Scientometrics (2021) 126:6891–6915
https://doi.org/10.1007/s11192-021-04049-z

Introducing a novelty indicator for scientific research: validating the knowledge-based combinatorial approach

Kuniko Matsumoto[1] · Sotaro Shibayama[2] · Byeongwoo Kang[3] · Masatsura Igami[1]

Received: 18 November 2020 / Accepted: 17 May 2021 / Published online: 23 June 2021

Abstract

Citation counts have long been considered as the primary bibliographic indicator for evaluating the quality of research—a practice premised on the assumption that citation count is reflective of the impact of a scientific publication. However, identifying several limitations in the use of citation counts alone, scholars have advanced the need for multifaceted quality evaluation methods. In this study, we apply a novelty indicator to quantify the degree of citation similarity between a focal paper and a pre-existing same-domain paper from various fields in the natural sciences by proposing a new way of identifying papers that fall into the same domain of focal papers using bibliometric data only. We also conduct a validation analysis, using Japanese survey data, to confirm its usefulness. Employing ordered logit and ordinary least squares regression models, this study tests the consistency between the novelty scores of 1871 Japanese papers published in the natural sciences between 2001 and 2006 and researchers' subjective judgments of their novelty. The results show statistically positive correlations between novelty scores and researchers' assessment of research types reflecting aspects of novelty in various natural science fields. As such, this study demonstrates that the proposed novelty indicator is a suitable means of identifying the novelty of various types of natural scientific research.

Keywords Bibliometrics · Novelty · Reference combination · Validation

Introduction

Citation counts have long been considered as a primary bibliographic indicator for evaluating the quality of research—a practice premised on the assumption that citation count is reflective of impact. However, scholars have identified several limitations in the use of citation counts alone, advancing the need for more multifaceted evaluation methods (Baird & Oppenheim, 1994; Bornmann et al., 2012; MacRoberts & MacRoberts,

✉ Kuniko Matsumoto
k-matsumoto@nistep.go.jp

[1] Research Unit for Science and Technology Analysis and Indicators, National Institute of Science and Technology Policy (NISTEP), Tokyo, Japan

[2] School of Economics and Management, Lund University, Lund, Sweden

[3] Institute of Innovation Research, Hitotsubashi University, Tokyo, Japan

1996; Nieminen et al., 2006). Multifaceted evaluations are particularly necessary insofar as the proper evaluation of the novelty of research is important to both researchers and science and technology policymakers, especially in the context of fostering disruptive science and technology.

While it is important to develop indicators for evaluations, such indicators should also be assessed. The introduction of new indicators in research evaluation may affect the behavior of researchers (Hicks et al., 2015), and it is therefore crucial to verify that the proposed indicators actually measure what they claim before their implementation, especially in the context of policymaking. This is similarly necessary when new indicators are introduced to scientometrics (Bornmann et al., 2019). Scholars have proposed and validated a range of new indicators for assessing the novelty of research (e.g., Bornmann et al., 2019; Tahamtan & Bornmann, 2018). Essentially, if indicators yield scores that largely agree with expert judgments, they can be considered as useful proxies for those judgments (Thelwall, 2017).

Among the various approaches proposed, the combinatorial novelty approach focuses on new combinations of knowledge sources (Uzzi et al., 2013; Wang et al., 2018). Assuming that novelty develops from unusual combinations of pre-existing knowledge (Kaplan & Vakili, 2015; Mednick, 1962; Simonton, 2003), combinatorial novelty indicators are usually measured through units comprising paired reference papers, journals, or keywords (e.g., Lee et al., 2015; Wang et al., 2017, 2018; Dahlin & Behrens, 2005; Trapido, 2015; Uddin & Khan, 2015). Units comprising pairs of reference papers have an advantage in measuring recombination novelty over other units. The use of paired reference papers is advantageous insofar as it discerns more elaborate and unusual combinations of existing knowledge than other approaches, even though the measure is more computationally intensive. However, despite its potential to measure the elaborate novelty of research, studies of novelty indicators using paired reference papers as a measure remain scarce. To the best of our knowledge, only a few papers can be found in this category, of which Dahlin and Behrens (2005) was one of the first to present the indicator. They proposed a novelty indicator using paired reference papers to measure the novelty of patents, which was then adopted by Trapido (2015) to measure the novelty of scientific publications. Nonetheless, even these novelty indicators using paired reference papers have not been widely used or discussed in previous studies focused on novelty, especially about scientific publications.

Addressing this gap, this study improves the novelty indicator proposed by Dahlin and Behrens (2005) to allow a broader range of natural sciences to be applied and verifies its utility using survey data. The indicators proposed by Dahlin and Behrens (2005) define recombination novelty as the degree of citation dissimilarity between a focal patent and prior arts in the same domain. Dahlin and Behrens' (2005) approach has some difficulties in measuring the novelty of scientific papers in various fields, including how to define the same domain of a focal paper and the data availability used to define the same domain. This study thus proposes a new way of identifying papers that fall into the same domain of focal papers using bibliometric data only. This new approach has two advantages it reduces the limitations related to data availability and it enables the measurement of the novelty of articles in any field. Using survey data from Japanese research conducted in various fields of the natural sciences, this study also demonstrates the utility of this approach through a validation analysis on whether novelty indicator scores are consistent with researchers' subjective judgments of novelty. In doing so, this study verifies whether the indicator measures researchers' subjective novelty, what kind of novelty it measures, and what field-specific features it possesses. As such, this study demonstrates the value of a new indicator able to serve as a proxy for researchers' subjective judgments of novelty and applicable to any field.

The rest of this paper is organized as follows. "Literature review" reviews the relevant literature on novelty indicators; "Proposed measure of novelty" section describes the novelty indicator analyzed in this study; "Data collection and validation methods" section describes the data collection process and the method used to verify the new indicator; "Results and discussion" section presents the results of the validation analysis; and "Conclusion" section discusses the implications of these results and suggests avenues of future research.

Literature review

Scholars in various disciplines have defined novelty as the unprecedented recombination of pre-existing knowledge components (Nelson & Winter, 1982; Romer, 1994; Schumpeter, 1939; Wang et al., 2017). Studies on combinatorial novelty have made steady progress, especially in respect to technological invention. Several studies have demonstrated that new combinations of existing knowledge present a potential watershed of new and valuable discoveries (Fleming, 2001; Uzzi et al., 2013; Verhoeven et al., 2016).

