

理工科硕士生数据素养现状及影响因素的实证研究

李舸鸣, 刘继安, 戚 佳

(中国科学院大学 公共政策与管理学院, 北京 100049)

摘要:基于社会生态系统理论,从个体和学校视角出发,实证分析理工科硕士生数据素养的现状及其影响因素。结果表明:理工科硕士生数据素养的总体水平较强,在数据意识和数据伦理方面表现得更好,对数据的获取、分析、存储和展示等技能掌握得较差,且个体的年级、学习态度、学术交流、学习经验和学校的教学培训会对数据素养产生显著影响。据此,建议学生要加强对数据素养的认识,提高学习主动性,学校要开展多层次、多样化数据素养教育。

关键词:数据素养;理工科硕士生;影响因素;科技创新

中图分类号:G64 **文献标志码:**A **文章编号:**1671-1807(2022)04-0264-07

互联网的快速发展及大数据的广泛应用使科学研究范式向“数据密集型科学”转变,主要表现为以数据为基础开展科学研究^[1]。在数据密集型的科研环境下,科学活动越来越依赖于对数据的管理与利用,数据不再仅仅用于支持和辅助科学研究,还能够驱动甚至是引领科学研究^[2]。数据素养作为数据管理与利用的核心概念,已经成为从事科学研究活动的必备技能^[3]。与此同时,建设数字中国是中国未来发展的一个重要目标,亟须一批擅长使用数据解决科技问题的创新人才。理工科研究生作为高层次人才的生力军,对数据的敏感度以及所掌握的数据管理与利用能力对于推动科技创新和社会发展至关重要。而硕士阶段是训练研究生科研能力的主要阶段,正是培养数据素养的黄金时期,所以研究数据素养的现状及其影响因素对于理工科硕士生有目的地提升数据素养、高校更好地开展数据素养教育具有重要意义。

目前国内外关于数据素养的研究已经取得了一定成果,但仍然存在某些局限性。①以往研究对象主要以高校科研人员为主,聚焦于硕士生的研究相对较少;②虽然部分学者对数据素养的影响因素进行了实证研究,但大多数研究仅围绕学生或学校的单一视角展开,并没有结合两种视角系统性地探讨数据素养的影响因素。因此,在前人理论分析以及在校硕士生访谈结果的基础上,使用问卷调查法

对理工科硕士生的数据素养现状进行探究,并从个体层面和学校层面探讨数据素养的影响因素,从而基于实证结果为理工科硕士生数据素养的培养提出可行性建议。

1 文献综述与研究假设

1.1 数据素养理论内涵

数据素养的概念最早出现并应用于美国教育界,是大数据密集时代背景下对信息素养概念的内涵延伸。信息素养是一种能力,强调对信息的反思性探索、正确地理解和评价以及规范地创造和使用^[4],数据是信息的一种类型,可以将其视为信息素养的一部分^[5],但是信息素养更强调对信息的查找、评估和应用,而数据素养除了关注查找和评估能力,更注重对数据的管理、分析和应用^[6],且数据处理的难度要远大于其他信息类型^[7]。

目前学界对数据素养尚未达成统一的定义,但绝大多数研究都基于数据生命周期视角来阐释数据素养的理论内涵。有学者将数据素养简单地视为科研人员开展科学研究所需掌握的数据技能。Jian 等和 Prado 等认为数据技能包含数据的收集、处理、评估及使用能力^[8-9]。秦小燕等和 Maybee 等则指出数据的存储与交流也是数据技能不可或缺的组成部分^[10-11]。然而,数据素养并不只是数据管理与利用的能力合集,还强调科研人员要具备数据意识,即能够认识到数据是一种重要的科研资

收稿日期:2021-12-22

基金项目:国家社科基金教育科学“十三五”规划国家一般课题(BIA170171)。

作者简介:李舸鸣(1996—),女,辽宁抚顺人,中国科学院大学公共政策与管理学院,硕士研究生,研究方向为高等教育管理;通信作者刘继安(1968—),女,北京人,中国科学院大学公共政策与管理学院,教授,博士,研究方向为高等教育管理;戚佳(1999—),女,重庆人,中国科学院大学公共政策与管理学院,硕士研究生,研究方向为高等教育管理。

