APPENDIX B AND C

Informing Research Choices: Indicators and Judgment

The Expert Panel on Science Performance and Research Funding

Council of Canadian Academies
Conseil des académies canadiennes

Science Advice in the Public Interest

Appendix B
International Case Studies

- Introduction to the International Case Studies
- Australia
- China
- Finland
- Germany
- The Netherlands
- Norway
- Singapore
- South Korea
- United Kingdom
- United States

1 Introduction to the International Case Studies

Many countries have experience with national research assessment initiatives relevant to this assessment. Wherever possible, the Panel has considered and included all such relevant experience in its deliberations. The Panel identified 10 countries, however, for more detailed analysis. The countries selected satisfied one or more of the following four general criteria:

- **Knowledge-powerful countries:** countries that have demonstrated sustained leadership and commitment at the national level to fostering science and technology and/or supporting research and development in the natural sciences and engineering (NSE).
- **Leaders in science assessment and evaluation:** countries that have notable or distinctive experience at the national level with use of science indicators or administration of national science assessment initiatives as related to research funding allocation.
- **Emerging science and technology leaders:** countries considered to be emerging "knowledge-powerful" countries and in the process of rapidly expanding support for science and technology, or playing an increasingly important role in the global context of research in the NSE.
- **Relevance to Canada:** countries known to have special relevance to Canada and the sponsor of this assessment because of the characteristics of their systems of government or the nature of their public research funding institutions and mechanisms.

For each of the 10 countries selected, the Panel developed a case study. While the Panel's findings are not exclusively based on these 10 case studies, the evidence collected for them played an important role in informing its conclusions.

Research for the case studies began with a comprehensive literature and document review to gather basic information about research assessment activities at the field level The Panel then carried out a series of interviews with representatives from research funding agencies and other relevant organizations. The interviews were used both to validate findings that had emerged from the literature and to build a more nuanced understanding of recent developments in research assessment. The Panel also created a short, online questionnaire to gather data on specific types of indicator use in research funding agencies.

The Panel's research focused on national research assessment practices (i.e., those carried out by a central government agency with a national scope) conducted *at the level of nationally aggregated research fields*. Other types of research assessment exercises (i.e., those targeted at individual researchers or institutions or those focused on evaluation of certain government R&D funding programs) are discussed where relevant. As a general rule, however, the Panel focused only on those initiatives involving assessment of research fields or disciplines, or those that could be used to support assessment at that level. (The national research assessment practices of the United Kingdom and Australia focus on institutional/group level assessments but can also be used to support conclusions about national performance at the field level).

The case studies are included in this appendix for reference purposes, and as valuable sources of information for those interested in looking into research assessment practices of these 10 countries further. Each case includes a description of the national research funding context and national priorities, relevant research funding allocation processes, practices used for evaluation of research disciplines and any related indicators, and the Panel's general observations or lessons learned . While the Panel's main conclusions from this evidence are presented and discussed in the body of the report, the case studies provide additional, country-specific insights.

2 Australia

Australia — Key Statistics

- Gross Expenditures on R&D (GERD), 2007 (PPP): $15.3 billion
- GERD as a % of GDP: 2.06%
- Total Researchers (Full-Time Equivalent, 2007): 87,140
- # of Researchers per Million People (2007): 4,224
- # of Scientific Publications (2008): 28,313

Source: UNESCO Science Report, 2010

2.1 NATIONAL RESEARCH FUNDING CONTEXT

Australia has a dual-support research funding system and provides funding both to institutions and projects. Institutional funding is delivered as block grants from the Department of Industry, Innovation, Science, Research and Tertiary Education (DIISR). In 2012 total funding distributed through these grants will equal A$1.63 billion (DIISR, 2011a). Several of these schemes are "performance-based" including the Joint Research Engagement (JRE) grants[1] and the Research Infrastructure Block Grants (RIBG). The objective of the JRE is "to maintain and strengthen Australia's knowledge base and research capabilities by developing an effective research and research training system in the higher education sector." Project-based research funding is provided by the Australia Research Council (ARC) and distributed through individual grants evaluated on the basis of peer review. The most significant ARC funding program is the National Competitive Grants Program (NCGP), which includes support for the social sciences, arts, and humanities as well as the natural sciences and engineering. ARC now allocates approximately A$680.4 million in research funding annually (ARC, 2010a).

2.2 NATIONAL RESEARCH PRIORITIES

The government of Australia has identified four thematic national research priorities that currently guide public research funding (DEST, n.d.; ARC, 2011a). These priorities are focused on key national challenges:

- "An environmentally sustainable Australia"
- "Promoting and maintaining good health"

1 The JRE program replaced the Institutional Grants Scheme (IGS) in 2010.

- "Frontier Technologies for Building and Transforming Australian Industries"
- "Safeguarding Australia"

The government has also released a national innovation strategy that identifies seven National Innovation Priorities (Commonwealth of Australia, 2009).[2] These involve setting priorities and strengthening coordination; improving skills and expanding research capacity; increasing innovation in business, government, and the community sector; and boosting collaboration — domestic and international – across the system (Commonwealth of Australia, 2009).

2.3 RESEARCH FUNDING ALLOCATION

The research funding allocation process for block grants to universities differs from the allocation process for project-based research funding. DIISR operates six block grant schemes associated with university research (DIISR, 2011a). Each scheme uses a different allocation formula, and some of the formulas specify performance indicators. Of these, the JRE is perhaps the most significant. In its allocation formula, the JRE uses a performance index based on research income, student load, and research publications, with a safety net to prevent a drop below 95 per cent of the previous year's funding level. Resource allocation for project-based funding through ARC, in contrast, is based primarily on peer review, with different funding programs focusing on discovery research and linkage (innovation, applied research, and university partnerships). ARC funds a selection of research centres and maintains a Special Research Initiatives program, designed to fund collaborative research activities and research in response to unforeseen opportunities (ARC, 2011a). It is also responsible for administering Australia's new national research evaluation system, Excellence in Research for Australia (ERA). ERA is not currently tied to specific funding allocation processes or decisions (see below).