Previous studies have proposed indicators focusing on new combinations of knowledge sources to assess the novelty of research. Combinations of knowledge are usually measured with units comprising pairs of reference papers, journals, or keywords. In this respect, a reference paper pair constitutes a more detailed information unit than that of keywords or journals. Therefore, using paired reference papers has the advantage of measuring more elaborate and unusual combinations of existing knowledge than approaches using paired journals or keywords. Several scholars have explored this approach. Dahlin and Behrens (2005) introduced an indicator to quantify how unusual combinations of references (i.e., combinations of knowledge sources) in a focal publication overlap with pre-existing combinations in the knowledge domain. The authors used this indicator to measure technological radicalness in patent data on tennis rackets. Adopting this measure, Trapido (2015) tested whether authors' past recognition for highly novel work results in positive audience evaluations of their new work. Meanwhile, Uzzi et al. (2013) examined the atypicality of referenced journal pairs in publications, showing that a paper rated highly for both novelty and conventionality is more likely to be highly cited. Uddin et al. (2015) used a usual and unusual keyword combination method to explore new domains of knowledge and multidisciplinary domains. Nonetheless, despite its potential to measure novelty, research on such indicators remains relatively limited, especially related to the approach using paired reference papers. As such, the development of an indicator generally applicable to a variety of disciplines will facilitate the accumulation of knowledge on novelty indicators measured with pairs of reference papers.

Moreover, while various novelty indicators have been proposed and used for empirical analysis, it is vital to check whether such indicators truly measure what they claim to measure, particularly as the use of new indicators to evaluate research may influence the behavior of researchers (Hicks et al., 2015). Several recent empirical studies have validated novelty indicators. Based on interviews with the authors of breakthrough papers in scientometrics, Tahamtan and Bornmann (2018) found that creative ideas are not necessarily inspired by past publications. Meanwhile, using peer-review data from biomedical studies, Bornmann et al. (2019) validated the novelty indicators proposed by Uzzi et al. (2013) and Wang et al. (2017), indicating that those of Uzzi et al. (2013) reflect peer-review assessments, while those of Wang et al. (2017) do not. Essentially, there is a lack of research verifying novelty indicators based on reference paper pairs. It is also necessary to verify the validity of the reference paper approach, as scholars have done in respect to other journal combination approaches.

Proposed measure of novelty

To measure the novelty of individual scientific papers, this study adopts a novelty indicator based on the combination-based novelty measure proposed by Dahlin and Behrens (2005). To assess the novelty of patents, Dahlin and Behrens (2005) proposed quantifying the degree of citation similarity between a focal patent and prior arts in the same technological domain to capture unusual knowledge recombination. The authors referred to the degree of citation similarity as the overlap score (OS). The overlap score, OS_{ij}, of two patents, *i* and *j*, is computed as the number of patents cited by both *i* and *j* divided by the sum of the unique patents cited by *i* or *j* (Dahlin & Behrens, 2005). The authors adopted the International Patent Classification (IPC) categories at the three-digit level to define the same domain of patents. Adopting the indicator to measure the novelty of scientific publications in electrical engineering, Trapido (2015) used bibliometric and non-bibliometric information to define this domain.

Two methodological considerations regarding the calculation of novelty indicators need to be taken into account when applying Dahlin and Behrens' (2005) approach to a broader range of natural science papers. First, researchers must consider which data are applicable for defining the domains of natural science in a consistent manner (i.e., data selection). Second, researchers must define the domains used to compare citation similarity because the fields of scientific papers are not classified into detailed units like those provided for under the IPC (i.e., definition of domains).

In respect to the first methodological concern, this study introduces a means of defining the same domain using only bibliometric information. Bibliometric data refer to that gained through a common data source such as Clarivate Analytics' Web of Science (WoS) and Elsevier's Scopus covering a wide range of countries, fields, and periods. By contrast, non-bibliometric data (e.g., a researcher's curriculum vitae) are not located in a large or common database equivalent to WoS or Scopus, making them more difficult to collect. Accordingly, non-bibliometric data are ill-suited to measuring the novelty of papers across multiple countries, fields, and periods. Therefore, bibliometric data are preferable for calculating the novelty indicator in a country, field, and period without restrictions on the acquisition of non-bibliographic information.

In regard to the definition of domains, this study defines papers in the same domain in accordance with the following two conditions. Same-domain papers are those that (i) co-cite at least one of the references of the focal paper and those whose (ii) field classification completely matches that of the focal paper in terms of the smallest unit in the bibliometric database. This study adopts the subject categories used by WoS, from which the focal papers used in the validation were extracted. All records in the WoS core collection are assigned one or more of the 254 subject categories (Clarivate Analytics, 2020). If the focal paper cites a paper commonly cited in various research fields, the paper that is less relevant to the focal paper is likely to be included in the same domain by the first condition. By applying the second condition, papers less relevant to the focal paper are removed from the same domain. Figure 1 illustrates the relationship between the focal paper and same-domain paper.

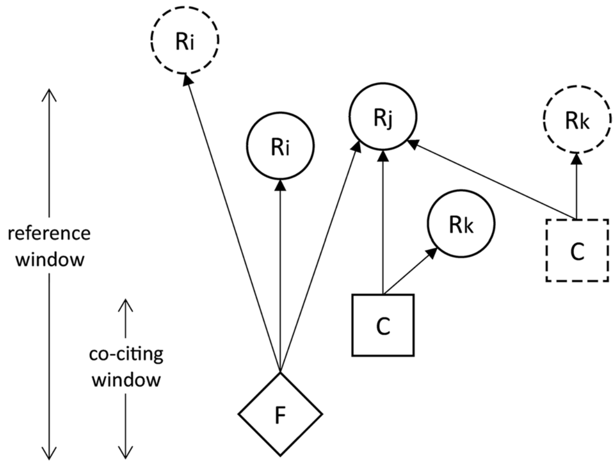

Fig. 1 Simplified illustration of the relation between the focal paper and same-domain paper. Figure 1 presents three citation networks comprising a focal paper (diamond), reference papers (circles), and co-citing reference papers matching the field classification of the focal paper (rectangles). References may be cited by the focal paper only (i), by both the focal paper and its co-citing reference paper (j), or by its co-citing reference paper only (k). The dotted-line circles and rectangles are out of the citation window and thus not considered in the novelty measurement. The overlap score (OS) is defined as the set of papers cited by both the focal paper and its co-citing references (N[Rj]) divided by the set cited (N[Ri] + N[Rj] + N[Rk])

In this study, the overlap score, OS_{ij}, between focal paper *i* and same-domain paper *j* is defined as the set of papers cited by both *i* and *j* ($[\text{Ref}]_i \cap [\text{Ref}]_j$) divided by the set that either *i* or *j* cites ($[[\text{Ref}]_i \cup [\text{Ref}]_j$):

$$\text{OS}_{ij} = \frac{[\text{Ref}]_i \cap [\text{Ref}]_j}{[\text{Ref}]_i \cup [\text{Ref}]_j}$$

The novelty score of a focal paper, *i*, is calculated by subtracting the mean overlap score for the same-domain papers from 1 as follows:

$$\text{Novelty}(i) = 1 - \frac{1}{n}\sum\nolimits_{j=1}^{n}\text{OS}_{ij}$$

The resulting score has a value ranging between 0 and 1, wherein 0 indicates completely identical citation patterns with the same-domain papers and 1 indicates completely dissimilar citation patterns. More dissimilar citation patterns reflect novelty, which is defined as the recombination of pre-existing knowledge components in an unprecedented manner (Nelson & Winter, 1982; Romer, 1994; Schumpeter, 1939; Wang et al., 2017).

