

文本可读性与学术论文的影响力:基于图书情报学的实证研究

刘宇 伍丹炜 叶继元

摘要 哪些因素会影响学术论文的被引次数是文献计量学领域的一个经典研究议题。目前的研究主要关注论文的内容特征和形式特征与被引次数之间的关系,鲜有研究从文本可读性视角切入这一议题。文本可读性影响读者对文本内容的理解和知识吸收,是一个关乎知识传播效率和研究成果认可度的重要因素。本研究在控制论文知识品质和权威性的基础上,使用文本可读性R值等五个变量研究论文的文本可读性对被引次数的影响。以中文图书情报学知名期刊发表于2016—2020年的论文为研究样本,研究发现论文的文本可读性R值、是否采用复合式标题、是否使用公式和表格对被引次数有显著影响,而是否使用图对被引次数没有显著影响。研究验证了中文情境下文本可读性对论文影响力的实质性作用,研究结果对科研人员改善自身的中文学术写作以及提高研究成果影响力具有重要参考价值。表3。参考文献83。

关键词 学术评价 文本可读性 被引次数 引文分析 学术影响力

分类号 G250.252

Text Readability and the Impact of Academic Articles: An Empirical Study Based on Library and Information Science

LIU Yu, WU Danwei & YE Jiyuan

ABSTRACT

It is a classical research issue in bibliometrics that what factors affect the citation frequency of academic articles. Previous studies have mainly focused on the relationship between the content or format characteristics of articles and the citations they received, while few studies shed light on the issue from the perspective of readability, which is an integral component of the content characteristics of an academic article. A critical factor concerning the efficiency of knowledge dissemination and the recognition of the research is to what extent readability affects a reader's understanding and absorbing of the information in the article. This study aims to examine the impact of text readability of an article on the citations it received.

The articles published from 2016 to 2020 in the well-known journals of Chinese library and information science were used as the research sample. The readability of the research sample was measured by the R-value algorithm, which is proposed by JING Xiyu of National Kaohsiung Normal University. The R-value is the most scientific readability indicator in analyzing Chinese text since it was designed to evaluate the readability of Chinese textbooks for primary and secondary school. Meanwhile, the fact that an article uses or does not use a compound title, formulas, tables, and figures was used as four complementary variables to

通信作者:叶继元,Email:yejiyuan@nju.edu.cn,ORCID:0000-0002-4232-8923 (Correspondence should be addressed to YE Jiyuan,Email:yejiyuan@nju.edu.cn,ORCID:0000-0002-4232-8923)

measure the readability of the research sample. Additionally, the quality and authority characteristics of articles were under control to investigate the impact of readability on the citations of articles. The co-authorship and the number of references were used as the quality indicator of an article. The journal prestige, author's job title, and institutional hierarchy were used to control the halo effect of citations received for an article. The publication age was used to control the cumulative effect of citations over time.

It has been found that the readability R-value of articles, whether or not using a compound title, and the presence or absence of formulas and tables in the text has a significant impact on citation frequency, while the presence or absence of figures has no significant impact on citation frequency. The study verified that readability does play a substantial role in the impact of an article in the Chinese scientific communication context. Furthermore, the research data shows that authoritative factors, time effects, and references number have significant impact on citation frequency in the Chinese scientific context, and the impact effect is relatively robust. However, co-authorship has a very limited effect on increasing citations. The results indicate that it is important for Chinese researchers to improve their academic writing skills for better readability and peer recognition as well as endeavor to publish results in higher-level journals to gain more attention from the academic community. 3 tabs. 83 refs.

KEY WORDS

Academic evaluation. Text readability. Citation frequency. Citation analysis. Academic impact.

自加菲尔德开发科学引文索引(SCI)以来,引文分析逐渐成为学术评价的一种重要方法。被引次数以及基于被引次数衍生的文献计量指标作为量化科研成果和科研主体学术影响力的指标,被广泛应用于职称晋升、基金申请等学术资源配置活动中^[1]。尽管论文影响力不仅仅表现为被引次数,但被引次数无疑是反映影响力的一个最易获取的简明指标。在这种情况下,什么样的论文被引次数多或者哪些因素会影响论文的被引次数,成为科研人员普遍关心的问题,也是文献计量研究的一个经典议题。现有研究大多从论文的外部形式特征出发,研究论文的点击次数、下载次数、作者职称及就职机构等因素与论文被引次数之间的关系,也有少量研究从文章主题、研究取向等内容特征出发考察论文被引的影响因素,但是,鲜有研究注意到文本可读性这一因素对论文被引的影响。本文从文本可读性这一内容特征出发,考察学术论文的可读性对其被引次数的影响效应,为学术界深入理解引用行为和引文分析提供新的理论解释。

1 研究综述

人们通常认为引用源于被引文献对施引作者的知识启迪,因此通过引文分析可以描述科学知识的扩散轨迹^[2],被引次数的多寡取决于论文质量的高低。实际上,作者的引用动机非常复杂,还受到很多社会性因素的影响。究竟哪些因素会对一篇论文的被引次数产生重要影响?文献计量学的相关研究显示影响论文被引次数的因素繁多^[3,4]。本文主要从内容特征和形式特征两个层面对已有研究所探寻的主要因素进行梳理。

1.1 论文内容特征与被引

理想状态下,被引次数只与研究成果的质量有关。很多研究指出,“论文质量是决定被引次数的主要因素之一,它可以预测论文未来的学术影响力”^[5,6]。如何衡量论文的质量是验证这一命题需要解决的首要问题,而论文质量却是一个难以量化的概念。长久以来,同行评议

是评价学术成果质量的通行方法,然而,当这一方法脱离匿名和个人独立判断等严肃、公正的条件约束时,专家的评估结果会受到外界社会因素的干扰^[7],这一缺陷在人情社会中更是表现得尤为突出^[8]。为了弥补同行评议的缺陷,以论文发表期刊的影响因子作为论文质量的替代性度量成为文献计量评价的常态^[9]。这种处理方式会陷入迭代逻辑的陷阱,即以整体的被引次数解释成员的被引次数,而整体被引次数是由成员被引次数加总求平均算得。鉴于此,有少量研究根据研究样本的特点提取能够表征内容品质的变量,常见的测量包括论文在期刊目次中的排位^[10]、是否获得期刊年度优秀论文奖^[10]、统计分析的使用是否恰当^[11]。

除了论文的内容品质之外,论文所研究的主题、论文的体裁和研究取向也是文本内容的重要特征,对论文的被引次数存在影响。主题代表学术论文研究内容的核心要素,是激发读者阅读兴趣的重要因素。主题通常从属于某一学科,不同学科在知识传播途径、学者引用偏好等方面存在诸多差异^[12,13],同样的差异可能还存在于同一学科的不同研究领域^[10]。随着时间的推移,学科差异会导致不同研究主题在被引次数上的差异,小学科的研究主题被引用的机会往往较小^[14,15]。随着中国社会科学开始强调“问题导向的学术研究”,由不同学科关注同一研究问题而形成的跨学科研究或交叉学科研究越来越受到重视^[16]。有研究显示,在学科多样性和学科差异性上分值更高的论文,被引次数也更高^[17]。此外,创新是学术研究的本质和生命力所在,主题新颖的研究往往能够吸引学者的注意力,学术成果主题新颖性与被引次数之间存在显著的相关关系^[18]。

很多研究发现论文的体裁和研究取向会影响其被引次数。在体裁上,大量研究发现综述型论文的被引次数会高于研究型论文^[19-21]。在研究取向上,经济学界的实证论文被引率高于理论论文^[22],社会学界的方法类论文比理论论文和实证论文得到更多被引^[23]。但是,也有研