源^[12-13],从而更好地使用数据解决科学问题。同时,为了保证科研成果的客观性和真实性,数据素养也要求科研人员要在生产和使用数据的过程中遵守相关的道德伦理及行为准则^[14]。综上所述,数据素养是指科研人员具有数据意识,在符合数据伦理的基础上使用数据技能解决科学问题的能力,其中数据技能包括数据的获取、分析、存储与展示4个方面。

1.2 理论框架与数据素养的影响因素

1.2.1 理论框架

社会生态系统理论指出,个体的行为会受到生存环境的影响,包括宏观系统、中观系统和微观系统3个部分。宏观系统指社会经济制度和文化等,中观系统指家庭、学校等小规模群体,微观系统指个人的生物、心理或社会特征^[15]。基于此,从中观系统和微观系统出发,探究个体特征和学校环境对硕士生数据素养所产生的影响。

1.2.2 个体层面的影响因素

以往的研究发现,由于男性对计算机表现出的兴趣明显高于女性,且更容易辨别信息的有效性,因此男性要比女性更擅长获取和评价信息^[16]。而数据作为信息的一种表现形式,男性可能比女性具有更强的数据素养。但是,郭倩等在研究科研人员的科学数据素养时发现,男性和女性的科学数据素养不存在显著差异^[17]。由此看来,性别是否会对理工科硕士生数据素养产生影响还有待论证,所以提出以下研究假设。

H1:性别对理工科硕士生数据素养有显著影响,男性的数据素养高于女性。

有学者认为不同学科对学生的数据素养有不同的要求,使得不同学科学生的数据素养表现出差异性。岳敏敏等关于研究生数据素养现状的研究结果显示,理工科学生和人文社科学生的数据能力存在明显差异^[18]。同样地,郝媛玲等在对高校图书情报人员的访谈中发现,学科背景是影响个体数据素养最主要的因素之一,统计学和数学等理学专业背景的学生对数据更为敏感,且对数据知识和方法的应用更为熟练^[19],所以对数学基础要求更高的理科类学生的数据素养要高于工科类学生的数据素养。基于此,提出以下研究假设。

H2:学科对理工科硕士生数据素养有显著影响,理科类学生的数据素养高于工科类学生。

代沁泉等针对 iSchool 院校本科生和研究生数据素养问卷调查的结果显示,高年级学生在数据意

识、数据的获取、挖掘、存储、利用等技能的得分要高于低年级学生^[20],所以年级会对学生的数据素养产生影响。同时,余维杰等基于研究生数据素养现状的调查研究也证实,相较于低年级学生而言,高年级学生在不同的科研阶段均表现出较强的数据素养^[21]。据此,提出以下研究假设。

H3:年级对理工科硕士生数据素养有显著影响,高年级学生的数据素养高于低年级学生。

除客观因素外,个体的学习态度和学术交流等主观因素也会影响数据素养。学习态度是指个体主动提升数据素养的意向^[22]。许胜江认为主观情绪是造成硕士生缺失数据能力的原因之一^[23]。毕达天和曹冉通过实证研究证明积极的学习态度会使科研人员具有更强的内在动机,促使其更加主动地学习数据知识和技能,从而提升个人的数据素养^[24]。学术交流是指科研人员在遇到数据问题时,能够获得的身边人的有益帮助和指导。迟玉琢的研究表明学术交流对数据素养能力产生显著正向影响^[25]。综上所述,提出如下假设。

H4:学习态度对理工科硕士生数据素养有显著正向影响。

H5:学术交流对理工科硕士生数据素养有显著正向影响。

在围绕信息素养影响因素的研究中有学者发现,高中阶段选修过信息类课程的学生,在大学阶段会表现出更强的信息素养^[26]。这是因为学习者会将某一领域的先前学习经验带入到新的学习阶段,从而促使知识发生迁移,帮助学习者构建新的知识体系^[27]。由于信息素养和数据素养存在许多共同之处,所以硕士阶段以前的学习经验可能也会对理工科学生的数据素养产生影响,提出如下研究假设。