ARC undertakes an initial allocation of research funding at the executive committee level across high-level discipline groupings. In the first step of this process, approximately 80 per cent of the funding is divided among disciplines based on two factors: the demand for research funding in each discipline (expressed as amount requested in current grant applications); and historical funding levels, which provide information on the relative research costs across disciplines. In the second step, the executive committee negotiates allocation of the remaining funding to proposals ranked by expert committees. This is not a formulaic process; it takes into account general consideration of overall level of quality of proposals

2 According to interviews with ARC personnel, these national priorities have little impact on actual funding allocation in the agency due to their breadth and generality.

in each cluster, demonstrated need for funding, and possibly other policy priorities or considerations (e.g., over- or underproduction of highly qualified personnel (HQP) by field). (Margaret Sheil, personal communication, June 15, 2011).

2.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

Australia has a substantial history of activity in national research evaluation. Quantitative indicators have been used in allocation of performance-based institutional research funding for nearly 20 years (OECD, 2010). Concerns about funding allocation based on simple measures of publication counts led to calls for a more nuanced approach to research evaluation.[3] In 2004 the government responded with the development of the Australian Research Quality Framework (RQF) initiative, which was intended to be a comprehensive, panel-based (i.e., peer review) process for evaluating research performance in Australian universities. A change in government in 2007 led to abandonment of the process, with the incoming government claiming it to be "poorly designed, administratively expensive, and relies on an 'impact' measure that is unverifiable and ill-defined." As a result, the RQF process was never implemented.

As an alternative, the government has developed the ARC-administered ERA initiative, which evaluates research performance at the institution/research group level across the country. The process is founded on expert review, but makes use of a selection of quantitative indicators as well as peer review of selected research outputs. ERA does not evaluate research impact, as was planned for the now defunct RQF; rather, its goals are to:

1. Establish an evaluation framework that gives government, industry, business and the wider community assurance of the excellence of research conducted in Australia's higher education institutions.
2. Provide a national stocktake of discipline-level areas of research strength and areas where there is opportunity for development in Australia's higher education institutions.
3. Identify excellence across the full spectrum of research performance.

3 See Donovan's (2008) discussion of the RQF, part of the movement towards richer, more nuanced qualitative approaches to evaluation aiming to gauge the wider social and economic benefits of research. The RQF was developed in response to the academic community in Australia advocating for the government to allocate university block funding on the basis of discipline-based peer review of research quality rather than the existing metrics-based formula (DEST, 2004 cited in Donovan, 2008).

4. Identify emerging research areas and opportunities for further development.
5. Allow for comparisons of Australia's research nationally and internationally for all discipline areas.

(ARC, 2011b).

A trial of the ERA process was carried out in 2009 for selected disciplines, and the first full assessment was completed in 2010; (the results, released in January of 2011, can be found at http://www.arc.gov.au/era/era_2010/outcomes_2010.htm). Although the ERA is not currently tied to research funding allocation, the government has announced its intention to include ERA results in the distribution of block grant funding to universities in the future, though specific funding formulas have not yet been released. While the ERA is targeted at evaluation of research fields (units) within institutions, assessment results can potentially be aggregated to provide a comparative evaluation of national research performance across disciplines.

2.5 USE OF SCIENCE INDICATORS

Beginning in 1990, block grants to universities have been allocated partially on the basis of student numbers and research funding awards (Geuna & Martin, 2003). The research portion of this component was calculated from the "Research Quantum," a value initially based solely on research funding. In 1995 the formula for that component was redefined as a Composite Index, comprising measures of research input and of research output (e.g., publication counts) (Geuna & Martin, 2003). This system of performance-based allocation has gone through various iterations over the years. Currently, DIISR administers six institutional research block grant funding schemes, with funding formulas based, to differing degrees, on quantitative indicators (see DIISR, 2011a). The general types of indicators used by DIISR for these schemes are:

- research student load
- research student total completions
- research income
- research publications
- previous program payments

(DIISR, 2011a)

Universities are required to submit data on these indicators to DIISR (data on students are collected and managed separately by the Department of Education, Employment, and Workplace Relations or DEEWR). Perhaps the most significant of these block grant schemes is the JRE. JRE funding is currently distributed to universities based on the JRE performance index, which is based on research income (60 per cent), student load (30 per cent), and research publications (10 per cent) (DIISR, 2011b). Although schemes such as the JRE include measures of publication output or volume in their allocation formulas, there is no attempt to include any assessment of research quality or impact in these allocations.

While the RQF initiative was never implemented, the government commissioned a substantial amount of work during its development, which may provide useful insights into the methodologies considered. The evaluation process was to be based on an expert review model with 13 assessment panels organized by cluster of related disciplines. Each panel was to conduct both an assessment of research quality (based on peer review of selected research outputs submitted for evaluation) and an evaluation of research impact (Donavan, 2008). Evaluation of research impact was intended to include consideration of social, economic, environmental, and cultural dimensions as assessed by evidence statements submitted from participating research groups and universities, with a final assessment provided based on a five-point scale.

The current ERA process is based on expert review, though it involves significant use of quantitative indicators to inform expert judgment. Eight research evaluation committees (RECs), organized thematically by research cluster, were convened to review research performance by field and sub-field across Australian research institutions.[4] The evaluation committees used four broad categories of indicators in their assessments:

- Indicators of research quality: Research quality was considered on the basis of ranked outlets, citation analyses, ERA peer review, and peer-reviewed Australian and international research income.

4 ERA definitions of research fields are based on the two-digit and four-digit fields of research as defined in the Australian and New Zealand Standard Research Classification system. A complete list of these fields is provided in the *ERA 2010 National Report* on ARC's website at http://www.arc.gov.au/era/era_2010/outcomes_2010.htm

- Indicators of research volume and activity: Research volume and activity was considered on the basis of total research outputs, research income, and other research items within the context of the profile of eligible researchers.
- Indicators of research application: Research application was considered on the basis of research commercialization income and other applied measures.
- Indicators of recognition: Research recognition was considered on the basis of a range of esteem measures.

