

家庭社会经济地位对儿童语言 能力的影响研究

侯晓舟, 常婉凝, 陈娜娜

[摘要]提升儿童语言能力是国家语言能力教育建设的重大命题之一。文章以CFPS数据2018年调查问卷为数据来源,围绕家庭社会经济地位对儿童语言能力教育的影响问题,利用逻辑斯蒂回归、决策树和随机森林三种统计方法,对其进行了实证分析。分析结果显示,家庭社会经济地位中家庭收入、父母学历均对儿童语言能力教育存在正相关关系且具有显著性影响。基于此结果,本文提出提高儿童语言能力教育的针对性建议。

[关键词]儿童语言能力; 家庭社会经济地位; 随机森林方法

一、引言

自Chomsky率先提出语言能力这一概念后,诸多学者围绕这一领域展开了多角度的研究,并将其概念和研究内容逐步延展。随着中国逐步走向国际舞台中心,国家语言能力教育建设成为研究热点。近年来,在我国开启全面建设社会主义现代化国家和促进全体人民共同富裕的新征程中,语言文字事业贡献卓著,同时也对共同富裕目标的实现产生了积极作用,《国家语言文字事业“十三五”发展规划》更是将“国家语言能力”收纳其中。正如李宇明(2018)所言,国家语言能力的提升能有效发挥语言在全球治理中的作用与价值。语言能力建设是服务国家发展战略、建设中华民族共同体与人类命运共同体的“软基建”工程(文秋芳和杨佳,2020)。因此,对语言能力教育问题进

[收稿日期] 2024-01-21

[基金项目] 河北省高等学校人文社会科学研究项目(语言文字研究专项)“基于逐步多元回归的我省学校学生语言能力影响因素研究”(YWZX2022008)。

[作者简介] 侯晓舟,河北大学外国语学院,电子邮箱地址:houxiaozhou@hbu.edu.cn;常婉凝,河北北方学院,电子邮箱地址:727519822@qq.com;陈娜娜(通讯作者),河北大学教育学院,电子邮箱地址:522946793@qq.com。

行研究的必要性毋庸置疑。

李宇明(2011)曾提出,一个国家的语言能力包括五个方面:语种能力、国家主要语言的国内外地位、公民语言能力、拥有现代语言技术的能力以及国家语言生活管理水平能力。本文将选取其中一个能力,即公民个人语言能力进行细化研究。儿童语言能力作为公民语言能力的重要组成部分之一,其影响因素一直是语言能力领域和语言习得领域的研究重点。本文以北京大学中国家庭追踪调查数据库(CFPS)2018年调查问卷为数据来源,从统计语言学视角出发,利用逻辑斯蒂回归、决策树和随机森林三种统计方法,对家庭社会经济地位对儿童语言能力的影​​响及其影响机制进行了分析,并基于推动语言教育发展的视角,为提高儿童语言能力提出了针对性建议,希冀能够对我国国家语言能力的建设与提升以及语言教育的发展产生一定推动作用。

二、文献综述

诸多学者对国家语言能力建设及提升问题进行过深入研究(魏晖,2015;文秋芳,2016;李宇明,2018等)。目前,我国国家语言能力建设已在国家语言治理能力、核心能力和战略能力等方面取得了较大进展,但仍存在一些不足。因此,有必要继续对国家语言能力提升问题进行研究。魏晖指出,根据李宇明对国家语言能力的分类,公民语言能力是国家语言能力的核心基础,对国家语言能力建设有支撑作用(魏晖,2015)。而儿童语言能力属于公民语言能力的重要组成部分,对儿童语言习得以及儿童语言能力提升问题进行研究具有重要价值。同时,儿童语言能力是反映儿童正常成长的一个重要因素(刘志军和陈会昌,2005),其重要性显而易见,因此学界对其影响因素进行了广泛研究。

研究发现,儿童语言能力发展受语言内部、外部因素制约。在语言外部因素方面,学界就家庭收入(Dahl and Lochner, 2012; Pace et al., 2017)、父母学历(李珂, 2016)、父母参与(Alyssa et al., 2005)、父母教养方式(李毅和谭婷, 2019; 刘志军和陈会昌, 2005)、儿童自我效能感(吉晖, 2019)、儿童健康(高振宇, 2016)以及学习资源(Jean et al., 2002; 李艳玮等, 2012; 张云运等, 2015)等因素对儿童语言能力产生的影响开展研究。除此之外,语言环境、学习动机、大脑发育等语言内部因素亦在儿童语言习得过程中发挥重要作用(郭翔飞, 2012; 刘颖, 2014)。本文围绕由家庭收入和父母学历决

定的家庭社会经济地位(任春荣, 2010)对儿童语言能力的影响展开分析研究。

诸多学者围绕家庭收入、父母学历对儿童语言能力进行了相关研究。研究表明, 家庭收入对儿童早期语言能力发展有显著的正向预测作用(Dahl and Lochner, 2012; Pace et al., 2017), 收入较高的家庭相对来说有更大几率为儿童提供良好语言习得环境以及充足的学习资源, 为有效推动儿童语言能力的发展提供物质保障。就父母学历而言, 在一定程度上可以反映父母的文化素养及其受教育程度, 是影响儿童语言能力的直接因素(李珂, 2016)。受教育程度以及文化素养较高的父母更有可能为儿童提供优良的文化环境和语言环境, 儿童接触到丰富的语言资源或其他学习资源的机会就更多。

此外, 父母作为教育计划方针的制定人和决策者时所进行的父母参与, 对儿童的教育计划进行制定与管理程度能够影响儿童学习动机(Alyssa et al., 2005)。学习动机是学习者为了实现学习目标而表现出的努力程度、学习态度和渴望程度(Gardner, 2010), 是儿童语言习得中重要的个体差异因素之一。父母作为儿童学习动机的外部影响因素之一, 其较高的参与度在有效激活儿童学习动机之时, 儿童习得语言的注意力以及付出努力的程度也随之提升。

