

Italian Research Quality Assessment Practices and Implications: A Postprint

Authors: Geng Haiying

Date: 2025-12-24T16:07:02+00:00

Abstract

Purpose/Significance: This study analyzes the Italian Research Quality Evaluation (VQR 2020-2024) system to reveal its characteristics, challenges, and implications for China, thereby providing references for constructing an evaluation system with Chinese characteristics and for research evaluation practices.

Method/Process: Employing literature analysis methods, this paper examines the evaluation objects and purposes, dimensions and criteria, selection and composition of review experts, and evaluation methods and processes of the Italian Research Quality Evaluation (VQR 2020-2024) system. Based on this analysis, it summarizes the system's features and challenges, and proposes recommendations in light of China's research evaluation practices.

Results/Conclusions: The Italian Research Quality Evaluation is conducted for performance-based funding of research institutions, establishing different evaluation criteria for research outputs, knowledge valorization activities, internationally competitive projects, and research infrastructures. The evaluation process is primarily based on peer review, gradually developing characteristics such as third-party agency evaluation under government leadership, expert-based peer review, continuously expanding evaluation dimensions, and timely meta-evaluation. However, the system also faces challenges including a limited number of evaluated outputs and high evaluation costs.

Full Text

The user wants me to translate a specific section of a Chinese academic paper about Italy's Research Quality Evaluation system. Let me carefully read through the requirements and the text.

Key requirements: 1. Preserve all math - but I don't see any math in this section 2. Preserve all citations - I see [?], [?], [?, ?] which must be kept exactly 3. Preserve figure/table markers - none in this section 4. Maintain academic

tone 5. Flowing, readable prose 6. Merge fragmented text 7. Skip garbage text - none apparent here 8. Use markdown with ## for sections 9. No Chinese characters left at all 10. Author names: Romanize using pinyin (but the author is already in English: Geng Haiying) 11. Journal names: Use English title if available 12. Institution names: Use official English names when known 13. Technical terms: Use standard English equivalents

The section structure is: - Title - Author name and affiliation - Abstract (with [Objective/Significance], [Method/Process], [Result/Conclusion] sections) - Keywords - Classification Number - DOI - Main body text

Let me translate this carefully, ensuring no Chinese characters remain and maintaining academic tone.

First, the title: “## Practice and Enlightenment of Research Quality Evaluation in Italy” - this seems fine, though “Enlightenment” might be better as “Implications” or “Lessons” in academic English. Let me think about the best translation. “启示” can be “implications,” “lessons,” or “insights.” “Enlightenment” is a bit philosophical. “Implications” or “Lessons Learned” might be more appropriate. I’ll go with “Implications” for a more academic tone.

Author: “Geng Haiying” - already in pinyin Affiliation: “Chinese Academy of Social Sciences Evaluation Studies, Beijing 100732, China” - this looks like an official name, I’ll keep it as is.

Abstract: [Objective/Significance] - This is a standard format in Chinese academic papers. I should keep this structure but translate the content. “This study analyzes Italy’s Research Quality Evaluation (VQR 2020-2024) system, revealing its characteristics, challenges, and implications for China. It aims to provide references and insights for constructing an evaluation system with Chinese characteristics and advancing research evaluation practices.”

[Method/Process] - Keep the structure “Using a literature analysis approach, the study examines the evaluation targets and objectives, aspects and criteria, selection and composition of review experts, and methods and processes of Italy’s VQR system. Based on this analysis, it summarizes the system’s features and challenges, while proposing recommendations tailored to China’s context.”

[Result/Conclusion] - Keep the structure “Italy’s Research Quality Evaluation is conducted to allocate performance-based funding to research institutions, applying distinct evaluation criteria for research outputs, knowledge valorization activities, internationally competitive projects, and research infrastructure. The process primarily relies on peer review, gradually evolving into a government-led third-party evaluation model. Key characteristics include peer review based on experts, continuous expansion of evaluation dimensions, and timely meta-evaluations. However, challenges such as limited evaluation coverage and high costs persist.”

Keywords: “Italy; VQR 2020-2024; Research quality evaluation; Research institutions; Institution evaluation” - keep as is

Classification Number: G311 - keep as is DOI: 10.31193/SSAP.J.ISSN.2096-6695.2025.04.07 - keep as is

Now the main body:

“Research evaluation is a critical component of a nation’s scientific research investment decision-making, ensuring efficient use of public research funds while serving as an important means to supervise research quality. It plays a key role in regulating research behavior, mobilizing researcher enthusiasm, and promoting scientific management of research projects and funding [?].”

This is good academic English. I’ll keep the citation.

“China’s research evaluation system was initiated alongside the establishment of competitive funding programs [?] and rapidly became a focal point of research system reform. Since the 1990s, with the development of bibliometrics and scientometrics, quantitative indicators have played an important role in research evaluation, but this has also led to the misuse of such indicators.”

Good translation.

“Over the past decades, China has also emphasized quantitative indicators in research evaluation, giving rise to tendencies such as “simplistic quantification,” “emphasizing quantity over quality,” “one-size-fits-all approaches,” and the “five-only” problem (over-reliance on specific metrics).”

The “five-only” problem is a specific Chinese policy term. I should keep the quotes and explain it as “over-reliance on specific metrics” which is what the original does.