To measure novelty, two citation windows must be set: the reference window, which comprises the range of publication years of the papers cited by the focal paper, and the co-citing window, which comprises the range of publication years of the same-domain papers. This prompts the question of how long the optimal period for the two citation windows should be. First, regarding the reference window, citation context studies that aim to characterize cited works have reported that highly cited older papers are likely to be cited to provide historical rationales or context (Ahmed et al., 2004; Oppenheim & Renn, 1978). Therefore, the longer the reference window, the greater the number of old reference papers with less relevance to the focal paper's core topic, increasing the possibility that papers less relevant to the focal paper will be included in the same domain. Therefore, this study suggests that shorter reference windows are preferable. Second, the longer the co-citing window, the greater the number of cited papers, including older and less relevant publications. From the perspective of the life cycle of the research topic, the more the publication years deviate, the greater the possibility of including papers that are less relevant to the focal paper in the same domain. Therefore, the co-citing window should be short rather than long.

This study seeks to verify the novelty measure using four patterns of citation windows, namely, two types of reference windows and two types of co-citing windows. More specifically, the reference window types consist of all years observed before the focal paper's year of publication and the 10 years before the focal paper's publication. The co-citing window types comprise the same period for the reference window and the three years before the focal paper's publication.

Data collection and validation methods

Survey data

This study uses survey data of Japanese researchers to validate the novelty indicator. The survey was jointly conducted by the Institute of Innovation Research, Hitotsubashi University and the National Institute of Science and Technology Policy between late 2009 and early 2010 (Nagaoka et al., 2010). The survey data have been utilized for various studies focusing on the research knowledge creation process (Igami et al., 2015; Lee et al., 2015; Murayama et al., 2015; Walsh & Lee, 2015).

The details of the survey samples are as follows (Nagaoka et al., 2010). The population of the survey was articles and letters in the WoS database (Science Citation Index Expanded, SCIE) of Clarivate Analytics, which primarily contains research papers related to the natural sciences. Reviews were excluded from the population. The time window of the papers for the survey was from 2001 to 2006. The bibliographic information and number of citations as of the end of December 2006 were used to identify possible focal papers. Two sets of possible focal papers were selected from the population:

Highly cited papers (approximately 3000 in each survey): Top 1% of highly cited papers in each journal field[1] and in each year; at least one organization of authors should be located in Japan. All highly cited papers in the time window were selected.

Normal papers (approximately 7000 in each survey): Randomly selected papers in each journal field and in each year from the population of the survey, excluding the above highly cited papers; at least one organization of authors should be located in Japan.

The corresponding authors or equivalents of approximately 10,000 possible focal papers were investigated and identified as survey targets. If multiple papers were assigned to a single author, one paper was randomly selected as the focal paper, while priority was given to highly cited papers in the selection process.

As a result, 7652 survey targets were identified for the survey. Of those, there are 1932 scientists whose focal paper is the highly cited paper and 5720 scientists whose focal paper

[1] The journal field refers the 22 scientific fields in the Essential Science Indicators (ESI) of Thomson Reuters.

is the normal paper. Of the 7562 survey targets, 2081 responses were received, making a response rate of 27%. The response rate is 29.3% for highly cited papers and 26.5% for normal papers.

The survey included a self-assessment question on the focal paper's types of research outputs. Respondents were asked to rate the relevance of types of research outputs on a five-point scale ranging from 1 = "Not relevant" to 5 = "Highly relevant."

This study adopted questions on the types of research relevant to theory, phenomenon, method, and material. Among the research types in Table 1, "new_theory" and "valid_theory" (research types related to theory), "new_phenom" and "under_phenom" (research types related to phenomenon), "new_meth" and "imprv_meth" (research types related to research method), and "new_mat" and "imprv_mat" (research types related to function, mechanism, or material) were paired. Research types with shortened research-type names that contain the word "new" reflect aspects of novelty in research. By comparing the novelty indicator in pairs of research types, this study verified whether the indicator is consistent with the researchers' subjective judgments.

Bibliometric data for the measurement of novelty

This study retrieved the data used in the novelty score computation from Clarivate Analytics' WoS XML format. More specifically, it accessed the following files: Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), Arts and Humanities Citation Index (AHCI), Conference Proceedings Citation Index: Science (CPCI-S), and Conference Proceedings Citation Index: Social Sciences and Humanities (CPCI-SSH). Data were extracted from material published between 1981 and the end of 2018.

This study adopted eight research fields aggregated from 19 ESI journal fields (all the fields except for Economics and Business, Social Science, and General): (1) Chemistry, (2) Materials Science, (3) Physics and Space Science, (4) Computer Science and Mathematics, (5) Engineering, (6) Environment/Ecology and Geosciences, (7) Clinical Medicine and Psychiatry/Psychology, and (8) Basic Life Sciences (Agricultural Sciences, Biology and Biochemistry, Immunology, Microbiology, Neuroscience and Behavior, Pharmacology and Toxicology, and Plant and Animal Science). The journals classified in the "multidisciplinary" field in the 22 ESI fields include a wide range of research results from life sciences to physical sciences. Many focal papers identified in the multidisciplinary were included in journals such as *PNAS* (*Proceedings of the National Academy of Sciences*), *Nature*, and *Science*. Therefore, the focal papers in the multidisciplinary field were reclassified into a scientific field based on the references of the paper[2].

In the validation analysis, this study used the data of papers for which a valid answer was obtained for the self-assessment question, while excluding those for which novelty scores could not be calculated and outlier novelty scores (i.e., papers for which the novelty scores are zero)[3]. A total of 1871 field observations were recorded

[2] The reclassification procedures of multidisciplinary field papers were as follows: (i) collecting the references of a focal paper in the multidisciplinary field; (ii) identifying the scientific field of each reference, where a field was identified based on the scientific fields of a journal; (iii) finding the most frequent scientific field in the references of the focal paper, except for multidisciplinary fields; and (iv) using the most frequent scientific field as the scientific field of the focal paper.

[3] These correspond to focal papers without reference papers or having no same-domain papers. For these focal papers, the novelty scores are not calculable or become zero (the latter case is rare in our study; there are only two observations).

Table 1 Research types

	Types reflecting novelty	Types reflecting the validation or improvement of existing knowledge
Theory	new_theory Developing a new hypothesis or theory	valid_theory Supporting or rejecting an existing hypothesis or theory
Phenomenon	new_phenom Discovering an unknown phenomenon or material	under_phenom Understanding of a phenomenon
Method	new_meth Developing a new research method	imprv_meth Improving an existing research method
Material et al	new_mat Creating a new function, mechanism or material	imprv_mat Improving an existing function, mechanism or material

Table 2 Descriptive statistics of novelty across all the fields

Citation window Reference window Co-citing window	Window 1 All 3 years	Window 2 All All	Window 3 10 years 3 years	Window 4 10 years All
Mean	0.948	0.950	0.933	0.933
SD	0.051	0.048	0.065	0.064
Min	0.471	0.500	0.333	0.333
Max	0.996	0.996	0.998	0.995
Obs	1871	1871	1871	1871

as follows: Chemistry, 262; Materials Science, 142; Physics and Space Science, 365; Computer Science and Mathematics, 56; Engineering, 100; Environment/Ecology and Geosciences, 102; Clinical Medicine and Psychiatry/Psychology, 217; Basic Life Sciences, 593; and Unknown, 34. This study only conducted a by-field validation analysis in fields with more than 150 papers to ensure more robust results. The following five fields were thus subject to the by-field analysis: Chemistry, Materials Science, Physics and Space Science, Clinical Medicine and Psychiatry/Psychology, and Basic Life Sciences.

