 4 和 社会 3	n/a
残差			
聚间	语言 1 读写 1- 读写 3, 读写 5, 读写 7	社会 1 社会 3	创造性 1- 创造性 2 创造性 4

计算 1-计算 3

计算 5-计算 7

表 5

进入学前班 准备性清单中项目的组内相关 (ICCs)

指标内容	教师	
	N	ICC
读写 4: 识别书面字母, 特别是自己名字中和熟悉词语中的字母	84	0.135
语言 5: 交流情感和需要	84	0.162
自我 5: 遵守班级常规	84	0.167
读写 5: 能将字母和发音联系起来	84	0.168
语言 6: 专心倾听别人的说话	84	0.173
读写 6: 鉴别一些词汇的起始音	84	0.178
语言 2: 复述所听故事中的内容	84	0.179
数学 4: 认识简单形状	84	0.192
语言: 参与对话	84	0.198
自我 2: 与同伴互动共同游戏	84	0.198
语言 4: 用至少 5 个词语的句子说话	84	0.210
自我 4: 寻求同伴或成人的帮助来解决冲突	84	0.218
自我 3: 运用语言表达自己的感受或识别冲突	84	0.241
身体 4: 完成任务, 如走迷宫、跳绳或用剪刀剪开	84	0.248
语言 3: 理解简单的包含两个步骤的语言指令	84	0.249
身体 3: 用书写工具写字或画画	84	0.255
数学 2: 数数时表现出一一对应	84	0.268
创造性 3: 愿意参加音乐活动	84	0.281
读写 2: 理解文字能传播意义	84	0.287
数学 7: 理解事件顺序	84	0.290
数学 1: 能数到 10	84	0.291
读写 7: 表现出初步的书写	84	0.294
创造性 1: 通过画、涂、造型或建造来表征经验	84	0.300
自我 1: 参加自我选择的的活动	84	0.300
数学 5: 识别模式	84	0.303
身体 1: 跑、跳和平衡	84	0.308
创造性 2: 参与装扮性游戏	84	0.321
读写 3: 独立地探索书籍	84	0.332
身体 2: 踢球或抛球, 爬楼梯、跳舞	84	0.332
数学 6: 能按照大小、形状、用途或者其他特点给物体分类	84	0.346

读写 1: 会拿书并从前到后翻页	84	0.362
数学 3: 能用各种常见物品测量物体	84	0.372

讨论

学前班准入技能清单是为了给学前班初期提供学生技能考察的框架。本研究呈现的分析给老师评判学生进入学前班时的准备水平提供了一种视角。数据显示,在评估学生进入学前班时的技能时,教师应该运用总体上的对学生技能的评价方案而不是运用原始工具所呈现的六个领域的结构。具体而言,教师的评判是以探索性因素分析的三个因素为中心:学生的学业准备性、社会准备性和非学业活动的准备性。这一发现可能是教师对进入学前班时学生技能的理解的结果,或者是由于工具的结构所导致。可能同一个工具在同年稍后的时间运用时,教师对学生的能力有了更复杂的理解,也许可能产生不同的因素结构。又或者,结构可能是等级量表的产物。有可能这一 3 分的顺序量表更倾向于鼓励对学生技能的总体上的评估。

从实证性因素分析中可以看出,教师对有相似内容的指标会判定相似的等级。很显然,全州对学前班准入技能清单的实施要求老师对同一领域内不同的技能指标判定不同的等级。这一现象的显著例子就是身体和运动领域,它包括与精细动作技能相关的两个指标和与大动作技能相关的两个指标。实证性因素分析的结果可能意味着,按照这种形式处理的话,一些原始指标的最初意义可能会丢失。修正指数意味着这两个语言指标的相关误差——“用至少 5 个词语的句子说话”和“交流情感和需要”。这两个指标的大致意思意味着内容相似,并且相关误差也意味着教师对每一个指标判定了相似的等级。用来制定指标的课程框架指出,儿童可能能够用手势或者声音交流他或她的感受而不用语言。这一解释在术语“交流”中并不清晰。其他领域可能覆盖了技能体系。可能“将字母与发音联系起来”这一指标并不需要与“鉴别一些词汇的起始音”这一指标放在一起。不能鉴别起始音的学生不会将声音与字母联系起来。

ICC 给我们看待教师对单一指标的理解提供了视角。较低的 ICC 在本文中有两种解释。首先,可能是学生在这些指标上有非常相似的等级。或者,可能是教师之间对这些指标的解释具有一致性。有较低 ICC 的两个相似指标是“将字母与声音联系起来”(ICC = .168, $M = 1.82$, $SD = .76$)和“鉴别一些词汇的起始音”(ICC