父母教养方式与儿童语言习得有显著相关性(Raviv et al., 2004; 袁言云等, 2020; 罗世兰等, 2021)。父母对儿童进行教育和培养的方式不同, 产生的作用不同。研究发现, 父母的教养方式, 尤其是父亲的严格教养方式, 与儿童的语言表达发展有显著正相关关系(陈会昌和刘志军, 2005), 而父亲温暖与理解的教养方式与儿童认知能力的相关程度最大(张茜洋等, 2017), 说明在这类教养方式下, 父母为儿童提供的充足的语言刺激极大地促进了儿童认知能力和语言能力发展。儿童自我效能感亦受父母教养方式影响(袁言云等, 2020), 自我效能感是儿童对自己能否在一定水平上完成某一活动所具有的能力判断、信念或主体自我把握与感受(夏征农和陈至立, 2015), 父母采取积极的教养方式时, 儿童在语言习得过程中提升自我效能感的可能性更高。儿童自我效能感对语言能力发展有正向影响作用(石雷山等, 2013; 吉晖, 2019), 这说明父母教养方式在很大程度上影响了儿童语言能力。

此外, 充足的学习资源可更直接影响儿童语言能力发展(Jean et al., 2002)。学习资源是包括信息、人员、资料、设备和技术等在内的可用于学习的资源总和(夏征农和陈至立, 2015), 其丰富程度直接影响儿童所处语言习得环境的优质程度及其进行语言习得的时长。同时, 学习资源作为家庭收入

与儿童语言能力发展之间的中介因素，还可间接影响儿童语言能力(李艳玮等，2012；张云运等，2015)。研究发现，儿童大脑的健康发育尤其影响儿童语言能力的发展，儿童健康状况与儿童语言能力呈正相关关系(Bradley and Corwyn, 2002；高振宇，2016)，即儿童健康状况越好，其语言习得能力越高。

通过对文献的梳理和总结可以发现，学界对儿童语言能力及其影响因素的研究成果颇多，但也存在一些不足之处：或是只注重理论分析，缺少与统计方法以及数据的结合；或是虽然结合了统计方法，但仅局限于最小二乘法等传统方法。如今，随着新技术的发展，使用随机森林等数据挖掘技术有助于克服传统计量研究方法解释性强、预测性弱的短板，对家庭社会经济地位对儿童语言能力的研究具有方法上的改进作用。因此，本文将采取随机森林等统计方法，基于统计数据分析家庭社会经济地位对儿童语言能力的影

三、研究设计

(一)数据来源

本文数据来源于北京大学2018年中国家庭追踪调查数据(CFPS)，并依照将“儿童”界定为“18岁以下的任何人”的联合国《儿童权利公约》界定，选取0—18岁的人群作为研究对象。通过数据清理、集成和变换后，最终从家庭经济问卷、个人自答问卷和少儿父母代答问卷^①中确定了1054个家庭分析样本。在模型分析过程中，为减少训练样本分布对试验结果的影响，本文选取70%的样本数据作为模型的训练集，30%的样本数据作为模型的测试集。

(二)变量界定

1. 被解释变量：交谈能力、语文成绩和普通话水平

在推动国家语言能力发展的过程中，不断提升公民语言能力已成为其基础建设，其中，普及普通话(文秋芳，2016)以及提升语文素养与能力(陆俭

^① 其中，少儿父母代答是指在调查中，针对未成年人在认知和表达能力上可能存在局限，或因各种原因不能完成个人访问的，CFPS为采集他们与受访家庭成员之间的关系和他们个人基本信息，会以父母代答的方式采集家庭中所有成员的信息，确保信息的准确性和完整性。

明, 2018)是提升公民个人语言能力的重要组成部分, 儿童语言能力作为公民语言能力的一部分, 其普通话水平和语文能力具有重要研究价值。因此, 本文将“交谈语言”“语文成绩”“普通话水平”作为计量模型中的被解释变量, 以此来反映儿童的语言能力。其中, 衡量普通话水平的词组测试题来自2010年CFPS认知能力测试模块中的词组测试部分, 主要为认字题, 考察字词能力(谢宇等, 2014)。

2. 解释变量: 家庭收入和父母学历

家庭社会经济地位包括家庭收入和父母学历(任春荣, 2010), 本文选用这两个因素作为解释变量。其中, 在本文选用的调查问卷中, 家庭收入的衡量标准为家庭在过去一年的总收入, 父母学历按父母接受访问时所获得的最高学历为依准。

3. 控制变量

文章将学习资源(Jean et al., 2002; 李艳玮等, 2012; 张云运等, 2015)、父母参与(Alyssa et al., 2005)、父母教养方式(李毅和谭婷, 2019; 刘志军和陈会昌, 2005)、儿童健康状况(高振宇, 2016)以及自我效能感(吉晖, 2019)这5个变量作为控制变量, 以过去12个月受访者是否有教育培训支出、父母是否主动与孩子沟通和交流、孩子不如意时父母的处理方式、受访儿童的健康状况、受访儿童对自己学业情况打分所得分数分别反映学习资源、父母参与、父母教养方式、儿童健康状况以及自我效能感这5个变量。

变量对应的测量问题及说明如表1所示。

表1 变量说明

变量	具体问题	变量说明
被解释变量	交谈语言	访问过程中主要使用以下哪种语言 方言取0, 普通话取1
	语文成绩	就您所知, 孩子上学期间平时的语文成绩如何 中、差取0, 优良取1
	普通话水平	词组测试题 0至17分取0, 18至34分取1

续表

变量		具体问题	变量说明
解释变量	家庭收入	过去 12 个月家庭总收入(元/年)	包括经营性收入、工资性收入、租金收入、政府补助或他人的经济支持等
	父母学历	访问时,已完成(毕业)的最高学历是什么	从未上过学取 1,文盲/半文盲取 2,小学取 3,初中取 4,高中/中专/技校/职高取 5,大专取 6,大学本科取 7,硕士取 8
控制变量	学习资源	过去 12 个月教育培训支出(元/年)	若受访者没有该项支出,取 0
	父母参与	父母主动与孩子沟通和交流	十分不同意取 1,不同意取 2,中立取 3,同意取 4,十分同意取 5
	父母教养方式	孩子成绩不如意时的处理方式	告诉这个孩子要更加努力地学习取 1,更多地帮助这个孩子取 2,联系他/她的老师取 3,限制孩子的活动取 4,体罚这个孩子取 5,责骂这个孩子取 6,不采取任何措施取 7
	儿童健康状况	受访儿童的健康状况	不健康取 1,一般取 2,比较健康取 3,很健康取 4,非常健康取 5
	自我效能感	受访者给自己的学业情况打分	非常不满意取 1,不满意取 2,较为满意取 3,满意取 4,非常满意取 5

(三)研究方法

本文采用描述性统计分析、逻辑斯蒂回归模型、决策树模型和随机森林模型对儿童语言能力影响因素进行分析。描述性统计分析主要是对各变量的分布进行统计描述和比较,分析儿童语言能力的总体特征。对比运用逻辑斯蒂回归分析方法、决策树模型和随机森林模型进行拟合验证,通过比较模型的预测精度,更加精准地探索家庭社会经济地位对儿童语言能力的影