究发现论文体裁和取向对被引没有显著影响,影响论文被引次数的是论文主题^[10]。

1.2 论文形式特征与被引

“文本形式特征的提取远远简单于内容特征”^[24],大量实证研究聚焦在论文的形式特征对其被引的影响。这些形式特征虽然不能直接反映出论文内容的品质,但是可以作为论文内容品质的信号,这种信号作用被称为可见性机制^[25]。根据论文外部形式特征的附属主体差异,本研究将其分为论文层面、期刊层面、作者层面三个层次。

(1) 论文层面的形式特征

按照现代学术规范,社会科学的研究论文遵循IMRD的“洋八股”结构^[26],由标题、摘要、关键词、引言、方法、研究发现、讨论及参考文献等固定部分组成。研究显示影响论文被引次数的结构性要素有标题类型、标题长度以及标题中是否含有特定符号^[27-29],论文摘要中专业词汇频次^[30]和摘要长度^[29],论文的篇幅^[29,31],研究设计的类型^[31,32],参考文献的数量^[29,33,34]及其国际化程度^[35]等。

(2) 期刊层面的形式特征

尽管“以刊评文”受到学术界诟病^[36],且同一期刊内发表的论文之间也存在巨大差异^[37],但是有大量的实证研究发现,期刊的影响因子或学术声望显著地正向影响论文的被引次数^[9,38,39]。在以重复发表论文为研究样本而达到有效控制论文质量差异的情况下,期刊的声望和影响力更是显著地决定了论文的被引次数^[40-42]。

(3) 作者层面的形式特征

对于有一定学术积累的读者而言,本领域专家的名字就是学术成果品质的重要象征之一,很多研究人员都有跟踪本领域重要专家的习惯。研究表明,论文的被引次数与作者声望显著正向相关^[38,43]。作者声望常用作者的学历、职称、毕业单位或工作单位的声望、是否为高被引学者等因素进行测量^[39,43]。也有一些研究发现作者人数对论文的被引次数有显著影响^[14,31,44]。

2 文本可读性:内容特征的拓展

论文被引的影响因素研究主要从形式特征视角展开,少量从内容特征视角展开的相关研究则重点关注质量、主题、体裁等要素。值得注意的是,研究成果被广泛传播与阅读是论文被引用的前提。高质量的论文不仅要有严谨的研究方法和创新性的研究发现,文本内容的文字表述是否易于受众阅读和理解,也是影响论文被引用的重要因素。大量讨论学术写作的文献都告诉读者,学术文本的写作要直接、简练、清晰,形成“透明玻璃”的风格;但是,绝大多数学者都不能否认学术文本中充斥着大量晦涩难懂的文字,有时候“晦涩变成了扮演高深莫测的‘化妆品’,一些作者依靠虚张声势来掩盖内容和信心的双重贫乏”^[45]。周雪光强调“只有认真写作、锤炼的文字,才值得认真阅读”^[46]。阅读是引用的前提,然而,目前对学术文本可读性与其影响力之间关系的研究尚不多见。

文本可读性也称文本难度,是指“文本材料所包含的影响读者有效使用的所有要素及其之间相互作用的总和”^[47],主要用来评价阅读材料能被读者阅读和理解的难易程度^[48]。20世纪20年代,美国教育界为应对中学生生源的变化、控制中学教材难度而开展关注教材可读性的研究,用以帮助教师为学习者选择合适的阅读材料^[49]。英文文本的可读性测量主要基于词汇、句法、篇章等文本特征,构建可读性度量公式^[50],如著名的蓝思(Lexile)分级阅读主要使用句长和词频两个变量来衡量读物的阅读难度^[51]。到20世纪80年代,有超过200个可读性公式被构建出来,比较具有代表性的有Dale-Chall Formula、Gunning FOG Index、SMOG Grading、FORCAST Formula、Flesch-Kincaid Formula等;美国教育部和国防部也建立以可读性公式为基

础的可读性评估体系,用来评估教材和国家政策文本。目前,英文文本的可读性研究主要通过有监督的机器学习构建可读性等级分类模型(分类法),或通过构建比较器、人工标注得到文本的两两相对难度,利用相对难度排序算法对文本可读性排序(排序法)^[50]。目前,已有研究团队开发出可以在线使用的可读性测试工具,如Coh-Metrix^[52]①、TextEvaluator^[53]②。

汉语文本可读性研究起步较晚。1971年,Yang针对繁体文本研制出了包含难词比、完整句子数和平均笔画数三个变量的可读性公式^[54],被认为是第一个汉语可读性公式。目前,汉语文本可读性研究仍集中在可读性公式的研制上^[50],如郭望皓、蔡建永等的研究^[55,56]。截至2020年,针对汉语的可读性公式有近30个^[49],其中最权威的是中国台湾学者荆溪昱研发的可读性公式。他以中国台湾中小学12个年級的中文教材为研究样本,将每个学期作为一个可读性级别,用文章总字数、平均句长、常用字占比、诗歌文体、文言文文体五个指标拟合教材可读性等级^[57]。2016年中国台湾学者宋曜廷团队开发出了针对繁体中文的可读性在线评测系统CRIE^[58]③。

文本可读性研究不仅在语言学、教育学、计算机文本处理领域受到广泛关注,在管理学等领域也开始得到应用。逯东等考察了上市公司年报可读性与融资约束之间的关系,发现上市公司年报的可读性越低,其面临的融资约束越高^[59]。柯青等研究了在线健康信息可读性对用户认知负荷和信息加工绩效的影响,发现可读性通过影响总注视时间、注视点个数进而显著影响眼跳距离,并建议提高信息可读性以减轻用户认知负荷^[60]。

文本可读性对学术成果被引的影响已经引起一些学者的关注。Snizek等研究发现在研究

① <http://cohmetrix.com>

② <https://textevaluator.ets.org/textevaluator>

③ <http://www.chinesereadability.net/CRIE/?LANG=CHT#>

伯基特淋巴瘤 (Burkitts-Lymphoma) 的论文中,摘要可读性与论文被引正相关,但在研究重夸克势 (Heavy Quark Potential) 的论文中,两者负相关^[61]。Hartley 等以心理学经典论文为研究样本,比较经典论文与对照论文 (同刊同期论文) 的可读性 (使用 Flesch 算法测量) 对其被引次数和同行重要性排序的影响,研究发现可读性越高的论文重要性排序越高,但可读性与被引次数之间没有显著关系^[62]。Lei 等以 *Scientometrics*、*Journal of Informetrics*、*Research Policy*、*Research Evaluation* 四种期刊在 2003—2012 年间出版的论文为样本,使用 FRE 和 SMOG 两种可读性指标测量论文摘要、全文的可读性与被引次数之间的关系,研究发现可读性和被引次数之间的相关关系不显著^[63]。Wesel 等以 FRES 可读性公式测量社会学、普通内科医学和应用心理学论文的摘要和全文的可读性,研究发现在社会学中全文可读性显著负向影响被引次数,在普通内科医学中摘要可读性显著负向影响被引次数,在应用心理学中全文可读性显著负向影响被引次数^[29]。Ante 等以技术领域论文为样本的实证研究也得到了相似结论^[64]。

早在 1990 年代初,国内即有学者注意到文本可读性对学术论文传播的重要性^[65]。目前国内关于文本可读性的研究主要集中在计算机领域,并以可读性公式研发为主,鲜有学者关注到文本可读性与学术论文影响力之间的关系。国外为数不多的相关研究的主流结论是文本可读性与被引之间负相关或相关性不显著,但是也有极少数研究得出可读性与被引之间正相关的结论。因此,中文情境下学术文本的可读性与论文影响力之间的关系究竟如何,是一个尚待探索的重要议题。