“In response, China has issued a series of reform documents on research evaluation, proposing to adhere to categorical evaluation, focus on the quality, contribution, and impact of landmark achievements, and prioritize indicators such as originality of research outcomes, transformation benefits, and satisfaction with science and technology services. These policies have also promoted a representative work evaluation system and emphasized the role of peer review mechanisms in the evaluation process. While these documents have pointed the way for reform, unresolved issues persist in actual evaluation practices.”

Good translation.

“From the late 20th century to the early 21st century, to maximize the 促进作用 of public resources in scientific research, European countries such as the United Kingdom, France, and the Netherlands established national-level research evaluation systems.”

“促进作用” - “promotional effect” or “effectiveness”. “Maximize the effectiveness of public resources in scientific research” sounds better.

“Against this backdrop, Italy also launched government-led national research evaluation activities targeting universities and research institutions. To date, Italy has completed four such evaluations, with results informing government

research policy formulation. Italy's national research evaluation system has significant influence both domestically and across Europe, continuously adjusting in line with international trends in research evaluation reform."

Good.

"Chinese scholars have introduced its development history, evaluation criteria and methods, and reform characteristics based on earlier versions [?, ?]. However, each Italian evaluation cycle introduces substantial changes, and the fourth evaluation currently underway features major adjustments. As of March 2025, searches of Chinese databases such as CNKI reveal no domestic research on this fourth evaluation."

CNKI is a known database, I can keep it as is.

"Therefore, this paper reviews Italy's fourth research quality evaluation system as a case study, aiming to provide references for constructing a Chinese-characteristic evaluation system and advancing research evaluation practices."

Let me review the entire translation for flow and accuracy. I need to make sure there are absolutely no Chinese characters left. Let me scan through the original text again to catch any I might have missed.

Original text has: - 进 (in 促进作用) - I translated this as "effectiveness" - 用 (in 作用) - part of "effectiveness" - 作 (in 作为) - translated as "as" or part of other words - 促 (in 促进) - translated as "promote" or "promotional"

Wait, the user said "Chinese characters found: 进, 用, 作, 促" - these appear to be part of larger words. I need to make sure my translation doesn't contain any Chinese characters. Let me double-check my translation.

Actually, looking at the original text more carefully, I see these characters are part of words like "促进作用" (promotional effect), "作为" (as/serve as), etc. My translation should not contain any of these Chinese characters.

Let me re-read my translation to ensure it's clean:

Practice and Implications of Research Quality Evaluation in Italy

Geng Haiying

(Chinese Academy of Social Sciences Evaluation Studies, Beijing 100732, China)

Abstract:

[Objective/Significance] This study analyzes Italy's Research Quality Evaluation (VQR 2020-2024) system, revealing its characteristics, challenges, and implications for China. It aims to provide references and insights for constructing an evaluation system with Chinese characteristics and advancing research evaluation practices. [Method/Process] Using a literature analysis approach, the study examines the evaluation targets and objectives, aspects and criteria,

selection and composition of review experts, and methods and processes of Italy's VQR system. Based on this analysis, it summarizes the system's features and challenges, while proposing recommendations tailored to China's context. [Result/Conclusion] Italy's Research Quality Evaluation is conducted to allocate performance-based funding to research institutions, applying distinct evaluation criteria for research outputs, knowledge valorization activities, internationally competitive projects, and research infrastructure. The process primarily relies on peer review, gradually evolving into a government-led third-party evaluation model. Key characteristics include peer review based on experts, continuous expansion of evaluation dimensions, and timely meta-evaluations. However, challenges such as limited evaluation coverage and high costs persist.

Keywords: Italy; VQR 2020–2024; Research quality evaluation; Research institutions; Institution evaluation

Classification Number: G311

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2025.04.07

Research evaluation is a critical component of a nation's scientific research investment decision-making, ensuring efficient use of public research funds while serving as an important means to supervise research quality. It plays a key role in regulating research behavior, mobilizing researcher enthusiasm, and promoting scientific management of research projects and funding [?]. China's research evaluation system was initiated alongside the establishment of competitive funding programs [?] and rapidly became a focal point of research system reform. Since the 1990s, with the development of bibliometrics and scientometrics, quantitative indicators have played an important role in research evaluation, but this has also led to the misuse of such indicators.

Over the past decades, China has also emphasized quantitative indicators in research evaluation, giving rise to tendencies such as “simplistic quantification,” “emphasizing quantity over quality,” “one-size-fits-all approaches,” and the “five-only” problem (over-reliance on specific metrics). In response, China has issued a series of reform documents on research evaluation, proposing to adhere to categorical evaluation, focus on the quality, contribution, and impact of landmark achievements, and prioritize indicators such as originality of research outcomes, transformation benefits, and satisfaction with science and technology services. These policies have also promoted a representative work evaluation system and emphasized the role of peer review mechanisms in the evaluation process. While these documents have pointed the way for reform, unresolved issues persist in actual evaluation practices.

