

Особенности «российского Хирша»: предикторы цитируемости научных статей в РИНЦ

А. А. Марголис, В. В. Пономарева, М. Г. Сорокова

Статья поступила
в редакцию
в мае 2019 г.

Марголис Аркадий Аронович

кандидат психологических наук, доцент, временно исполняющий обязанности ректора Московского государственного психолого-педагогического университета. E-mail: margolisaa@mgppu.ru

Пономарева Виктория Викторовна

начальник отдела информационного обеспечения и компьютеризации Фундаментальной библиотеки Московского государственного психолого-педагогического университета. E-mail: ronomarevavv@mgppu.ru

Сорокова Марина Геннадьевна

доктор педагогических наук, кандидат физико-математических наук, профессор кафедры прикладной математики факультета информационных технологий Московского государственного психолого-педагогического университета. E-mail: sorokovamg@mgppu.ru

Адрес: 127051, Москва, ул. Сретенка, 29.

Аннотация. Статья посвящена исследованию предикторов цитируемости российских научных публикаций по психологии в Российском индексе научного цитирования (РИНЦ). Рассмотрены четыре группы показателей: формальные атрибуты статьи (12 параметров), параметры видимости статьи на интернет-порталах *eLibrary* (3 параметра) и *PsyJournals* (2 параметра), определяющие доступность текста статьи для потенциального читателя, и атрибуты авторского способа научно-цитирования (3 параметра). Особое

внимание уделено атрибутам цитирования как качественным характеристикам способа работы автора над научным текстом и выстраивания диалога с другими исследователями. Методом многомерной статистики — факторного и кластерного анализа — статистически подтверждена связь ряда изучаемых параметров с цитируемостью в РИНЦ. Для каждой из четырех групп методом множественного регрессионного анализа выявлены показатели, наиболее существенные для прогнозирования цитируемости, и проведен сравнительный анализ их предиктивной способности. Показано, что самыми информативными для прогнозирования цитируемости являются параметры видимости (доступности для читателя) статьи, менее важными — атрибуты статьи, а самыми слабыми — атрибуты типа научного цитирования. Метод логистической регрессии позволил найти параметры-предикторы и определить точность предсказания принадлежности статей к группам высоко- и низкоцитируемых: для параметров видимости на *PsyJournals* и *eLibrary* она составляет 77,3 и 72,9%, а для атрибутов статьи и атрибутов цитирования — 69,9 и 60,9% соответственно. Если в статье мало диалогических (интертекстуальных) цитирований, она с высокой вероятностью будет низкоцитируемой в РИНЦ, но большое их количество не гарантирует высокой цитируемости. Даны рекомендации авторам по повышению цитируемости их статей. Выборка составила

Авторы выражают признательность за предоставление данных и сотрудничество ООО «Научная электронная библиотека», Управлению информационными и издательскими проектами, студентам ФГБОУ ВО МГППУ.

662 статьи из шести российских научных психологических журналов, индексируемых одновременно в РИНЦ и в базах *Web of Science* и *Scopus*.

Ключевые слова: Российский индекс научного цитирования, российские научные публикации по психоло-

гии, цитируемость, видимость статьи, авторский способ научного цитирования, репутация автора, авторитетность журнала.

DOI: 10.17323/1814-9545-2020-1-230-255

В национальном проекте «Наука» на 2018–2024 гг.¹ запланирован значительный рост количества российских научных журналов, включенных в международные базы данных *Web of Science* и *Scopus* (400 на 31 декабря 2021 г. и 500 на 31 декабря 2024 г.). Кроме того, поставлена задача увеличить количество публикаций российских ученых в ведущих мировых научных журналах, входящих в первый и второй квартиль по индексам научного цитирования.

Российские исследователи, работающие в сфере психологических наук, в настоящее время публикуются преимущественно в отечественной научной периодике и оцениваются с позиций Российского индекса научного цитирования (РИНЦ). Для того чтобы российские научные журналы вошли в ведущие мировые индексы цитирования и заняли там высокие позиции, они должны публиковать качественные статьи, которые интересны как российским, так и иностранным ученым и потенциально могут стать высокоцитируемыми и в России, и за рубежом. Решить задачу увеличения доли российских научных публикаций в высокорейтинговых научных изданиях, индексируемых в международных базах, невозможно без выяснения того, что определяет успех статьи в настоящее время и ее цитируемость в системе РИНЦ. Для ответа на эти вопросы мы сосредоточились на изучении факторов, влияющих на цитируемость статей в отечественных психологических журналах, входящих не только в РИНЦ, но и в *Web of Science Core Collection (Emerging Sources Citation Index – ESCI, Social Sciences Citation Index – SSCI)* и *Scopus*. Выводы из этого исследования при сопоставлении с факторами, определяющими цитируемость в международных базах, позволят понять, что необходимо учесть авторам научных публикаций и редколлегиям российских научных журналов для достижения целей национального проекта «Наука» в части увеличения доли российских научных публикаций, индексируемых в *WoS* и *Scopus*.

¹ Паспорт национального проекта «Наука». <http://static.government.ru/media/files/vCAoi8zEXRVsuy2Yk7D8hvQbpbUSwO8y.pdf>

1. Возможные предикторы цитируемости статей

Научная коммуникация, предполагающая «дистрибуцию знания» [Gieryn, 2002], состоит из двух этапов: создание элемента нового знания и ввод его в систему существующего научного знания. Одним из способов осуществления такой коммуникации является создание текстов статей [Чернявская, 2016; Баженова, 2009], которое предполагает диалог автора с научными сторонниками и оппонентами. Публикация в профильных научных журналах позволяет оперативно донести результаты исследования до научного сообщества. В лингвистических работах по анализу научных текстов используется понятие интертекстуальности как чередования высказываний, принадлежащих разным авторам: с одной стороны, это высказывания автора создаваемого текста, с другой — высказывания иных авторов в форме цитат из референтных источников [Данилевская, 2006; 2009]. Иными словами, научный текст является интертекстом благодаря наличию в нем сопоставления авторской позиции с позициями других исследователей.

Выделяют два основных типа интертекстуальных отсылок [Мишанкина, 2010]: эксплицитные (прямое и косвенное цитирование) и имплицитные (фоновая ссылка). Доминирование того или иного типа цитирования обусловлено степенью освоенности концепции, ее включенности в систему научного знания. При этом косвенное цитирование может осуществляться в виде пересказа [Валгина, 2003], т. е. выделения основной мысли публикации, или же скрытого использования фрагмента чужого текста, включенного в авторский без отсылок.

Интертекстуальное цитирование наряду с формальным и неинтертекстуальным составляют атрибуты цитирования статьи, которые мы рассматриваем как возможные предикторы цитируемости статей в РИНЦ.

Интертекстуальное цитирование — наличие в тексте прямой или косвенной цитаты, анализа содержания цитаты и соответствующего источника в списке литературы. Интертекстуальное цитирование подразумевает аналитическую работу над источником, рефлексию, критическую оценку представленного в нем материала, при этом достоверность обрабатываемой информации должна подтверждаться наличием в тексте цитаты или рассуждений, заключений с указанием источника.

Формальное цитирование — наличие источника в списке литературы без упоминания его в тексте.

Неинтертекстуальное цитирование — отсутствие позиции автора статьи по отношению к точке зрения авторов цитируемых источников. В этом случае в тексте статьи нет прямой или косвенной цитаты, авторской критической оценки цитируемого текста. Для неинтертекстуального цитирования характерны отсылки, в которых перечисляется сразу много источников, например [5; 7; 8; 9; 25]. Конечно, использование неинтертекстуального

цитирования оправданно при описании актуальности и новизны исследования, но оно часто встречается и в публикациях, направленных на искусственное повышение библиометрических показателей: такие статьи отличаются небольшим объемом текста, несоразмерным ему списком литературы и отсылками, в которых перечисляются почти все источники из списка литературы, например [1–72].

Наряду с атрибутами цитирования мы оценивали еще два вида библиометрических показателей с точки зрения их связи с цитируемостью статьи в Российском индексе научного цитирования. Вторую группу показателей составили атрибуты статьи — количественные характеристики, косвенно свидетельствующие о квалификации и репутации автора, его профессиональном кругозоре, осведомленности в материале и сфере его научной коммуникации, а также об авторитетности журнала, в котором опубликована статья. К третьей и четвертой группам отнесены альтметрики — показатели видимости статьи (ее доступности для скачивания в виде полного текста или аннотации) на порталах *eLibrary.ru* и *PsyJournals*: число просмотров аннотаций и число скачиваний или загрузки полных текстов статей, а также включение их в подборки в личных кабинетах пользователей.