3 研究设计

3.1 变量测量

(1) 解释变量:可读性

现有的中文可读性计算公式中,中国台湾

学者荆溪昱提出的公式最为权威^[66]。如前所述,荆溪昱使用的语料文本是中国台湾中小学的语文教材,不同年级语文教材之间的阅读难度差异明确且极少产生争议,这一特点极大提高了语料文本可读性测量的可靠性,使得文本可读性的拟合公式具有较高的权威性。使用其他语料产生的文本可读性公式,可读性本身会因为读者自身知识背景差异而不同,因而会削弱可读性公式的公认度和权威性。因此,本文使用荆溪昱的公式计算样本论文的文本可读性,作为核心解释变量,该公式表达如下^[67]:

$$R = 8.76105604 + 0.00272438 \times X1 + 0.07866782 \times X2 - 8.94311010 \times X3 + 0.42920182 \times X4 + 3.23677141 \times X5$$

其中, $X1$ 为文本总字数, $X2$ 为平均句长, $X3$ 为常用字比, $X4$ 为诗歌文体, $X5$ 为文言文文体。由于研究样本中不存在诗歌文体和文言文文体, $X4$ 、 $X5$ 取值为 0; 公式中常用字是指中国台湾小学语文教材中的 495 个常用繁体字,不适用于大陆的文本,因此本文将 $X3$ 的内容更换为国家《通用规范汉字表》中的一级汉字。最终计算得到 R 值, R 值越大表示文本越不易理解、可读性越低。

标题是论文的重要组成部分,好的标题往往能起到画龙点睛的效果,吸引读者的注意力。复合式标题是指由冒号、破折号、问号等标点符号连接主副两部分所构成的标题。尽管英语学术界对复合式标题的使用存在一定的争议,但是其使用范围日益扩大却是不争的事实^[68]。在汉语学界,复合式标题具有强调、修辞功能以及强化文章学术性的作用^[69]。有研究显示图书采用带有冒号的复合式标题可以显著提升其被引次数^[70]。鉴于复合式标题的修辞功能,本研究设置“复合式标题”这一二分类变量作为可读性的辅助测量。

除了文本特征外,一些非文本特征也是影响科学论文可读性的不可忽视的因素,如图、表及公式等数学工具的使用可以更加逻辑清晰地展示变量的关系和数据的特征^[71]。Snizek 等研

究发现,在 DNA 研究领域,没有图表(Figures)的论文被引次数显著低于有图表的论文^[61]。Nieri 等研究发现,在移植治疗领域,论文有图像(Image)可以显著增加其被引次数^[72]。Ayres 等研究发现,没有公式的法学领域论文的被引次数显著高于有公式的论文^[73]。因此,本研究设计图、表、公式三个二分类变量作为可读性的辅助测量。

(2) 控制变量

经典的科学社会学认为知识启迪导致文献被引,而科学知识社会学则强调权威尊崇在引用行为中的驱动性作用^[41,74]。因此,本文从论文的知识品质和权威性两个角度设置了五个控制变量。

严格意义上讲,论文的质量需要通过历史检验才能最终确定,对当前研究成果知识品质的判断要么依赖专家的判断,要么通过文献外部特征信号予以测量。本研究选择“是否合作”“参考文献数”这两个变量作为论文知识品质的信号性测量。一方面,科研合作可以突破个体学者的认知和理性的局限,Glänzel 研究发现国际合作论文、国内合作论文相比独作论文都有着更高的被引次数^[75]。另一方面,参考文献体现了作者知识来源的广度和深度,进而影响论文的知识品质,Liskiewicz 等发现参考文献数显著正向影响论文的被引次数^[34]。

在权威性方面,本研究从期刊、作者职称和作者所属机构三个方面进行测量。①发表期刊是否为权威期刊是国内学者普遍关心的问题。期刊的权威性和影响力对论文被引有着极为关键的作用^[9]。本研究根据样本期刊是否为“FMS 管理科学高质量期刊推荐列表”的 T1 级将其分为权威和核心两类。②职称是专业技术人员头上的“紧箍咒”^[76],也是最为常用的测量专业技术人员学术地位和权威性的指标。由于我国历来有重视第一作者的出版传统,本研究根据《各专业职称等级对照表》划分论文第一作者职称等级,即正高、副高、中级、初级;不在对照表中的头衔和没有明确职称信息的作者归为“其他”类。考虑到研究生在科研和论文发表中

的作用,本研究单独设置“学生”类。③第一作者单位级别。我国高校作为科研的主力军具有明显的梯队性,梯队的级别可以有效衡量高校的社会认可度和权威等级,本研究将其分为 985 高校、211 高校、普通高校,将高校图书馆归入不同级别的高校,将研究院、其他类型图书馆等机构列为“其他”。

此外,考虑到发表时间越久的论文越容易累积更高的被引次数,本文还设置了“出版年龄”(用 2022 减去具体出版年求得)这一控制变量。

3.2 数据收集与分析方法

本研究的因变量是论文的影响力,学术界比较公认的客观、简明测量论文影响力的指标就是被引次数^[1]。被引是一个累积的过程,有研究发现一篇论文发表超过 5 年之后比较适宜用被引次数评估其影响力^[77]。本文以 CNKI 期刊全文数据库为数据源,选择《中国图书馆学报》《情报学报》《图书情报工作》《图书情报知识》这四种被 FMS 收录的期刊作为样本。同时,考虑到《大学图书馆学报》曾于 2012 年入选“国家社科基金第一批资助学术期刊目录”,在图书情报学界占据重要地位,因此将其与 FMS 期刊 T2 级期刊等同,也纳入研究样本。2021 年 8 月 18 日在 CNKI 数据库上使用自编 Python 工具抓取上述五种期刊在 2016—2020 年间发表的论文,剔除会议报道等非研究性记录,采集样本论文的作者信息、期刊信息、全文文本及被引次数,并计算样本论文正文的 R 值,最终得到 3 850 个研究样本。

已有大量研究显示^[78],被引次数作为典型的计数变量,符合负幂分布,一般使用泊松回归进行研究^[79]。泊松回归要求数据满足等离散现象,即事件的条件均值等于条件方差。本研究中被引次数不满足这一条件(均值=15.27,方差=19.206),方差大于均值说明数据具有一定的聚焦性,有过离散现象存在,因此负二项回归模型更适用于此次分析。本研究使用 SPSS 24 的广义线性模型拟合负二项回归。

4 数据分析与发现

4.1 数据描述

表1所示为本研究所用的分类变量的描述性统计信息。在控制变量中,将近80%的论文是合作完成的,权威期刊论文占20%略多,第一

作者中正高、副高职称的数量最多且相对均衡,初级职称的作者较少;第一作者的单位主要是高校,其中985高校的占比最高。在预测变量中,没有复合式标题和没有公式的论文较多,占比均约为75%,大部分论文都有图和表,占比都接近75%。