From the late 20th century to the early 21st century, to maximize the effectiveness of public resources in scientific research, European countries such as the United Kingdom, France, and the Netherlands established national-level research evaluation systems. Against this backdrop, Italy also launched government-led national research evaluation activities targeting universities and research institutions. To date, Italy has completed four such evaluations,

with results informing government research policy formulation. Italy's national research evaluation system has significant influence both domestically and across Europe, continuously adjusting in line with international trends in research evaluation reform. Chinese scholars have introduced its development history, evaluation criteria and methods, and reform characteristics based on earlier versions [?, ?]. However, each Italian evaluation cycle introduces substantial changes, and the fourth evaluation currently underway features major adjustments. As of March 2025, searches of Chinese databases such as CNKI reveal no domestic research on this fourth evaluation. Therefore, this paper reviews Italy's fourth research quality evaluation system as a case study, aiming to provide references for constructing a Chinese-characteristic evaluation system and advancing research evaluation practices.

This looks good. No Chinese characters remain. The citations are preserved. The structure is maintained. The language is academic and fluent.

Wait, let me check the title again. The original is "Practice and Enlightenment of Research Quality Evaluation in Italy". I changed "Enlightenment" to "Implications". Let me reconsider. "Enlightenment" in Chinese academic context often means "lessons learned" or "implications". "Implications" is more standard in English academic writing. "Lessons" could also work. I think "Implications" is the best choice.

Let me also check "Chinese-characteristic" - this is a standard term in Chinese policy discourse. In English, it might be better as "with Chinese characteristics" which is what I used in the abstract. But in the last sentence, I wrote "Chinese-characteristic evaluation system". Let me make it consistent. I should use "evaluation system with Chinese characteristics" throughout.

Looking at the last sentence: "Therefore, this paper reviews Italy's fourth research quality evaluation system as a case study, aiming to provide references for constructing a Chinese-characteristic evaluation system and advancing research evaluation practices."

I should change "Chinese-characteristic evaluation system" to "evaluation system with Chinese characteristics" for better flow and consistency with the abstract.

Let me also check "knowledge valorization activities" - "valorization" is a term used in European research policy context, so it's appropriate here.

"representative work evaluation system" - this is a direct translation of "代表作制度". In English academic context, this might be better as "representative works evaluation system" or "representative output evaluation system". But "representative work evaluation system" is understandable and keeps the original##
1 Italy's Research Quality Evaluation (VQR 2020-2024) System

Italy's research evaluation activities are organized and implemented by the National Agency for the Evaluation of Universities and Research Institutes (Agenzia nazionale di valutazione del sistema universitario e della ricerca, ANVUR). The evaluation system is called "Valutazione della Qualità della Ricerca" (VQR).

ANVUR has completed three VQR cycles: the first from 2011–2013, evaluating research activities from 2004–2010 with results published in summer 2013; the second from 2015–2017, evaluating 2011–2014 activities with results published in February 2017; and the third from 2020–2022, evaluating 2015–2019 activities with results published in July 2022. The fourth VQR evaluation launched at the end of 2023, assessing research activities from 2020–2024 with results planned for 2026, referred to as “VQR 2020–2024.” Based on the fundamental elements of the evaluation system and analyses of national research evaluation frameworks [?], this study examines the basic elements of VQR 2020–2024 according to relevant documents published by ANVUR [?].

1.1 Evaluation Objects and Objectives

VQR evaluates research institutions, specifically including Italian universities, public research institutions supervised by the Ministry of University and Research, and other private research institutions that request evaluation and are willing to pay evaluation fees. The evaluation can target entire universities or research institutions, or their subordinate departments, but does not focus on individual researchers.

VQR assesses research work to guide public funding allocation, reward high-performing institutions, and promote continuous improvement. Italian law explicitly stipulates that VQR is an external evaluation that assesses the effectiveness and efficiency of all publicly funded research activities on the basis of fairness, independence, professionalism, transparency, and openness. All institutions receiving public funding must undergo evaluation, and results serve as a basis for allocating university funds. However, evaluation results only apply to a small portion of total funding—the so-called “Quota Premiale” (premium quota)—making performance-based funding’s impact much smaller than in other countries (particularly the UK) [?]. In summary, VQR has three main purposes: (1) promoting research excellence and innovation to enhance national competitiveness; (2) improving public funding efficiency to meet national strategic development needs; and (3) promoting knowledge sharing and utilization, strengthening collaboration, and enhancing socio-economic benefits. VQR 2020–2024 comprehensively elaborates on these purposes, including 11 specific aspects such as promoting and enhancing national research quality, encouraging broader researcher participation, and motivating researchers to engage in internationally competitive projects [?].

1.2 Evaluation Dimensions and Criteria

1.2.1 Evaluation Dimensions VQR 2020–2024 includes four evaluation dimensions: representative research outputs, knowledge valorization activities, internationally competitive projects, and research infrastructure.

Research outputs must be published during 2020–2024. Institutions submit representative works, with total submissions not exceeding 2.5 times the number

of researchers. VQR evaluates diverse output types, including journal articles, monographs and book chapters, patents, and other forms of results.

Following VQR 2015–2019’s formal evaluation of institutions’ third mission activities, VQR 2020–2024 continues to assess knowledge valorization (third mission/societal impact). Knowledge valorization refers to the process of creating economic and social value by collaborating with different sectors to transform data, technical skills, and research results into sustainable knowledge-based products, services, solutions, and policies. This dimension is evaluated through case studies submitted by institutions, with each institution required to submit at least one case study and a maximum number linked to institutional size. VQR 2020–2024 identifies five themes and specific fields related to case studies (Table 1), with each case study addressing up to three themes and detailing activities’ impacts on economy, society, culture, health, and environment during 2020–2024, particularly highlighting differences before and after interventions.