Validation methods

Using Japanese self-assessment data, this study tested the convergent validity of a novelty indicator based on unusual knowledge combinations from two perspectives. First, this study verified the consistency between the novelty score and self-assessments of the research types by eliminating the influence of field-specific features using the data of all the fields. Second, this study analyzed each field to clarify the field-specific features.

As researchers' self-assessments of the research types were ordinal, this study adopted ordered logit models to verify the validity of the novelty indicator. This study also adopted the ordinary least squares (OLS) regression model—a representative linear regression model (Zdaniuk, 2014)—to test the robustness of the results. It then used the novelty scores measured by the four citation windows, namely, two reference windows (1981 to the focal paper's publication year and the 10 years before the focal paper's publication) and two co-citing windows (the same period for the reference window and the three years before the focal paper's publication). The regression models and results are detailed in the next section.

Table 3 Descriptive statistics of novelty in the individual fields

Citation window Reference window: 10 years Co-citing window: 3 years	Chemistry	Materials science	Physics and space science	Clinical medicine and psychiatry/psychology	Basic life sciences
Mean	0.942	0.924	0.926	0.940	0.953
SD	0.045	0.051	0.068	0.064	0.049
Min	0.717	0.725	0.389	0.400	0.333
Max	0.991	0.998	0.982	0.988	0.995
Obs	262	142	365	217	593

Results and discussion

Descriptive statistics

Novelty scores

Table 2 presents the summary statistics of the novelty scores of all the fields for each of the four citation windows[4]. As expected, the results indicate that a longer reference window leads to a higher novelty score. More specifically, a longer reference window increases the likelihood that papers less relevant to the focal paper will be included in the same domain, resulting in focal papers having a higher novelty score (i.e., a lower overlap score). Table 3 presents the summary statistics of the novelty scores of each field for the 10-year reference window and three-year co-citation window. According to the results, the focal papers in Basic Life Sciences (0.953) have the highest mean score, while those in Materials Science (0.924) and Physics and Space Science (0.926) have the lowest scores[5].

Relevant degrees of research type by researchers' self-assessment

Table 4 presents the summary statistics of the degree of relevance for each research type. According to the results, respondents give higher scores to "under_phenom" (3.718) and "new_theory" (3.502) and lower scores to "imprv_mat" (2.370) and "new_mat" (2.795). As Table 5 shows, the results reveal the similarities and differences between the fields in respect to the degree of the relevance of each research type. Regarding the similarities by field, "under_phenom" receives higher scores, while "imprv_mat" receives lower scores. Differences by field are shown for "new_mat" and "new_theory." In the fields of Chemistry

[4] As shown in Tables 2 and 3, our novelty scores are close to 1 and their variances are small. Previous research indicators (i.e., those used by Dahlin and Behrens (2005) and Trapido (2015)), which are the basis of our indicators, also have similar features. The small variation in the scores may make it difficult to interpret whether novelty is high or low, especially for the practical use of the indicators. On this point, applying methods such as standardization would help interpret the indicators. Figure 2 is one such example where we adopted percentile representation for the horizontal axis.

[5] This tendency is also confirmed in the other citation windows.

Table 4 Descriptive statistics of the degree of the relevance of research types across all the fields

Research type	new_theory	valid_theory	new_phenom	under_phenom	new_meth	imprv_meth	new_mat	imprv_mat
Mean	3.502	3.303	3.095	3.718	3.097	2.904	2.795	2.370
SD	1.279	1.259	1.451	1.198	1.338	1.247	1.518	1.272

The minimum and maximum measures and number of observations are the same for all research types. The minimum relevance is 1, while the maximum is 5. The results are based on 1871 observations

Table 5 Descriptive statistics of the degree of the relevance of research types in each field

	Field	new_theory	valid_theory	new_phenom	under_phenom	new_meth	imprv_meth	new_mat	imprv_mat
Mean	Chemistry	3.294	3.015	3.485	3.599	3.553	3.084	3.557	2.729
	Materials science	3.183	3.070	3.254	3.732	3.338	2.859	3.676	3.148
	Physics and space science	3.479	3.622	3.121	3.814	3.381	3.088	2.677	2.362
	Clinical medicine and psychiatry/psychology	3.516	3.290	2.700	3.525	2.585	2.594	2.410	2.230
	Basic life sciences	3.678	3.325	3.298	3.840	2.777	2.757	2.828	2.298
SD	Chemistry	1.257	1.184	1.333	1.186	1.230	1.254	1.426	1.259
	Materials science	1.330	1.270	1.376	1.123	1.388	1.164	1.308	1.232
	Physics and space science	1.315	1.264	1.423	1.188	1.275	1.206	1.443	1.278
	Clinical medicine and psychiatry/psychology	1.251	1.160	1.449	1.167	1.289	1.274	1.531	1.259
	Basic life sciences	1.224	1.247	1.429	1.148	1.297	1.228	1.444	1.178

The minimum and maximum are the same for all research types and fields. The minimum degree of relevance is 1, while the maximum is 5. The following numbers of observations were recorded for each field: Chemistry, 262; Materials science, 142; Physics and space science, 365; Clinical medicine and psychiatry/psychology, 217; and Basic life sciences, 593

Table 6 Ordered logit regression results of the convergent validity between researchers' self-assessment of research types and novelty indicators in the four citation window patterns across all the fields

	Independent variables = Novelty			
Citation window	Window 1	Window 2	Window 3	Window 4
Reference window	All	All	10 years	10 years
Co-citing window	3 years	All	3 years	All
Dependent variables (Research type)				
new_theory	0.078	0.096 *	0.107 **	0.118 **
	(1.70)	(2.00)	(2.67)	(2.87)
valid_theory	0.097 *	0.085	0.106 **	0.083 *
	(2.29)	(1.89)	(2.58)	(2.02)
new_phenom	0.097 **	0.11 **	0.12 ***	0.139 ***
	(2.91)	(3.14)	(3.73)	(4.22)
under_phenom	0.063	0.083	0.091 *	0.102 *
	(1.43)	(1.95)	(2.06)	(2.30)
new_meth	0.028	0.047	0.071 *	0.085 *
	(0.79)	(1.31)	(2.02)	(2.39)
imprv_meth	−0.053	−0.044	−0.037	-0.035
	(−1.39)	(−1.12)	(−0.93)	(−0.84)
new_mat	0.065	0.07	0.1 **	0.104 **
	(1.74)	(1.75)	(2.99)	(2.95)
imprv_mat	−0.037	−0.035	−0.031	−0.034
	(−0.98)	(−0.90)	(−0.87)	(−0.92)

All coefficients are standardized. Robust statistics are shown in parentheses. The models control for publication year and field with dummy variables. A total of 1871 observations were collected for all the research types and window patterns

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

and Materials Science, respondents give higher scores to "new_mat." In Physics and Space Science, Clinical Medicine and Psychiatry/Psychology, and Basic Life Sciences, respondents give higher scores to "new_theory." The means of the novelty scores by research type relevance degree and correlation coefficients between novelty score and each research type relevance degree are listed in the Appendix.