表1 分类变量的描述统计

		测量赋值	个案数	百分比(%)
控制变量	合作	否(0)	779	20.2
		是(1)	3 071	79.8
	权威期刊	否(0)	2 961	76.9
		是(1)	889	23.1
	职称	其他(1)	72	1.9
		学生(2)	821	21.3
		初级(3)	69	1.8
		中级(4)	784	20.4
		副高(5)	1 009	26.2
		正高(6)	1 095	28.4
	单位	其他(1)	454	11.8
		普通高校(2)	914	23.7
		211高校(3)	723	18.8
		985高校(4)	1 759	45.7
预测变量	复合式标题	否(0)	2 841	73.8
		是(1)	1 009	26.2
	公式	否(0)	2 890	75.1
		是(1)	960	24.9
	图	否(0)	993	25.8
		是(1)	2 857	74.2
	表	否(0)	1 028	26.7
		是(1)	2 822	73.3

表2是因变量被引次数、预测变量R值和控制变量参考文献数、出版年龄的描述性统计信息。被引次数的极差较大,呈现出明显的负幂分布;R值比较接近正态分布,而且

不同论文的R值离散度较小;参考文献数的极差较大,也呈现出负幂分布的特征;出版年龄呈现出轻微负偏态势,表明每年发文量略有上升趋势。

表 2 连续变量的描述统计

		个案数	最小值	最大值	平均值	标准差	偏度	峰度
因变量	被引次数	3 850	0	248	15.27	19.21	4.42	33.19
预测变量	R 值	3 850	11.30	73.82	30.73	6.98	1.01	2.57
控制变量	参考文献数	3 850	0	199	31.31	19.61	1.94	7.07
	出版年龄	3 850	2	6	4.06	1.42	-0.61	-1.31

4.2 数据分析

表 3 为论文被引频次影响因素的负二项回归模型。模型一纳入了出版年龄、是否合作和参考文献数量三个变量,考察论文的发表年龄和知识品质对被引次数的影响;模型二增加了论文是否发表在权威期刊、第一作者职称和第一作者单位级别三个变量,用于考察权威性因素对被引次数的影响;模型三进一步加入 R 值、论文是否使用复合式标题、是否有公式、是否有图、是否有表等五个预测变量,用于检验本研究的核心命题——文本可读性对被引次数的影响。Omnibus 检验显示三个模型均显著有效(似然比卡方值均显著);随着模型中不断纳入新的自变量,三个模型的拟合优度皮尔逊卡方值的变化值在不断减小,说明模型的整体解释力在不断上升并收敛。

模型一显示论文出版年龄和参考文献数量会显著影响其被引频次,是否合作对论文被引频次影响不显著。具体来说,出版年龄越大的论文被引次数越高($\beta = 0.221, p = 0.000$),出版年龄每增加一个单位时,论文的被引次数增加一个单位的优势比的概率会增加 24.8% (1.248-1),这说明论文的被引次数存在明显的累积效应。参考文献数量多的论文容易获得更多的被引($\beta = 0.007, p = 0.000$),参考文献每增加一个单位时,论文的被引次数增加一个单位的优势比的概率会增加 0.7% (1.007-1)。

模型二显示在控制出版年龄和知识品质因素的情况下,三个权威性因素均显著影响论文的被引次数。具体来看,相对于权威期刊论文来说,非权威期刊论文的被引次数增加一个单

位的优势比的概率会减少 22.7% (0.773-1);相对正高来说,第一作者为副高、中级、初级和学生的论文的被引次数增加一个单位的优势比的概率分别会减少 25.1% (0.749-1)、31.3% (0.687-1)、32.4% (0.676-1)、26.2% (0.738-1)。相对于在 985 高校任职的第一作者来说,在 211 高校、普通高校任职的第一作者,其论文的被引次数增加一个单位的优势比的概率分别会减少 14.1% (0.859-1)、16.9% (0.831-1)。

模型三显示在控制出版年龄、知识品质和权威性因素的情况下,可读性要素中的 R 值($\beta = 0.012, p = 0.010$)、标题是否为复合式标题($\beta = -0.089, p = 0.021$)、是否有公式($\beta = 0.104, p = 0.021$)、是否有表格($\beta = 0.087, p = 0.034$)显著影响论文的被引次数,而论文中是否有图对被引次数没有显著影响。具体来说,R 值越高论文的被引次数越高,R 值每增长一个单位时,论文的被引次数增加一个单位的优势比的概率会增加 1.2% (1.012-1),即越难读(R 值越高)的论文被引次数反而越高。相对于采用了复合式标题的论文来说,没有采用复合式标题的论文增加一个单位的被引次数的优势比的概率会减少 8.9% (0.911-1),即采用复合式标题会显著增加论文的被引次数。相对于有公式的论文来说,没有公式的论文,其被引次数增加一个单位的优势比的概率会增加 10.9% (1.109-1),即公式的使用会降低论文的被引次数。相对于有表格的论文来说,没有表格的论文,其被引次数增加一个单位的优势比的概率会增加 9.1% (1.091-1),即表格的使用会降低论文的被引次数。

表 3 被引频次影响因素的负二项回归模型

参数	模型一				模型二				模型三			
	β	瓦尔德卡方	显著性(p)	Exp(β)	β	瓦尔德卡方	显著性(p)	Exp(β)	β	瓦尔德卡方	显著性(p)	Exp(β)
(截距)	1.576	610.309	0.000	4.834	2.012	709.156	0.000	7.481	1.609	185.833	0.000	4.998
出版年龄	0.221	313.379	0.000	1.248	0.238	355.236	0.000	1.269	0.247	372.541	0.000	1.280
[合作=0]	-0.067	2.666	0.109	0.935	-0.052	1.488	0.229	0.950	-0.094	4.412	0.036	0.910
[合作=1]	0	.	.	1	0 ^a	.	.	1	0 ^a	.	.	1
参考文献数量	0.007	62.187	0.000	1.007	0.006	44.114	0.000	1.006	0.004	20.457	0.000	1.004
[权威期刊=0]					-0.258	39.719	0.000	0.773	-0.252	35.052	0.000	0.777
[权威期刊=1]					0 ^a	.	.	1	0 ^a	.	.	1
[职称=1其他]					-0.110	0.752	0.386	0.896	-0.157	1.527	0.217	0.855
[职称=2学生]					-0.303	38.942	0.000	0.738	-0.265	28.796	0.000	0.767
[职称=3初级]					-0.391	8.969	0.003	0.676	-0.351	7.719	0.007	0.704
[职称=4中级]					-0.375	56.676	0.000	0.687	-0.351	48.847	0.000	0.704
[职称=5副高]					-0.290	39.285	0.000	0.749	-0.279	35.888	0.000	0.757
[职称=6正高]					0 ^a	.	.	1	0 ^a	.	.	1
[单位=1其他]					-0.065	1.382	0.240	0.937	-0.066	1.412	0.235	0.936
[单位=2普通]					-0.185	17.472	0.000	0.831	-0.163	13.390	0.000	0.849
[单位=3“211”]					-0.153	10.984	0.001	0.859	-0.133	8.242	0.004	0.876
[单位=4“985”]					0 ^a	.	.	1	0 ^a	.	.	1
R 值									0.012	19.237	0.000	1.012
[复合式标题=0]									-0.089	5.351	0.021	0.911
[复合式标题=1]									0 ^a	.	.	1
[公式=0]									0.104	5.862	0.015	1.109
[公式=1]									0 ^a	.	.	1
[图=0]									-0.047	1.255	0.263	0.954
[图=1]									0 ^a	.	.	1
[表=0]									0.087	4.510	0.034	1.091
[表=1]									0 ^a	.	.	1
(标度)	1 ^b				1 ^b				1 ^b			
(负二项)	1 ^b				1 ^b				1 ^b			
Omnibus	似然比卡方	自由度	显著性(p)		似然比卡方	自由度	显著性(p)		似然比卡方	自由度	显著性(p)	
	345.719	3	0.000		530.644	12	0.000		572.469	17	0.000	
拟合优度	皮尔逊卡方	4 766.449			皮尔逊卡方	4 145.281			皮尔逊卡方	4 036.048		