四、实证结果分析

(一)描述性分析

各变量的描述性统计分析如表 2 所示。总体而言,当前国内儿童语言能力整体处于中等水平,交谈语言、语文成绩、普通话水平三个变量的均值均高于 0.5。家庭收入均值为 64766.866 元,高于 2018 年的人均可支配收入,

说明大部分家庭的经济状况良好；父母学历均值为 3.885，调查样本表明父母学历层次大多处于初中水平；学习资源均值为 8592.733 元，表明大多数家庭都愿意为儿童语言能力发展进行投资，提供充足的学习资源以改善和提高其能力；父母参与均值为 3.733，说明大部分父母对儿童语言习得过程较为了解且参与程度较高；父母教养方式均值为 1.931，说明大部分父母愿意采取积极科学的教养方式；儿童健康状况整体较好，均值为 3.926；儿童自我效能感的均值为 3.34，表明大部分儿童对自己的学业表现较为满意，自信水平较高。

表 2 各变量描述性分析

	变量	均值	标准差	极小值	极大值	样本量
被解释变量	交谈语言	0.544	0.498	0	1	1054
	语文成绩	0.568	0.496	0	1	1054
	普通话水平	0.597	0.491	0	1	1054
解释变量	家庭收入	64766.866	61270.915	0	560000	1054
	父母学历	3.885	1.424	1	8	1054
控制变量	学习资源	8592.733	11375.706	0	100000	1054
	父母参与	3.773	0.796	1	5	1054
	父母教养方式	1.931	1.632	1	7	1054
	儿童健康状况	3.926	0.901	1	5	1054
	自我效能感	3.340	0.921	1	5	1054

(二)家庭社会经济地位对儿童语言能力的影响分析

1. 逻辑斯蒂回归模型分析结果

本文通过多重共线性检验，检验了模型中各解释变量之间的高度相关性，结果显示各解释变量的方差膨胀因子值均小于 10，说明各变量之间不存在多重共线性，即不存在伪回归问题。各模型的逻辑斯蒂回归分析结果，如表 3 所示。

表 3 各变量描述性分析

	模型一 交谈语言	模型二 语文成绩	模型三 普通话水平
家庭收入	0.00002*** (0.00)	0.00002*** (0.00)	0.00001*** (0.00)

续表

	模型一 交谈语言	模型二 语文成绩	模型三 普通话水平
父母学历	0.2591*** (0.69)	0.5467*** (0.09)	0.6264*** (0.08)
学习资源	0.00002* (0.00)	0.0002*** (0.00)	0.0001*** (0.00)
父母参与	-1.000*** (0.13)	1.0910*** (0.15)	0.2853* (0.12)
父母教养方式	-0.3780*** (0.06)	-0.2806*** (0.06)	-0.2775*** (0.06)
儿童健康状况	-0.0764 (0.10)	0.1137 (0.11)	0.7735*** (0.11)
自我效能感	0.5014*** (0.10)	1.2860*** (0.14)	0.2571* (0.10)
截距	1.3080* (0.69)	-12.0500*** (1.11)	-7.4150*** (0.85)

注：*、**、***分别表示 $p < 0.1$ 、 $p < 0.05$ 、 $p < 0.01$ ；括号内为标准误。

从表3中可知，家庭收入、父母学历、自我效能感、学习资源均可以显著促进儿童交谈语言、语文成绩和普通话水平的提升。家庭收入和父母学历与儿童交谈语言、语文成绩及普通话水平成正相关关系。父母学历每提高1个水平，儿童交谈语言提高0.2591个水平，语文成绩提高0.5467个水平，普通话水平提高0.6264个水平，说明父母学历显著影响儿童语言能力。另外，由数据分析可知，家庭收入对交谈语言与语文成绩的影响高于其对普通话水平的影响。综上所述，家庭社会经济地位与儿童语言能力存在正相关关系，且存在显著影响。

此外，儿童自我效能感和学习资源与交谈语言、语文成绩、普通话水平成正相关关系，但儿童自我效能感对语言能力有显著的积极影响，而学习资源影响程度却很小。父母教养方式与三个被解释变量均成负相关关系，父母参与和儿童健康状况与交谈语言成负相关关系，而与语文成绩和普通话水平则成正相关关系。

2. CART 决策树模型分析结果

决策树是通过一系列规则对数据进行分类的过程。决策树使用类似于

一棵树的结构来表示类的划分，树的构建可以看成是变量选择的过程，内部节点则表示树选择了某几个变量作为划分。本文经过对儿童交谈语言、语文成绩和普通话水平分别进行模型建立并生成 CART 决策树后，可以直观地看到各棵树的深度、节点、事件发生的概率和包含的样本数量，以此分析各自的影响因素及影响情况。

(1)交谈语言的 CART 决策树模型决策树

交谈语言的 CART 决策树模型决策树共分为 4 层，分类结果见图 1。决策树的第一层为家庭收入，说明家庭收入与交谈语言的关联性最强，家庭收入高于 49250 元的儿童使用普通话进行交谈的概率为 76.5%，而低于 49250 元的儿童使用普通话的概率只有 29.1%。在家庭收入高于 49250 元的分组中，如果父母参与程度较高，儿童使用普通话进行交谈的可能性将增加为 83.07%。在家庭收入低于 49250 元的分组中，与交谈语言关系最密切的也是父母参与，说明父母在儿童活动中较高的参与程度能够促进儿童更主动地进行交流沟通，有利于增加儿童使用普通话进行交谈的机会。在父母参与程度较低的家庭中，自我效能感对交谈语言的影响较大，说明对儿童自身而言，对自己学业表现的自信程度较高也有利于交谈水平的提高。家庭社会经济地位越高，家庭收入也就越高，父母便能拥有更多的时间与儿童相处并参与儿童的学业活动(陈晶晶, 2022)，交谈水平也就随之提高。与逻辑斯蒂回归模型的结果相同，家庭收入对儿童交谈能力存在显著的正相关关系。由此