VQR 2020–2024 adds evaluation of research infrastructure and internationally competitive projects. Research infrastructure evaluation targets research institutions and voluntarily participating private institutions, which may submit a research infrastructure report on a voluntary basis. VQR 2020–2024 lists research infrastructure types (Table 2) covering not only digital infrastructure, energy, environment, health and nutrition, physics and engineering, but also social sciences and humanities.

Internationally competitive project evaluation focuses on projects successfully applied for and launched during 2020–2024 that meet three conditions: (1) funding of at least €50,000; (2) funding from non-Italian public or private agencies (e.g., European Commission, UN, WHO, ESA, NASA, NIH, NSF); and (3) allocation through peer review based on international competition. This evaluation aims to guide researchers toward international projects and enhance global influence.

1.2.2 Evaluation Criteria VQR 2020–2024 establishes different criteria for research outputs, knowledge valorization (third mission/societal impact), and research infrastructure (Table 3).

Based on these criteria, research outputs, case studies, or research infrastructure are rated as “Excellent,” “Good,” “Standard,” “Satisfactory,” or “Unsatisfactory,” corresponding to scores of 1, 0.8, 0.5, 0.2, and 0 for institutional research quality evaluation.

1.3 Review Experts

VQR review experts include Evaluation Expert Group (Gruppi di Esperti della Valutazione/Evaluation Expert Groups, GEV) members and external reviewers. VQR 2020–2024 comprises 17 disciplinary fields, one interdisciplinary knowledge valorization field, and one research infrastructure field. The 17 disciplinary fields include 10 STEM fields and 7 social sciences and humanities fields, each

with its own GEV. GEV members are selected through open recruitment, ANVUR lottery, or direct appointment. For the 17 disciplinary fields, ANVUR announces selection criteria and qualification requirements in advance, with eligible candidates entered into GEV candidate pools. For each field, 75% of GEV members are selected by lottery from the candidate pool, while 25% are directly appointed by ANVUR based on qualifications and experience. For knowledge valorization evaluation, ANVUR establishes an interdisciplinary review group appointed from publicly registered qualified candidates. Additionally, ANVUR appoints 10 GEV members for research infrastructure based on specific requirements. Beyond GEV members, GEVs may engage external experts selected from ANVUR-provided lists and approved by the ANVUR council. Review expert lists are published by ANVUR within 30 days after the final VQR report release. Table 4 shows VQR 2020–2024 evaluation fields and GEV expert numbers.

1.4 Evaluation Methods

Each field's GEV develops evaluation method documents detailing field scope, GEV organization, evaluation methods, and procedures. VQR 2020–2024 primarily uses peer review to evaluate research outputs, case studies, and research infrastructure. The following focuses on research output evaluation methods.

Informed Peer Review: This approach combines citation metrics and other quantitative indicators with peer review, mainly used for evaluating research outputs published in STEM and economics journals indexed in WoS and Scopus. Each output is reviewed and scored by two experts based on evaluation criteria and bibliometric indicators. Key metrics include citation counts and journal impact indicators such as 5-year impact factor (IF5) and Article Influence Score (AIS) from WoS, and CiteScore, SNIP, and Scimago Journal Rank (SJR) from Scopus. For example, in mathematics and computer science, ANVUR provides two citation data tables for each discipline category, output type (article, letter, review), and publication year (2020–2024): (1) an output citation data table showing citation thresholds for ranking in global citation distribution percentiles (0%–10%, 10%–35%, 35%–60%, 60%–80%, 80%–100%); and (2) a journal citation data table containing journal impact indicator values and percentiles relative to all journals in the discipline and year. GEVs use this citation data as reference when assigning ratings.

Pure Peer Review: This method is used for humanities research outputs or other fields where citation-based peer review is inappropriate.

The user wants me to translate a specific section from Chinese to English. The section is “### 1.5 Evaluation Process”. Let me analyze the content:

The text appears to be already partially in English, but with some Chinese characters mixed in (“分” , “充”). The user says “These appear to be: mixed technical content and proper nouns”. Let me look at the text carefully:

1.5 Evaluation Process

Research outputs are typically evaluated by two GEV members based on disciplinary expertise.

[Figure 1: see original paper]

****Phase 1****: Reviewers work anonymously and independently. The GEV coordinator assigns each

****Phase 2****: Reviewers communicate through the platform to determine final ratings. In the s

I can see the Chinese characters “分” and “充” in the text. Let me identify where they are: - “分” appears in “分” (which seems to be part of “分” but I need to see the context) - “充” appears in “充分” (which means “sufficient” or “adequate”)

Looking at the sentence: “Modifying results requires 充分 justification.” - here “充分” is the Chinese word that needs to be translated to English. The character “分” is part of “充分” .

So the task is to: 1. Translate the Chinese characters/words to English 2. Keep the rest of the English text as is 3. Maintain the structure (section heading, figure marker, phase headings) 4. Ensure no Chinese characters remain

The Chinese word “充分” means “sufficient”, “adequate”, or “full” . In academic context, “sufficient justification” or “adequate justification” would be appropriate.

Let me re-read the entire section to ensure I understand it correctly and identify any other Chinese characters.