(ARC, 2011b; also ARC, 2008a, 2008b).[5]

The ERA methodology applies different indicators and assessment processes by discipline, depending on the extent to which submitted outputs are included in indexed peer-reviewed journals. In disciplines where the majority of outputs are not indexed journal articles, a separate, peer-review process (referred to as "ERA peer review") was used to evaluate selected research outputs (primarily for fields in the arts and humanities and social sciences).[6] Disciplines with a mix of different outputs (i.e., in the range of 50 per cent coverage in indexed journals) were given a choice as to whether their assessment should be based on bibliometric analysis or peer review. In all cases, disciplines opted for citation-based analysis. With respect to citation-based indicators, a number of benchmarks were developed to guide evaluation committees, based on such indicators as relative citation impact (normalized by field) compared against Australian institutions and world averages and distribution of articles based on world centile thresholds. (For detailed information on the citation benchmarking methodology, see ARC, 2010b.)

ARC also went to significant lengths to develop a ranked journal outlet system for the ERA. This system categorized approximately 2,000 journals across all fields into four tiers (A*, A, B, C) based on the perceived quality of the journal, ostensibly providing a way to assess research quality in submitted journal articles that was not based on citations. The ranked journal outlet system was used in the ERA 2010 but was abandoned by the government in Summer 2011. According to press reports, this system was eliminated primarily out of concerns that ranked journal outlets were being abused by research managers in the university sector, rather than concerns about how the system was actually used in the ERA (Rowbothom, 2011).

5 The full list of indicators used in the ERA, as well as their definitions, can be found in the *ERA 2010 National Report* on ARC's website at http://www.arc.gov.au/era/era_2010/outcomes_2010.htm

6 A list of which ERA indicators were applied to which research fields and sub-fields is contained in the ERA Discipline Matrix, available at http://www.arc.gov.au/era/era_2012/key_documents_2012.htm

The ERA also employed a "low-volume threshold," intended to ensure that research units and institutions with insufficiently large research output were not assessed. For disciplines using citation-based indicators, this cut-off was defined as at least 50 apportioned indexed journal articles (ARC, 2011b). For disciplines relying on peer review, the cut-off was set at less than 30 apportioned research outputs over a six-year period. In practice, these cut-offs resulted in a number of institutions not being assessed in the full set of research fields (see the ERA 2010 results in ARC, 2011b).

Final assessments for each field of research were provided by the respective committees based on the "ERA rating scale." This five-point rating scale was designed to assess research performance against the "world standard." For example, a rating of 5 indicates: "The Unit of Evaluation profile is characterized by evidence of outstanding performance well above world standard presented by the suite of indicators used for evaluation;" whereas a rating of 1 implies: "The Unit of Evaluation is characterized by evidence of performance well below world standard presented by the suite of indicators used for evaluation" (ARC, 2011b).

2.6 OBSERVATIONS AND LESSONS LEARNED

Australia has substantial depth of experience from which to draw fin research evaluation and use of indicators in research funding allocation. Strong evidence suggests that the introduction of performance-based allocations for research block grants in the early 1990s had a significant impact on Australia's output of research publications. Butler (2003) examined trends in output of journal articles from Australian universities in detail, and found that output rose significantly in the mid-1990s. The fact that this trend was widely shared across fields, and yet did not extend to research institutions outside of the higher education system, strongly suggests that the increase was in response to use of publication counts in block grant funding formulas (Butler, 2003; OECD, 2010). Butler, however, also found that, while output of articles increased over the period, the impact of Australian research as measured by average relative citations declined. This finding raised questions about "the wisdom of a policy that rewards quantity, with scant regard for quality" (Butler, 2003), and led to an increasing emphasis on incorporating assessment of research quality into Australia's national research evaluation and funding schemes in more recent years.

Other impacts of Australia's use of performance indicators for research funding have been noted as well. There is some evidence that use of publication counts in determining research funding has led to instances of game playing among researchers and institutions. For example, universities have made token efforts to include at least one international participant in conferences to be classified as "international conferences," and professors have formed journal editorial boards composed solely of former graduate students (OECD, 2010). Such examples, however, appear to be relatively rare. On a more positive note, the establishment of new data collection procedures and data repositories has been another result of the introduction of performance-based funding schemes. The need for consistent, high-quality data on students and publications has led to large investments in data infrastructure, which has been seen as an unexpected, though largely positive, consequence (OECD, 2010).

Concern that Australia's formula for distributing institutional block grants prioritized *quantity* of outputs over *quality* led directly to development of the RQF. Although never implemented, research and analysis commissioned in the development of RQF methodology may contain useful insights. For example, various expert advisory groups and working groups were formed to examine questions about what methodologies should be used by the RQF. Some of these findings may be germane to national research evaluation initiatives in other countries. For example, in considering evaluation of research impact (defined as economic, social, environmental, and cultural impacts of research), one working group concluded that quantitative indicators were underdeveloped and should not be used for determining impact evaluation ratings. Instead, it was recommended that research groups develop and submit text-based impact statements describing the impacts of their research, which would then be assessed based on a standardized rating scale. (See Donovan (2008) for a discussion of various methodological recommendations made by government advisory groups in development of the RQF.)

Finally, the ERA is now a potential source of insights for other countries in the process of developing similar initiatives. There are several areas where this initiative may already point to interesting lessons. First, in providing Australia's research disciplines with the ability to choose between citation-based and peer-review based indicators, the ERA provides a useful model for analyzing applicability of different types of indicators across disciplines. The ERA Discipline Matrix may therefore be an important piece of evidence in assessing relevance of various indicators by field.[7] Second, use of the ERA five-point scale has allowed for comparisons of perceived research quality across disciplines. Such comparisons have now become widespread, both at the national level and internally within

7 See Key 2012 Documents on the Australian Research Council (ARC) website for a copy of the Discipline Matrix at http://www.arc.gov.au/era/era_2012/key_documents_2012.htm

universities. Third, initial assessment results from the ERA appear to have been relatively widely accepted by the Australian research community. If this is the case, it provides an early vindication for Australia's new model of research assessment.