需要指出的是,在没有纳入可读性因素时,是否为合作论文对其被引次数没有显著影响(如模型一、模型二所示);当模型纳入可读性要素时,是否为合作论文对被引次数有显著影响($\beta = -0.094, p = 0.036$)。这说明在出版年龄、权威性、可读性以及参考文献数量相同的情况下,相对于合作论文来说,非合作论文的被引次数增加一个单位的优势比的概率会减少 9% (0.910-1)。

5 讨论

论文、作者和期刊等多个层面的内容特征和形式特征均可能影响其被引次数,根据科学社会学和科学知识社会学对引用动机的理论解释^[74],这些特征要么属于反映论文知识品质的信号,要么属于反映论文权威属性的信号。本研究在综合考虑知识品质和权威性两类因素的基础上,加入了反映文本可读性的变量,检验可读性对论文被引次数的影响。

5.1 知识品质层面

数据显示,在不考虑权威性和可读性的情况下,论文的出版年龄与参考文献数量都显著地正向影响论文的被引次数,是否为合作论文对论文被引影响不显著。但是,随着权威性因素和可读性因素的加入,是否为合作论文对论文被引次数的影响变得显著。这一特征使我们警惕一些固有观念——“团队论文的引用量通常比独立作者论文高”^[80]。从理论上说,合作有利于突破个体作者的认知局限,提高论文的知识品质;但是,作者数量的增加也意味着论文的社会资本和人际网络的增加。因此,发现合作论文拥有更高被引次数的研究,无法有效地排除社会资本因素对论文被引的影响,从而无法得到合作对论文被引的净效应。这是本研究中“合作”这一变量的显著性在不同模型中发生变化的根本原因,这种变化也表明合作这一要素对论文被引次数的影响效用要远远弱于时

间、权威性和可读性等因素。参考文献数量对论文被引的正向影响与主流的实证研究结果相一致^[43]。此外,本研究的数据也在一定程度上证实 Didegah 等研究结果的合理性,即“合作会提高被引率”的论断只在某些特定条件下成立^[81]。

5.2 权威性层面

测量权威性的三个变量对论文的被引次数均有显著影响,而且变量不同取值的显著性在两个模型中保持一致。这一特点充分说明了科学知识社会学对引用行为的解释效力,即学者在引用文献时具有明显的权威尊崇心理^[41],通过引用权威文献来增强自身研究的合法性和说服力。国内外已有不少实证研究发现,论文所发表期刊的权威性对其被引次数有显著的正向影响,如 Peng、谢娟等使用影响因子作为期刊权威性的测量指标^[9,39]。然而,影响因子本身是论文集合被引次数的均值,这样的测量方式存在自我循环论证的问题。本研究的样本数据来自国内图书情报学界具有较高公认度的知名期刊,对期刊权威性的测量依据 FMS 期刊评价体系,采用二分法测量,这种测量方法在一定程度上弱化了自我循环论证。因此,本研究在期刊权威性的测量上拓展了已有研究的测量方法,研究结论也进一步验证了权威性因素对论文被引次数影响效应的稳健性。此外,本研究还使用第一作者职称与第一作者所在机构的等级作为权威性的测量指标,数据显示职称和机构等级对论文被引次数有显著的正向影响。这一研究结论与 Bornmann 等对化学期刊论文^[43]、Long 等对管理学期刊论文^[82]的研究结论相一致。需要注意的是,三种权威性因素叠加会进一步强化引用行为的马太效应,导致论文被引次数分布的差异更加巨大,这是表 2 中被引次数分布的峰度远高于其他变量的原因。

5.3 可读性层面

文本可读性对论文被引次数的影响是本研

究关注的核心议题。虽然 Lei 等以英文情报学论文为研究样本发现,可读性对论文的被引次数没有显著影响^[63];但是,本研究的数据显示,R 值越高,论文被引次数就越高,即越难阅读的论文越容易被引用。这一结论和 Wesel 等以社会学和心理学领域、Ante 等以技术领域论文为样本的研究所得到的结论相一致^[29,64]。阅读难度高(即 R 值高、需要精读)的论文反而更容易被引,这乍看起来有悖于人们对可读性认知的常识——通常来说,人们都会认为可读性越高(即越容易阅读)的论文,被引次数会越高。

产生这一现象的主要原因在于,如何有效测量学术论文文本的可读性仍然属于有待探索的前沿课题。虽然自然语言处理技术的进步为可读性的自动分析提供了多种思路和方法^[50],但是,文本可读性测量与阅读者的认知过程、认知特点紧密结合^[49],使得这一问题更为复杂。目前使用的可读性公式高度依赖以中小学教科书为基础的语言学习性语料库,鲜有基于学术文本语料库研发的可读性公式。本研究采用荆溪显提出的可读性测量公式,主要由总字数、平均句长及常用字占比三个因素决定,也具有同样的问题。也就是说,现有可读性公式往往适用于语言学习类文本,诸如教科书等,而不一定适用于以知识创新为核心的学术文本。语言学习类文本的阅读主要以审美为导向,附带有知识学习性质;而学术文本的阅读主要以知识为导向,阅读学术文本时读者需要调用大量的储备知识,才能对学术文本的知识内容进行理解和吸收。创新性较高的论文往往使用专业精准的文字表达,陈述知识创新的内容及复杂的研究工作,读者阅读这些文本就需要调用更多的知识储备。因此,学术文本的可读性越低(即 R 值越高),知识含量可能更高,读者在精读此类文本时知识收获更大,进而被引次数也会更高。

在文本可读性层面的四个辅助性指标中,论文是否采用复合式标题、是否使用公式、是否使用表格显著影响论文的被引次数,而是否使用图对被引次数的影响不显著。首先,使用复

合式标题会显著提升论文的被引次数。复合式标题体现了论文作者的文字驾驭能力和语言技巧,会增加论文的语言美感、提升论文的可读性,更容易吸引读者的注意力,从而促进论文的被引。虽然魏瑞斌以图书情报学英文期刊论文为样本的研究发现,标题中的冒号对被引的影响并不明显^[28],但是,本研究的结果显示在中文语境中复合式标题会显著正向影响被引次数。这一特征与人们的日常认知吻合度更高,通常来说复合式标题可以展示作者更高的文字表达功底,深入浅出的学术文本便于知识传播和被引。其次,尽管公式、图、表作为形式化的信息表达工具在图书情报学论文中已经得到比较广泛的采纳和应用^[83],但是本研究的结果显示,使用公式、表格会显著降低论文的被引次数,而使用图对论文的被引次数影响不显著。这一结论与人们的日常认知也比较吻合,图能够以更加直观的视觉感受展示数据的核心特征,而人们对公式和表格所表达的信息进行理解时需要更多的逻辑推理和认知参与,这也是大众媒体更多使用图、较少使用表、基本不使用公式的原因所在。本研究的结果与 Ayres 等以法学论文为样本的研究结果相一致^[73],但是与 Snizek、Nieri 等以自然科学为样本的研究结果不吻合^[61,72]。这也表明图书情报学的学术表达方式与社会科学更为吻合,与自然科学仍然有显著区别。

6 结语

被引次数是衡量论文影响力乃至学术价值的一个最常用指标,深入研究被引次数的影响因素对科研人员自身的发展和科研管理工作都具有重要的现实意义。本研究从内容角度强调文本可读性作为独立因素对论文被引次数的影响,并以 R 值、是否采用复合式标题和文本中是否有图、表以及公式五个方面测量文本可读性。在控制了出版年龄、论文知识品质和权威性因素的情况下,研究数据显示文本可读性对学术论文的被引频次有显著影响,具体来说学术文