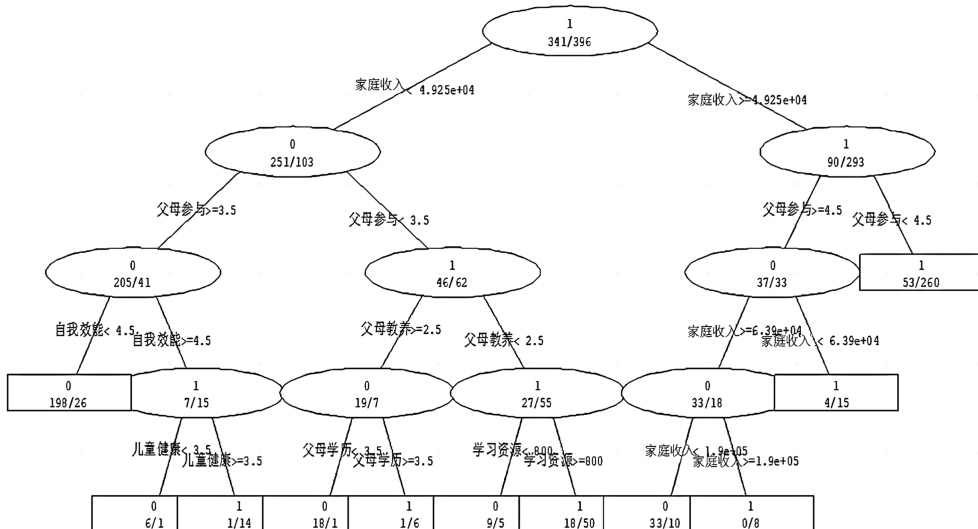


图 1 基于 CART 决策树模型的交谈语言影响因素

可知，家庭社会经济地位亦与儿童交谈能力存在显著的正相关关系，即家庭社会经济地位越高，家庭收入越高，儿童的语言习得能力就越强。

(2)语文成绩的 CART 决策树模型决策树

语文成绩的 CART 决策树模型决策树共分为 5 层，分类结果见图 2。决策树的第一个最佳分组变量是学习资源，说明学习资源对语文成绩的相关性最高。当家庭对学习资源的投入高于 7250 元时，88.55% 儿童的语文成绩较为优秀。语文成绩 CART 决策树的第二层说明儿童健康状况及自我效能感对语文成绩的影响较为密切。正如前文所述，家庭社会经济地位包含了家庭成员的学历水平，而父母学历越高，便越有可能为儿童提供更多更好的学习资源(李珂，2016)，更有利于儿童语文成绩的提高。结合逻辑斯蒂回归模型结果可知，学习资源、儿童健康及其自我效能感均对儿童语文成绩存在显著影响，即学习资源越充分、儿童越健康、自我效能感越强，儿童语文成绩就越高。

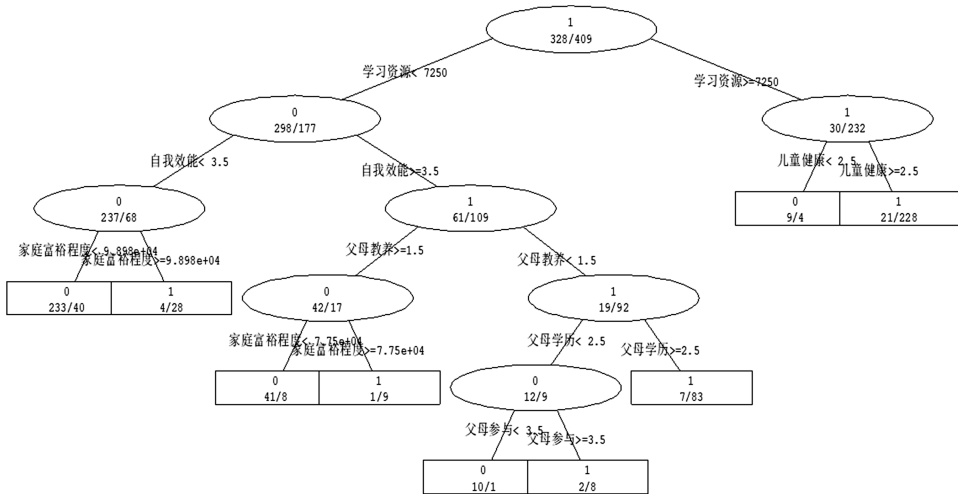


图 2 基于 CART 决策树模型的语文成绩影响因素

(3)普通话水平的 CART 决策树模型决策树

普通话水平的 CART 决策树模型决策树共分为 6 层，分类结果见图 3。决策树的第一层为父母学历，在父母学历较高的分组中，92.13% 的儿童普通话水平较高，而在父母学历较低的分组中，普通话水平较高的儿童仅占 47.6%，表明父母学历对儿童普通话水平的影响十分显著，受教育水平较高的父母可以为儿童语言表达能力提供更多帮助。当父母学历较低时，若儿童可以获得较丰富的学习资源，儿童普通话水平也会得到显著提高，概率提高

为 76.22%，说明学习资源也会对儿童普通话水平产生影响。由此可知，父母学历与学习资源均会对儿童普通话水平有正相关关系且存在显著影响，此结论与逻辑斯蒂回归模型结果相同。此亦说明家庭社会经济地位对儿童普通话水平有正相关关系且存在显著影响，即家庭社会经济地位越高，父母学历越高，儿童的普通话水平越高。

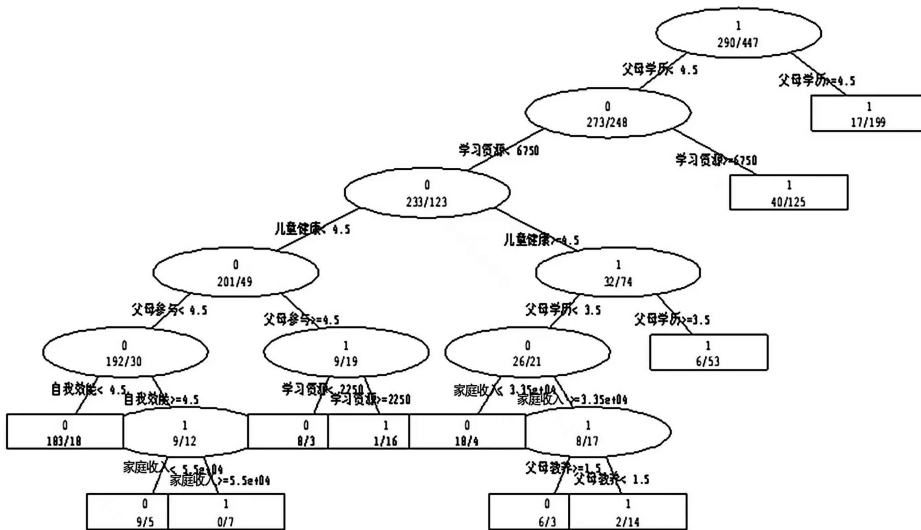


图 3 基于 CART 决策树模型的普通话水平影响因素

3. 随机森林模型分析结果

随机森林是一种分类和预测方法，已成为主流机器学习模型之一。在分析家庭社会经济地位对儿童语言能力的影响方面，随机森林不仅可以较好地处理模型中的异常值和噪声，具有很好的容忍度，而且可以进行较为精准的预测，具有很高的准确性。将全部影响因子作为解释变量后，分别对交谈语言、语文成绩和普通话水平进行随机森林模型的建立，并依据基尼指数计算每个因子对分类树每个节点上观测值的异质性的影响，对因子的重要性情况进行排序，从而分析出各模型的主要影响因素。因子的重要性随基尼指数的增大而增加，具体结果如图 4 所示。