2.7 REFERENCES

ARC (Australian Research Council) (2008a). ERA Indicator Principles. Canberra, Australia: ARC.

ARC (Australian Research Council) (2008b). ERA Indicator Descriptors. Canberra, Australia: ARC.

ARC (Australian Research Council) (2010a). ARC Annual Report 2009–10. Canberra, Australia: Australian Research Council.

ARC (Australian Research Council) (2010b). ERA 2010: Citation Benchmark Methodology. Canberra, Australia: ARC.

ARC (Australian Research Council) (2011a). ARC Profile. Retrieved May 9, 2011, from http://www.arc.gov.au/about_arc/arc_profile.htm

ARC (Australian Research Council) (2011b). Excellence in Research for Australia 2010: National Report. Canberra, Australia: ARC.

ARC (Australian Research Council) (2011c). The Excellence in Research for Australia (ERA) Initiative. Retrieved May 9, 2011, from: http://www.arc.gov.au/era/default.htmButler, L. (2003). Explaining Australia's increased share of ISI publications -- The effects of a funding formula based on publication counts. *Research Policy, 32*(1), 143–155.

Commonwealth of Australia (2009). Powering Ideas — An Innovation Agenda for the 21st Century. Retrieved April 2012, from http://www.innovation.gov.au/Innovation/Policy/Documents/PoweringIdeas.pdf

DEST (Department of Science and Technology) (n.d.). The National Research Priorities and their Associated Priority Goals Retrieved May 9, 2011, from http://www.dest.gov.au/NR/rdonlyres/AF4621AA-9F10-4752-A26F-580EDFC644F2/2846/goals.pdf

DIISR (Department of Innovation Industry Science and Research) (2011a). Research Block Grants Retrieved May 2011, from http://www.innovation.gov.au/Research/ResearchBlockGrants/Pages/default.aspx

DIISR (Department of Innovation Industry Science and Research) (2011b). The Process for Determining Joint Research Engagement (JRE) Scheme Grant Amounts. Canberra, Australia: DIISR.

Donovan, C. (2008). The Australian Research Quality Framework: A live experiment in capturing social, economic, environmental, and cultural returns of publicly funded research *Reforming Evaluation of Research: New Directions for Evaluation, 2008*(118), 47–60.

Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva, 41*(4), 277–304.

OECD (Organization for Economic Co-operation and Development) (2010a). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, Paris, France.

Research), D. D. o. I. I. S. a. (2009). Powering Ideas: An Innovation Agenda for the 21st Century (pp. 68). Canberra, Australia: DIISR.

Rowbotham, J. (2011, May 30). End of an ERA: Journal Rankings Dropped, *The Australian* Retrieved May 2012, from http://www.theaustralian.com.au/higher-education/end-of-an-era-journal-rankings-dropped/story-e6frgcjx-1226065864847

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

3 China

China — Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP):	$102.4 billion
• GERD as a % of GDP:	1.44%
• Total Researchers (Full-Time Equivalent, 2007):	1,423,380
• # of Researchers per Million People (2007):	1,071
• # of Scientific Publications (2008):	104,968
• World Share of Exports in High-Technology Products (2007):	18%

Source: UNESCO Science Report, 2010

3.1 NATIONAL RESEARCH FUNDING CONTEXT

Many government departments and agencies fund research and development activities in China, including central, regional, and local governments. With respect to discovery research, China's institutional setup involves a number of important actors including the Ministry of Science and Technology (MOST), the National Natural Science Foundation of China (NNSFC), the National Development and Reform Commission (NDRC), the Ministry of Finance (MOF), and the Chinese Academy of Sciences (CAS) (OECD, 2008).

MOST is responsible for designing and implementing China's S&T and innovation policies, and for administering many national R&D funding programs such as the National Programme for Key Technology R&D, the National Programme for High-tech R&D (also known as the 863 Programme), and the National Programme for Key Basic R&D (also known as the 973 Programme).[8] Of these, the 973 Programme was created explicitly to fund early-stage discovery research.

The NNSFC is the primary funder of project-based discovery research in China, funding research through individual grant applications analogous to research councils in other countries. The NNSFC funds research through a variety of funding programs (e.g., General Program, Key Program, Major Program).

8 See OECD (2008) for full profiles and descriptions of these programs.

CAS is the government's primary vehicle for funding intramural R&D, and administers over 100 government-run research institutes and laboratory facilities. The most significant funding program at CAS is the Knowledge Innovation Program (KIP), which was created in 1998 with the following objectives: to build CAS into a leading academic institution and comprehensive R&D centre; and to turn it into a scientific research base of international standing, an incubator of talented S&T personnel, and a springboard for development of China's high-tech industries (OECD, 2008). CAS is also a major beneficiary of the government's other R&D funding programs (in 2002 CAS received approximately 20 per cent of the NNSFC's total funding) (OECD, 2008). The 2011 budget allocated approximately $29.6 billion in funding for all S&T, of which $3.9 billion was earmarked for support of discovery research. The largest share of that amount (46 per cent) will go to research funding through the NNSFC (Stone, 2011).