Являются ли наиболее цитируемыми лучшие по качеству статьи — на этот вопрос нет однозначного ответа [García, Rodríguez-Sánchez, Fdez-Valdivia, 2019]. Установлено, что на цитируемость влияют много факторов, среди которых доступность работы для научного сообщества (например, через *Google Scholar*, *OrcidID*, *ResearchGate* или *Twitter*) и репутация журнала, но особенно важен результат рецензирования рукописи: более качественные рецензии в журнале повышают качество публикации. Показано, что статьи авторов, связанных с многопрофильными научными центрами, а также публикации результатов исследований, получивших дополнительное финансирование, имеют более высокую скорость цитирования [Youtie, 2014]. К числу факторов, положительно коррелирующих с цитируемостью, или являющихся ее предикторами, зарубежные исследователи относят параметры авторитетности журнала, определенные характеристики авторов статьи и ссылок в данной статье, а также число авторов и количество источников в списке литературы [Didegah, Thelwall, 2013; Borsuk et al., 2009], раннюю цитируемость статьи и объем ее цитирований в первые два года [Yu et al., 2014]. В зарубежных изданиях появился целый ряд публикаций, посвященных исследованию связи альтметрик с библиометрическими показателями научных статей (например, [Haustein, Costas, Larivière, 2015; Sugimoto, Larivière, 2016]). В них, в частности, показано, что вероятность появления в Твиттере наиболее высока для новостных сообщений о научных результатах и для редакционных мате-

риалов, а мейнстрим-медиа сфокусированы на научных статьях и также на редакционных материалах [Haustein, Costas, Larivière, 2015]. Выявлена положительная связь некоторых альтернативных ресурсов *Mendeley* с качеством статьи и ее цитируемостью [Bornmann, Haunschild, 2018].

Основное внимание в российских публикациях уделяется оценке эффективности научной деятельности при помощи количественных наукометрических показателей. Актуальность этих исследований авторы обосновывают необходимостью использования наукометрических характеристик наряду с экспертными оценками при принятии управленческих решений на всех этапах руководства наукой [Кабакова, 2014], и в частности при распределении финансирования [Левин, 2017]. По мнению авторов, наукометрические показатели являются ключевыми для оценки результативности деятельности вуза как научной организации и его сотрудников [Воробьев, Мишенева, 2018], а также для сравнения вузов при проведении рейтингования [Гришакина, 2016]. Возросший интерес к наукометрическим показателям вызван возможностью автоматизации процесса оценивания с использованием программных средств авторитетных баз данных — *Web of Science*, *Scopus*, *Google Scholar*, *PubMed*, *Chemical Abstracts* и др., а также доступностью информации и быстротой проверки в экспресс-оценивании публикационной деятельности ученых [Кабакова, 2014].

В отечественных работах математические методы используются для выявления особенностей динамики и взаимосвязи различных библиометрических характеристик научной деятельности отдельного автора и целой организации. Выявлена положительная корреляция между индексом Хирша автора и общим числом процитированных статей этого автора [Федоров, Попов, 2014]. С помощью модели простой линейной регрессии показана прямая связь рейтингового балла *QS University Rankings: BRICS* с количеством цитирований статьи [Шакирова, 2016]. На основании совокупности четырех библиометрических показателей организации — число статей в журналах; число статей в журналах, входящих в *Web of Science* или *Scopus*; число цитирований статей за пять лет; число авторов статей в журналах — с помощью метода кластерного анализа научно-исследовательские учреждения удалось распределить на четыре группы с разной динамикой и объемом цитирований [Крючков, Артамонова, 2016].

Особенностями нашего исследования являются анализ в качестве возможных предикторов цитируемости большой совокупности библиометрических показателей статей (20 характеристик), в том числе особой группы атрибутов цитирования статьи — параметров, отражающих способ работы автора с источниками и демонстрирующих диалогичность научного текста.

Объектом исследования являются количественные и качественные показатели цитирования и видимости (доступности) психологических статей, опубликованных в 2014–2015 гг. и представленных в научной электронной библиотеке *eLibrary.ru* и на портале психологических изданий *PsyJournals.ru*, а также их цитируемости в РИНЦ.

Предмет исследования — связь атрибутов статьи, атрибутов особенностей авторского цитирования и параметров видимости статей на порталах *eLibrary* и *PsyJournals* с цитируемостью научных публикаций.

Цель исследования состоит в выявлении предикторов цитируемости статей среди атрибутов статьи, количественных и качественных показателей цитирования источников авторами этих статей и параметров их видимости на порталах *eLibrary* и *PsyJournals*, а также сравнительный анализ их предиктивной способности.

Для достижения этой цели необходимо решить следующие задачи:

- 1) проведение библиометрического анализа для определения атрибутов статьи, атрибутов авторского цитирования и параметров видимости научных публикаций;
- 2) сравнение полярных по цитируемости в РИНЦ групп статей по обобщенным параметрам — факторам;
- 3) оценка взаимосвязи параметров видимости (альтметрик) *PsyJournals* с цитируемостью в РИНЦ;
- 4) сравнение действенности двух наборов характеристик — атрибутов статьи и атрибутов авторского цитирования, с одной стороны, и параметров видимости на *eLibrary* и *PsyJournals*, с другой, — как критериев классификации статей на более активно и менее активно цитируемые в РИНЦ;
- 5) прогнозирование показателей цитируемости по параметрам статьи и сравнительный анализ их предиктивной способности.

Для исследования были отобраны шесть высококачественных психологических журналов (табл. 1): три ведущих журнала МГППУ из 13 издаваемых им, два старейших журнала по психологии и журнал «Психология. Журнал Высшей школы экономики», занимающий 2-е место в рейтинге *Science Index* РИНЦ за 2017 г. Все журналы входят в базу *Web of Science Core Collection*.

Выборку составили 662 статьи: 315 статей за 2014 г. ($N_1 = 315$) и 347 статей за 2015 г. ($N_2 = 347$). При применении различных методов анализа объем выборки сокращается из-за особенностей работы программы SPSS версий 23 и 25 с пропущенными данными. Для 200 статей предоставлены данные просмотров аннотации и скачиваний PDF по сведениям портала психологических изданий *PsyJournals.ru*.

2. Описание исследования

2.1. Выборка

Таблица 1. **Выбранные для исследования журналы**

Название журнала	Импакт-фактор РИНЦ 2017 г.	Место в рейтинге Science Index 2017 г.	Ядро РИНЦ	Web of Science Core Collection	Scopus
Психологическая наука и образование	2,099	1	Входит	ESCI	–
Культурно-историческая психология	0,614	14	Входит	ESCI	Индексируется
Экспериментальная психология	0,682	16	Входит	ESCI	–
Вопросы психологии	1,181	3		SSCI	–
Психология. Журнал Высшей школы экономики	1,092	2	Входит	ESCI	Индексируется
Психологический журнал	1,801	5	Входит	SSCI	Индексируется

Портал психологических изданий *PsyJournals.ru* является российским тематическим репозиторием научных и научно-практических периодических и продолжающихся изданий по широкому спектру актуальных направлений психологической науки и образования². Он содержит более 95 периодических изданий с архивной коллекцией статей более чем за 20 лет. На портале размещены в полнотекстовой форме более 10 тыс. статей. В базе данных представлены сведения о более чем 8 тыс. авторов. Ежедневно портал посещают до 7 тыс. читателей.

2.2. Исследуемые параметры

В исследовании проверялась следующая гипотеза: цитируемость статьи в РИНЦ зависит от 4 типов факторов, согласующихся между собой, — формальных атрибутов статьи, атрибутов авторского способа научного цитирования, параметров видимости статьи на интернет-порталах *eLibrary* и *PsyJournals*, и она может быть спрогнозирована по совокупности этих параметров.

Перечислим исследуемые параметры, на основании которых предположительно можно прогнозировать цитируемость статьи, выраженную числом цитирований в РИНЦ (CITrisc) в базе *eLibrary*.