Regression results

Results of all the fields

Table 6 presents the results of the ordered logit regression models for the four citation windows. This study uses an ordered logit regression model with robust standard errors, one independent variable, and two dummy variables. More specifically, the dependent variable is the degree of the relevance of each research output type (e.g., "new_theory" and "valid_theory") as rated on a five-point scale ranging from 1: "Not relevant" to 5: "Highly relevant," while the independent variable is the novelty score. Dummy variables are used to control for publication year and field. As noted, this study categorizes eight research types into four pairs: theory ("new_theory" and "valid_theory"); phenomenon ("new_phenom" and "under_phenom"); research method ("new_meth" and "imprv_meth"); and function, mechanism, or material ("new_mat" and "imprv_mat") (Table 1). This study discusses the results in Table 4 in the context of each research type category.

First, regarding the research types related to theory, the results reveal statistically significant positive correlations between the novelty scores and survey assessment results for "new_theory" and "valid_theory" in the regression models with short reference windows (i.e., windows 3 and 4). The comparison of the coefficients of "new_theory" and "valid_theory" in the short reference window regression models reveals that "new_theory" has a higher coefficient than "valid_theory."

Second, in respect to the research types related to phenomenon, the results reveal statistically significant positive correlations between the novelty scores and survey assessment results for "new_phenom" in all the regression models. However, in the case of "under_phenom," there are only slightly significant positive correlations between the novelty scores and regression models of the short reference windows. The comparison of the coefficients of the two research types reveals that "new_phenom" has a higher coefficient than "under_phenom" in all the regression models.

Third, regarding the research types related to method, the results reveal slightly significant positive correlations between the novelty scores and "new_meth" in the short reference window regression models. Moreover, "new_meth" has a higher coefficient than "imprv_meth" in all the regression models.

Fourth, in respect to the research types related to function, mechanism, or material, the results reveal statistically significant positive correlations between the novelty scores and survey assessment results for "new_mat" in the short reference window regression models. Additionally, "new_mat" has a higher coefficient than "imprv_mat" in all the regression models.

As the results indicate, the novelty scores with the short reference windows appear to have convergent validity with the self-assessments of research types reflecting aspects of novelty in research (thus containing the word "new"). Furthermore, research types expected to reflect aspects of novelty in research tend to have higher coefficients in the regression models. As such, the indicator produces scores that strongly reflect researchers' evaluations of novelty. Meanwhile, in respect to the co-citing window, the results are similar regardless of whether the window is long or short. These results are partially confirmed by the OLS regression model[6]. The robustness of the results is confirmed by the consistency between those of the ordered logit and OLS regression models. The results thus confirm that the proposed novelty indicator is a useful proxy for researchers' evaluations of novelty. Considering the validation results and computational costs, this study argues that it is more practical if both the reference and the co-citing windows are short.

Figure 2 shows the predicted value of researchers' self-assessments for each research type against the novelty indicators in a 10 year-reference window and a three-year

[6] The ordered logit and OLS regression models use the same dependent and independent variables with robust standard errors.

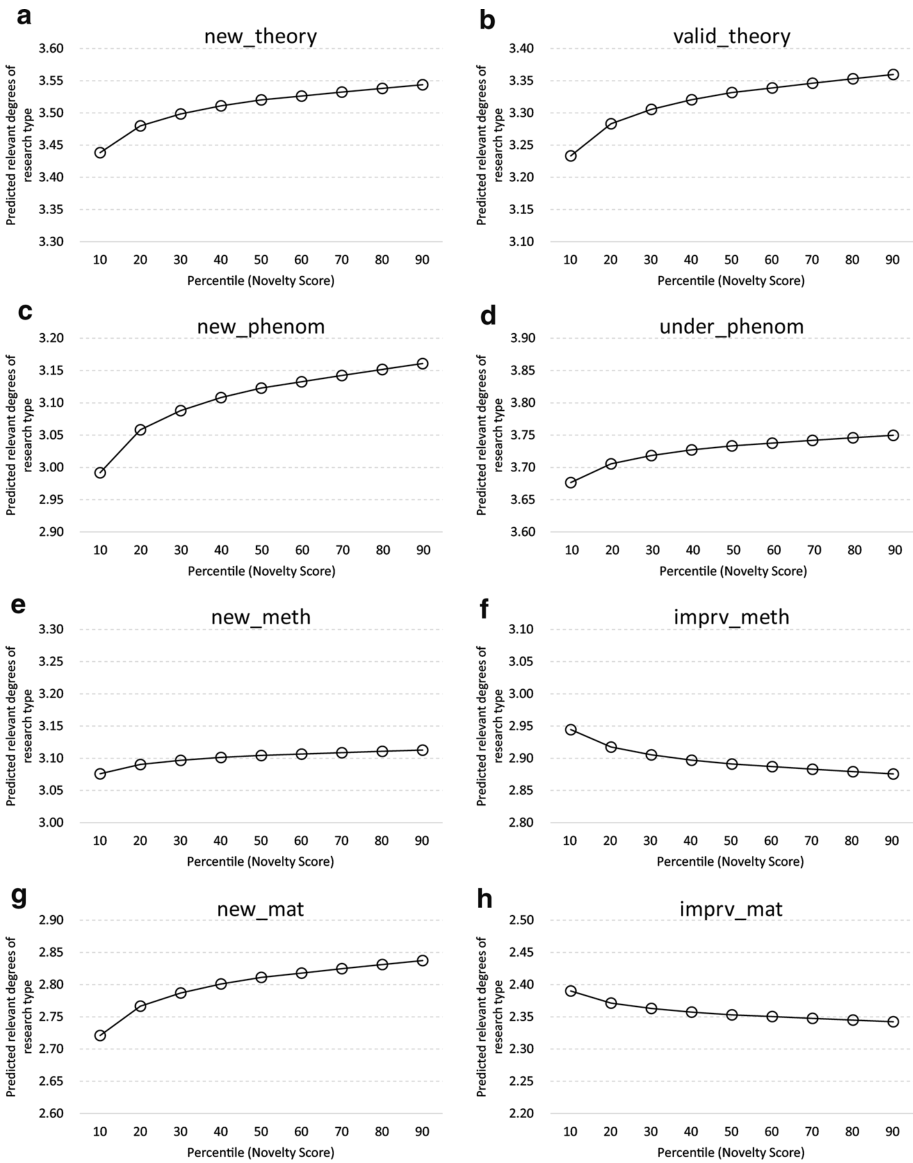

Fig. 2 Prediction of researchers' self-assessments of each research-type. Figure 2 presents the novelty scores of all the fields with a 10-year reference window and three-year co-citation window. The predicted degree of relevance for each research type comprises the expected values measured by the marginal effects of each degree's predicted probability. Note, using dummy variables, the "new_meth" and "imprv_mat" models only control for the field to avoid errors in the calculation of the marginal effects; all the other models control for publication year and field using dummy variables

co-citation window. Figure 2 reveals that with the exception of "new_meth," a higher novelty score leads to a higher degree of relevance in research types reflecting aspects of novelty. Contrary to our expectation, the degree of the relevance of "valid_theory" increases with an increase in its novelty score. This finding suggests that unusual combinations of