3.2 NATIONAL RESEARCH PRIORITIES

Research priorities in China have been established in a series of national long-term S&T plans. The latest, adopted in 2006, is the Medium to Long Term Plan for the Development of Science and Technology (MLP). The development of the MLP involved more than 2,000 scientists, engineers, and corporate executives who were mobilized into a program of "strategic research" to identify critical opportunities in 20 areas considered to be of central importance to China's future (Cao *et al.*, 2006). This plan, intended to guide the country's development until 2020, establishes a comprehensive set of national S&T priorities including 68 national S&T goals; 11 key areas of research important to the economy (e.g., energy, agriculture): 16 special research projects (e.g., large-scale integrated circuit manufacturing technologies, manned space flights); 8 cutting-edge technology areas (e.g., biotech, advanced manufacturing); 8 science challenges (e.g., cognitive science, deep structure of matter); and 4 major new research programs (protein research, nanoscience, growth and reproduction, and quantum modulation research) (Wilsdon & Keeley, 2007; Cao *et al.*, 2006). The MLP also includes a national commitment to raise national R&D investment to 2.5 per cent of GDP by 2020, and to become one of the world's top five countries in terms of patents and citations.

3.3 RESEARCH FUNDING ALLOCATION

All national R&D funding is guided by the overarching MLP. In some cases, science megaprojects are explicitly identified in this plan to be funded by MOST or other participating government agencies or research facilities. While the bulk of funding for R&D programs comes from the central government, many programs also receive support from local governments, commercial banks, and private companies (OECD, 2008). The NNSFC receives funding directly with annual budget appropriations from the Ministry of Finance. The majority of the NNSFC's research funding is then allocated through traditional, grant-based application processes based on peer review (OECD, 2008). Funding programs themselves are created by the central administration in the NNSFC, presumably guided by national R&D priorities such as those identified in the MLP. The grant evaluation process is based on peer review and uses a panel system, composed of 92 disciplinary panels as well as external experts and referees (personal communication, Yonghe Zheng, August 31, 2011).

The largest program at the NNSFC is "General Programs," which receives about half of the NNSFC budget. Every year, its budget is allocated by the Bureau of Planning among field-based departments: Mathematical and Physical Sciences, Chemical Science, Life Sciences, Earth Sciences, Engineering and Materials Sciences, Information Sciences, and Management Sciences. In 2009 the Life Sciences department was divided to provide a separate department for Health Sciences. The allocation process is designed to create a stable funding environment, with only minor reallocations year to year, and is based on four factors: funding levels by department from the previous year; a prediction of the current year's application demand by department; deliberations of the decision-makers, taking into account policy considerations (about two per cent of the budget); and the "health of interdisciplinary research fund." Policy considerations can include national emergency needs such as research in response to natural disasters or infectious outbreaks, or national research priorities. The director of the Bureau of Planning consults with leaders of the field-based departments and eminent researchers to get a sense of their needs. The Bureau of Planning also conducts some policy analysis and reports to the president of the NNSFC. Final approval on budget allocation is obtained at the meeting with the president.

3.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

There is no known system of national research evaluation by research discipline or field in China. Government R&D programs are evaluated by the National Centre for Science and Technology Evaluation (NCSTE), a government agency created in 1997 specifically to evaluate government R&D funding programs. The NCSTE has now undertaken evaluations of several major R&D funding programs, including the 873 Programme and the Knowledge Innovation Program at CAS. The NCSTE is composed of roughly 25 staff specialized in management consulting, public policy research, technology-economy analysis, and system engineering (OECD, 2008).

The NCSTE has also formulated China's S&T Evaluation Standards — a national set of standards to govern S&T/R&D program evaluation. This document provides standardized definitions of key evaluation concepts and procedures as well as recommendations for methodologies and guidelines covering relationships between evaluators and the organizations they are evaluating. These standards are now also the basis for training in S&T evaluation (OECD, 2008).

The NNSFC undertakes both *ex ante* and *ex post* evaluations relating to its funding programs. Again, initial project proposals are evaluated under a peer review system. Each proposal is initially evaluated by a group of experts or referees that pre-selects around 30 per cent of the proposals. Its evaluations are then passed on to evaluation panels of 8 to15 individuals that make the final selection for that discipline. Mid-term and *ex post* evaluations of funded projects are also carried out with grant recipients reporting progress using standard indicators (e.g., patents) to demonstrate project impacts and successes (OECD, 2008). The NNSFC is also involved in evaluation of government research facilities, such as those of the CAS, and uses expert reviews as well as *in situ* evaluations to assess these facilities and provide recommendations on their management and administration (OECD, 2008). The NNSFC has been considering undertaking field-based evaluations in the future. The main goal of the evaluations would be to find ways to improve management of funding processes (such as how best to fund basic sciences as compared to the more applied sciences such as engineering) rather than to inform budget allocation across fields (Yonghe Zheng, personal communication, July 20, 2011).

3.5 USE OF SCIENCE INDICATORS

While the government does not appear to undertake systematic evaluations of national research performance by discipline or field, several instances of indicator use in China are worth highlighting.

First, the government has enshrined certain indicators as national S&T priorities, such as increasing national R&D investment equal to 2.5 per cent of GDP by 2020 and being one of the world's top five countries in terms of patent applications and journal article citations. MOST is responsible for generally collecting and reporting S&T-related data, and publishes an annual *S&T Statistics Data Book*, which includes basic statistics on R&D funding, research outputs, HQP, etc.[9]

9 See the 2007 China *Science & Technology Statistics Data Book* on the Ministry of Science and Technology of the People's Republic of China's website at http://www.most.gov.cn/eng/statistics/2007/200801/P020080109573867344872.pdf

Second, with respect to evaluation of government programs, the NCSTE has developed standardized approaches and indicators for use in its evaluations. According to the OECD (OECD, 2008), these consist of:

- a five-step generic evaluation procedure, including *ex ante*, mid-term and *ex post* evaluations;
- three main evaluation techniques, used systematically and preferably in combination:
  - Peer review: standard "purely scientific" peer reviews always used for *ex ante* evaluation; "mixed" peer review including socio-economic impact assessment associating economists, finance experts, marketing managers and future users, sometimes used for both *ex ante* and *ex post* evaluation;
  - Case studies: mostly used in *ex post* evaluation;
  - Performance indicators: generally in the form of multi-criteria "scoring-type" approaches or cost-benefit analysis; however, there seems to be a tendency to focus more on inputs and involvement than on outputs;
- Data collection through questionnaires and interviews and workshops for discussing results;
- Systematic surveys of project participants.