Атрибуты статьи (всего 12):

- 1) число источников в списке литературы (NS) — количество позиций в списке цитируемой литературы без учета дублирования (по данным *eLibrary*). Для публикаций в журналах,

² <http://psyjournals.ru>

- предоставляющих на портал *eLibrary* неполную информацию о статьях, данные уточнялись на сайтах издательств или в печатных версиях изданий;
- 2) число международных источников (NSi) — количество процитированных в статье источников, в названии которых отсутствует кириллица. По результатам содержательного анализа было учтено количество переводных изданий и проведена корректировка;
 - 3) число национальных источников (NSn) — количество процитированных в статье источников, в названии которых присутствует кириллица. Не учитывались позиции, представленные в транслитерированном виде, так как они уже были учтены в русскоязычном варианте;
 - 4) число современных источников (NSm) — количество процитированных в статье источников, опубликованных за предшествующие пять лет по отношению к году публикации статьи. Уточнено по результатам содержательного анализа;
 - 5) средневзвешенный импакт-фактор источников (IFs) — среднее по показателю «двухлетний импакт-фактор РИНЦ» для журналов, указанных в списке литературы анализируемой статьи и индексируемых в РИНЦ;
 - 6) средневзвешенный индекс Хирша авторов источников (Has) — среднее по показателю «индекс Хирша по публикациям в РИНЦ» для авторов всех источников с наличием описаний на *eLibrary*, входящих в список литературы анализируемой статьи. Для расчета использовались данные формы «Анализ публикационной активности автора» на *eLibrary*. В случае если у всех авторов источников из списка литературы анализируемой статьи отсутствовал профиль на *eLibrary*, параметру присваивалось нулевое значение;
 - 7) средневзвешенное количество цитирований источников в РИНЦ (CSrisc) — среднее по полю «Цитирования в РИНЦ» в информации о публикации для всех идентифицированных источников из списка литературы к анализируемой статье. Если среди источников в списке литературы не было идентифицированных публикаций, параметру присваивалось нулевое значение;
 - 8) средневзвешенное количество цитирований источников в WOS (CSwos) — среднее по полю «Количество цитирований» в списке найденного *Web of Science Core Collection* (WOS) для каждого источника из списка литературы к анализируемой статье, название которого указано латиницей. Если среди источников нет проиндексированных в WOS, параметру присваивалось нулевое значение;
 - 9) максимальный индекс Хирша (Hmax) — максимальное значение показателей по полю «Индекс Хирша по публикациям в РИНЦ» по данным формы «Анализ публикационной актив-

- ности автора» анализируемой статьи. Если автор один, указывался индекс Хирша этого автора;
- 10) число авторов (N_a) — число авторов, заявленных в статье;
 - 11) количество страниц (NP) определялось по указанному в информации о публикации диапазону страниц в поле «Страницы»;
 - 12) импакт-фактор журнала (IF) — двухлетний импакт-фактор РИНЦ журнала, в котором была опубликована анализируемая статья.

Атрибуты особенностей авторского цитирования (всего 3):

- 1) число фактов формального цитирования (C_f) — количество источников, не упоминающихся в тексте статьи (нет отсылок в тексте) при их наличии в списке литературы;
- 2) число фактов неинтертекстуального цитирования (C_{nit}) — количество источников, на основе которых в тексте строится «абстрактное» обобщение без конкретики, с указанием нескольких отсылок (3 и более). Если источник цитировался несколько раз, и один из них рефлексивно, формальное цитирование не засчитывалось;
- 3) число фактов интертекстуального цитирования (C_{it}) — количество источников, процитированных хотя бы один раз с обобщением или приведением цитаты, ее анализом и правильным оформлением.

Кодирование цитирований проводила проинструктированная группа студентов под руководством преподавателей. Здесь возможны случайные ошибки кодирования из-за субъективности мнений, как и при любом измерении, но не может быть системных искажений благодаря большой численности кодировщиков.

Параметры видимости (доступности публикации) на *eLibrary.ru* (всего 3):

- 1) число просмотров *eLibrary* (V_e) определялось по полю «Просмотры» раздела «Альтметрики» в информации о публикации для анализируемой статьи на портале *eLibrary*;
- 2) число загрузок (уникальные пользователи) *eLibrary* (D_e) — количество уникальных пользователей (авторов), загрузивших статью, определялось по полю «Загрузки» раздела «Альтметрики» в информации о публикации для анализируемой статьи. Учитывалось значение, указанное в скобках;
- 3) подборки *eLibrary* (S_e) — количество подборок, в которые была включена статья, определялось по полю «Включено в подборки» раздела «Альтметрики» в информации о публикации для анализируемой статьи на портале *eLibrary*. Учитыв-

валось число персональных подборок пользователей *eLibrary*, которые включают данную публикацию.

Параметры видимости на *PsyJournals.ru* (всего 2):

- 1) число просмотров аннотаций *PsyJournals* (VApj);
- 2) число скачиваний полных текстов статей в формате PDF (Dpj).

Поскольку количество исследуемых параметров достаточно велико, использовались преимущественно методы многомерной статистики, позволяющие выделить обобщенные характеристики статей, предположительно взаимосвязанные с цитируемостью в РИНЦ.

Метод эксплораторного факторного анализа применялся для выделения групп переменных, по-разному описывающих единое содержание, отраженное в названиях факторов. Далее изучались различия между высоко- и низкоцитируемыми статьями по этим обобщенным характеристикам.

Для сравнения действенности четырех групп параметров как критериев классификации статей на подвыборки более цитируемых и менее цитируемых в РИНЦ и оценки их взаимной согласованности применялся кластерный анализ. Каждая группа рассматривалась как совокупность параметров кластеризации и анализировалась по единому алгоритму: выборка статей разбивалась методом *k*-средних на два кластера, так чтобы средние значения параметров максимально различались в этих кластерах; проверялась значимость различий между ними по параметрам кластеризации для качественного описания этих кластеров, а затем по цитируемости в РИНЦ. Кроме того, группы статей из обоих кластеров сравнивались по остальным исследуемым параметрам.

Для прогнозирования показателей цитируемости в РИНЦ по параметрам статей каждая из четырех групп параметров рассматривалась по отдельности. С точки зрения особенностей математической обработки данных эта задача предполагала два возможных способа решения:

- предсказание показателя цитируемости в РИНЦ статей по совокупности параметров публикации и выявление среди них наиболее существенных для такого прогноза. Эта задача решалась методом множественного регрессионного анализа;
- рассмотрение двух полярных по уровню цитируемости в РИНЦ групп статей: низкоцитируемых (не более одного цитирования) и высокоцитируемых (8 и более цитирований); прогнозирование принадлежности статьи с большой вероятностью (более 0,5) к одной из этих групп по совокупности

2.3. Методы исследования

параметров публикаций; определение показателей, наиболее существенных для предсказания, и точности прогноза. Эта задача решалась методом логистической регрессии.

3. Анализ результатов
3.1. Сравнение полярных по цитируемости в РИНЦ групп статей по обобщенным параметрам

Проведен эксплораторный факторный анализ с извлечением факторов методом главных компонент и ортогональным Вари-макс-вращением. Факторизуемость корреляционной матрицы проверялась с помощью критериев Кайзера—Майера—Олкина (КМО) и Бартлетта. Критерий КМО = 0,661 подтверждает достаточную адекватность выборки, критерий сферичности Бартлетта ($\chi^2 = 4447,12$, $df = 153$, $p < 0,001$) также свидетельствует о возможности описания корреляционной матрицы с помощью эксплораторного факторного анализа.

Выделены шесть факторов, в совокупности объясняющих 66,14% общей дисперсии, что благоприятно. Факторы 1-й и 2-й, а также с 3-го по 6-й являются примерно равнозначными по вкладу в общую объясненную дисперсию (F1 — 17,4%, F2 — 13,9%, F3 — 10,7%, F4 — 8,8%, F5 — 7,9%, F6 — 7,2%) и допускают отчетливую содержательную интерпретацию. Для интерпретации факторов выбирались переменные, имеющие максимальные нагрузки на этот фактор (по модулю не менее 0,4), одновременно превосходящие факторные нагрузки данной переменной на все остальные факторы. В табл. 2 переменные, использованные для интерпретации, расположены по убыванию нагрузок.

Для проверки различий между группами статей с разным уровнем цитируемости в РИНЦ по выделенным факторам сначала были рассчитаны квартили по переменной CITrisc: нижний квартиль P25 = 1, медиана P50 = 3, верхний квартиль P75 = 8. В соответствии с ними статьи были распределены на три группы: с низкой (не более одного цитирования), средней (от 2 до 7 цитирований) и высокой цитируемостью (8 цитирований и выше).