The section is about an evaluation process, likely for research outputs. It mentions: - GEV members (Group of Evaluation?) - VQR evaluation process - Two phases of review - External experts - Consensus groups - Rating levels

The only Chinese characters I can see are in “充分” . Let me check if there are any others...scanning through the text again: - “分” - appears as part of “充分” - “充” - appears as part of “充分”

That’ s it. The rest is already in English.

So I need to: 1. Keep the heading “### 1.5 Evaluation Process” 2. Keep the first paragraph as is (it’ s already English) 3. Keep the [Figure 1: see original paper] marker 4. Keep the **Phase 1** and **Phase 2** headings 5. In the Phase 2 paragraph, replace “充分” with an appropriate English word

The sentence is: “Modifying results requires 充分 justification.”

This should become: “Modifying results requires sufficient justification.” or “Modifying results requires adequate justification.” or “Modifying results requires full justification.”

In academic context, “sufficient justification” is most appropriate and natural.

Let me also check if there are any other issues. The text seems to flow well otherwise. The user wants me to “Produce FLOWING, READABLE prose” and

“MERGE fragmented text into complete, well-formed sentences and paragraphs” . However, this section appears to be already well-structured in English, except for that one Chinese word.

I should also check for any other potential issues: - “GEV” - this is an acronym, likely for “Gruppo di Valutazione” (Italian) or similar. I should keep it as is. - “VQR” - another acronym, likely “Valutazione della Qualità della Ricerca” (Italian). Keep as is. - The structure is already good with clear paragraphs and phase headings.

The only change needed is translating “充分” to “sufficient” .

Wait, let me double-check the original text again to make sure I haven’t missed anything. The user provided the section to translate, and it appears to be already mostly in English with just that one Chinese word “充分” .

Actually, looking more carefully at the user’s instructions: “Chinese characters found: 分, 充” and “These appear to be: mixed technical content and proper nouns” . This suggests that these characters might be part of technical terms or proper nouns. But in the context of “充分 justification” , it’s clearly just the Chinese word for “sufficient” .

Let me also check if “分” appears elsewhere. Scanning the text: - “two GEV members” - no - “maximum of 50 outputs” - no - “Figure 1” - no - “Phase 1” - no - “1-10 scale” - no - “five rating levels” - no - “consensus group” - no - “Phase 1### 2 Characteristics of Italy’s Research Quality Evaluation

Under government leadership, Italy’s research quality evaluation has gradually developed characteristics including third-party agency evaluation, expert-based peer review, continuously expanding evaluation dimensions, and timely meta-evaluation.

2.1 Evaluation Subject: Third-Party Agency Under Government Leadership

ANVUR, the implementing organization for Italy’s research quality evaluation, is a specialized independent evaluation agency established at the national level. From ANVUR’s establishment to VQR implementation, all activities proceed under corresponding laws and regulations, demonstrating strong government leadership. Simultaneously, ANVUR enjoys organizational, administrative, and financial autonomy, ensuring strong independence.

2.2 Evaluation Method: Expert-Based Peer Review

VQR employs peer review for research outputs, knowledge valorization activities, and research infrastructure, with research output evaluation based on representative works. VQR requires institutions to submit selected outputs from the evaluation period—for example, VQR 2020–2024 limits total submissions to 2.5

times the number of active researchers. Limiting submission quantity facilitates peer review feasibility while guiding institutions and researchers toward quality over quantity. Additionally, VQR allows representative works to take various forms depending on disciplinary characteristics, including high-level papers, patents, monographs, creative designs, databases, and other output types. For peer review, each disciplinary GEV develops specific evaluation criteria, reference bibliometric indicators, and decides whether to use informed or pure peer review, reflecting the concept of categorical evaluation.

2.3 Evaluation Standards: Continuously Expanding Dimensions

VQR 2020–2024 evaluation dimensions include research outputs, knowledge valorization, research infrastructure, and internationally competitive projects, covering academic contributions, non-academic contributions, and research support. This demonstrates Italy's expanding research quality evaluation toward multiple dimensions, emphasizing research's role in promoting economic, social, and cultural development beyond academia. Research infrastructure serves as the foundation for scientific research, high-quality output production, and academic innovation, significantly influencing the overall research environment. Including it in the evaluation framework encourages institutions to prioritize infrastructure development.

2.4 Evaluation Feedback: Post-Evaluation Meta-Evaluation Mechanism

Many changes in VQR 2020–2024 stem from meta-evaluation recommendations from VQR 2015–2019. To improve and refine the process, ANVUR conducted three levels of meta-evaluation after VQR 2015–2019 [?]: First, ANVUR's in-depth analysis of evaluation results through multi-dimensional examination of submitted outputs and case studies to understand overall evaluation object characteristics and provide data analysis foundations for subsequent external expert reviews. Second, ANVUR's survey of evaluation procedures to assess satisfaction among participants (evaluated institution staff and GEV members) regarding service quality and data platform usability, informing future process improvements. Third, ANVUR's engagement of international experts to externally evaluate VQR's purposes, methods, and indicators, analyzing strengths and weaknesses and proposing improvement recommendations for future cycles. ANVUR's meta-evaluation work provides direction for subsequent evaluations, with many international expert suggestions incorporated into new cycles. Through meta-evaluation, ANVUR can promptly grasp VQR's effectiveness, reflect on the evaluation process itself, strengthen evaluation management, and improve future work.