Third, the Knowledge Innovation Program at CAS is evaluated under a set of evaluation procedures and guidelines. Originally this system focused on quantity of research outputs; however, since 2002 there has been an increased emphasis on quality. The features of this system include self-evaluations by researchers; peer review evaluation system adapted to discovery versus applied research; coordinated evaluation of the research institute (each institute receives an overall mark such as excellent, good, pass, etc.); cross-disciplinary evaluations (e.g., a mathematics institute evaluates a physics institute); evaluations oriented towards supporting decision-making; and prizes for outstanding research achievements (OECD, 2008).

Finally, universities and research institutes in China often have an explicit performance assessment and reward system based on publication output. Universities frequently provide financial bonuses to scientists and academics based on journal publications, with larger rewards typically provided for publications in more prominent (or more highly cited) journals, or journals included in Thomson Reuters' *Science Citation Index* (Shao & Shen, 2011; Qiu, 2010). These incentive structures are implemented at the discretion of individual universities, but are motivated by emphasis on these performance indicators originating in the central government. In many cases, the associated financial rewards or bonuses are scaled to the impact factor of the journal published. For example, Table 1 outlines the publication bonus system for Zhejiang University. Incentive systems of this type may vary across universities depending on the specifics of the policy. In general, however, they have resulted in a strong institutional emphasis on measures of publication output and journal impact factors across China.

Universities and research institutes often have an explicit performance assessment and reward system based on publication output. The NNSCF does not employ indicators in this manner for any of its funding decisions (Yonghe Zheng, personal communication, July 20, 2011).

Table 1

Financial Bonuses for Faculty Publications at Zhejiang University

Journal Classification	Monetary Reward (RMB)	C$***
Nature or *Science* (first author); decreased by 50% according to the sequence of authors	200,000	29,245
Journals in the *Science Citation Index* (first author)		
IF* ≥ 10	14,000	2,047
5 ≤ IF < 10	5,000	731
3 ≤ IF < 5	4,000	585
1 ≤ IF < 3	3,000	439
IF < 1	2,000	292
**EI journals (first author)	1,800	263
ISTP (first author)	600	88

* IF=Impact Factor; **EI= Elsevier Index; ***C$ figures calculated based on the exchange rate for May 2, 2011.
Source: Shao & Shen, 2011

3.6 OBSERVATIONS AND LESSONS LEARNED

China's experiences with S&T evaluation differ from many other nations. There is no systematic evaluation of research disciplines at the national level, and therefore no relevant body of experience with indicators or assessment of this type. One of the prominent conclusions of an extensive review of the Chinese innovation system conducted by the Organisation for Economic Co-operation and Development (OECD) is that China's national evaluation system for R&D/S&T remains underdeveloped (OECD, 2008). No professional evaluation of government investment in research and development was undertaken before 1994, and the practice of evaluation is still not fully integrated into China's research funding activities. The need for more fully developed, systematic, and transparent S&T evaluation activities at the national level is repeatedly emphasized in this review (OECD, 2008). Due to the relatively recent adoption of formal S&T evaluation practices, Chinese experience in this regard may hold limited lessons for other countries.

In one respect, however, China does have a significant source of experience with science performance indicators. The institutionalized practice of providing scientists and researchers with financial bonuses based on publications has had significant ramifications in China. On the one hand, China's record of scientific publications has shown tremendous growth in the past decade. This growth cannot be entirely attributed to university bonus policies, but these types of incentives no doubt played a role in catalyzing research productivity.

There is also evidence to suggest that these policies have resulted in a range of unintended behavioural changes among Chinese researchers. Studies suggest that a substantial market for ghost-writing papers on nonexistent research has emerged, with illicit websites providing services such as fictional research papers, bypassing peer review for payment, and forging copies of legitimate Chinese or international journals (Qiu, 2010). Anonymous surveys have found that as many as one in three researchers working in Chinese universities admit to having plagiarized, falsified, or fabricated data (Qiu, 2010). Discoveries of falsified data or research have led to several high-profile retractions from journals in recent years, as well as calls for policy reform in the higher education sector (Qiu, 2010; Cyranoski, 2010).

A separate issue, identified by Shao and Shen (2011), is that policies that incentivize publication in highly cited, international journals may be inhibiting development of Chinese scientific journals. This may bias researchers against working on regionally specific issues and/or publishing in Chinese language journals. The Chinese experience with publication-based bonuses may therefore argue for caution when considering a direct linkage between financial incentives for researchers and indicators based on research outputs.

Finally, China's recent record of S&T planning, such as that undertaken for the MLP, may hold significant lessons for development of national research priorities. The result is a comprehensive national strategy document intended to guide Chinese investment in R&D in the coming years, a period in which the level of research investment is expected to continue to grow dramatically (see Cao *et al.*, 2006; Wilsdon & Keeley, 2007). The process of developing the MLP may hold significant lessons for other countries in their own research priority-setting mechanisms, though application of these lessons may be complicated by both differences in scale and in the nature of the political process.

3.7 REFERENCES

Cao, C., Suttmeier, R. P., & Simon, D. F. (2006). China's 15-year science and technology plan. *Physics Today, 59*(12), 38–43 Retrieved May 2012, from http://levin.suny.edu/pdf/physics%20Today-2006.pdf

Cyranoski, D. (2010). China debates university reform. *Nature, 464*(7287), 336–337 Retrieved May 2012, from http://www.nature.com/news/2010/100316/full/464336a.html

MOST (Ministry of Science and Technology for China) (n.d.). S&T Statistics Retrieved March, 2012, from http://www.most.gov.cn/eng/statistics/2007/index.htm

OECD (Organization for Economic Co-operation and Development) (2008). OECD Reviews of Innovation Policy: China. Paris, France: Organization for Economic Co-operation and Development.