为探索家庭社会经济地位对儿童语言能力的影响，对各解释变量进行了随机森林回归拟合与验证，图 4 显示了各因子重要性的排序。

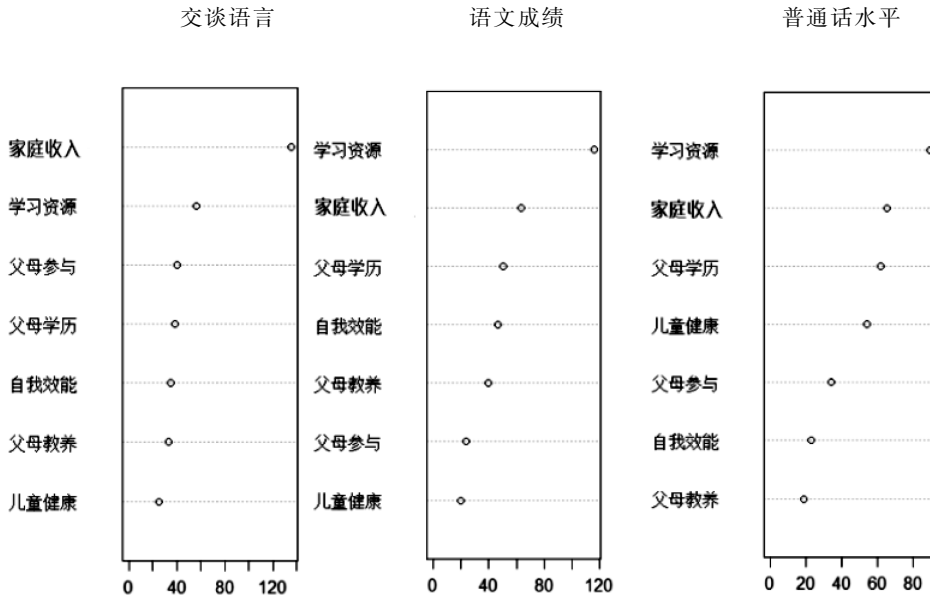


图4 基于随机森林模型的儿童语言能力影响因素重要性分析

由图4可知，儿童交谈语言重要性排序在前4位的影响因子分别是家庭收入、学习资源、父母参与和父母学历。其中家庭收入最重要，且重要性远超其他因子，基尼指数接近140%；学习资源也会对交谈语言产生影响，这与逻辑斯蒂回归模型的分析结果趋于一致。其次是父母参与和父母学历，基尼指数分别为48%和37%，这说明家庭收入越高、父母学历越高，即社会经济地位越高的家庭，对儿童使用普通话进行交谈的影响越大。若在这样的家庭中，学习资源越充足、父母参与积极度越高，则儿童就越有可能使用普通话进行交谈。

语文成绩指标依据准确性的重要性排名依次为：学习资源、家庭收入、自我效能感、父母学历、父母教养方式、父母参与和儿童健康状况。其中，学习资源是影响儿童语文成绩最主要的因子，其影响显著高于其他因子，基尼指数接近于120%，说明增加儿童的教育培训等方面的学习资源支出有利于语文成绩的提高；其次为家庭收入，这与CART决策树模型的分析结果完全一致，说明家庭收入较高且可提供丰富学习资源的家庭，即家庭社会经济地位越高，儿童语文成绩也就越高。

在运用随机森林模型对儿童普通话水平进行特征变量选择后，根据重要性指数排名，从高到低依次是学习资源、家庭收入、父母学历、儿童健康、父母参与、自我效能和父母教养方式。其中，学习资源对于儿童普通话水平

的影响程度最高,基尼指数为85%;其次是家庭收入、父母学历和儿童健康状况,三者的重要性基本相同,但显著高于其他因子,基尼指数均在55%以上。基尼指数排在后三位的依次是父母参与、自我效能感和父母教养方式,基尼指数低于35%,说明对于儿童普通话水平的影响较小。由此可知,学习资源、家庭收入和父母学历均为儿童普通话水平的主要影响因素,因此社会经济地位越高的家庭,即家庭越富裕、父母学历越高的家庭,儿童普通话水平越高。

根据上述分析,可知家庭收入、父母学历、学习资源和父母参与对儿童交谈水平、普通话水平、语文成绩具有很大影响力,而家庭收入、父母学历又与家庭社会经济地位密不可分(任春荣,2010;李珂,2016;陈晶晶,2022)。因此可知,良好家庭社会经济地位对儿童语言能力存在显著性作用,即家庭社会经济地位越高,儿童语言能力越好。

4. 模型结果的比较分析

本文运用逻辑斯蒂回归模型、CART决策树模型和随机森林模型,分别对儿童交谈语言、语文成绩和普通话水平的影响因素进行实证分析,对三者取平均数进而评判儿童语言能力。在此基础上,通过各模型的预测准确率和AUC值对三大模型进行比较分析。

根据模型变量选择结果,分别对逻辑斯蒂回归模型、CART决策树模型和随机森林模型进行预测准确率的计算,结果如表4所示。对比可知,随机森林模型的预测准确率明显高于逻辑斯蒂回归模型和CART决策树模型,且拟合效果较好。无论是在训练样本、测试样本的准确率方面,还是在全样本模型的预测中,随机森林模型的正确判别率均比CART决策树模型和逻辑斯蒂回归模型高。在儿童语言能力预测方面,随机森林模型对训练样本(70%)和测试样本(30%)的预测准确率分别高于CART决策树模型13%和6%左右;在全样本模型的预测中,随机森林模型拟合的准确率为96.17%,而逻辑斯蒂回归模型和CART决策树模型对于儿童语言能力的预测准确率分别为76.97%、85.55%,随机森林模型的预测率分别高于二者20%和10%左右。由此可知,逻辑斯蒂回归模型、CART决策树模型和随机森林模型的预测率呈现逐渐上升的态势,随机森林模型在预测能力和分析家庭社会经济地位对儿童语言能力的影响方面表现最佳,模型识别精度显著高于其他两种模型,具有一定的预测优势和应用价值,可为家庭社会经济地位对儿童语言能力影响的预测和决策提供参考。