H6:相关学习经验对理工科硕士生数据素养有显著正向影响。

1.2.3 学校层面的影响因素

高校是培养学生数据素养的重要主体,所提供的教学培训和教学资源会对学生的数据素养产生影响。教学培训是指学生所接受的提升数据素养的专业训练,经过专业训练后,学生的数据素养会得到提高^[28]。而教学资源是指学校开展数据素养教育的配套设施,包括用于数据管理和利用的数据库及软件库等^[29]。郭倩和李建霞的研究已经证实学生数据技能的培养离不开学校教学资源的投入^[30]。基于此,提出如下假设。

H7:相关教学培训对理工科硕士生数据素养有显著正向影响。

H8:相关教学资源对理工科硕士生数据素养有显著正向影响。

综上所述,研究框架如图 1 所示。

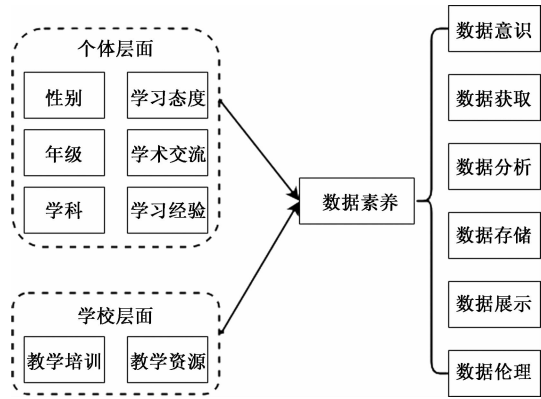


图 1 研究框架

2 研究设计与数据收集

2.1 研究方法

采用问卷调查方法,在文献调研的基础上编制理工科硕士生数据素养现状及影响因素的调查问卷。运用 SPSS26.0 和 Amos23.0 对问卷结果进行数据分析:SPSS26.0 用于变量的信度检验、探索性因子分析、描述性统计分析及回归分析;Amos23.0 用于变量的验证性因子分析,对问卷质量进行评估。

2.2 问卷设计

基于上文提出的研究框架,参考和借鉴国内学者编制的量表,并结合专家意见对题项加以调整和修改,形成初步的调查问卷。随后,将调查问卷发放给 46 名同学进行预测试,根据测试结果及反馈意见再次修改问卷,最终得到正式的调查问卷。

调查问卷由 4 个部分组成。第 1 部分为调查对象的基本信息,包括性别、学科和年级。第 2 部分关于理工科硕士生的数据素养现状,采用李克特五级量表,从数据意识、数据获取、数据分析、数据存储、数据展示及数据伦理 6 个维度测量数据素养,其中,数据意识、数据存储和数据伦理的题项改编自张晓阳等^[31]的问卷,数据获取、数据分析和数据展示的题项改编自隆茜^[13]的问卷。第 3 部分是对个体主观因素的测量,采用李克特五级量表的形式,其中,学习态度和学术交流的题项根据迟玉琢^[25]的问卷进行修订,学习经验为自编题项。第 4 部分是对学校层面影响因素的测量,教学资源的题项改编自沈玖玖等^[29]的问卷,采取李克特五级量表的形式;教

学培训为自编题项,指学生参与教学培训的程度,分别赋值 0~8。

2.3 数据来源

U 大学是一所依托国家科研机构办学,以理工学科为主的研究型大学,涵盖全部理学一级学科和工学的大部分学科,且研究生规模大,目前有 2.6 万多名在校硕士生,故在 U 大学中抽样选取硕士生作为调查对象具有一定的代表性和可操作性。

由于 U 大学二年级及以上研究生分散在各研究所,因受疫情的影响,现场调研存在一定难度,所以依托问卷星平台,使用分层抽样和方便抽样相结合的方法,以“线上+线下”结合的方式发放调查问卷。线上问卷在学科类别内按滚雪球方式发放,先发放给相关领域熟悉的同学和老师,再请他们帮忙发放给其他同学的方式,共得到 337 份问卷;线下问卷通过在工位、宿舍楼等随机寻找调查对象的方法,共得到 223 份问卷。两种形式共发放问卷 560 份,回收问卷 560 份,剔除作答时间过短、非目标院校、非目标专业等无效问卷 58 份,最终保留有效问卷 502 份,有效回收率为 89.6%。