Qiu, J. (2010). Publish or perish in China. *Nature, 463*, 142-143Retrieved May 2012, from http://www.nature.com/news/2010/100112/full/463142a.html

Shao, J., & Shen, H. (2011). The outflow of academic papers from China: Why is it happening and can it be stemmed? *Learned Publishing, 24*(2), 95–97.

Stone, R. (2011). China bets big on small grants, large facilities. *Science, 331*(6022), 1251.

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

Wilsdon, J., & Keeley, J. (2007). China: The Next Science Superpower? In C. Leadbeater & J. Wilsdon (Eds.), *The Atlas of Ideas*. London, UK: Demos Retrieved May 2012, from http://www.demos.co.uk/files/China_Final.pdf

4 Finland

Finland — Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP):	$6.7 billion
• GERD as a % of GDP:	3.47%
• Total Researchers (Full-Time Equivalent, 2007):	40,879
• # of Researchers per Million People (2007):	7,707
• # of Scientific Publications (2008):	8,328

Source: UNESCO Science Report, 2010

4.1 NATIONAL RESEARCH FUNDING CONTEXT

In 2009, 82 per cent of funding for universities in Finland came from public sources. Base university funding accounted for 50 per cent (about €547 million), while the remaining 32 per cent came from external public funding sources (about €343 million) (Statistics Finland, 2009). The primary contributor from external public funds was the Academy of Finland, operating under the Ministry of Education and Culture, which provided 16 per cent of total university funding or €172 million. Another major contributor, the Finnish Funding Agency for Technology and Innovation (TEKES), under the Ministry of Employment and the Economy, funds innovation R&D projects in companies, universities, and research institutes. In 2007 this organization provided 9 per cent of total university funding (€99 million).

The current structure for research funding in Finland encourages applied research. The government's major policy documents have placed scientific research primarily in a technological and economic context. In 2008 TEKES received about twice as much government R&D funding as the Academy of Finland (Ministry of Employment and the Economy, 2009). A recent review of the state of science and research in Finland, however, recommended increasing support of discovery research (Academy of Finland, 2010a).

4.2 NATIONAL RESEARCH PRIORITIES

In 2006 Finland's Research and Innovation Policy Council approved the development of six new public-private partnerships (Strategic Centres for Science, Technology and Innovation) in areas that are important for the Finnish economy. TEKES is the main funder of the centres; they focus on energy and the environment, metal products and mechanical engineering, the forest cluster, information and communication industry and services, built environment innovations, and health and well-being[10] (TEKES, 2011).

The Finnish government launched the current Innovation Strategy in 2008 (OECD, 2010a); it includes measures to encourage innovation in non-technological business areas such as the services sector, but does not specify any priority areas for fields of scientific research. The Academy of Finland supports research mainly through competitive project-based funding, but also funds thematic research programs (e.g., climate change), individual research posts, funding for foreign professor-level researchers invited to work in Finland, and Centres of Excellence (Academy of Finland, 2010a). While quality is the primary criterion for project funding, decisions may favour research that serves the centres. Although researchers involved in the centres are eligible for the Academy's competitive project grants, no designated budget is earmarked for the centres.

4.3 RESEARCH FUNDING ALLOCATION

The Ministry of Education uses formula-based allocation for core university funding (see Table 1). This method, first implemented in 1998, was revised in 2010 (OECD, 2010b) partly as a result of recommendations and opinions from national- and discipline-level evaluations (see Discipline Assessment and Evaluation). The current formula incorporates a performance-based allocation for education, research, and researcher education, and also takes into account education and science policy objectives (Table 1). Once received, allocation of funds within universities is at the discretion of the institution but remains aligned with the outcomes agreed upon with the Ministry of Education (OECD, 2010b). Further reform of the core funding system is expected in 2012, following a recent international review of the Finnish innovation system and further recommendations from discipline evaluations.

10 More detail on TEKES and a complete list of its key businesses and research areas can be found at http://www.tekes.fi/en/community/Innovation_environment/356/Innovation_environment/1254

Table 1

Finnish core funding for universities

Funding Component		% of Total	Basis of Allocation
Formula-based core funding	Research & researcher education	45	Extent of activities (75%)
			Quality and effectiveness (25%)
	Education	41.3	Extent of activities (85%)
			Quality and effectiveness (15%)
Other education and science policy considerations		25.0	Strategic development (25%)
			Education and discipline structure (75%)

Source: OECD, 2010c

The Academy of Finland decides on allocation of funding among its four research councils. Factors taken into account in these decisions include number of researchers, broadness of the field (number of disciplines under the Research Council), importance of the field (needs of society), and cost of research including infrastructure and salary needs (Sirpa Nuotio, personal communication, May 12, 2011; Academy of Finland, 2009). The Research Councils are highly independent once they receive funding from the Board. There is no predetermined allocation among fields; instead, funding is allocated on a competitive basis with scientific quality the most important criterion. Finland recognizes that, as a small country, it needs to develop specialized areas of strength to maintain an international competitive advantage. As such, it does not attempt to provide funding equally to all disciplines, and there is no guarantee that all fields of research will be funded every year (Sirpa Nuotio, personnel communication, May 12, 2011). Cost of research is dealt with by varying grant sizes by field in response to need. Once the highest quality proposals have been identified, the Research Councils consider several other factors. Individual proposals are considered in terms of relevance to society and innovativeness,[11] among other factors (Academy of Finland, 2008b). Targeted project calls prioritize specific fields of research, such as those relevant to the Strategic Centres for Science, Technology and Innovation (Academy of Finland, 2008a). In addition, decisions are informed by science policy priorities of the Finnish government, guidelines and provisions prepared by the Board,

10 A recent Academy report found that one-fifth of funded projects could be identified as breakthrough, and the success rate of breakthrough proposals is higher than average (Häyrynen, 2007).

plans of action of the individual Research Councils, and expert recommendations from evaluations of individual fields (see Discipline Assessment and Evaluation). For example, the Research Council for the natural sciences and engineering has opened targeted calls for proposals for energy efficiency (2010), mechanical engineering (2009), and embedded systems in computer science (2008), partly as a result of comprehensive evaluations of those fields (Sirpa Nuotio, personal communication, May 12, 2011).