本可读性越低(即R值越高、越难阅读、需要精读),被引次数会越高,使用复合式标题的论文会得到更高的被引,使用公式和表格会降低论文的被引次数,是否使用图对被引次数的影响不显著。这一研究结果并非意味着科研人员在学术写作时要以艰深晦涩作为自己的表达风格,恰恰相反,研究结果提示科研人员在进行学术写作时,要在严谨精确表达的基础上,尽可能有效地提升论文修辞水平和文本可读性,只有经过认真锤炼的学术文本才能深入浅出地传递观点,才能最大程度地激发读者的阅读兴趣,获

得广泛的传播和同行的认可。

本研究在以下两个方面存在一定缺陷。一是文本可读性的计算依赖于计算机学科尤其是自然语言处理领域的研究成果,可读性公式的开发以语言学习语料而非学术文本语料为依托。因此,随着可读性公式这一关键议题被深入研究和新公式的出现,本研究的某些发现可能会有所变化。二是本研究以图书情报学领域论文为研究样本,因此,研究结论比较适用于社会科学领域,对于文学、历史学等传统特征深厚的学科可能并不适用。

致谢:本文系国家社会科学基金一般项目“中国人文社会科学领域的学术诚信观念与行为研究”(项目编号:20BTQ019)的研究成果。

参考文献

- [1] HAMERMESH D S. Citations in economics: measurement, uses, and impacts[J]. *Journal of Economic Literature*, 2018, 56(1): 115-156.
- [2] 刘玉仙,武夷山. 关于引文本质的思考[J]. *科学学研究*, 2015, 33(12): 1779-1786, 1882. (LIU Y X, WU Y S. On essence of citations[J]. *Studies in Science of Science*, 2015, 33(12): 1779-1786, 1882.)
- [3] TAHAMTAN I, SAFIPOUR A A, AHAMDZADEH K. Factors affecting number of citations: a comprehensive review of the literature[J]. *Scientometrics*, 2016, 107(3): 1195-1225.
- [4] 牟象禹,龚凯乐,谢娟,等. 论文被引频次的影响因素研究——以国内图书情报领域为例[J]. *图书情报知识*, 2018(4): 43-52. (MOU X Y, GONG K L, XIE J, et al. Contributing factors of citation: an empirical study of library and information science in China[J]. *Documentation, Information & Knowledge*, 2018(4): 43-52.)
- [5] JABBOUR C J C, JABBOUR A B L S, DE OLIVEIRA J H C. The perception of Brazilian researchers concerning the factors that influence the citation of their articles: a study in the field of sustainability[J]. *Serials Review*, 2013, 39(2): 93-96.
- [6] BUELA-CASAL G, ZYCH I. Analysis of the relationship between the number of citations and the quality evaluated by experts in psychology journals[J]. *Psicothema*, 2010, 22(2): 270-276.
- [7] RENNIE D. Let's make peer review scientific[J]. *Nature*, 2016, 535(7610): 31-33.
- [8] 李剑鸣. 自律的学术共同体与合理的学术评价[J]. *清华大学学报(哲学社会科学版)*, 2014, 29(4): 73-78. (LI J M. Self-disciplined academic community and reasonable academic evaluation[J]. *Journal of Tsinghua University(Philosophy and Social Sciences)*, 2014, 29(4): 73-78.)
- [9] PENG T-Q, ZHU J J H. Where you publish matters most: a multilevel analysis of factors affecting citations of Internet studies[J]. *Journal of the American Society for Information Science and Technology*, 2012, 63(9): 1789-1803.
- [10] STREMER S, VERNIERS I, VERHOEF P C. The quest for citations: drivers of article impact[J]. *Journal of Marketing*, 2007, 71(3): 171-193.

- [11] NIEMINEN P, CARPENTER J, RUCKER G, et al. The relationship between quality of research and citation frequency[J]. *BMC Medical Research Methodology*, 2006, 6(1): 1-8.
- [12] RADICCHI F, CASTELLANO C. Testing the fairness of citation indicators for comparison across scientific domains: the case of fractional citation counts[J]. *Journal of Informetrics*, 2012, 6(1): 121-130.
- [13] 张垒. 高被引论文的特征因素及其对影响因子贡献研究[J]. *中国科技期刊研究*, 2015, 26(8): 880-885. (ZHANG L. Research on the characteristic factors of highly cited papers and their contribution to the impact factor[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2015, 26(8): 880-885.)
- [14] MIETTUNEN J, NIEMINEN P. The effect of statistical methods and study reporting characteristics on the number of citations: a study of four general psychiatric journals[J]. *Scientometrics*, 2003, 57(3): 377-388.
- [15] FILION K B, PLESS I B. Factors related to the frequency of citation of epidemiologic publications[J]. *Epidemiologic Perspectives & Innovations*, 2008, 5(1): 1-7.
- [16] 渠敬东. 学科之术与问题之学[J]. *开放时代*, 2022(1): 49-52. (QU J D. The art of discipline and the study of problems[J]. *Open Times*, 2022(1): 49-52.)
- [17] CHEN S, QIU J, ARSENAULT C, et al. Exploring the interdisciplinarity patterns of highly cited papers[J]. *Journal of Informetrics*, 2021, 15(1): 101-124.
- [18] 逯万辉, 苏金燕, 余倩. 学术成果主题新颖性与学术引用的相关关系研究[J]. *情报资料工作*, 2018(6): 68-73. (LU W H, SU J Y, YU Q. Research on correlation of academic achievement theme novelty and academic citation[J]. *Information and Documentation Services*, 2018(6): 68-73.)
- [19] VANCLAY J K. Factors affecting citation rates in environmental science[J]. *Journal of Informetrics*, 2013, 7(2): 265-271.
- [20] 朱大明. 研究型论文与综述型论文引文量的对比分析[J]. *编辑学报*, 2010, 22(1): 33-34. (ZHU D M. Comparative analysis of citations between research papers and review papers[J]. *Acta Editologica*, 2010, 22(1): 33-34.)
- [21] FROSCHE D L, SAXBE D, TOMIYAMA A J, et al. Assessing the scholarly impact of health psychology: a citation analysis of articles published from 1993 to 2003[J]. *Health Psychology*, 2010, 29(5): 555-562.
- [22] JOHNSTON D W, PIATTI M, TORGLER B. Citation success over time: theory or empirics?[J]. *Scientometrics*, 2013, 95(3): 1023-1029.
- [23] PERITZ B C. Are methodological papers more cited than theoretical or empirical ones? The case of sociology[J]. *Scientometrics*, 1983, 5(4): 211-218.
- [24] 黄萃, 吕立远. 文本分析方法在公共管理与公共政策研究中的应用[J]. *公共管理评论*, 2020, 2(4): 156-175. (HUANG C, LV L Y. The application of text analysis to public management and policy research[J]. *China Public Administration Review*, 2020, 2(4): 156-175.)
- [25] WANG M, YU G, YU D. Mining typical features for highly cited papers[J]. *Scientometrics*, 2011, 87(3): 695-706.
- [26] 彭玉生. “洋八股”与社会科学的规范[J]. *社会学研究*, 2010(2): 180-210. (PENG Y S. The structure of empirical social research[J]. *Sociological Studies*, 2010(2): 180-210.)
- [27] JAMALI H, NIKZAD M. Article title type and its relation with the number of downloads and citations[J]. *Scientometrics*, 2011, 88(2): 653-661.
- [28] 魏瑞斌. 论文标题特征与被引的关联性研究[J]. *情报学报*, 2017, 36(11): 1148-1156. (WEI R B. Study