3 Challenges Facing Italy' s Research Quality Evaluation

Italy' s VQR aims to effectively promote research improvement through evaluation practices. Evidence suggests that the worst-performing institutions have made the greatest progress [?]. Additionally, performance-based funding should promote greater international cooperation. While Italy' s research internationalization has continuously improved, whether this stems from VQR or represents a broader global phenomenon remains unclear [?]. In fact, since VQR' s implementation, it has faced criticism from Italy' s scientific community. Scholars have summarized VQR 2011-2014' s defects, including: evaluation covering only a relatively small proportion of scientific publications by participating researchers; erroneous and untimely use of journal metrics to evaluate individual papers; and conceptually misleading bibliometric indicator normalization and aggregation standards [?]. These issues directly affect the research ecosystem, fostering risk-averse cultures that inhibit innovative, unconventional, unpredictable, or interdisciplinary research while encouraging “publish or perish” mentalities that may prioritize quantity over quality [?]. Evaluation also exacerbates regional disparities, undermining balanced regional development goals [?]. Consequently, ANVUR continuously adjusts evaluation criteria and methods in response to criticism and international reform trends, yet still faces significant challenges.

Limited Evaluation Coverage: VQR' s requirement for institutions to submit relatively few outputs (2-3 per researcher) aims to reduce reviewer burden and evaluation costs while guiding focus toward quality. However, submitting only 2-3 outputs per researcher over a five-year evaluation period limits the assessment of institutional productivity, enabling identification of institutions producing small quantities of high-quality work but potentially failing to recognize those with both high output and high quality—effectively identifying less excellent institutions but possibly missing truly excellent ones [?].

Societal Impact Evaluation Issues: Social sciences impact is often too indirect, non-linear, and diffuse to be fully captured through the third mission [?]. Moreover, constructing “societal impact” in social sciences requires embedding multiple strategies into research design, yet universities rarely promote these strategies, preferring fields with direct commercial benefits [?]. Thus, societal impact evaluation itself faces disciplinary differences.

Expert Selection Issues: VQR 2004-2010 and VQR 2011-2014 used “open recruitment + ANVUR appointment,” but research showed excessive connections among review experts based on knowledge structure networks, violating fairness principles [?]. VQR 2015-2019 adjusted to “open recruitment + ANVUR lottery,” but issues emerged regarding review group composition structure [?]. VQR 2020-2024 further adjusted to a combined “open recruitment + ANVUR appointment and lottery” approach, striving to balance fairness and representation. Therefore, the fairness of expert selection procedures and representativeness of expert composition remain unresolved practical issues that VQR continues to explore.

Evaluation Costs: Like the UK' s Research Excellence Framework (REF), VQR faces high evaluation costs [?]. VQR 2004–2010 cost €105.7 million (including CINECA platform operation), while VQR 2011–2014 cost €147 million [?], with costs continuously rising. VQR 2015–2019 and VQR 2020–2024' s exclusive use of peer review will undoubtedly further increase costs—a common issue in peer review-based evaluations.

4 Implications for China' s Research Evaluation

While Italy' s research quality evaluation system is not perfect and has many problems, its organizational implementation, periodic conduct, and continuous improvement approach offer valuable insights. Based on Italy' s practices and China' s current evaluation status and reform trends, several implications emerge for China' s research institution evaluation practices.

4.1 Establish and Improve Institutional Systems, Develop Third-Party Research Evaluation Agencies

Third-party evaluation refers to comprehensive evaluation activities conducted by professional evaluation institutions with independent legal status that have no affiliation or interest relationship with commissioners and evaluatees, following established procedures and standards [?]. Third-party evaluation offers advantages of independence, objectivity, and fairness [?], compensating for defects in internal and self-evaluation while effectively avoiding overlapping rights in policy formulation, implementation, and evaluation [?]. ANVUR, Italy' s research quality evaluation implementing agency, is a specialized independent national-level evaluation institution with no affiliation to evaluated institutions, ensuring evaluation fairness to some extent.

Since the late 20th century, Chinese ministries have commissioned research institutions to conduct independent evaluations of government-funded S&T plans, projects, institutions, and policies to optimize resource allocation and improve efficiency and quality [?]. Subsequently, China established several third-party evaluation agencies conducting evaluations in S&T, public policy, and government performance domains. However, these agencies face practical dilemmas, including insufficient legal norms, lack of independence guarantees, unclear qualification standards, and inadequate procedural rules [?], requiring effective measures to promote independent development and establish standardized, professional evaluation systems [?]. By contrast, ANVUR' s institutional nature, purpose, activities, and organizational structure are clearly defined by law, with VQR evaluation elements (content, criteria, methods) and procedures explicitly stipulated by corresponding decrees.

Drawing from Italy' s experience and based on China' s national conditions, government departments should strengthen top-level design of research evaluation and intensify efforts to cultivate third-party professional research evaluation agencies. First, **legislative guarantees:** Accelerate legal construction for

third-party evaluation, formulate robust laws and regulations to create favorable institutional environments, and legislatively define basic rights and obligations of all parties in third-party evaluation activities, particularly the fundamental rights required for agencies to conduct work. Second, **enhance evaluation capacity**: Clarify qualification standards for third-party evaluation institutions and issue guiding evaluation norms. Third, **improve procedural rules**: Refine evaluation procedures and streamline evaluation supervision processes [?].