Оценка различий между этими группами по факторам F1–F6 проведена методом непараметрического однофакторного анализа Краскела—Уоллиса. Обнаружены достоверные различия по факторам F1 «Количественные показатели источников» ($p = 0,035$, $p < 0,05$), F2 «Видимость статей на eLibrary» ($p = 0,000$, $p < 0,001$), F3 «Весомость источников и журнала» ($p = 0,000$, $p < 0,001$), F5 «Качественные показатели источников» ($p = 0,009$, $p < 0,01$) и F6 «Показатели авторского коллектива статьи» ($p = 0,000$, $p < 0,001$). По фактору F4 «Показатели текста» различий нет ($p = 0,315$). Дополнительно исследовались попарные различия по каждому из факторов F1, F2, F3, F5 и F6 по критерию Манна—Уитни.

Особый интерес, на наш взгляд, представляют два факта. Во-первых, есть достоверные различия между полярными

Таблица 2. Интерпретация факторов

Фактор	Нагрузка	Переменная
Фактор 1 Количественные показатели источников	0,817	Число современных источников (NSm)
	0,760	Число фактов неинтертекстуального цитирования (Cnit)
	0,726	Число международных источников (NSi)
	0,741	Число фактов формального цитирования (Cf)
	0,638	Число источников в списке литературы (NS)
Фактор 2 Видимость статей на eLibrary	0,827	Число загрузок (уникальные пользователи) eLibrary.ru (De)
	0,778	Число просмотров eLibrary.ru (Ve)
	0,752	Подборки eLibrary.ru (Se)
Фактор 3 Весомость источников и журнала	0,731	Средневзвешенный импакт-фактор источников (IFs)
	0,465	Импакт-фактор журнала (IF)
	0,449	Число национальных источников (NSn)
Фактор 4 Показатели текста	0,666	Количество страниц (NP)
	0,526	Число фактов интертекстуального цитирования (Cit)
Фактор 5 Качественные показатели источников	0,820	Средневзвешенное число цитирований источников РИНЦ (CSrisc)
	0,686	Средневзвешенный индекс Хирша авторов источников (Has)
Фактор 6 Показатели авторского коллектива статьи	0,794	Число авторов (Na)
	0,567	Максимальный индекс Хирша (Hmax)

группами с высоким и низким уровнем цитируемости в РИНЦ по всем указанным факторам, причем во всех случаях значения фактора в «высокой группе» по CИTrisc достоверно выше, чем в «низкой группе».

Во-вторых, по F2 «Видимость статей на eLibrary» и F6 «Показатели авторского коллектива статьи» значимые различия выявлены между всеми тремя группами: показатели в средней группе по цитируемости в РИНЦ достоверно превышают показатели в группе низкоцитируемых статей, а показатели в группе высокоцитируемых статей достоверно выше показателей в двух других группах.

Для разбиения этой выборки статей на три группы по обеим альтметрикам *PsyJournals* были вычислены квартили: VArj

3.2. Оценка взаимосвязи параметров видимости *PsyJournals* с цитируемостью в РИНЦ

Таблица 3. **Оценка различий между тремя группами статей с разным уровнем альтметрик *PsyJournals* по цитируемости в РИНЦ методом однофакторного анализа по Краскелу — Уоллису**

Альтметрики <i>PsyJournals</i>	Уровень альтметрики <i>PsyJournals</i> . Среднее по цитируемости (M)			Статистика критерия Краскела — Уоллиса	Уровень значимости <i>p</i>
	Низкий (N1 = 50)	Средний (N2 = 100)	Высокий (N3 = 50)		
D _{pj}	3,18	6,11	20,02	32,462	0,000
V _{Apj}	2,54	6,10	20,68	37,744	0,000

Таблица 4. **Сравнение высоко-, средне- и низкоцитируемых в РИНЦ групп статей по альтметрикам *PsyJournals* с помощью критерия Краскела — Уоллиса**

Альтметрики <i>PsyJournals</i>	Уровни CiTrisc. Среднее по цитируемости (M)			Статистика Краскела — Уоллиса	Уровень значимости <i>p</i>
	Низкий: 0–1 (N1 = 65)	Средний: 2–7 (N2 = 81)	Высокий: 8 и более (N3 = 54)		
D _{pj}	644,14	911,06	1149,33	42,854	0,000
V _{Apj}	999,06	1252,42	1581,67	40,172	0,000

(P25 = 874,25, P50 = 1100,5, P75 = 1477,5), D_{pj} (P25 = 543,0, P50 = 715,0, P75 = 1054,25). Различия по цитируемости в РИНЦ между группами статей с низкими, средними и высокими показателями альтметрик *PsyJournals* проверялись по критерию Краскела — Уоллиса. Результаты представлены в табл. 3, где указаны средние значения CiTrisc (M) для каждой из трех групп, различающихся уровнем альтметрик. Для обеих альтметрик получены достоверные различия ($p < 0,000$).

Попарные сравнения по Манну — Уитни также выявляют достоверные различия (от $p < 0,01$ до $p < 0,001$) между всеми группами статей по средней цитируемости в РИНЦ для обеих альтметрик *PsyJournals*, причем показатели цитируемости достоверно выше на каждом следующем более высоком уровне по сравнению с предыдущим.

Далее решалась обратная задача — проверить различия между группами с разным уровнем цитируемости в РИНЦ по обоим альтметрикам *PsyJournals* по критерию Краскела — Уоллиса (табл. 4): различия достоверны в обоих случаях. В табл. 4 приведены средние значения (M) по соответствующей альтметрике *PsyJournals* для каждой из трех групп статей.

Попарные различия между этими группами статей по показателям альтметрик *PsyJournals* оценивались по критерию Манна—Уитни: различия достоверны во всех случаях ($p < 0,01$ или $p < 0,001$), причем показатели по каждой альтметрике *PsyJournals* достоверно выше на каждом следующем более высоком уровне цитируемости по сравнению с предыдущим. Этот результат особенно важен, так как он получен на выборке объемом всего 200 статей, в то время как аналогичный результат для F2 «Видимость статей на *eLibrary*» получен на выборке втрое большего объема — более 600 статей.

Связь альтметрик *PsyJournals* с CИTrisc исследовалась также с помощью коэффициента ранговой корреляции Спирмена. Выявлена статистически значимая ($p < 0,001$) прямая связь средней силы между CИTrisc и переменной V_{Apj} ($\rho = 0,472$, $p < 0,001$) и переменной D_{pj} ($\rho = 0,484$, $p < 0,001$). Это означает, что чем выше значения показателей по обоим альтметрикам *PsyJournals*, тем выше цитируемость статей в РИНЦ.

Таким образом, связь альтметрик *PsyJournals* с цитируемостью можно считать статистически доказанной: чем выше видимость статей на *PsyJournals*, тем выше цитируемость в РИНЦ.

Для решения этой задачи общая выборка статей методом кластерного анализа k -средними разбивалась на два кластера сначала по совокупности параметров видимости на *eLibrary* (всего 3), затем — на *PsyJournals* (всего 2), далее по совокупности атрибутов цитирования (всего 3) и, наконец, по совокупности атрибутов статьи (всего 12). Во всех четырех случаях в один из кластеров входили статьи с более высокими показателями по параметрам кластеризации (и по большинству оставшихся показателей), которые и достоверно лучше цитировались, чем статьи из другого кластера. Следовательно, эти четыре типа параметров согласуются друг с другом: при их повышении цитируемость в РИНЦ также имеет тенденцию к увеличению.