Table 7 Ordered logit regression results of the convergent validity between researchers' self-assessment of research types and novelty indicators in the five fields

	Independent variables = Novelty				
Citation window	Chemistry	Materials science	Physics and space science	Clinical medicine and psychiatry/psychology	Basic life sciences
Reference window: 10 years					
Co-citing window: 3 years					
Dependent variables (Research type)					
new_theory	0.136	−0.041	0.017	0.039	0.392 ***
	(1.50)	(−0.22)	(0.20)	(0.45)	(5.25)
valid_theory	0.045	0.122	0.137	−0.088	0.127 *
	(0.43)	(0.89)	(1.87)	(−1.12)	(2.13)
new_phenom	0.175	0.147	0.102	0.198	0.089 *
	(1.53)	(0.98)	(1.75)	(1.89)	(1.98)
under_phenom	−0.09	−0.214	0.07	0.063	0.135 *
	(−0.77)	(−1.49)	(0.78)	(0.65)	(2.11)
new_meth	0.114	0.014	0.041	0.017	0.051
	(1.26)	(0.10)	(0.56)	(0.22)	(0.88)
imprv_meth	−0.01	−0.485 ***	0.009	−0.047	−0.099
	(−0.13)	(-3.50)	(0.11)	(−0.54)	(−1.04)
new_mat	0.208 *	0.177	0.012	0.353 **	0.086
	(2.45)	(1.00)	(0.25)	(2.78)	(1.84)
imprv_mat	−0.051	−0.343 ***	−0.069	0.197 *	−0.042
	(−0.59)	(−3.59)	(−1.17)	(2.18)	(−0.76)
Number of papers	262	142	365	217	593

All coefficients are standardized. Robust statistics are shown in parentheses. The models control for publication year using a dummy variable. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

knowledge are required to support or reject an existing hypothesis or theory. Although the regression coefficients are not statistically significant, the degrees of the relevance of "imprv_meth" and "imprv_mat" decrease with increasing novelty scores. This suggests that research focused on "improving an existing research method" or "improving on an existing function, mechanism, or material" tends to rely on the knowledge base similar to that of publications in the same domain.

Results of each field

Table 7 presents the results of the ordered logit regression models testing for the convergent validity between researchers' self-assessments of research type and novelty indicators in the five fields. The models adopt a 10-year reference window and a three-year co-citation window. The results reveal that with the exception of "new_theory," most research types expected to reflect aspects of novelty in research have higher coefficients in all the fields than their counterparts.

Moreover, the convergent validity between the novelty scores and self-assessment of research types differs from one field to another. In this respect, Materials Science is the only field in which there is a negative correlation between research types and novelty scores. The results reveal the statistically significant negative correlations between the novelty scores and survey assessment results for "imprv_meth" and "imprv_mat." Meanwhile, in the fields of Clinical Medicine and Psychiatry/Psychology, there is a significant positive correlation between the novelty scores and survey assessment results related to function, mechanism, or material, although "new_mat" possesses a correlation significantly stronger than that of "imprv_mat." The field of Basic Life Sciences possesses the most research types that correlate with novelty scores, with "new_theory" evidencing particularly strong positive correlations. However, the results do not confirm any strong correlation between novelty score and research type in the fields of Chemistry and Physics and Space Science.

The robustness of this study's results is confirmed by the consistency between the results of both the ordered logit and the OLS regression models. This study also confirms the consistency between the results for other citation window patterns. The results thus suggest that the research types in which novelty scores have statistically significant correlations match the representative research types of each field. For example, Chemistry and Materials Science research is likely to focus on the research types related to function, mechanism, or material. As shown in Table 5, respondents give higher scores to "new_mat" in the two fields. Similarly, research on Basic Life Sciences is likely to focus on the research types related to theory. Further research is necessary to interpret the results in the fields of Physics and Space Science and Clinical Medicine and Psychiatry/Psychology. These results may be due to the dispersal of research types that are likely to be the focus.

Conclusion

Novelty is an integral aspect of outstanding research. Over the last two decades, scholars have proposed a number of indicators to identify and measure novelty in research using various types of bibliographic information and methods. However, as new indicators may affect researchers' behavior (Hicks et al., 2015), the development of new indicators must include an investigation of their validity (Thelwall, 2017).

This study improves the novelty indicator proposed by Dahlin and Behrens (2005), which quantifies the degree of citation similarity between a focal paper and a pre-existing same-domain paper, to enable application to a broader range of natural sciences. In respect to the novelty measurement method, this study introduces a means of defining the same domain using only bibliometric information, making it possible to measure combinatorial novelty using reference paper pairs in various research fields, countries, and periods providing the bibliographic data are available. This study thus advances the utility of this indicator for large-scale novelty analysis, including comparative and time-series analyses.

This study also tests the convergent validity of the novelty indicator from two points of view by using researchers' self-assessments of the degree of the relevance of each research type based on data obtained from a survey of Japanese researchers. First, this study analyzes data covering multiple fields in the natural sciencesto clarify the commonalities between the novelty indicators and researchers' self-assessments across natural science disciplines. The analysis of the natural sciences finds the convergent validity between the novelty score and self-assessments of research types reflecting aspects of novelty in research, namely, "developing a new hypothesis or theory," "discovering an unknown phenomenon or material," "developing a new research method," and "creating a new function, mechanism or material." This study finds that these research types have higher coefficients in the regression models than their counterparts (i.e., research types that reflect aspects of validation or improving existing knowledge). The robustness of this study's results is confirmed by the consistency between the results of the ordered logit and OLS regression models. Second, this study conducts validation analyses of each field to clarify the field-specific characteristics. In this respect, the results reveal that for all five fields, with the exception of research types related to new theory, all research types expected to reflect aspects of novelty in research have higher coefficients in the regression models than their counterparts. This study also observes different results for each field in respect to the convergent validity between the novelty score and self-assessments of research types.