4.2 Recognize Diverse Outputs, Implement Multi-Dimensional Evaluation Standards

Italy' s VQR offers insights through its diverse representative outputs and expanded evaluation dimensions. First, diversified scientific research activities produce diversified outputs. Disciplines range from natural sciences and engineering to medicine and humanities/social sciences. Research types include basic research, applied research, technological development, social welfare research, S&T management services, and experimental techniques in S&T fields, as well as theoretical research, applied policy research, and artistic performance creation in humanities and social sciences. Different disciplines and research types exhibit significant differences in research models and output forms. Research outputs include not only papers, monographs, patents, and policy consultation reports, but also national/industry standards, reference books, popular science works, encyclopedia entries, internal research reports, datasets, software, and artistic products—all important research outputs that can serve as representative works in evaluation. Second, research impact is multi-dimensional, encompassing both academic contributions and contributions to socio-economic development, policy formulation, and science popularization. Third, research is influenced by various foundational support conditions, making research infrastructure another necessary evaluation component.

Therefore, research evaluation should fully consider disciplinary fields and research types, formulate differentiated evaluation criteria, employ differentiated methods, and implement multi-dimensional, multi-level categorical evaluation systems. Although Chinese policy documents repeatedly emphasize categorical evaluation, phenomena such as “paper-only” and “one-size-fits-all” approaches persist. Drawing from Italy' s experience, China' s research evaluation should emphasize: First, **countering “paper-only”** by expanding output recognition scope and equally treating diverse output types. Second, **enriching impact evaluation dimensions** by comprehensively measuring academic and non-academic contributions to guide diversified research development. Third, **strengthening systematic evaluation** by including not only research outputs but also research environment/culture, infrastructure, and student training—particularly at the institutional level where systematic evaluation should be emphasized. Fourth, **developing distinctive criteria and methods** based on evaluation purposes, objects, disciplinary fields, and research types.

4.3 Emphasize Post-Evaluation Feedback, Conduct Timely Meta-Evaluation

The direct purpose of evaluation is to serve management by providing key information for decision-making. Meta-evaluation evaluates the evaluation itself. Conducting meta-evaluation reflects on evaluation quality, summarizes successful experiences, and corrects errors and biases in evaluation processes, helping to scientize and standardize evaluation activities and improve quality and applicability [?]. Meta-evaluation originated in education and has been discussed in research evaluation circles [?], but China's current research evaluation system lacks meta-evaluation mechanisms [?]. Meta-evaluation can be internal self-evaluation or external third-party evaluation, and can evaluate evaluation purposes, indicators, methods, results, or evaluation effects/impacts.

Italy's VQR practice suggests several meta-evaluation approaches. First, **conduct self-meta-evaluation**: Evaluation agencies should conduct multi-dimensional in-depth analysis of evaluation results internally to understand overall evaluation object characteristics, verify consistency between results and expectations, and provide data analysis foundations for subsequent external meta-evaluation. Second, **conduct thematic meta-evaluation**: Agencies should conduct specialized evaluations of specific process aspects, such as expert selection, comparison of qualitative and quantitative methods, and data platform support. Third, **conduct comprehensive meta-evaluation**: Agencies can engage third-party or external experts to conduct meta-evaluation of the entire evaluation process, diagnosing strengths and weaknesses and proposing improvement recommendations. Fourth, **evaluate evaluation results application or effects**: Third-party agencies should analyze positive and negative impacts of evaluation results—essentially evaluating or reflecting on research evaluation policy effects—to help relevant institutions adjust research evaluation policies.

Footnote:

A university's "third mission" refers to activities beyond teaching and research where universities, leveraging existing human, teaching, and research resources, engage in multi-dimensional interaction and cooperation with industry and government to conduct technology consulting or transfer, collaborative research, university spin-off creation, and other knowledge innovation and commercialization activities related to but distinct from teaching and research. See Xia Qinghua et al., "The Connotation and Understanding of University's 'Third Mission'," *China Terminology*, 2011(4).

References:

- [?] Xing Jianjun. Italy's Research Quality Evaluation System: Development, Role, and Characteristics [?]/Sun Yanhong. *Italy Development Report (2022-2023)*. Beijing: Social Sciences Academic Press, 2023: 158-178.
- [?] Qiu Junping, Zhang Yuchen, Zhou Zifan. Eight Relationships That Must

Be Handled Properly in Reconstructing China's Research Evaluation System in the New Era [?]. *Journal of Library Science in China*, 2021, 47(1): 47-60.

[?] Chen Kun, Li Xiaoxuan, Yang Guoliang. Reform of Italy's Research Evaluation System [?]. *China Science and Technology Forum*, 2015(2): 148-154.

[?] Cheng Yanlin, Song Qiuhui, Chen Jiani, et al. Italy's Research and Third Mission Evaluation and Its Implications for China [?]. *World Science and Technology Research and Development*, 2022, 44(4): 557-566.

[?] Zhang Shuguang, Chen Lu. Evolution, Mechanism, and Reform Characteristics of Italian University Research Evaluation [?]. *International and Comparative Education*, 2024, 51(9): 59-75.

[?] Jiang Ying. Research on National Research Evaluation Systems in Research Performance Funding Systems [?]. *Journal of University of Chinese Academy of Social Sciences*, 2024, 44(12): 5-22.