3 数据分析

3.1 信效度检验

3.1.1 信度检验

李克特五级量表常用克隆巴赫系数(Cronbach' α),即内部一致性系数来检验信度,该系数越大说明量表的信度越高。吴明隆提出,在社会科学领域中,使用者既要提供总量表的信度系数,也要提供各测量变量的信度系数:针对测量变量的量表, $\alpha \geq 0.50$ 为可以接受, $\alpha \geq 0.70$ 为适中, $\alpha \geq 0.80$ 为良好, $\alpha \geq 0.90$ 为优秀;针对整体量表, $\alpha \geq 0.60$ 为可以接受, $\alpha \geq 0.70$ 为适中, $\alpha \geq 0.80$ 为良好, $\alpha \geq 0.90$ 为优秀^[32]。采用 SPSS26.0 对数据素养、学习态度、学术交流及教学资源 4 个测量变量及整体量表进行信度检验,整体量表的 Cronbach' α 为 0.924,各层面量表的 Cronbach' α 均大于 0.7,说明问卷信度较好。

3.1.2 效度检验

首先检验问卷的内容效度。问卷建立在充分的文献梳理基础之上,且经过了专家评估和预测试的两轮修改和调整,确保了题目分布的合理性,从而保证问卷具有良好的内容效度。

其次,由于学习态度、学术交流、教学资源以及数据素养 4 个变量均由量表题进行测量,且包含部分修改题项以及自主设计题项,因此在构建回归模型之前有必要进行因子分析。通过 KMO 和 Bart-

lett 球型检验,4 个变量的 KMO 值均大于 0.6,说明可以进行因子分析^[33]。对学习态度、学术交流、教学资源而言,探索性因子分析结果显示各题项的因子载荷均在 0.7 以上,分别能够解释总方差的 66.32%、71.95%、68.40%,说明学习态度、学术交流和教学资源构念具有良好的结构效度;对数据素养而言,探索性因子分析结果显示该变量为六维结构,各题项的因子载荷均在 0.5 以上,共解释总方差

的 76.53%,说明数据素养构念具有良好的结构效度。再次,为了检验 4 个变量的区分效度和收敛效度,使用 Amos23.0 进行验证性因子分析,结果表明:各变量 AVE 值均在 0.4 以上、CR 值均在 0.7 以上,说明量表的收敛效度可以接受^[34];各变量之间的相关性显著,相关系数均小于 0.5,且所有相关系数都小于 AVE 的平方根,说明量表具有良好的区分效度^[35],结果见表 1。

表 1 区分效度和收敛效度

变量	数据素养	学习态度	学术交流	教学资源	AVE	CR	AVE 平方根
数据素养	0.531				0.531	0.941	0.728
学习态度	0.105	0.493			0.493	0.741	0.773
学术交流	0.100	0.227	0.598		0.598	0.815	0.702
教学资源	0.072	0.166	0.182	0.564	0.564	0.789	0.751

3.2 描述性统计分析

3.2.1 样本基本概况

整体而言,调查对象在性别、学科以及年级上的分布较为平均,见表 2。从性别来看,男性学生占 51.0%,女性学生占 49.0%;从学科来看,理学学生占 53.4%,工学学生占 46.6%;从年级来看,硕士一年级占 40.6%,硕士二年级占 33.7%,硕士三年级占 25.7%。

表 2 样本基本概况

题项	选项	频率	占比/%	总计
性别	男	256	51.0	502
	女	246	49.0	
学科	理学	268	53.4	502
	工学	234	46.6	
年级	一年级	204	40.6	502
	二年级	169	33.7	
	三年级	129	25.7	

3.2.2 数据素养及影响因素的得分概况

理工科硕士研究生的数据素养水平通过学生自评的数据素养得分进行衡量。数据素养共有 6 个维度,取值为每个维度的平均值,满分为 5 分。从表 3 中可以看出,理工科硕士生数据素养整体水平较好。学生认为自己在数据意识和数据伦理方面表现得更加出色,在数据的获取、分析、存储、展示等数据技能方面表现得相对较差,其中数据分析的能力最为薄弱。

就影响因素的得分情况而言,从个体主观因素来看,理工科硕士生认为自己针对数据知识和技能开展的学术交流较为丰富,且有一定的学习经验,但是学习态度不够积极;从学校层面因素来看,理