Поясним на примере, как мы рассуждали далее при сравнении действенности каждой из четырех групп параметров как критериев классификации статей на более активно и менее активно цитируемые научным сообществом. Предположим, имеется некоторое множество объектов — в нашем случае научных статей — и требуется найти критерии разделения их на «хорошие» и «плохие», в нашем случае на более цитируемые и менее цитируемые. Например, на олимпиаде школьников нужно разработать набор задач, чтобы отделить группу интеллектуалов — победителей олимпиады от остальных учащихся. Если набор олимпиадных задач позволит это сделать, то средний показатель интеллекта в группе победителей будет очень высоким, а в группе остальных участников — существенно ниже, т. е. разность средних будет велика. Если же задачи не очень удач-

3.3. Сравнение действенности четырех групп параметров как критериев классификации статей на более цитируемые и менее цитируемые в РИНЦ и оценка их взаимной согласованности

ные, то среди победителей будут как интеллектуалы, так и менее способные, а часть одаренных школьников окажется среди проигравших. В этом случае средние показатели интеллекта в группах победителей и проигравших будут ближе друг к другу, т. е. разность между ними будет меньше. Аналогично можно рассуждать, сравнивая распределения победителей и проигравших по уровням интеллекта, что дает более общую картину. Если среди победителей по сравнению с проигравшими значительно преобладает доля школьников с высоким интеллектом, а среди проигравших — наоборот, с относительно низким, то набор олимпиадных задач можно считать удачным критерием. Чем меньше контраст долей, тем хуже эти задачи выполняют функцию критерия для выявления талантливых учащихся.

Обращает на себя внимание соотношение средних значений цитируемости в РИНЦ в «высоком» и «низком» кластерах: по видимости на *PsyJournals* разность средних самая большая (22,8 против 5,1, Манн — Уитни, $p < 0,001$), по видимости на *eLibrary* она несколько меньше (14,6 против 5,7, Манн — Уитни, $p < 0,001$), а самая маленькая эта разность для двух кластеров, выделенных по атрибутам цитирования (7,8 против 5,5, Манн — Уитни, $p < 0,001$), а также по атрибутам статьи (7,5 против 5,3, Манн — Уитни, $p < 0,001$).

Сравним также соотношение процентных долей статей из «высокого» и «низкого» кластеров в полярных группах по цитируемости в РИНЦ, т. е. в «низкой» (не более одного цитирования) и в «высокой» группе (8 и более цитирований).

Для кластера более видимых на *eLibrary* статей соотношение долей в «высокой» группе по цитируемости составляет 52,3% против 20,7% в кластере менее видимых, а в «низкой» группе — 9,3% против 35,3% от соответствующих выборок. Видимость на *PsyJournals* дает примерно те же пропорции: для кластера более видимых на *PsyJournals* статей соотношение долей в «высокой» группе по цитируемости составляет 52,4% против 20,3% в кластере менее востребованных, а в «низкой» группе — 7,1% против 39,2% от выборок. В обоих случаях различия достоверны (χ^2 , $p < 0,001$).

Совокупность атрибутов статьи не дает такого контраста в пропорциях высоко- и низкоцитируемых статей в кластерах с лучшими и худшими атрибутами: для «высокой» группы по цитируемости соотношение составляет 28,8% против 21,4%, а для «низкой» группы по цитируемости — 24,2% против 34,6% от соответствующих кластеров. Другими словами, пропорции (процентные доли) статей из обоих кластеров — с лучшими и худшими атрибутами статьи — в полярных категориях по цитируемости в РИНЦ ближе к 1, а уровень значимости различий ниже (χ^2 , $p < 0,05$).

При сравнении распределений статей с более высокими и более низкими атрибутами цитирования контраст также оказы-

ваются не очень сильным: пропорции практически те же, а различия достоверны на том же уровне значимости (χ^2 , $p < 0,05$). Статьи с более высокими атрибутами цитирования чаще являются высокоцитируемыми в РИНЦ (31,3% против 21,8% от выборок), в то время как для низкоцитируемых статей пропорции составляют 21,2% против 34,2% от тех же кластеров.

24,2% публикаций с лучшими атрибутами статьи и 21,2% с лучшими атрибутами цитирования, тем не менее, являются низкоцитируемыми в РИНЦ, т. е. имеют не более одного цитирования. И наоборот, 21,4% статей с худшими атрибутами статьи и 21,8% статей с худшими атрибутами цитирования цитируются 8 и более раз. Получается, что группировка статей по совокупности атрибутов статьи или атрибутов цитирования обеспечивает цитируемость в РИНЦ в меньшей степени (или с большей частотой ошибки), чем группировка по видимости на *eLibrary* или *PsyJournals*.

Таким образом, более действенной с точки зрения прогнозирования цитируемости статьи в РИНЦ среди исследованных групп параметров является видимость на *eLibrary* и *PsyJournals*. Это означает, что более высокой цитируемости научных статей с большей уверенностью можно ожидать с ростом значений альтметрик *eLibrary* или *PsyJournals*, чем при возрастании показателей атрибутов статьи и атрибутов цитирования. Вместе с тем видимость статьи сама может зависеть от ее цитируемости, поэтому речь идет не о каузальности, а о взаимосвязи.

Чтобы проверить, какие факторы являются максимально значимыми для прогнозирования цитируемости в РИНЦ, был проведен множественный регрессионный анализ с пошаговым отбором для каждой из четырех групп факторов по отдельности. Тем самым внутри этих групп выявлялись переменные, которые можно считать наиболее существенными предикторами цитируемости. Зависимой переменной стала *CITrisc*. Все выделенные далее предикторы для всех четырех групп факторов положительно коррелируют с *CITrisc*, и, следовательно, их возрастание ассоциировано с повышением цитируемости.

Перед применением множественного регрессионного анализа проведен корреляционный анализ по Пирсону и Спирмену. Выявлены только слабые и средние значимые корреляции между некоторыми переменными. Наиболее выражена прямая связь по Спирмену ($p < 0,001$) у переменной *CITrisc* с переменными *Ve* ($\rho = 0,404$), *De* ($\rho = 0,336$), *Se* ($\rho = 0,316$) и *Hmax* ($\rho = 0,334$). Корреляции по Пирсону дают примерно ту же картину. Отсутствие корреляций, близких к единице, между кандидатами в предикторы дает основания применить множественный регрессионный анализ.

3.4. Прогнозирование показателей цитируемости *CITrisc* по параметрам статьи

Для двух групп показателей, слабее других ассоциированных с повышением цитируемости, получены следующие результаты. Среди атрибутов статьи (всего 12) выделены 3 предиктора, совокупность которых объясняет 17,6% дисперсии зависимой переменной CITrisc (R^2 равен 0,176) и значимо влияет на нее (ANOVA, $p < 0,001$). Предикторами являются в порядке убывания их важности для предсказания, о которой говорит величина стандартизованного коэффициента β : Hmax ($\beta = 0,317$, $p < 0,001$), IF ($\beta = 0,197$, $p < 0,001$) и NS ($\beta = 0,102$, $p < 0,01$). Для группы атрибутов цитирования (всего 3) был выделен только один, но в содержательном плане самый важный предиктор — число фактов интертекстуального цитирования (Cit). Он объясняет лишь 1,4% дисперсии зависимой переменной, но значимо влияет на нее (ANOVA, $p < 0,01$). Величина коэффициента $\beta = 0,119$ ($p < 0,01$).

Множественный регрессионный анализ параметров видимости на *eLibrary* (всего 3) позволил выделить 2 предиктора — De ($\beta = 0,383$, $p < 0,001$) и Ve ($\beta = 0,091$, $p < 0,05$), объясняющих 19,6% дисперсии переменной CITrisc, что на 2% больше, чем объясняют предикторы — атрибуты статьи, и значимо влияющих на нее (ANOVA, $p < 0,001$). Наконец, при включении в анализ параметров видимости на *PsyJournals* (всего 2) обе переменные оказались предикторами: VApj ($\beta = 0,363$, $p < 0,001$) и Dpj ($\beta = 0,179$, $p < 0,05$), объясняющими самый большой процент (24,7%) дисперсии переменной CITrisc и значимо влияющими на нее (ANOVA, $p < 0,001$).

Далее приводятся результаты логистической регрессии с пошаговым включением в уравнение переменных-предикторов по методу максимального правдоподобия для этих же четырех групп параметров. Зависимая переменная Y (бинарная): «низкая группа» по CITrisc (не более одного цитирования, $N = 200$), «высокая группа» (8 и более цитирований, $N = 145$). Во всех случаях B-коэффициенты регрессионных уравнений оказались положительными, следовательно, выделенные параметры повышают вероятность попадания статьи в высокую группу по цитируемости в РИНЦ. Чем больше коэффициент B_i для данной переменной X_i , тем больший вклад она вносит в предсказание того, что статья будет отнесена к высокой группе.