[?] ANVUR. VQR 2020-2024 Related Documents [?]. [?]. <https://www.anvur.it/en/research/evaluation-research-quality/evaluation-procedures>.

[?] Bonaccorsi A. Two Decades of Experience in Research Assessment in Italy [?]. *Scholarly Assessment Reports*, 2020, 2(1): 2-19.

[?] ANVUR. VQR 2020-2024 Research Quality Evaluation Announcement [?]. [?]. https://www.anvur.it/sites/default/files/2024-11/Bando-VQR-2020-2024_{31}ottobre.pdf.

[?] ANVUR. VQR 2020-2024 Knowledge Valorization Case Submission Methods [?]. [?]. <https://www.anvur.it/sites/default/files/2024-11/Documento-modalita-di-conferimento-dei-casi-studio%20VdC-VQR-2020-2024.pdf>.

[?] ANVUR. VQR 2020-2024 Infrastructure Report Submission Methods [?]. [?]. <https://www.anvur.it/sites/default/files/2024-11/Documento-modalita-di-conferimento-infrastrutture-di-ricerca-VQR-2020-2024.pdf>.

[?] ANVUR. Critical Analysis of VQR 2015-2019 Results [?]. [?]. https://www.anvur.it/sites/default/files/2025-02/ANVUR_{Critical}-Analysis-of-VQR-2015-2019-Results.pdf.

[?] ANVUR. VQR 2015-2019 Expert Review Report [?]. [?]. https://www.anvur.it/sites/default/files/2025-02/Expert-Review-Panel_{Report}-on-VQR-2015-2019.pdf.

[?] Abramo G, D' Angelo C A. The Different Responses of Universities to Introduction of Performance-Based Research Funding [?]. *Research Evaluation*, 2021, 30(4): 514-528.

[?] Abramo G, D' Angelo C A. The Impact of Italian Performance-Based Research Funding Systems on the Intensity of International Research Collaboration [?]. *Research Evaluation*, 2023, 32(1): 47-57.

[?] Franceschini F, Maisano D. Critical Remarks on the Italian Research Assessment Exercise VQR 2011-2014 [?]. *Journal of Informetrics*, 2017, 11(2): 337-357.

[?] Buoso E. Automated Decisions, Research Quality Assessment and Step-Based Salary System: Potential Dangers for Academic Freedom in Italy [?]. [?]. Seckelmann M, Violini L, Fraenkel-Haeblerle C, et al. *Academic Freedom Under Pressure?* Springer, 2021: 153-162.

[?] Grisorio M J, Prota F. Italy's National Research Assessment: Some Unpleasant Effects [?]. *Studies in Higher Education*, 2020, 45(4): 736-754.

- [?] Muhonen R, Tellmann S M. Challenges of Reporting Societal Impacts for Research Evaluation Purposes—The Case of Sociology [?]/Engels T C E, Kulczycki E. *Handbook on Research Assessment in the Social Sciences*. Edward Elgar Publishing, 2022: 335-349.
- [?] Aiello E, Donovan C, Duque E, et al. Effective Strategies That Enhance the Social Impact of Social Sciences and Humanities Research [?]. *Evidence & Policy*, 2021, 17(1): 131-146.
- [?] Baccini A, Re C. Is the Panel Fair? Evaluating Panel Compositions Through Network Analysis. The Case of Research Assessments in Italy [?]. *Scientometrics*, 2025, 130(4): 2093-2135.
- [?] Geuna A, Piolatto M. Research Assessment in the UK and Italy: Costly and Difficult, but Probably Worth It (At Least for a While) [?]. *Research Policy*, 2016, 45(1): 260-271.
- [?] Cheng Yanlin. How to Ensure Independence of Third-Party Evaluation [?]. *China Science and Technology Forum*, 2017(7): 14-18.
- [?] Cheng Yanlin, Zhang Wei. Third-Party Evaluation in China: Characteristics, Types, and Development Strategies [?]. *China Science and Technology Forum*, 2022(9): 139-148.
- [?] Fang Yan. On Several Relationships in Public Domain S&T Evaluation [?]. *Science and Society*, 2011, 1(3): 72-83, 61.
- [?] Fang Yan. Actively Exploring Scientific and Rational S&T Evaluation Systems [?]. *Qiushi*, 2012(4): 53-54.
- [?] Tan Hualin, Wu Ang. Dilemmas and Institutional Improvement of Third-Party Evaluation of S&T Achievements in China [?]. *Jinan Journal (Philosophy & Social Sciences Edition)*, 2018, 40(9): 32-40.
- [?] Zhu Junwen, Liu Li, Zhu Jiani. *International S&T Policy Development Report: S&T Evaluation Volume* [?]. Shanghai: Shanghai Jiao Tong University Press, 2015: 279.
- [?] Zhu Shaoqiang, Tang Lin, Ke Qing. Meta-Evaluation Mechanism for Academic Evaluation [?]. *Journal of Chongqing University (Social Science Edition)*, 2010, 16(3): 81-85.
- [?] Yu Liping, Pan Yuntao, Wu Yishan. Research on the Impact of Meta-Evaluation on Improving S&T Evaluation Quality [?]. *Science Research Management*, 2012, 30(4): 512-516.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.