Finally, we show three directions for future research. Firstly, this study confirms that Dahlin and Behrens' (2005) novelty indicators can be used to identify the novelty of publications in various types of research across multiple fields in the natural sciences. This study's findings also suggest that the research type suitable for identifying the novelty of papers may differ by field, indicating the need for further, comprehensive field comparison studies to expand extant knowledge on novelty indicators. Secondly, previous empirical studies on the validation of novelty indicators have faced limitations in terms of data and scope. For example, both Tahamtan and Bornmann (2018) and Bornmann et al. (2019) noted data limitations, including those related to field. This study is similarly limited, as the data used are restricted to Japan. Future research should validate proposed novelty indicators using data from a range of fields, countries, and years to enhance our understanding of the characteristics of such indicators. Lastly, our indicator measures the degree of unusual knowledge recombination as research novelty. Given the structure of the novelty indicators, several factors influence the degree of unusual knowledge recombination. For example, unusual knowledge recombination could be attributable to the field dissimilarity of references, oldness/newness of references, and regional differences in references. Field dissimilarity would be associated with the interdisciplinary of research and the latter two would be associated with the absorptive capacity of new knowledge and regionalization of knowledge. We need to carry out more research in the future to better understand the origin of novelty.

Appendix

See Tables 8, 9, 10 and 11

Table 8 Means of novelty across all the fields by the degree of the relevance of research types

Reference window	Co-citing window	The degree of relevance	Means of novelty across all the fields							
			new_theory	valid_theory	new_phenom	under_phenom	new_meth	imprv_meth	new_mat	imprv_mat
All	3 years	1	0.936	0.934	0.936	0.936	0.949	0.951	0.943	0.947
		2	0.945	0.950	0.948	0.940	0.953	0.950	0.953	0.954
		3	0.950	0.948	0.949	0.950	0.948	0.950	0.945	0.949
		4	0.950	0.950	0.953	0.949	0.946	0.945	0.956	0.943
		5	0.950	0.952	0.955	0.951	0.944	0.939	0.948	0.945
All	All	1	0.938	0.937	0.938	0.938	0.951	0.953	0.945	0.949
		2	0.948	0.953	0.949	0.943	0.955	0.951	0.956	0.956
		3	0.951	0.951	0.952	0.952	0.951	0.953	0.948	0.952
		4	0.953	0.952	0.954	0.951	0.949	0.949	0.958	0.946
		5	0.952	0.953	0.957	0.954	0.947	0.941	0.950	0.947
10 years	3 years	1	0.917	0.916	0.915	0.914	0.931	0.936	0.924	0.930
		2	0.929	0.935	0.932	0.920	0.938	0.936	0.935	0.942
		3	0.935	0.933	0.934	0.937	0.934	0.936	0.930	0.936
		4	0.936	0.936	0.941	0.934	0.932	0.931	0.945	0.927
		5	0.937	0.939	0.943	0.938	0.931	0.922	0.938	0.932
10 years	All	1	0.916	0.918	0.914	0.912	0.930	0.935	0.924	0.930
		2	0.930	0.936	0.932	0.922	0.939	0.936	0.936	0.942
		3	0.933	0.932	0.934	0.936	0.933	0.936	0.930	0.935
		4	0.937	0.935	0.941	0.934	0.932	0.931	0.944	0.927
		5	0.937	0.938	0.944	0.938	0.932	0.922	0.938	0.932
Obs		1	181	223	410	136	301	340	601	667
		2	239	257	252	167	342	341	238	369
		3	424	480	367	363	450	539	319	413
		4	514	553	435	627	431	460	369	319
		5	513	358	407	578	347	191	344	103

Table 9 Correlation coefficients between novelty across all the fields and the degree of the relevance of research types

Reference window	Co-citing window	Correlation coefficients							
		new_theory	valid_theory	new_phenom	under_phenom	new_meth	imprv_meth	new_mat	imprv_mat
All	3 years	0.064	0.080	0.127	0.072	−0.048	−0.059	0.056	−0.023
All	All	0.074	0.074	0.136	0.084	−0.042	−0.056	0.059	−0.020
10 years	3 years	0.079	0.085	0.147	0.092	−0.013	−0.050	0.097	−0.009
10 years	All	0.083	0.072	0.161	0.098	−0.009	−0.049	0.099	−0.009
Obs		1871	1871	1871	1871	1871	1871	1871	1871

Table 10 Means of novelty in each field by the degree of the relevance of research types

Field	The degree of relevance	Means of novelty in each field															
		new_theory		valid_theory		new_phenom		under_phenom		new_meth		imprv_meth		new_mat		imprv_mat	
Chemistry	1	0.939	(28)	0.937	(38)	0.925	(32)	0.934	(19)	0.931	(20)	0.944	(37)	0.932	(39)	0.948	(59)
	2	0.933	(42)	0.942	(42)	0.943	(29)	0.947	(29)	0.941	(31)	0.950	(44)	0.937	(23)	0.935	(54)
	3	0.945	(70)	0.944	(85)	0.941	(55)	0.947	(56)	0.940	(69)	0.933	(81)	0.931	(46)	0.945	(69)
	4	0.939	(69)	0.941	(72)	0.947	(72)	0.942	(92)	0.942	(68)	0.941	(60)	0.948	(61)	0.936	(59)
	5	0.949	(53)	0.942	(25)	0.944	(74)	0.936	(66)	0.946	(74)	0.948	(40)	0.948	(93)	0.944	(21)
Materials science	1	0.911	(21)	0.917	(24)	0.910	(21)	0.950	(8)	0.919	(19)	0.940	(22)	0.906	(14)	0.938	(16)
	2	0.927	(24)	0.918	(20)	0.923	(23)	0.947	(12)	0.928	(23)	0.931	(29)	0.899	(14)	0.945	(28)
	3	0.939	(32)	0.930	(37)	0.924	(31)	0.920	(29)	0.923	(31)	0.933	(50)	0.929	(25)	0.923	(39)
	4	0.925	(38)	0.921	(44)	0.936	(33)	0.924	(54)	0.931	(29)	0.903	(29)	0.938	(40)	0.909	(37)
	5	0.913	(27)	0.938	(17)	0.922	(34)	0.916	(39)	0.921	(40)	0.890	(12)	0.923	(49)	0.915	(22)
Physics and space science	1	0.911	(39)	0.913	(33)	0.915	(73)	0.906	(23)	0.917	(39)	0.925	(50)	0.930	(120)	0.931	(134)
	2	0.932	(47)	0.909	(39)	0.926	(53)	0.921	(33)	0.931	(54)	0.919	(55)	0.927	(48)	0.931	(67)
	3	0.929	(84)	0.922	(71)	0.930	(73)	0.931	(60)	0.924	(83)	0.928	(118)	0.904	(76)	0.916	(81)
	4	0.932	(90)	0.936	(112)	0.926	(89)	0.928	(122)	0.930	(107)	0.938	(97)	0.941	(72)	0.924	(64)
	5	0.923	(105)	0.930	(110)	0.935	(77)	0.927	(127)	0.927	(82)	0.908	(45)	0.929	(49)	0.932	(19)
Clinical medicine and psychiatry/psychology	1	0.948	(20)	0.951	(21)	0.926	(68)	0.941	(20)	0.938	(58)	0.943	(63)	0.927	(100)	0.931	(94)
	2	0.930	(28)	0.946	(31)	0.939	(35)	0.887	(18)	0.947	(50)	0.931	(36)	0.939	(25)	0.939	(28)
	3	0.940	(43)	0.940	(58)	0.952	(38)	0.949	(49)	0.927	(52)	0.947	(56)	0.948	(25)	0.946	(55)
	4	0.937	(72)	0.935	(78)	0.947	(46)	0.945	(88)	0.950	(38)	0.938	(50)	0.953	(37)	0.953	(31)
	5	0.945	(54)	0.938	(29)	0.946	(30)	0.939	(42)	0.940	(19)	0.919	(12)	0.960	(30)	0.952	(9)