- on the relationship between the title characteristics and citation[J]. Journal of the China Society for Scientific and Technical Informaiton, 2017, 36(11): 1148-1156.)
- [29] VAN WESEL M, WYATT S, HAAF T J. What a difference a colon makes: how superficial factors influence subsequent citation[J]. Scientometrics, 2014, 98(3): 1601-1615.
- [30] IBANEZ A, LARRANAGA P, BIELZA C. Predicting citation count of *Bioinformatics* papers within four years of publication[J]. Bioinformatics, 2009, 25(24): 3303-3309.
- [31] ANTONIOU G A, ANTONIOU S A, GEORGAKARAKOS E I, et al. Bibliometric analysis of factors predicting increased citations in the vascular and endovascular literature[J]. Annals of Vascular Surgery, 2015, 29(2): 286-292.
- [32] ANNALINGAM A, DAMAYANTHI H, JAYAWARDENA R, et al. Determinants of the citation rate of medical research publications from a developing country[J]. SpringerPlus, 2014, 3(1): 1-6.
- [33] 姜磊, 林德明. 参考文献对论文被引频次的影响研究[J]. 科研管理, 2015, 36(1): 121-126. (JIANG L, LIN D M. Research on the impact of reference on citation[J]. Science Research Management, 2015, 36(1): 121-126.)
- [34] LISKIEWICZ T, LISKIEWICZ G, PACZESNY J. Factors affecting the citations of papers in tribology journals[J]. Scientometrics, 2021, 126(4): 3321-3336.
- [35] 龚凯乐, 谢娟, 成颖, 等. 引文国际化对论文被引的影响研究[J]. 图书情报知识, 2020(2): 102-112. (GONG K L, XIE J, CHEN G Y, et al. Research on the citation impact of reference internationalization[J]. Documentation, Information & Knowledge, 2020(2): 102-112.)
- [36] 朱剑. 重建学术评价机制的逻辑起点——从“核心期刊”、“来源期刊”排行榜谈起[J]. 清华大学学报(哲学社会科学版), 2012, 27(1): 5-15, 159. (ZHU J. Legitimate starting point for reconstructing academic evaluation system: based on various kinds of journal ranking[J]. Journal of Tsinghua University(Philosophy and Social Sciences), 2012, 27(1): 5-15, 159.)
- [37] KAMAT P V. Most cited versus uncited papers. What do they tell us?[J]. ACS Energy Letters, 2018, 3: 2134-2135.
- [38] VAN DALEN H, HENKENS K. What makes a scientific article influential? The case of demographers[J]. Scientometrics, 2001, 50(3): 455-482.
- [39] 谢娟, 成颖, 李江, 等. 文化资本与论文影响力的关系[J]. 情报学报, 2019, 38(9): 943-953. (XIE J, CHENG Y, LI J, et al. The relationship between cultural capital and the academic impact of journal articles[J]. Journal of the China Society for Scientific and Technical Informaiton, 2019, 38(9): 943-953.)
- [40] LARIVIERE V, GINGRAS Y. The impact factor's matthew effect: a natural experiment in bibliometrics[J]. Journal of the American Society for Information Science and Technology, 2010, 61(2): 424-427.
- [41] 刘宇, 张永娟, 齐林峰, 等. 知识启迪与权威尊崇: 基于重复发表的引文动机研究[J]. 图书馆论坛, 2018, 38(4): 49-57. (LIU Y, ZHANG Y J, QI L F, et al. Knowledge enlightenment or authority adoration: an empirical study of citation motivation based on duplicate publications[J]. Library Tribune, 2018, 38(4): 49-57.)
- [42] KNOTHE G. Comparative citation analysis of duplicate or highly related publications[J]. Journal of the American Society for Information Science and Technology, 2006, 57(13): 1830-1839.
- [43] BORNMAN L, SCHIER H, MARX W, et al. What factors determine citation counts of publications in chemistry besides their quality?[J]. Journal of Informetrics, 2012, 6(1): 11-18.

- [44] 马荣康,李真真. 高被引还是零被引:基于论文被引的最佳科研合作规模研究——来自 *Financial Times* TOP 45商学院期刊的证据[J]. 情报学报,2020,39(11):1182-1190. (MA R K, LI Z Z. High citation or zero citation: exploring the optimal scale of research cooperation based on the citation of scientific publication—evidence from the *Financial Times* TOP 45 journals[J]. Journal of the China Society for Scientific and Technical Information, 2020, 39(11): 1182-1190.)
- [45] 刘军强. 写作是门手艺[M]. 桂林:广西师范大学出版社,2020:40. (LIU J Q. The craft of writing[M]. Guilin: Guangxi Normal University Press, 2020: 40.)
- [46] 周雪光. 只有认真写作、锤炼的文字,才值得认真阅读[EB/OL]. (2022-04-13) [2022-06-24]. https://mp.weixin.qq.com/s/xy_NH8b-6SeaPmHrNEBMhw. (ZHOU X G. Only carefully written and tempered words are worth reading carefully[EB/OL]. (2022-04-13) [2022-06-24]. https://mp.weixin.qq.com/s/xy_NH8b-6SeaPmHrNEBMhw.)
- [47] DALE E, CHALL J S. The concept of readability[J]. Elementary English, 1949, 26(1): 19-26.
- [48] 陈练文,李信,赵超烨. 学术文本可读性和复杂度评价研究[J]. 数字图书馆论坛,2018(5): 62-66. (CHEN L W, LI X, ZHAO C Y. Analysis of the readability and complexity of academic texts[J]. Digital Library Forum, 2018(5): 62-66.)
- [49] 王蕾. 文本可读性公式研究发展阶段及特点[J]. 语言教学与研究,2022(2): 29-40. (WANG L. Development stages and characteristics of studies on readability formula[J]. Language Teaching and Linguistic Studies, 2022(2): 29-40.)
- [50] 吴思远,蔡建永,于东,等. 文本可读性的自动分析研究综述[J]. 中文信息学报,2018,32(12): 1-10. (WU S Y, CAI J Y, YU D, et al. A survey on the automatic text readability measures[J]. Journal of Chinese Information Processing, 2018, 32(12): 1-10.)
- [51] STENNER A J. Measuring reading comprehension with the Lexile framework[M]//FISHER J R W P, MASSENGILL P J. Explanatory models, unit standards, and personalized learning in educational measurement. Singapore: Springer, 2022: 63-88.
- [52] GRAESSER A C, MCNAMARA D S, KULIKOWICH J M. Coh-Metrix: providing multilevel analyses of text characteristics[J]. Educational Researcher, 2011, 40(5): 223-234.
- [53] SHEEHAN K M, KOSTIN I, NAPOLITANO D, et al. The TextEvaluator tool: helping teachers and test developers select texts for use in instruction and assessment[J]. The Elementary School Journal, 2014, 115(2): 184-209.
- [54] YANG S J. A readability formula for Chinese language[D]. Wisconsin: University of Wisconsin-Madison, 1971.
- [55] 郭望皓. 对外汉语文本易读性公式研究[D]. 上海:上海交通大学,2010. (GUO W H. Research on the formula of legibility of Chinese as a foreign language[D]. Shanghai: Shanghai Jiaotong University, 2010.)
- [56] 蔡建永. 汉语二语文本可读性公式研究[D]. 北京:北京语言大学,2020. (CAI J Y. Research on the readability formula of Chinese second language texts[D]. Beijing: Beijing Language and Culture University, 2020.)
- [57] 荆溪晏. 中文国文教材的适读性研究:适读年级值的推估[J]. 教育研究资讯,1995(5): 113-127. (JING X Y. A study on readability of Chinese textbooks: estimation of readability grade values[J]. Educational Research & Information, 1995(5): 113-127.)
- [58] SUNG Y T, CHANG T H, LIN W C, et al. CRIE: an automated analyzer for Chinese texts[J]. Behavior Research Methods, 2016, 48(4): 1238-1251.
- [59] 逯东,宋昕倍. 媒体报道、上市公司年报可读性与融资约束[J]. 管理科学学报,2021,24(12): 45-61. (LU