工科硕士生对学校所提供的教学资源比较满意,但是参与教学培训的程度不高。

表 3 数据素养及影响因素得分

变量		平均值	最小值	最大值	标准差
数据素养	数据意识	4.02	2.00	5.00	0.62
	数据获取	3.76	2.00	5.00	0.63
	数据分析	3.24	1.00	5.00	0.81
	数据存储	3.88	1.33	5.00	0.71
	数据展示	3.76	1.00	5.00	0.71
	数据伦理	4.39	2.00	5.00	0.60
	数据素养	3.84	1.96	4.94	0.52
影响因素	学习态度	3.47	1.00	5.00	0.69
	学术交流	3.93	1.33	5.00	0.68
	学习经验	3.77	1.00	5.00	0.96
	教学培训	3.67	0.00	8.00	1.76
	教学资源	3.48	1.00	5.00	0.85

3.3 多元线性回归分析

首先对数据素养和不同层面的影响因素进行多重共线性检验,结果显示各变量的方差膨胀因子平均值为 1.281,最大方差膨胀因子值为 1.510,容差均大于 0.1,说明变量间的多重共线性问题较小,满足做回归分析的假定条件。基于此,采用多元线性回归分析方法检验个体客观因素、个体主观因素以及学校层面因素是否对理工科硕士生的数据素养产生显著影响,构建模型(1)至模型(4)(表 4)。模型(1)检验性别、学科、年级对数据素养产生的影响;模型(2)在控制个体客观因素的基础上,检验学习态度、学术交流、学术经验对数据素养产生的影响;模型(3)在控制个体客观因素的基础上,检验教学培训与教学资源对数据素养产生的影响;模型(4)检验个体层面和学校层面因素对数据素养产生的共同影响。

表 4 不同因素对数据素养的影响

影响因素		模型(1)		模型(2)		模型(3)		模型(4)	
		β	SE	β	SE	β	SE	β	SE
性别(男性为参照组)		-0.037	0.041	-0.008	0.032	-0.053	0.039	-0.015	0.032
学科(理学为参照组)		-0.044	0.041	0.011	0.032	-0.079	0.040	-0.010	0.032
年级(一年级为参照组)	二年级	0.377***	0.047	0.208***	0.038	0.331***	0.046	0.202***	0.038
	三年级	0.520***	0.051	0.280***	0.043	0.476***	0.050	0.282***	0.043
学习态度				0.160***	0.027			0.142***	0.028
学术交流				0.239***	0.027			0.223***	0.027
学习经验				0.367***	0.019			0.360***	0.019
教学培训						0.168***	0.012	0.038**	0.010
教学资源						0.168***	0.024	0.083	0.021
R^2		0.249	0.549	0.315	0.557				
F		41.094***	85.939***	37.942***	68.858***				
ΔR^2		0.249***	0.301***	0.066***	0.008***				
ΔF		41.094***	109.760***	24.023***	4.642**				

注：* 表示 $P<0.05$ ；** 表示 $P<0.01$ ；*** $P<0.001$ 。

模型(1)的结果显示,在个体客观层面的影响因素中,性别($\beta=-0.037, P>0.05$)和学科($\beta=-0.044, P>0.05$)对理工科硕士生数据素养的影响在 5%的显著水平上不显著,而年级对数据素养产生的影响显著,可以解释数据素养 24.9%的变异。与一年级学生相比,二年级($\beta=0.377, P<0.001$)和三年级($\beta=0.520, P<0.001$)学生的数据素养更强。模型(2)的结果显示,在控制个体客观因素后,个体层面因素能够解释数据素养 54.9%的变异,学习态度($\beta=0.160, P<0.001$)、学术交流($\beta=0.239, P<0.001$)、学习经验($\beta=0.367, P<0.001$)对理工科硕士生数据素养产生显著正向影响。模型(3)的结果显示,在控制个体客观因素后,纳入学校层面的影响因素使得回归模型能够解释数据素养 31.5%的变异,教学培训($\beta=0.168, P<0.001$)和教学资源($\beta=0.168, P<0.001$)对数据素养产生显著正向影响。