= .178, $M = 1.86$, $SD = .77$)。在这种情况下,相对较低的平均数和较高的标准差可能使我们相信,低 ICC 是解释一致性和学生差异性的结果。语言在这些指标上的特殊性也是我们感兴趣的。有最高 ICC 的指标,“运用各种日常物品测量物体”(ICC = .372),平均分(1.73)和标准差(.70)较低。在这种背景下,ICC 可能反映了教师之间解释的不一致性,或者所有的老师倾向于给自己的学生判定相似的等级。对于这一指标来说,语言是相当模糊的。“物体”和“日常物品”并没有给出定义。另外,关于询问学生进行这些活动的频率和准确性,也没有提供清楚的解释。

最后,在理解了这些数据来自于同一个城区同一个时间点的基础上,我们可以对单个的指标进行评论。在读写领域内,教师给学生对书的熟悉程度这一指标判定了最高的等级,而对他们的早期书写技能给予了最低的等级。所有指标的平均等级非常接近语言领域量表的中间分,语言领域中得分最高的指标与参与对话和用至少 5 个词语的句子说话有关。对计算领域来说,教师对“能够数到 10”打分最高,对事件顺序相关的指标打分最低。自我和社会,身体和运动以及创造性领域的所有平均等级均高于量表的中间分。这些较高的等级在因素分析中促进了学业指标和非学业指标的划分。在自我和社会领域,“参加自我选择的活动”这一指标平均等级最高。在整个工具中,与跑、跳和平衡相关的指标平均分最高,标准差最低。在创造性领域,与学生愿意参加音乐活动相关的指标得分最高。这些结果反映了一个学区的情况,不能推及到其他的样本群体。

启示

虽然我们的工作建立在工具结构有效性的基础上,研究结果可以用来了解对儿童进入学前班时准备状况更加详细的理解。领域和子领域的结构是工具开发的轮廓。我们的发现为儿童进入学前班时需测量的技能分类提供了基础。以清单中的指标为基础,对进入学前班时学生发展和教育的评价应该包括以下方面的评分或测查:

- 表达性语言
- 接受性语言
- 对故事的反应
- 对书籍的熟悉程度
- 对字母的熟悉程度

- 早期书写
- 数数
- 形状和模式
- 测量
- 精细动作技能
- 大动作技能
- 冲突解决
- 社会参与
- 参与自我选择的活动
- 创造性技能

以这些内容为基础的测查比领域水平的测查（如，语言、读写、计算等）更能够使老师对每个学生的发展提供一个更加详细的分析。但是，需要谨慎的是，康涅狄克州的清单指标为这一工具提供了原始的描述信息。需要老师和早期教育研究人员进行更多的工作来给每一方面的内容提供更多的描述和定义。

本研究还为设计幼儿评价量表的研究人员、评价人员和管理人员提供了结构上的指导。首先，为了达到一致性，在不同评价者运用工具时，专门的语言是必需的。如果“交流”包括非语言的动作，就应该记录下来。在本研究中，比较详细的指标测查出的等级也比较一致。而且，在以教师为主导的评分量表中，应排除重复的项目来减轻数据收集的负担。第二，本研究结果强调了评定量表中等级数量的问题。在本研究中，要求教师运用简单的 3 分等级量表来评估学生在每一个指标上的表现。另外，等级量表的目的是考察学生独立地和一致地表现他们特定技能的程度。对评估的可靠实施要求对单一内容测查的信息进行描述。经过这种改变，一个更加丰富的 4 分等级量表将能够提供更多的等级，这样就能进行更复杂的分析。这样的量表可能包括一个一致性的 4 分等级量表（几乎不，很少，大多时候，所有时候）和一个独立性的等级量表。

局限性

本研究有几个方面的局限。来自 83 个学校的 1600 个学生样本限制了能够使用的分析技术和本研究分析的功效。分析中比较明显的一些重复和矛盾可能来自于对指标水平信息的不恰当运用，例如，指标是用来描述领域内涵，并不是用于

指导或者定义单个学生的技能。最后，这些数据仅代表一个州的城市学生群体。在本研究中，97%的学生享受免费午餐或低价午餐，60%的学生是西班牙人，32%的学生是黑人。在该州 2006 年的数据中，27%的学生享受免费午餐或低价午餐，19%的学生是西班牙人，14%的学生是黑人（康涅狄克州教育部，2007）。这一差别值得引起注意，它限制了本研究结果的推广度。

致谢

本研究受到了康涅狄克州教育部的资助。本研究代表了作者的观点而非资助机构的观点。

参考文献

Browne, Michael W., & Cudeck, Robert. (1993). Alternative ways of assessing model fit. In Kenneth A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Connecticut State Department of Education. (2007). *Data bulletin: Kindergarten, 2006-2007*. Hartford: Connecticut State Department of Education.

Denton Flanagan, Kristin, & McPhee, Cameron. (2009, October). *The children born in 2001 at kindergarten entry: First findings from the kindergarten data collections of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)* (NCES 2010-005). Washington, DC: National Center for Education Statistics.

Fabrigar, Leandre R.; Wegener, Duane T.; MacCallum, Robert C.; & Strahan, Erin J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299.

Hains, Ann Higgins; Fowler, Susan A.; Schwartz, Ilene S.; Kottwitz, Esther; & Rosenkoetter, Sharon. (1989). A comparison of preschool and kindergarten teacher expectations for school readiness. *Early Childhood Research Quarterly*, 4(1), 75-88.