表4 模型预测率汇总

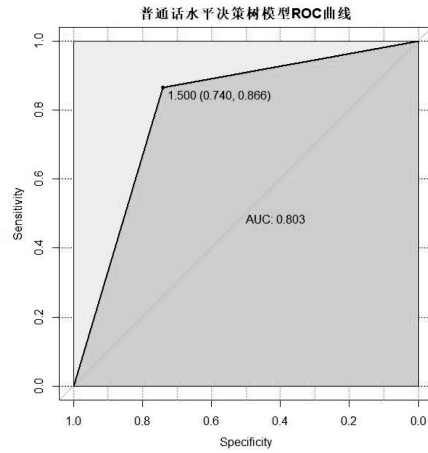
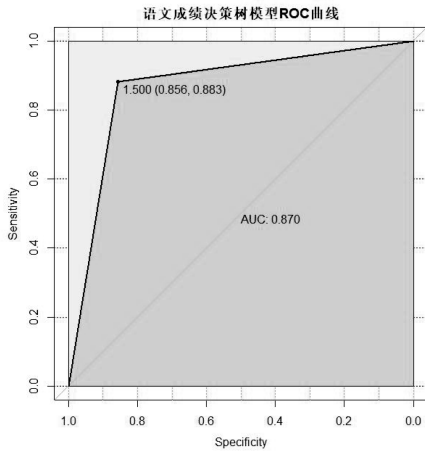
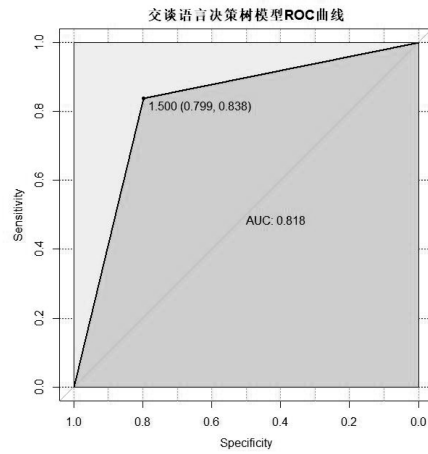
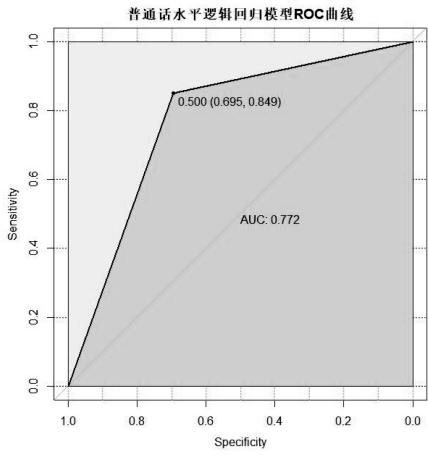
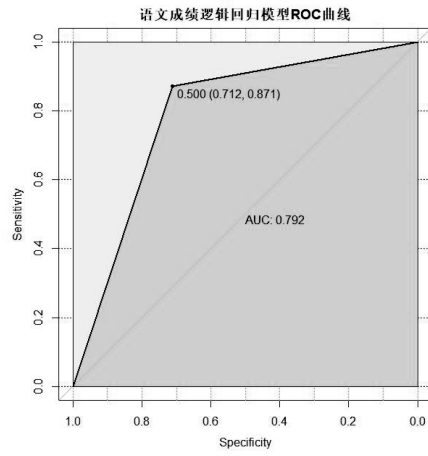
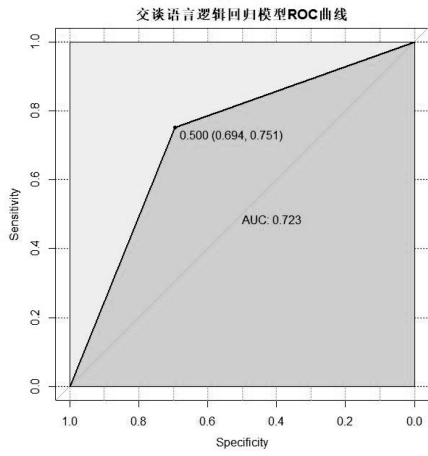
模型	交谈语言	语文成绩	普通话水平	语言习得能力
基准模型(逻辑斯蒂回归模型)	72.56%	79.81%	78.55%	76.97%
训练样本	85.07%	89.42%	85.48%	86.66%
CART 决策树模型	82.02%	87.07%	81.39%	83.49%
测试样本	82.56%	89.37%	84.72%	85.55%
总体	100%	100%	99.73%	99.91%
随机森林模型	87.38%	91.11%	89.23%	89.24%
测试样本	95.07%	97.06%	96.39%	96.17%
总体				

为进一步比较逻辑斯蒂回归模型、CART 决策树模型和随机森林模型的预测精度，分别测算了各模型的 AUC 值，结果如表 5 所示。从 AUC 来看，三大模型的 AUC 均大于 0.5，表明三个模型对家庭社会经济地位对儿童语言能力影响的分析均具有一定的预测作用。针对儿童语言能力，逻辑斯蒂回归模型的 AUC 为 0.762，CART 决策树模型的 AUC 为 0.830，随机森林模型的 AUC 为 0.869，随机森林模型的预测能力优于较逻辑斯蒂回归模型和 CART 决策树模型，与三大模型预测率的分析结果一致。进一步说明随机森林模型对家庭社会经济地位对儿童语言能力影响的识别具有一定的参考意义，且模型的拟合效果更好，预测能力更强，表现更为稳定。

表5 模型 AUC 值汇总

模型	交谈语言	语文成绩	普通话水平	语言习得能力
基准模型(逻辑斯蒂回归)	0.723	0.792	0.772	0.762
CART 决策树	0.818	0.87	0.803	0.830
随机森林	0.832	0.906	0.868	0.869