从模型(4)的结果来看, R^2 和 F 值的改变量均在 1%的显著水平上显著,同时纳入个体层面和学校层面影响因素的回归模型能够解释数据素养 55.7%的变异。此时,只有年级、学习态度($\beta=0.142, P<0.001$)、学术交流($\beta=0.223, P<0.001$)、学习经验($\beta=0.360, P<0.001$)以及教学培训($\beta=0.038, P<0.05$)对数据素养仍然产生显著正向影响,而教学资源($\beta=0.083, P>0.05$)对数据素养的影响在 5%的显著水平上不显著。此外,从标准化回归系数来看,年级、学习态度、学术交流及学习经验的标准化回归系数的绝对值较大,说明对数据素养有较高的解释力,而教学培训的标准化回归系数的绝对值相对较小,说明对数据素养的解

释力较低。综上所述,研究假设 H3~H7 得到验证,假设 H1、H2、H8 未通过验证。

4 结论与建议

4.1 研究结论

4.1.1 数据素养现状

从数据素养的描述性统计分析结果可以看出,理工科硕士生对个人数据素养的总体水平比较认可,认为自己的数据伦理能力最强,数据分析能力最弱。这是因为自翟天临、曹雪涛等学术不端事件发生以来,国内高校愈发重视对学生的数据伦理教育,专门开设了相关的必修课程,如 U 大学的《学术道德与学术写作规范》等。相比之下,针对学生数据技能的培养,尽管国内部分高校开设了数据处理和分析的课程,但是教学内容以面向全校学生的 Excel、CiteSpace 等基础数据分析软件为主^[36],而不同人群对数据技能的需求有所不同^[37],所以学校的数据素养教育体系要根据教学对象需求的差异性进一步调整和完善。

4.1.2 个体层面因素对数据素养的影响

就个体客观因素而言,只有年级会对理工科硕士生的数据素养产生显著影响,且年级越高的学生,数据素养越强。这是因为与低年级学生相比,高年级学生从事科学研究的时间更长,有机会得到更多的实践与锻炼,所以表现出更强的数据素养。相反,性别不会对理工科硕士生的数据素养产生影响,这进一步支持了郭倩等^[17]的研究结果;理工科硕士生所处的学科领域也不会对数据素养产生影响,这与郝媛玲等^[19]的访谈结论不一致,还有待扩大样本进一步论证。

就个体主观因素来看,学习态度、学术交流以及学习经验均会对理工科硕士生的数据素养产生

显著正向影响,即学生的学习态度越积极、学术交流越多、学习经验越丰富,其数据素养会越强。积极的学习态度会使学生对数据素养有更高的需求,促使其在学习数据知识和技能上投入更多的时间和精力,不断提升个人的数据素养;好的学术交流意味着学生身处数据素养较高的环境中,在遇到数据问题时,能够及时获得有益的指导和帮助,从而在解决问题的过程中掌握更多的数据知识和技能;而与缺少学习经验的学生相比,学习经验丰富的学生在面对数据问题时更有信心,也更容易掌握进阶的数据知识和技能,追求更高的数据素养。

4.1.3 学校层面因素对数据素养的影响

从多元线性回归分析的结果可以看出,当学校层面的影响因素单独作用于数据素养时,教学培训和教学资源均对理工科硕士生的数据素养产生显著正向影响,而当个体层面和学校层面的影响因素共同作用于数据素养时,只有教学培训对数据素养的影响达到显著水平。这与以往的研究结论不一致,可能的原因是:硕士生的科研学习主要以课题组为单位,学习资源更多地来源于课题组;而现有研究的调查对象大多为本科生,学校教学为本科生学习的主要方式,其学习资源更多依赖于学校的供给,据此猜测硕士生的数据素养可能更受课题组资源的影响,但是由于问卷对教学资源的测量主要针对学校层面,所以此结论还有待进一步检验。

回归分析的结果还显示,相比于学校层面的影响因素,个体层面的影响因素对数据素养有更高的解释力,可能的原因如下:根据社会学习理论,虽然个体的行为会受外界环境的影响,但更重要的是受自我引发的行为结果的影响^[38],也就是说学生通过教学培训习得的数据知识和技能,还要经过个体不断地练习与实践,才能真正将其转化为个人的数据素养。

4.2 对策建议

基于上述研究结论,提出以下建议:

从个体层面来看,“师傅领进门,修行靠个人”,理工科硕士生要加强对数据素养重要性的认识,提高学习的主动性。首先,端正学习态度。研究生学习具有较强的自主性,学生对数据素养的提升不能一味地依靠“填鸭式”教学,要意识到数据是科学研究必不可少的组成部分,主动地参加培训或讲座,学习相关的知识与技能。其次,增强与老师、师兄师姐、同学的交流与合作。通过与他人的交流与合作,个体能够获得相关的经验和有益的帮助,从而提升个人的数据素养。最后,坚持不断地学习与积