Harradine, Christine C., & Clifford, Richard M. (1996, April). *When are children ready for kindergarten? Views of families, kindergarten teachers, and child care providers*. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Hayton, James C.; Allen, David G.; & Scarpello, Vida. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191-205.

Heaviside, Sheila, & Farris, Elizabeth. (1993, September). *Public school kindergarten teachers' views on children's readiness for school* (NCES 93-410). Washington, DC: National Center for Education Statistics.

Hox, Joop J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Hu, Li-tze, & Bentler, Peter M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.

Kagan, Sharon Lynn; Moore, Evelyn; & Bredekamp, Sue (Eds.). (1995, June). *Reconsidering children's early development and learning: Toward common views and vocabulary*. Washington, DC: National Education Goals Panel.

Kaiser, Henry F. (1991). Coefficient alpha for a principal component and the Kaiser-Guttman rule. *Psychological Reports*, 68(3), 855-858.

Kamata, Akihito; Bauer, Daniel J.; & Miyazaki, Yasuo. (2008). Multilevel measurement model. In Ann A. O'Connell & D. Betsy McCoach (Eds.), *Multilevel analysis of educational data* (pp. 345-388). Charlotte, NC: Information Age.

Lin, Huey-Ling; Lawrence, Frank R.; & Gorrell, Jeffrey. (2003). Kindergarten teachers' views of children's readiness for school. *Early Childhood Research Quarterly*, 18(2), 225-237.

Love, John M. (2001, December). *Instrumentation for state readiness assessment: Issues in measuring children's early development and learning*. Princeton, NJ: Mathematica Policy Research.

Meisels, Samuel J. (1999). Assessing readiness. In Robert C. Pianta & Martha J. Cox (Eds.), *The transition to kindergarten* (pp. 39-66). Baltimore, MD: Paul H. Brookes.

Muthén, Linda K., & Muthén, Bengt O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Authors.

National Association for the Education of Young Children (NAEYC) & National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2009). *Where we stand on curriculum, assessment, and program evaluation*. Retrieved August 10, 2011, from <http://www.naeyc.org/files/naeyc/file/positions/StandCurrAss.pdf>

Pett, Marjorie A.; Lackey, Nancy R.; & Sullivan, John J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Rathbun, Amy, & West, Jerry. (2004). *From kindergarten through third grade: Children's beginning school experiences*. (NCES 2004-007). Washington, DC: National Center for Education Statistics.

Scott-Little, Catherine; Kagan, Sharon Lynn; & Clifford, Richard M. (Eds.). (2003). *Assessing the state of state assessments: Perspectives on assessing young children*. Greensboro: University of North Carolina, SERVE.

Scott-Little, Catherine; Kagan, Sharon Lynn; & Frelow, Victoria Stebbins. (2006). Conceptualization of readiness and the content of early learning standards: The intersection of policy and research? *Early Childhood Research Quarterly*, 21(2), 153-173.

Stedron, Jennifer, & Berger, Alexander. (2010, August). *NCSL technical report: State approaches to school readiness assessment*. Retrieved August 10, 2011, from <http://www.ncsl.org/documents/Educ/KindergartenAssessment.pdf>

Thompson, Bruce. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Wesley, Patricia W., & Buysse, Virginia. (2003). Making meaning of school readiness in schools and communities. *Early Childhood Research Quarterly*, 18(3), 351-375.

West, Jerry; Denton, Kristin; & Germino-Hausken, Elvira. (2000). *America's kindergartners*. (NCES 2000-070). Washington, DC: National Center for Education Statistics.

West, Jerry; Germino-Hausken, Elvie; & Collins, Mary. (1993, September). *Readiness for kindergarten: Parent and teacher beliefs* (NCES 93-257). Washington, DC: National Center for Education Statistics.

作者信息

Dr. Jessica Goldstein 是康涅狄克大学测量评估机构的助理教授。戈德斯坦博士的研究兴趣包括特殊群体大规模量表评估体系的有效性和学校绩效考核体系中学生成绩的多样化测量手法的运用。

联系方式

Jessica Goldstein
Assistant Professor in Residence
Measurement, Evaluation, and Assessment Program
Department of Educational Psychology
University of Connecticut
249 Glenbrook Road, Unit 2064
Storrs, CT 06269
Email: Jessica.Goldstein@uconn.edu

Dr. D. Betsy McCoach 是康涅狄克大学测量评估机构的副教授。她在多层线性建模、工具设计、因素分析和结构方程建模方面有着丰富的经验。她参与执笔的文章和书籍超过 50 篇。

联系方式
D. Betsy McCoach
University of Connecticut
249 Glenbrook Road, Unit 2064
Storrs, CT 06269-2064
Telephone: 860486-0183
Fax: 860-486-0180
Email: betsy.mccoach@uconn.edu

译者：中国华东师范大学 杨志艳
审校：中国华东师范大学 周欣