Для группы атрибутов статьи (всего 12) наиболее важными для предсказания стали 3 переменные, включенные в итоговое уравнение регрессии и перечисленные по убыванию существенности вклада: IF (регрессионный коэффициент $B_1 = 1,084$, $p < 0,001$), Hmax ($B_2 = 0,082$, $p < 0,001$), NSn ($B_3 = 0,025$, $p < 0,05$). Процент правильных предсказаний для статей из «низкой группы» по CITrisc составляет 80,5%, для «высокой группы» — 55,2%, а общая процентная доля — 69,9%, что весьма достойно. Для группы атрибутов цитирования (всего 3)

предиктором является только Cit ($B = 0,020$, $p < 0,01$). Точность предсказания для статей из «низкой группы» по CITrisc очень высокая — 92,0%, но для «высокой группы» очень мала — всего 17,9%. Общая процентная доля правильных предсказаний — 60,9%, что тоже благоприятно.

Анализ параметров видимости на *eLibrary* (всего 3) показывает, что важными для предсказания стали две переменные: De ($B1 = 0,081$, $p < 0,001$) и Ve ($B2 = 0,015$, $p < 0,001$). Процент правильных прогнозов для статей из «низкой группы» по CITrisc очень высок и составляет 87,4%, для «высокой группы» он несколько ниже — 55,6%, а общая процентная доля составила 72,9%, что выше, чем для атрибутов статьи и атрибутов цитирования. Наконец, для группы параметров видимости на *PsyJournals* (всего 2) обе переменные оказались важны для предсказания: Dpj ($B1 = 0,002$, $p < 0,01$) и VApj ($B2 = 0,001$, $p < 0,05$). Процент правильных предсказаний здесь самый высокий: для статей из «низкой группы» по CITrisc он составляет 87,7%, для «высокой группы» — 64,8%, а общая процентная доля — 77,3%, что весьма благоприятно.

Строго говоря, полученные результаты свидетельствуют о сопутствующих изменениях исследованных параметров, а не о причинно-следственной связи. Действительно, размещение статей на порталах является способом их продвижения и поэтому может увеличивать их видимость и цитируемость. С другой стороны, наиболее цитируемые статьи, имеющие ценность для научного сообщества, чаще скачиваются и привлекают внимание пользователей порталов, поэтому можно говорить о взаимном влиянии цитируемости и видимости публикаций. С этой точки зрения исследованные показатели отчасти можно рассматривать не только как предикторы, но и как параметры, влияющие на цитируемость.

Следующие показатели позволяют судить о профессиональном кругозоре автора статьи и его осведомленности в предметной области: число источников в списке литературы (NS), число международных источников (NSi), число национальных источников (NSn), число современных источников (NSm), число фактов интертекстуального цитирования (Cit). И даже такие показатели, как число фактов формального цитирования (Cf) и число фактов неинтертекстуального цитирования (Cnit), несмотря на то что они снижают качество статьи, все же свидетельствуют об определенной степени владения материалом, так как в список литературы включаются обычно наиболее известные работы. О значимости процитированных источников и репутации авторов свидетельствуют средневзвешенный импакт-фактор источников (IFs), средневзвешенный индекс Хирша авторов источников

4. Обсуждение результатов

(Has), максимальный индекс Хирша (Hmax), средневзвешенное число цитирований источников РИНЦ (CSrisc), средневзвешенное количество цитирований источников в WOS (CSwos), несмотря на то что автор статьи и не ориентируется на эти показатели при цитировании. Число авторов (Na) можно отнести к характеристикам профессиональной коммуникации, а импакт-фактор журнала (IF), в котором опубликована статья, — к показателям его авторитетности. Наконец, количество страниц (NP), хотя и ограничивается требованиями журнала, позволяет отсеять слишком короткие — а значит, скорее всего, малосодержательные публикации. Оставшиеся три параметра видимости на *eLibrary.ru* (число просмотров *eLibrary* (Ve), число загрузок (уникальные пользователи) *eLibrary* (De), подборки *eLibrary* (Se)) и два параметра на *PsyJournals* (число просмотров аннотаций *PsyJournals* (VApj) и число скачиваний полных текстов статей в формате PDF (Dpj)) также могут повышать цитируемость, так как увеличивают доступность публикации.

Сравнительный анализ действенности четырех групп параметров как критериев классификации статей на более цитируемые и менее цитируемые в РИНЦ показал, что самыми сильными являются параметры видимости на интернет-порталах *eLibrary* и *PsyJournals*, затем атрибуты статьи и самыми слабыми — атрибуты цитирования. Действительно, при сравнении статей из «высокого» и «низкого» кластеров по цитируемости в РИНЦ получены достоверные различия для всех четырех групп параметров, однако разность в средних значениях *CITrisc* наибольшая для параметров видимости на *PsyJournals* и на *eLibrary* и наименьшая — для атрибутов статьи и атрибутов цитирования. Сравнение распределений статей из двух кластеров по уровням цитируемости в РИНЦ подтверждает этот вывод: статьи из «высокого кластера» с большей вероятностью являются высокоцитируемыми в РИНЦ, а статьи из «низкого кластера» — низкоцитируемыми. Но если при кластеризации по параметрам видимости на *PsyJournals* и на *eLibrary* этот контраст велик и в группе высокоцитируемых в РИНЦ доля статей из «высокого кластера» примерно на 30% больше, чем доля статей из «низкого кластера», то в группах атрибутов статьи и атрибутов цитирования этот контраст значительно слабее: для них разница составляет не более 10%. Этот вывод особенно важен для выработки практических рекомендаций авторам по повышению цитируемости их статей. По-видимому, нужно стремиться сделать статьи более открытыми и доступными, возможно, не только через *eLibrary*, но и с помощью других ресурсов, например через порталы *PsyJournals*, *ResearchGate* и др.

Результаты факторного анализа подтверждают сделанный вывод. Кроме того, обратим внимание на фактор F6 «Показатели авторского коллектива статьи», значения которого повышают

ся с ростом цитируемости в РИНЦ. Этот фактор, включающий число авторов (N_a) статьи, отчасти отражает командный характер работы, поэтому следует стремиться к работе в коллаборациях и сотрудничеству с коллегами, возможно, из смежных научных направлений. Второй составляющей этого фактора является максимальный индекс Хирша (H_{max}), следовательно, к работе в исследовательских коллективах желательно привлекать авторитетных ученых. Этот вывод согласуется с требованиями финансирующих организаций при распределении грантов.

Прогнозирование показателей цитируемости в РИНЦ для каждой из четырех групп по отдельности также подтверждает большую действенность параметров видимости статьи по сравнению с ее атрибутами и атрибутами цитирования. Так, при применении множественного регрессионного анализа наибольший процент дисперсии зависимой переменной CIT_{risc} объясняют параметры видимости на *PsyJournals* (24,7%), затем — параметры видимости на *eLibrary* (19,6%), еще менее — из группы атрибутов статьи (17,6%), и самый маленький процент дисперсии объясняет единственный, но очень важный предиктор «число фактов интертекстуального цитирования» из группы атрибутов цитирования — (1,4%). О том же говорят результаты применения метода логистической регрессии, где особенно важна точность прогнозирования. Общий процент правильных предсказаний для группы параметров видимости *PsyJournals* самый высокий (77,3%), чуть ниже он для группы параметров видимости на *eLibrary* (72,9%), еще ниже — для атрибутов статьи (69,9%) и самый низкий, но все же вполне достойный, — для атрибутов цитирования (60,9%).

Предикторы всех четырех групп предсказывают принадлежность статей к «высокой группе» по цитируемости хуже, чем к «низкой»: здесь процент правильных прогнозов составляет 64,8, 55,6, 55,2 и всего лишь 17,9% соответственно. Последнее означает, что если в статье мало интертекстуальных цитирований, она с очень большой вероятностью будет низкоцитируемой в РИНЦ, но большое их количество не гарантирует высокой цитируемости в РИНЦ. На это же указывает согласованность всех четырех групп параметров: статьи с более высокими качественными и количественными показателями, как правило, имеют лучшую видимость и чаще цитируются. Однако и худшие по качеству статьи могут стать высокоцитируемыми, а статьи лучшего качества — остаться незамеченными научным сообществом.

Для повышения цитируемости статьи в РИНЦ важны еще три ее формальных параметра: импакт-фактор журнала, число источников в списке литературы и число национальных источников. Таким образом, авторитетность журнала является существенным фактором продвижения статьи. Список литературы должен формироваться только из источников, являющихся на-

учной основой исследования, чтобы, с одной стороны, обеспечить возможность интертекстуальности, а с другой — не делать его чрезмерно большим, что может провоцировать формальное цитирование. Число национальных источников важно в первую очередь для индексации в региональных базах цитирования.