累。数据素养的培养是一个循序渐进的过程,不会一蹴而就,学生要摆脱“老师让学才学,科研需要才学”的观念,在日常学习中就重视培养自己的数据素养,才能更好地使用数据解决科学问题,提升科研效率。

从学校层面来看,高校要根据教学对象的差异性开展多层次、多样化的数据素养教育。在教学形式上,课程修读是低年级学生的关注重点,所以面向该群体的数据素养教育要以必修课和选修课的形式为主,而高年级学生的关注重点是自己的研究课题,所以面向高年级学生的数据素养教育要具有更强的灵活性,可以把讲座和培训作为主要的教学形式。在教学内容上,面向低年级学生开设基础的数据素养课程,重点介绍数据的理论知识以及数据分析软件的应用等;面向高年级学生提供进阶的数据素养培训项目与实践机会,使其能够结合个人的研究方向,有针对性地提升个人的数据素养。在教学队伍上,图书馆作为高校提供数据素养教育的重要单元,可以与专业课教师和导师共同组成一支教学团队,由图书馆馆员教授基本的知识和方法,由专业课教师结合所学专业讲授具体的应用技能,由导师提供科研实践的机会,从而全方位地支持学校的数据素养教育。

参考文献

- [1] HEY T, TANSLEY S, TOLLE K. The fourth paradigm: Data-intensive scientific discovery[J]. Proceedings of the IEEE, 2011, 99(8): 1334-1337.
- [2] 李树深. 数据与计算是科技创新的巨大驱动力[J]. 数据与计算发展前沿, 2019, 1(5): 1-1.
- [3] 孟祥保, 常娥, 叶兰. 数据素养研究: 源起、现状与展望[J]. 中国图书馆学报, 2016, 42(2): 109-126.
- [4] ACRL. Framework for information literacy for higher education[EB/OL]. (2015-02-09) [2020-10-14]. <http://www.ala.org/acrl/standards/ilframework>.
- [5] SCHNEIDER R. Research data literacy[C]// European Conference on Information Literacy. Springer, 2013.
- [6] SCHIELD M. Information literacy, statistical literacy and data literacy[J]. IASSIST Quarterly, 2004, 28(2): 7-14.
- [7] THOMPSON K A, EDELSTEIN D M. A reference model for providing statistical consulting services in an academic library setting[J]. IASSIST Quarterly, 2005, 28(2): 35.
- [8] JIAN Q, D'IGNAZIO J. Lessons learned from a two-year experience in science data literacy education[C]// 31st Annual IATUL Conference. International Association of Scientific and Technological University Libraries. Purdue University, 2010.
- [9] PRADO J C, MARZAL M Á. Incorporating data literacy