Table 10 (continued)

Field	The degree of relevance	Means of novelty in each field															
		new_theory		valid_theory		new_phenom		under_phe-nom		new_meth		imprv_meth		new_mat		imprv_mat	
Basic life sciences	1	0.913	(44)	0.937	(62)	0.947	(103)	0.939	(34)	0.948	(125)	0.955	(120)	0.948	(161)	0.952	(196)
	2	0.945	(59)	0.959	(89)	0.951	(77)	0.952	(38)	0.956	(134)	0.955	(131)	0.950	(99)	0.958	(154)
	3	0.954	(129)	0.949	(156)	0.946	(109)	0.951	(127)	0.952	(152)	0.953	(160)	0.954	(109)	0.949	(136)
	4	0.953	(173)	0.955	(166)	0.957	(148)	0.951	(184)	0.949	(112)	0.950	(137)	0.956	(129)	0.950	(84)
	5	0.963	(188)	0.959	(120)	0.957	(156)	0.957	(210)	0.961	(70)	0.945	(45)	0.958	(95)	0.949	(23)

Novelty scores of each field adopt a 10-year reference window and a three-year co-citation window. The total numbers of observations are shown in parentheses

Table 11 Correlation coefficients between novelty in each field and the degree of the relevance of research types

Field	Correlation coefficients							
	new_theory	valid_theory	new_phenom	under_phenom	new_meth	imprv_meth	new_mat	imprv_mat
Chemistry	0.079	0.025	0.114	−0.038	0.083	−0.007	0.143	−0.044
Materials science	−0.008	0.089	0.091	−0.171	0.004	−0.278	0.137	−0.225
Physics and space science	0.022	0.109	0.088	0.055	0.028	0.002	0.013	−0.044
Clinical medicine and psychiatry/psychology	0.022	−0.071	0.130	0.084	0.010	−0.039	0.200	0.132
Basic life sciences	0.231	0.093	0.080	0.076	0.039	−0.055	0.077	−0.031

Novelty scores of each field adopt a 10 year reference window and a three-year co-citation window. The following numbers of observations were recorded for each field: Chemistry, 262; Materials science, 142; Physics and space science, 365; clinical medicine and psychiatry/psychology, 217; and Basic life sciences, 593

Acknowledgements We wish to thank Natsuo Onodera for his invaluable insights regarding the measuring of the novelty score.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by KM. The first draft of the manuscript was written by KM, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding None.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Ahmed, T., Johnson, B., Oppenheim, C., & Peck, C. (2004). Highly cited old papers and the reasons why they continue to be cited: Part II. The 1953 Watson and Crick article on the structure of DNA. *Scientometrics, 61*, 147–156.

Baird, L. M., & Oppenheim, C. (1994). Do citations matter? *Journal of Information Science, 20*(1), 2–15.

Bornmann, L., Schier, H., Marx, W., & Daniel, H. D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics, 6*(1), 11–18.

Bornmann, L., Tekles, A., Zhang, H. H., & Fred, Y. Y. (2019). Do we measure novelty when we analyze unusual combinations of cited references? A validation study of bibliometric novelty indicators based on F1000Prime data. *Journal of Informetrics, 13*(4), 100979.

Clarivate Analytics. (2020). Web of science core collection help. https://images.webofknowledge.com/images/help/WOS/hp_subject_category_terms_tasca.html. Accessed 16 October 2020.

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical? Defining and measuring technological radicalness. *Research Policy, 34*(5), 717–737.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science, 47*(1), 117–132.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429–431.

Igami, M., Nagaoka, S., & Walsh, J. P. (2015). Contribution of postdoctoral fellows to fast-moving and competitive scientific research. *The Journal of Technology Transfer, 40*(4), 723–741.

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal, 36*(10), 1435–1457.

Lee, Y.-N., Walsh, J. P., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy, 44*(3), 684–697.

MacRoberts, M., & MacRoberts, B. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435–444.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220–232.

Murayama, K., Nirei, M., & Shimizu, H. (2015). Management of science, serendipity, and research performance: Evidence from a survey of scientists in Japan and the US. *Research Policy, 44*(4), 862–873.

Nagaoka, S., Igami, M., Eto, M., & Ijichi, T. (2010). *Knowledge creation process in science: Basic findings from a large-scale survey of researchers in Japan*. IIR Working Paper, WP#10–08. Japan: Institute of Innovation Research, Hitotsubashi University.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Belknap Press of Harvard University Press.

Nieminen, P., Carpenter, J., Rucker, G., & Schumacher, M. (2006). The relationship between quality of research and citation frequency. *BMC Medical Research Methodology*. https://doi.org/10.1186/1471-2288-6-42

Oppenheim, C., & Renn, S. P. (1978). Highly cited old papers and reasons why they continue to be cited. *Journal of the American Society for Information Science, 29*, 225–231.

Romer, P. M. (1994). The origins of endogenous growth. *Journal of Economic Perspectives, 8*(1), 3–22.
Schumpeter, J. A. (1939). *Business cycles: A theoretical, historical and statistical analysis of the capitalist process*. McGraw-Hill Book Company.
Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: The integration of product, person and process perspectives. *Psychological Bulletin, 129*(4), 475–494.
Tahamtan, I., & Bornmann, L. (2018). Creativity in science and the link to cited references: Is the creative potential of papers reflected in their cited references? *Journal of Informetrics, 12*(3), 906–930.
Thelwall, M. (2017). Web indicators for research evaluation: A practical guide. *Synthesis Lectures on Information Concepts, Retrieval and Services, 8*(4), i1–i155.
Trapido, D. (2015). How novelty in knowledge earns recognition: The role of consistent identities. *Research Policy, 44*(8), 1488–1500.
Uddin, S., Khan, A., & Baur, L. A. (2015). A framework to explore the knowledge structure of multidisciplinary research fields. *PLoS ONE, 10*(4), e0123537. https://doi.org/10.1371/journal.pone.0123537
Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science, 342*(6157), 468–472.
Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy, 45*(3), 707–723.
Walsh, J. P., & Lee, Y. N. (2015). The bureaucratization of science. *Research Policy, 44*(8), 1584–1600.
Wang, J., Lee, Y.-N., & Walsh, J. P. (2018). Funding model and creativity in science: Competitive versus block funding and status contingency effects. *Research Policy, 47*(6), 1070–1083.
Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy, 46*(8), 1416–1436.
Zdaniuk, B. (2014). Ordinary least-squares (OLS) model. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research.*Springer.