- D, SONG X B. Media coverage, readability of listed companies' annual reports and financing constraints[J]. Journal of Management Sciences in China, 2021, 24(12): 45-61.)
- [60] 柯青, 丁松云, 秦琴. 健康信息可读性对用户认知负荷和信息加工绩效影响眼动实验研究[J]. 数据分析与知识发现, 2021, 5(2): 70-82. (KE Q, DING S Y, QIN Q. Health information readability affects users' cognitive load and information processing; an eye-tracking study[J]. Data Analysis and Knowledge Discovery, 2021, 5(2): 70-82.)
- [61] SNIZEK W E, OEHLER K, MULLINS N C. Textual and nontextual characteristics of scientific papers; neglected science indicators[J]. Scientometrics, 1991, 20(1): 25-35.
- [62] HARTLEY J, SOTTO E, PENNEBAKER J. Style and substance in psychology: are influential articles more readable than less influential ones?[J]. Social Studies of Science, 2002, 32(2): 321-334.
- [63] LEI L, YAN S. Readability and citations in information science: evidence from abstracts and articles of four journals (2003-2012)[J]. Scientometrics, 2016, 108(3): 1155-1169.
- [64] ANTE L. The relationship between readability and scientific impact: evidence from emerging technology discourses [J/OL]. Journal of Informetrics, 2022, 16(1) [2023-09-04]. <https://doi.org/10.1016/j.joi.2022.101252>.
- [65] 戴兵. 学术期刊可读性刍议[J]. 广西大学学报(哲学社会科学版), 1993(1): 116-120. (DAI B. A humble opinion on the readability of academic journals[J]. Journal of Guangxi University (Philosophy and Social Science), 1993(1): 116-120.)
- [66] 秦琴, 柯青, 丁松云. 中文在线健康教育信息可读性计算及应用实证——以食品安全领域为例[J]. 现代情报, 2020, 40(5): 111-121. (QIN Q, KE Q, DING S Y. An empirical study on readability calculation and application of Chinese online health education information—a case study of food safety[J]. Journal of Modern Information, 2020, 40(5): 111-121.)
- [67] 荆溪昱, 赵世范, 翁凌志. 中文文章适读性线上分析系统之发展研究[C]//科技教育课程改革与发展学术研讨会论文集. 2007: 47-57. (JING X Y, ZHAO S F, WENG L Z. The development of readability-analyzing system for Chinese articles[C]//Proceedings of the Symposium on Curriculum Reform and Development of Science and Technology Education. 2007: 47-57.)
- [68] 姜亚军. 复合式论文标题之争: 评述与启示[J]. 外语教学, 2010, 31(6): 29-32. (JIANG Y J. Combination essay title battle[J]. Foreign Language Education, 2010, 31(6): 29-32.)
- [69] 鞠衍清, 王芳. “A:B”型论文标题: 使用频度及分类——对《新华文摘》的调查[J]. 辽东学院学报(社会科学版), 2014, 16(6): 146-150. (JU Y Q, WANG F. Usage frequentness and classification of “A:B” type paper titles; a case study with Xinhua Digest[J]. Journal of Eastern Liaoning University (Social Sciences), 2014, 16(6): 146-150.)
- [70] 阮选敏, 吕冬晴, 成颖, 等. 题名标点对我国人文社会科学学术图书被引的影响研究[J]. 情报学报, 2019, 38(5): 458-472. (RUAN X M, LV D Q, CHENG Y, et al. Factors affecting citations of monographs in Chinese humanities and social sciences; the role of punctuation marks[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(5): 458-472.)
- [71] 洪永淼, 汪寿阳. 数学、模型与经济思想[J]. 管理世界, 2020, 36(10): 15-27. (HONG Y M, WANG S Y. Mathematics, model and economic thought[J]. Journal of Management World, 2020, 36(10): 15-27.)
- [72] NIERI M, CLAUSER C, FRANCESCHI D, et al. Randomized clinical trials in implant therapy: relationships among methodological, statistical, clinical, paratextual features and number of citations[J]. Clinical Oral Implants

- Research, 2010, 18(4): 419-431.
- [73] AYRES I, VARS F E. Determinants of citations to articles in elite law reviews[J]. Journal Legal Study, 2000, 29(S1): 427-450.
- [74] BORNEMANN L, DANIEL H D. What do citation counts measure? A review of studies on citing behavior[J]. Journal of Documentation, 2008, 64(1): 45-80.
- [75] GLANZEL W. National characteristics in international scientific co-authorship relations[J]. Scientometrics, 2001, 51(1): 69-115.
- [76] 罗旭. 如何破解职称“紧箍咒”[N]. 光明日报, 2017-01-05(15). (LUO X. How to crack the title “curse” [N]. Guangming Daily, 2017-01-05(15).)
- [77] YU T, YU G, LI P-Y, et al. Citation impact prediction for scientific papers using stepwise regression analysis[J]. Scientometrics, 2014, 101(2): 1233-1252.
- [78] SEGLEN P O. The skewness of science[J]. Journal of the American Society for Information Science, 1992, 43(9): 628-638.
- [79] 梁兴堃. 图情档研究中的回归分析: 基本原理[J]. 图书情报知识, 2021, 38(3): 154-164. (LIANG X K. Fundamental principles of regression analysis in library, information and archives management[J]. Documentation, Information & Knowledge, 2021, 38(3): 154-164.)
- [80] 王大顺, 巴拉巴西. 给科学家的科学思维[M]. 贾韬, 汪小帆, 译. 天津: 天津科学技术出版社, 2021: 160. (WANG D S, BARABÁSI A L. The science of science[M]. JIA T, WANG X F, trans. Tianjin: Tianjin Science and Technology Press, 2021: 160.)
- [81] DIDEGAH F, THELWALL M. Which factors help authors produce the highest impact research? Collaboration, journal and document properties[J]. Journal of Informetrics, 2013, 7(4): 861-873.
- [82] LONG R G, BOWERS W P, BARNETT T, et al. Research productivity of graduates in management: effects of academic origin and academic affiliation[J]. Academy of Management Journal, 1998, 41(6): 704-714.
- [83] 周晓英, 董伟, 朱小梅, 等. 图书馆学情报学高影响力论文特征及所反映的学科差异分析[J]. 中国图书馆学报, 2012, 38(4): 99-109. (ZHOU X Y, DONG W, ZHU X M, et al. Features and difference analysis of high-impact papers in information science and library science[J]. Journal of Library Science in China, 2012, 38(4): 99-109.)

刘宇 云南大学新闻学院教授, 博士生导师。云南 昆明 650500。

伍丹炜 云南大学历史与档案学院硕士研究生。云南 昆明 650091。

叶继元 南京大学信息管理学院特聘教授, 教育部社会科学委员会委员(管理学部委员), 博士生导师。江苏 南京 210023。

(收稿日期: 2023-01-26)