为更加直观地反映三大模型在儿童交谈语言、语文成绩和普通话水平影响因素方面的预测能力，本文绘制了各模型在各语言表达能力方面的 ROC 曲线图，如图 5 所示。结果同样表明，在儿童语言表达能力的各方面，随机森林的 ROC 曲线下方面积均高于逻辑斯蒂回归模型和 CART 决策树模型，这也说明了随机森林模型的拟合效果比逻辑斯蒂回归模型和 CART 决策树模型的拟合效果好。



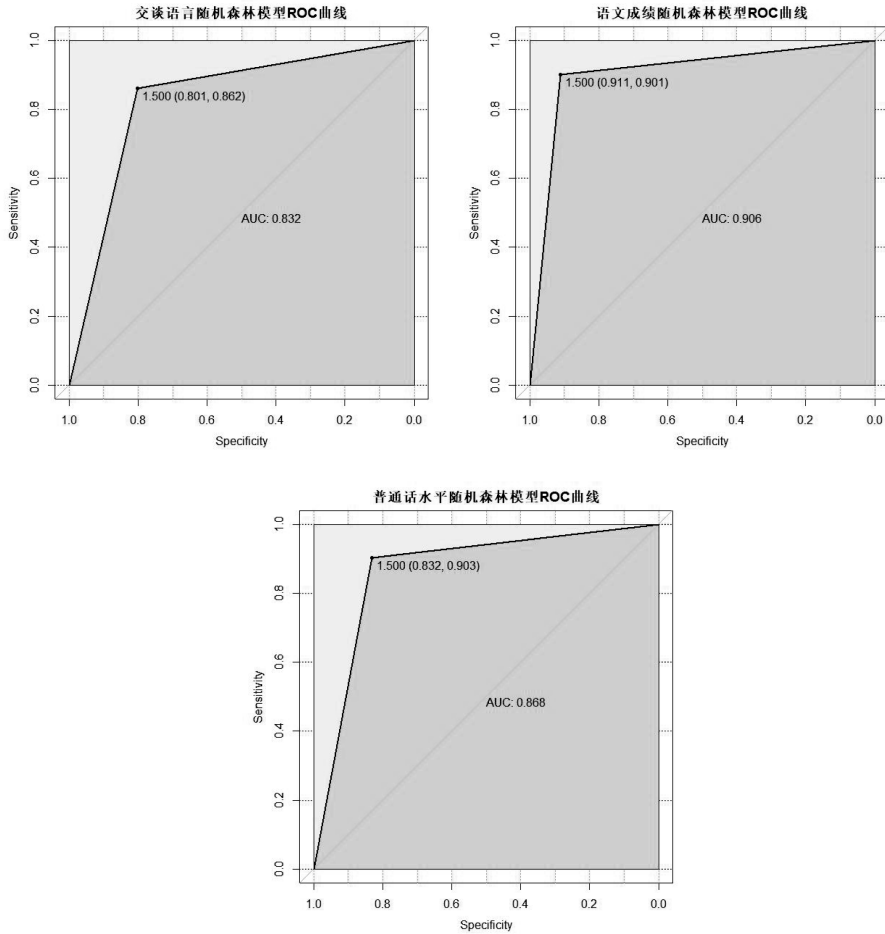


图5 逻辑斯蒂回归模型、CART决策树模型和随机森林模型的ROC曲线图

五、研究结论与对策建议

本文基于CFPS数据，利用逻辑斯蒂回归、决策树和随机森林三种方法，对家庭社会经济地位对儿童语言能力的影响进行研究，发现随机森林模型的识别精度明显高于其他二者，具有一定的预测优势和应用价值。具体而言，本文的主要研究结论如下：

第一，逻辑斯蒂回归模型结果表明家庭收入对儿童语言、语文成绩、普通话水平均存在正相关关系且有着显著性影响，该结论与CART决策树模型和随机森林数据的分析结果相同，且随机森林数据分析还说明了家庭收入是

儿童交谈语言中最重要的影响因素，其基尼指数接近140%。家庭收入通过影响语言环境，对儿童语言能力产生影响。在早期语言习得过程中，良好的语言环境能够刺激儿童认知能力的发展，推动语言能力的进步。因此，社会经济地位越高的家庭，其收入水平也越高，更有可能为儿童创设更好的语言习得环境并提供更加充足的学习资源(Jean Y. W. et al., 2002)，从而扩大儿童接触的语言材料的广度(张洁, 2019)，增加儿童接触语言的时间，进而促使儿童语言能力的提高。

第二，根据随机森林的数据显示，父母学历对儿童普通话水平的正向影响作用最大，其基尼指数达到了55%以上，说明父母的高学历背景能显著提升儿童的普通话水平。这一结论可由逻辑斯蒂回归模型结果佐证，逻辑斯蒂回归模型结果显示儿童交谈语言、语文成绩、普通话水平都和父母学历存在正相关关系，并有着显著影响。儿童在早期语言习得过程中深受语言环境的影响，语言环境包括家庭文化环境，而社会经济地位较高的家庭，父母拥有较高学历，文化素养相对较高，更有可能积极参与儿童语言习得过程并且注重教养方式(李珂, 2016)，更有可能为儿童创设浓厚的家庭文化环境和优质的语言习得环境，使儿童有机会接触到更丰富的语言材料和学习资源(张洁, 2019)，当优质的语言环境和丰富的学习资源与儿童的思维和认知发展要求相匹配时，有助于最大程度激发儿童内在驱动力(盖淑华, 2019)，促使其认知能力将外界语言信息进行整理加工，转化为自身语言知识，并进行语言运用。

第三，儿童语言能力还受到学习资源、父母参与、父母教养方式、儿童健康状况及自我效能感等因素的多重影响。

《国家中长期语言文字事业改革和发展规划纲要(2012—2020年)》在指导思想中明确提出了要“增强国家语言实力，提高国民语言能力”，尤其是近年来有关“汉语危机”的争论愈发激烈的情况下，更加需要加大传播民族文化，提升国际话语权。因此，提高国民和国家的语言能力愈发必要和迫切。而语言能力的提高，离不开语言教育的发展。实施国民语言教育是应对汉语应用能力下降、持续提高国民个体语言能力的需要。基于推动语言教育发展的角度，结合本文的研究发现，为提高儿童语言能力提出以下针对性建议。

首先，合理利用国家对于教育的财政投入。国家有力、高效、稳健的财政投入政策是对高质量教育的有力保障(柳倩和黄嘉琪, 2019)。在建立教育财政公平促进机制的同时，加大语言教育资源的投入并注重优质资源分配的公平性，以及强调高质量语言示范的提供。