- into information literacy programs:Core competencies and contents[J]. *Libri*,2013,63(2):123-134.
- [10] 秦小燕,初景利. 基于 ITE-KSA 结构的科学数据素养能力指标体系研究[J]. *图书与情报*,2019(1):115-124.
- [11] MAYBEE C, ZILINSKI L. Data informed learning: A next phase data literacy frame work for higher education [R/OL]. [2020-10-14]. https://docs.lib.purdue.edu/lib_fspr/95/.
- [12] 孟祥保,李爱国. 国外高校图书馆科学数据素养教育研究[J]. *大学图书馆学报*,2014,32(3):11-16.
- [13] 隆茜. 数据素养能力指标体系构建及高校师生数据素养能力现状调查与分析[J]. *图书馆*,2015(12):51-56,62.
- [14] 沈婷婷. 数据素养及其对科学数据管理的影响[J]. *图书馆论坛*,2015,35(1):68-73.
- [15] ZASTROW C, KIRST-ASHMAN K. Understanding human behavior and the social environment[M]. Cengage Learning,2006.
- [16] BARO E E, FYNEMAN B. Information literacy among undergraduate students in Niger Delta University[J]. *Electronic Library*,2009,27(4):659-675.
- [17] 郭倩,李建霞. 基于多元主体的高校数据素养教育生态模式构建研究[J]. *图书馆理论与实践*,2019(5):49-57.
- [18] 岳敏敏,董同强. 高校博士生数据素养发展现状及提升策略研究[J]. *图书馆学研究*,2021(17):2-9.
- [19] 郝媛玲,沈婷婷,高珊. 高校数据素养教育实践的思考和建设:基于哈佛大学案例和我国图书情报人员访谈的分析[J]. *图书情报工作*,2015,59(12):44-51.
- [20] 代沁泉,熊回香,陈琦,等. 中国内地 iSchool 院校学生数据素养评价研究及启示[J]. *图书馆学研究*,2020(17):16-27.
- [21] 余维杰,周娅莉,吴锦池. 我国研究生在科研活动中的数据素养现状研究:以双生命周期理论为视角[J]. *图书情报工作*,2020,64(7):84-93.
- [22] 杨雨琪. 国内高校大学生信息素养影响因素研究:元人种志方法的应用[J]. *新世纪图书馆*,2020(10):27-33.
- [23] 许胜江. 数据能力:硕士研究生亟需加强的基本功[J]. *学位与研究生教育*,2007(9):29-32.
- [24] 毕达天,曹冉. 科研人员数据素养影响因素分析:基于 SEM 及 fsQCA 方法[J]. *情报学报*,2021,40(1):11-20.
- [25] 迟玉琢. 科研人员科学数据能力影响因素实证研究[J]. *情报探索*,2019(7):7-16.
- [26] FABBI J L. Fortifying the pipeline: A quantitative exploration of high school factors impacting the information literacy of first-year college students[J]. *College & Research Libraries*,2015,76(1):31-42.
- [27] BRANSFORD J D, BROWN A L, COCKING R R. How people learn: Brain, mind, experience, and school [M]. National Academy Press,1999.
- [28] 张艳梅. 用户数据素养教育视角下的图书馆科学数据管理研究[J]. *图书与情报*,2015(4):139-141,109.
- [29] 沈玖玖,徐萍,艾文华. 高校数据素养教育对大学生数据素养的影响分析:以南昌大学为例[J]. *图书馆学研究*,2019(7):2-14.
- [30] 郭倩,李建霞. 聚焦安全伦理教育的高校学生数据素养能力实证研究[J]. *新世纪图书馆*,2019(5):14-19.
- [31] 张晓阳,李楣. 基于胜任特征的研究生数据素养能力测评量表研究[J]. *图书情报工作*,2017,61(8):89-95.
- [32] 吴明隆. 问卷统计分析实务:SPSS 操作与应用[M]. 重庆:重庆大学出版社,2010:244-245.
- [33] KAISER H F. An index of factorial simplicity[J]. *Psychometrika*,1974,39(1):31-36.
- [34] FORNELL C, LARCKER D F. Evaluating structural equation models with unobservable variables and measurement error[J]. *Journal of Marketing Research*,1981,18(3):337-346.
- [35] HAIR J F, BLACK W C, BABIN B J, et al. Multivariate data analysis: A global perspective[M]. Seventh Edition. Pearson,2010.
- [36] 宋甲丽,程结晶. 高校图书馆科学数据素养教育现状调查及建议[J]. *图书馆学研究*,2018(19):2-9.
- [37] 王维佳,曹树金,廖响赞. 大学图书馆开展数据素养教育的方法与策略[J]. *图书馆*,2017(1):89-95,103.
- [38] BANDURA A. Social learning theory [M]. Englewood Cliffs:Prentice Hall,1977.

An Empirical Study on the Postgraduates' Data Literacy and Its Influencing Factors in the Field of Science and Technology

LI Geming, LIU Ji'an, QI Jia

(School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Based on social ecosystems theory, postgraduates' data literacy and its influencing factors in the field of science and technology are analyzed from the perspective of individual and school. The results reveal that students' data literacy is good with better awareness and ethics, while with worse skills, such as acquisition, analysis, storage and presentation. Besides, individual's grade, learning attitude, academic communication, learning experience, universities' teaching and training have positive effects on data literacy. Therefore, suggestions are put forward as the following: students should enhance their understanding of data literacy and improve their initiative in learning data skills; school should provide multi-level and diversified data literacy education according to the different demands of teaching objects.

Keywords: data literacy; postgraduate of science and technology; influencing factor; scientific and technological innovation