5. Ограничения методов исследования При проведении факторного анализа использовались не все 20 переменных, а только 17 из них. Параметр «средневзвешенное количество цитирований источников в WOS» нарушал факторную структуру, поэтому здесь не рассматривался. Он использовался при применении других методов, например регрессионного анализа. Параметры видимости на *PsyJournals* исключались, так как выборка статей, размещенных на этом портале, была втрое меньшего объема ($N = 200$). Вопрос о влиянии этих параметров на цитируемость в РИНЦ исследовался отдельно методом кластерного анализа.

Другим ограничением является применение методов факторного анализа и множественной регрессии при отсутствии согласия переменных с нормальным распределением. Отчасти это связано с большим объемом выборки: на выборках объема $N > 200$ статистический критерий показывает различия даже при очень слабых отклонениях от нормальности. Кроме того, распределения показателей цитируемости всегда имеют выраженную левостороннюю асимметрию: подавляющее большинство статей мало цитируются, и лишь небольшая их доля становится особенно заметной. В зарубежных исследованиях эту проблему решают разными способами, например с помощью логарифмической нормализации переменных и применения более робастной ковариационной матрицы [Bornmann, Haunschild, 2018] или использования пуассоновской регрессионной модели [Didegah, Thelwall, 2013]. Вместе с тем распространена и практика не вполне строгого применения параметрических методов без обсуждения проблемы нормальности распределения данных [Yu et al., 2014; Vorsuk et al., 2009].

6. Выводы Успех статьи зависит как от атрибутов цитирования и атрибутов самой статьи, так и от ее видимости на интернет-порталах. Однако, как следует из результатов нашего исследования, параметры видимости на *eLibrary* и *PsyJournals* являются более существенными предикторами цитируемости статей в РИНЦ, чем атрибуты самой статьи: они позволяют с большей точностью предсказать принадлежность статьи к группе высокоцитируемых.

Интертекстуальные ссылки — единственный предиктор повышения цитируемости в РИНЦ среди атрибутов цитирования.

Отсутствие в статье диалога автора с членами научного сообщества с высокой вероятностью ведет к неудаче статьи. Вместе с тем большое количество интертекстуальных ссылок не гарантирует высокой цитируемости статьи. При этом нельзя недооценивать качество текста научной статьи: статья с формальными или неинтертекстуальными ссылками не будет принята в авторитетные научные журналы.

По результатам нашего исследования сформулируем кратко основные рекомендации авторам научных публикаций по повышению цитируемости. Ссылки в статье должны быть только интертекстуальными. Нужно стремиться сделать статьи более открытыми и доступными, возможно, не только через *eLibrary*, но и с помощью других ресурсов, например через порталы *PsyJournals*, *ResearchGate*. Предпочтительно работать в коллаборациях с коллегами, привлекая к сотрудничеству авторитетных специалистов. Нужно стремиться публиковать статьи в журналах с высоким импакт-фактором. Список литературы должен формироваться только из источников, являющихся научной основой исследования, быть достаточно большим (по данным зарубежных статей — в среднем от 10 до 30 источников для эмпирических исследований). Если исследование сфокусировано на российскую аудиторию, целесообразно включить достаточное число национальных источников, так как это важно для индексации в региональных базах цитирования.

Литература

1. Баженова Е. А. (2009) Научный текст в дискурсивно-стилистическом аспекте // Вестник Пермского университета. Российская и зарубежная филология. № 5. С. 24–32.
2. Валгина Н. С. (2003) Теория текста: учеб. пособие. М.: Логос.
3. Воробьев Н. Е., Мишенева Ю. И. (2018) Публикационная активность и оценка качества работы преподавателей высшей школы // Педагогика. № 2. С. 121–125.
4. Гришакина Е. Г. (2016) Публикационная активность российских исследователей: университетская наука // Наука. Инновации. Образование. № 4 (22). С. 137–151.
5. Данилевская Н. В. (2009) Научный текст в аспекте теории дискурса // Язык. Текст. Дискурс. № 7. С. 37–46.
6. Данилевская Н. В. (2006) Чередование старого и нового знания как механизм развертывания научного текста (Аксиологический аспект): автореф. дис. ... докт. филол. наук. Екатеринбург: Уральский государственный педагогический университет. <https://dlib.rsl.ru/01003273950>
7. Кабакова Е. А. (2014) Использование наукометрических показателей при оценке научной деятельности // Вопросы территориального развития. № 8 (18). <http://vtr.isert-ran.ru/article/1437>
8. Крючков Д. В., Артамонова Г. В. (2016) Статистический анализ публикационной активности научно-исследовательских медицинских учреждений (по базе данных РИНЦ) // Социальные аспекты здоровья населения. Т. 48. № 2. С. 9.

9. Левин В. И. (2017) Математические методы оценки вклада ученого в науку // Системы управления, связи и безопасности. № 1. С. 160–174.
10. Мишанкина Н. А. (2010) Лингвокогнитивное моделирование научного дискурса: автореф. дис. ... докт. филол. наук. Томск: Томский государственный университет. <https://dlib.rsl.ru/viewer/01004611456#?page=1>
11. Федоров П. П., Попов А. И. (2014) Взаимосвязь показателей цитирования российских ученых // Вестник Российской академии наук. Т. 84. № 3. С. 222–232.
12. Чернявская В. Е. (2016) Научно-исследовательская статья как вербализация нового научного результата (на материале лингвистики) // Жанры речи. № 1 (13). С. 56–64. doi:10.18500/2311-0740-2016-1-13-56-64.
13. Шакирова Ю. А. (2016) Особенности влияния наукометрических показателей на динамику рейтинговых позиций вузов // Высшее образование в России. № 3. С. 161–165.
14. Bornmann L., Haunschild R. (2018) Do Altmetrics Correlate with the Quality of Papers? A Large-Scale Empirical Study Based on F1000Prime Data // PLoS One. Vol. 13. No 5. e0197133. doi:10.1371/journal.pone.0197133.
15. Borsuk R. M., Budden A. E., Leimu R., Aarssen L. W., Lortie C. J. (2009) The Influence of Author Gender, National Language and Number of Authors on Citation Rate in Ecology // The Open Ecology Journal. Vol. 2. No 1. P. 25–28.
16. Didegah F., Thelwall M. (2013) Determinants of Research Citation Impact in Nanoscience and Nanotechnology // Journal of the American Society for Information Science and Technology. Vol. 64. No 5. P. 1055–1064.
17. García J. A., Rodríguez-Sánchez R., Fdez-Valdivia J. (2019) Do the Best Papers Have the Highest Probability of Being Cited? // Scientometrics. Vol. 118. No 3. P. 885–890.
18. Giere R. N. (2002) Scientific Cognition as Distributed Cognition // P. Carruthers, St. Stich, M. Siegal (eds) The Cognitive Basis of Science. Cambridge: Cambridge University. P. 285–299.
19. Haustein S., Costas R., Larivière V (2015) Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns // PLoS ONE. Vol. 10. No 3. e0120495. doi:10.1371/journal.pone.0120495.
20. Sugimoto C. R., Haustein S., Ke Q., Larivière V. (2016) Social Media Metrics as Indicators of Broader Impact // OECD Blue Sky III Forum on Science and Innovation Indicators (Ghent, Belgium, 19–21 September 2016). <https://www.oecd.org/sti/172%20-%20SugimotoOECDaltmetrics.pdf>
21. Youtie J. (2014) The Use of Citation Speed to Understand the Effects of a Multi-Institutional Science Center // Scientometrics. Vol. 100. No 3. P. 613–621.
22. Yu T., Yu G., Li P. Y., Wang L. (2014) Citation Impact Prediction for Scientific Papers Using Stepwise Regression Analysis // Scientometrics. Vol. 101. No 2. P. 1233–1252.