其次，学校应持续增强对于语言教育的重视度。学校应意识到语言能力

是接受一切教育的基础，因此要将发展儿童语言能力作为学校教育中的一项重要任务。教师作为专门的教育工作者，应树立正确的语言能力教育理念，重视语言教材的选择，使教学内容符合儿童的认知发展水平；注重语言环境的创立，促进儿童在具体情境中的语言运用和交际能力；同时，注重语言教育中儿童自主能力的培养，培养其学习过程中的自我效能感。

最后，家庭应发挥其在儿童语言教育方面的优势。父母是儿童最早获得语言经验的对象，应在早期教育中形成以儿童为中心的正确语言教育观，从儿童自身的认知水平出发，积极创设有利于儿童口语表达的情景和良好的语言学习氛围，为儿童提供高质量的语言输入。同时，政府、社会为促进家庭语言教育提供相应的语言规划理念，语言工作者提供相应的专业指导也尤为重要。

[参考文献]

- 陈晶晶，2022：《家庭社会经济地位与初中生学业成就的关系：未来取向的中介作用与领悟社会支持的调节作用》，《中国健康心理学杂志》第5期。
- 盖淑华，2019：《基础外语学习的“语言匹配输入假说”》，《外语教学与研究》第4期。
- 高振宇，2016：《社会经济地位对儿童大脑发展的影响机制及其教育启示》，《教育发展研究》第6期。
- 郭翔飞，2012：《儿童语言习得与儿童一般认知发展规律》，《外语学刊》第4期。
- 吉晖，2019：《家庭社会经济地位对儿童语言能力发展的影响分析》，《语言文字应用》第3期。
- 李珂，2016：《儿童语言学习理论的发展及其影响因素与策略》，《学前教育研究》第7期。
- 李艳玮、李燕芳、刘丽莎，2012：《家庭收入对儿童早期语言能力的影响作用及机制：家庭学习环境的中介作用》，《中国特殊教育》第2期。
- 李毅、谭婷，2019：《家庭经济资本和中小学生学习兴趣的关系：家庭文化资本的中介作用》，《心理与行为研究》第4期。
- 李宇明，2011：《提升国家语言能力的若干思考》，《南开语言学刊》第1期。
- 李宇明，2018：《语言在全球治理中的重要作用》，《外语界》第5期。
- 刘颖，2014：《环境语言输入与儿童语言习得相关性计量研究》，《语言教学与研究》第5期。
- 刘志军、陈会昌，2005：《7岁儿童语言表达的影响因素分析》，《心理科学》第5期。
- 柳倩、黄嘉琪，2019：《中国与OECD国家学前教育投入水平的比较研究》，《教育经济评论》第3期。
- 陆俭明，2018：《语文课程所担负的任务与责任——兼说提升国民个人语言能力的意义》，

《语文建设》第25期。

罗世兰、张大均、刘云艳, 2021:《家庭社会经济地位对幼儿良好行为习惯的影响: 父母教养方式与幼儿心理素质的中介作用》,《心理发展与教育》第1期。

任春荣, 2010:《学生家庭社会经济地位(SES)的测量技术》,《教育学》第5期。

石雷山、陈英敏、侯秀等, 2013:《家庭社会经济地位与学习投入的关系: 学业自我效能的中介作用》,《心理发展与教育》第1期。

魏晖, 2015:《国家语言能力有关问题探讨》,《语言文字应用》第4期。

文秋芳, 2016:《国家语言能力的内涵及其评价指标》,《云南师范大学学报(哲学社会科学版)》第2期。

文秋芳、杨佳, 2020:《提升国家语言能力, 助推两个共同体建设》,《语言文字应用》第4期。

夏征农、陈至立, 2015:《大辞海》, 上海: 辞书出版社。

谢宇、张晓波、李建新等, 2014:《中国民生发展报告》, 北京: 北京大学出版社。

袁言云、吴妙霞、王志航等, 2020:《家庭社会经济地位与儿童一般自我效能感的关系: 父母关爱与应对方式的链式中介作用》,《中国临床心理学杂志》第5期。

张洁, 2019:《国外贫困与儿童语言发展研究的回顾与展望》,《语言战略研究》第1期。

张茜洋、冷露、陈红君等, 2017:《家庭社会经济地位对流动儿童认知能力的影响: 父母教养方式的中介作用》,《心理发展与教育》第2期。

张云运、骆方、陶沙等, 2015:《家庭社会经济地位与父母教育投资对流动儿童学业成就的影响》,《心理科学》第1期。

Bradley, R. H. and R. F. Corwyn, 2002, "Socioeconomic Status and Child Development", *Annual Review of Psychology*, (53), 371-399.

Dahl, G. B. and L. Lochner, 2012, "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit", *American Economic Review*, 102(5), 1927-1956.

Gardner, R. C., 2010, *Motivation and Second Language Acquisition: The Social Educational Model*, New York: Peter Lang Publishing, Inc..

Gonzalez-DeHass, A. R., et al., 2005, "Examining the Relationship between Parental Involvement and Student Motivation", *Educational Psychology Review*, 17(2), 99-123.

Jean, Y. W., R. L. Miriam and B. Jeanne, 2002, "How Money Matters for Young Children's Development: Parental Investment and Family Processes", *Child development*, 73(6), 1861-1879.

Pace, A., R. Luo and K. Hirsh-Pasek, et al., 2017, "Identifying Pathways between Socioeconomic Status and Language Development", *Annual Review of Linguistics*, 3(1), 285-308.

Raviv, T., M. Kessenich and F.J. Morrison, 2004, "A Mediation Model of the Association between Socioeconomic Status and Three-year-old Language Abilities: The Role of Parenting Factors", *Early Childhood Research Quarterly*, 19(4), 528-547.

A Research on the Influence of Family Socioeconomic Status on Children's Language Competence

HOU Xiao-zhou¹, CHANG Wan-ning², CHEN Na-na³

(1: School of Foreign Studies, Hebei University; 2: Hebei North University;

3: College of Education, Hebei University)

Abstract: Improving children's language competence is a major proposition for the construction of national language competence. This paper selects the 2018 questionnaire of the CFPS database and uses logistic regression, classification and regression tree and random forest methods to conduct empirical analysis on the influence family socioeconomic status on children's language ability. The analysis results show that both of the family affluence and parental education have a positive correlation and significant impact on children's language competence. Based on the results, this paper puts forward some specific suggestions to improve children's language competence.

Key words: children's language competence; family socioeconomic status; random forest method

(责任编辑: 刘泽云 责任校对: 刘泽云 胡咏梅)