The “Russian Hirsch”: Predictors of Citation Usage of Scholarly Works in the RSCI

Arkady Margolis

Candidate of Sciences in Psychology, Associate Professor, Interim Rector of Moscow State University of Psychology and Education. E-mail: margolisaa@mgppu.ru

Authors

Viktoria Ponomareva

Head of the Information Support and Computerization Department, Fundamental Library, Moscow State University of Psychology and Education. E-mail: ponomarevavv@mgppu.ru

Marina Sorokova

Doctor of Sciences in Pedagogy, Candidate of Sciences in Physics and Mathematics, Professor, Department of Applied Mathematics, Faculty of Informational Technologies, Moscow State University of Psychology and Education. E-mail: sorokovamg@mgppu.ru

Address: 29 Sretenka Str., 127051 Moscow, Russian Federation.

The article investigates the predictors of citation rate in the Russian Science Citation Index (RSCI) for Russian publications in psychology. Four groups of indicators are analyzed: formal attributes of a publication (12 indicators), parameters of publication visibility on eLibrary (3 indicators) and PsyJournals (2 indicators) that define accessibility of publication to potential readers, and author-based citation parameters (3). Special attention is paid to citation parameters as qualitative characteristics of the author’s method of elaborating the scientific text and construing dialogue (in the form of citations) with other researchers. Relationship between the indicators analyzed and the RSCI citation rate is proven statistically using the multivariate statistical methods of factor analysis and cluster analysis. For each of the four groups, the strongest predictors of citation rate are identified by multiple regression analysis, which are then compared by their predictive power. It is shown that visibility (accessibility) indicators are the best predictors of citation rate, followed by formal publication attributes and, finally, citation type parameters as having the lowest predictive power. The method of logistic regression allows to identify the ultimate predictors of citation rate and measure their accuracy in predicting whether a publication is low- or highly cited, which is 77.3% and 72.9% for the indicators of visibility on PsyJournals and eLibrary (respectively), 69.9% for formal attributes, and 60.9% for citation parameters. A publication that has few in-text citations is very likely to have a low RSCI citation rate, yet a high number of in-text citations does not guarantee a high citation impact. Recommendations are provided for authors to increase their citation rates. The sample is represented by 662 publications in six Russian psychology journals, each indexed in the RSCI, Web of Science, and Scopus.

Abstract

Russian Science Citation Index, Russian publications in psychology, citation rate, publication visibility, author-based citation, author reputation, journal credibility.

Keywords

Bazenova E. (2009) Nauchny tekst v diskursivno-stilisticheskom aspekte [Discourse-Stylistic Approach to the Research of Scientific Texts]. *Vestnik*

References

- Permskogo universiteta. Rossiyskaya i zarubezhnaya filologiya*, no 5, pp. 24–32.
- Bornmann L., Haunschild R. (2018) Do Altmetrics Correlate with the Quality of Papers? A Large-Scale Empirical Study Based on F1000Prime Data. *PLoS One*, vol. 13, no 5, e0197133. doi:10.1371/journal.pone.0197133.
- Borsuk R. M., Budden A. E., Leimu R., Aarssen L. W., Lortie C. J. (2009) The Influence of Author Gender, National Language and Number of Authors on Citation Rate in Ecology. *The Open Ecology Journal*, vol. 2, no 1, pp. 25–28.
- Cherniavskaia V. (2016) Nauchno-issledovatel'skaya statya kak verbalizatsiya novogo nauchnogo rezultata (na materiale lingvistiki) [Scientific Result in Research Paper in Linguistics]. *Zhurnal rechi/Speech Genres*, no 1 (13), pp. 56–64. doi:10.18500/2311-0740-2016-1-13-56-64.
- Danilevskaya N. (2006) *Cheredovanie starogo i novogo znaniya kak mekhanizm razvertyvaniya nauchnogo teksta (Aksiologicheskii aspekt)* [Alternation of Old and New Knowledge as a Method of Scientific Text Elaboration (Axiological Aspect)] (PhD thesis abstract). Yekaterinburg: Ural State Pedagogical University. Available at: <https://dlib.rsl.ru/01003273950> (accessed 30 January 2020).
- Danilevskaya N. (2009) Nauchny tekst v aspekte teorii diskursa [Scientific text in Interdiscours Aspect]. *Yazyk. Tekst. Diskurs*, no 7, pp. 37–46.
- Didegah F., Thelwall M. (2013) Determinants of Research Citation Impact in Nanoscience and Nanotechnology. *Journal of the American Society for Information Science and Technology*, vol. 64, no 5, pp. 1055–1064.
- Fedorov P., Popov A. (2014) Vzaimosvyaz pokazateley tsitirovaniya rossiyskikh uchenykh [Relationship Among the Citation Indexes of Russian Scholars]. *Herald of the Russian Academy of Sciences*, vol. 84, no 3, pp. 222–232.
- García J. A., Rodríguez-Sánchez R., Fdez-Valdivia J. (2019) Do the Best Papers Have the Highest Probability of Being Cited? *Scientometrics*, vol. 118, no 3, pp. 885–890.
- Giere R. N. (2002) Scientific Cognition as Distributed Cognition. *The Cognitive Basis of Science* (eds P. Carruthers, St. Stich, M. Siegal), Cambridge: Cambridge University, pp. 285–299.
- Grishakina E. G. (2016) Publikatsionnaya aktivnost rossiyskikh issledovatelei: universitetskaya nauka [Publication Activity of Russian Researchers: Academic Science]. *Nauka. Innovatsii. Obrazovanie/Science. Innovation. Education*, no 4 (22), pp. 137–151.
- Haustein S, Costas R, Larivière V (2015) Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLoS ONE*, vol. 10, no 3, e0120495. doi:10.1371/journal.pone.0120495.
- Kruchkov D., Artamonova G. (2016) Statisticheskii analiz publikatsionnoy aktivnosti nauchno-issledovatel'skikh meditsinskikh uchrezhdeniy (po baze dannykh RISC) [Publishing Activity Statistics of Medical Research Institutes (Based on the Russian Science Citation Index Database)]. *Sotsialnye aspekty zdorovya naseleniya/Social Aspects of Population Health*, vol. 48, no 2, p. 9.
- Kabakova E. A. (2014) Ispolzovanie naukometricheskikh pokazatelei pri otsenke nauchnoi deyatelnosti [The Use of Scientometric Indicators in Evaluation of Scientific Activity]. *Voprosy territorial'nogo razvitiya/Territorial Development Issues*, no 8 (18). Available at: <http://vtr.isert-ran.ru/article/1437> (accessed 30 January 2020).
- Levin V. (2017) Matematicheskie metody otsenki vklada uchenogo v nauku [Mathematical Theory of Estimation of Contribution of Scientist to Science].

- Sistemy upravleniya, svyazi i bezopasnosti/Systems of Control, Communication and Security*, no 1, pp. 160–174.
- Mishankina N. (2010) *Lingvokognitivnoe modelirovanie nauchnogo diskursa* [Linguo-Cognitive Modelling of Scientific Discourse] (PhD thesis abstract). Tomsk: Tomsk State University. Available at: <https://dlib.rsl.ru/viewer/01004611456#?page=1> (accessed 30 January 2020).
- Shakirova Yu. (2016) Osobennosti vliyaniya naukometricheskikh pokazateley na dinamiku rejtingovykh pozitsiy vuzov [How Scientometric Indices Influence on the Dynamics of University Ranking Position]. *Vysshee obrazovanie v Rossii/Higher Education in Russia*, no 3, pp. 161–165.
- Sugimoto C., Haustein S., Ke Q., Larivière V. (2016) *Social Media Metrics as Indicators of Broader Impact*. Paper presented at OECD Blue Sky III Forum on Science and Innovation Indicators (Ghent, Belgium, 19–21 September 2016). Available at: <https://www.oecd.org/sti/172%20-%20SugimotoOEC-Daltmetrics.pdf> (accessed 30 January 2020).
- Valgina N. (2003) *Teoriya teksta: Uchebnoe posobie* [Theory of the Text: Study Guide]. Moscow: Logos.
- Vorobev N., Misheneva Yu. (2018) Publikatsionnaya aktivnost i otsenka kachestva raboty prepodavatelei vysshei shkoly [Publication Activity and Quality Assessment of Higher Education Teachers]. *Pedagogika*, no 2, pp. 121–125.
- Youtie J. (2014) The Use of Citation Speed to Understand the Effects of a Multi-Institutional Science Center. *Scientometrics*, vol. 100, no 3, P. 613–621.
- Yu T., Yu G., Li P. Y., Wang L. (2014) Citation Impact Prediction for Scientific Papers Using Stepwise Regression Analysis. *Scientometrics*, vol. 101, no 2, pp. 1233–1252.