4.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

The Academy of Finland dedicates a relatively high level of effort to evaluations of the research system, and conducts two types of field-level assessments. The first is a broad evaluation of the state and quality of the Finnish research system, conducted every three years, in which each field of research is subject to its own short evaluation. The second type is a field-specific assessment, conducted upon request and narrowly focused on a specified field of research in more depth. The main purposes of the evaluations are to provide feedback to the scientific community and policy-makers, and to track research trends by identifying emerging areas of interest and needs for future development (Sirpa Nuotio, personal communication, March 18, 2011). Neither of the two evaluation types systematically dictates funding decisions, but recommendations directed towards government ministries or the Academy may have indirect effects, such as guiding decisions for new Academy thematic research programs, targeted project calls (see examples given above), doctoral research programs, or research fellow positions (Sirpa Nuotio, personal communication, May 12, 2011). Foresighting activities are also carried out to identify areas of competence for the future.

Reviews of the state and quality of scientific research in Finland: As part of these national reviews, each Research Council provides an overview of the current state of science and research in each of its respective fields of expertise. Since there are four Research Councils, each with about 10 to 15 fields, about 40 to 60 fields are reviewed in this process. The analyses are informed by both expert opinion and bibliometrics (see Use of Indicators below). Previously, each Research Council approached the evaluations in a slightly different way, but a more systematic and consistent method that will apply to all fields is currently in development (Sirpa Nuotio, personnel communication, June 3, 2011). As it stands, in selecting projects for funding, the Academy of Finland applies five criteria: scientific quality and innovativeness of the research plan, competence of the applicant/research team, feasibility of the research plan, cooperation contracts for research, and significance of the research project for the promotion of professional careers in research and researcher training (Academy of Finland, 2010a).

Evaluation of disciplines and research fields: The Academy publishes on average two to three discipline-specific evaluations per year. These evaluations can be commissioned from within the Academy or requested by the research community or other science funding agencies and authorities (Academy of Finland, 2008b). The objective of these exercises is to evaluate the quality of the discipline and its sub-fields as compared with international standards (Academy of Finland, 2008c). The assessments are meant to provide feedback to scientists and research units, as well as ministries, funding agencies, and other stakeholders, to inspire discussion and debate, and to identify potential problem areas and areas of future potential. The evaluations also inform the national reviews described previously.

The evaluations are carried out by international panels of experts that consider the field as a whole together with its sub-fields, and also critically assess each research unit that carries out substantial and relevant research within the field (for example, the evaluation of chemistry research in Finland in 2005–2009 resulted in the evaluation of 41 research units in the country) (Academy of Finland, 2011). The expert panels are informed by both qualitative and quantitative data gathered from self-assessments of the relevant research units, as well as on-site visits/interviews with the units (Academy of Finland, 2008c, 2011). Examples of strategic insights from such discipline evaluations include strengthening coordination among research units in the field, allocating funding to fields that benefit most from the input, increasing the size of research units so that broad and complex research problems can be addressed and resolved, and encouraging greater mobility among researchers (Academy of Finland, 2008b).

4.5 USE OF SCIENCE INDICATORS

Institutional funding: Table 2 below presents a detailed breakdown of indicator weights used for the research portion[12] of the allocation of block funding to universities by the Ministry of Education (summary is presented in Table 1 above) (OECD, 2010c). Indicators of the extent of research activities (as opposed to quality) are given the highest weighting (75 per cent) and include teaching and research person-years and measures based on numbers of PhDs. Indicators of research quality account for the remaining 25 per cent, and include external research funding, publication counts, and a somewhat unique indicator called "internationalization of research" measured by the amount of international competition for research funding and the mobility of researchers and teachers.

12 Note that the research portion is 34 per cent of total university block funding.

Reviews of the state and quality of scientific research in Finland: Each evaluation canvasses expert opinion through workshops attended by leading researchers from the scientific community and, in some cases, industry and government representatives. The assessments are further informed by bibliometric analyses, and in general the indicators used include number and share of scientific publications, relative citation impact, highly cited publications, and international publishing cooperation (Academy of Finland, 2010a). Each field is benchmarked internationally using these indicators, which are meant to be proxies of scientific visibility, scientific impact, and leading edge research in the field being analyzed (Academy of Finland, 2010a). In the past the exact manner in which the field assessments

Table 2

Indicators of research and researcher education for core funding of universities

Funding Criteria		% of Total	Specific Indicators	% of Total
Extent of activities in research and researcher education (75%)		75	Teaching and research person-years (50%)	37.5
			Total no. of doctoral degrees determined in the agreement between the Ministry and the university (25%)	18.8
			No. of doctoral degrees completed at the university (25%)	18.8
Quality and effectiveness of research and researcher education (25%)	Nationally competed research funding (60)	15	Academy of Finland funding for the university (50%)	7.5
			Funding allocated to the university on basis of Academy's decisions on Centres of Excellence (30%)	4.5
			TEKES funding for the university (20%)	3.0
	Scientific publications (20%)	5	No. of refereed international publications (60%)	3.0
			No. of other scientific publications (40%)	2.0
	Internationalisation of research (20%)	5	Amount of internationally competed research funding (60%)	3.0
			Overall extent of teacher and researcher mobility (40%)	2.0
		100		100

Source: OECD, 2010c

were performed varied by Research Council. Methods are currently being developed to achieve evaluation consistency for all fields (Sirpa Nuotio, personal communication, June 3, 2011).

Evaluation of disciplines and research fields: For these evaluations, the Academy identifies relevant research units and invites them to fill out detailed self-assessment questionnaires that gather both qualitative and quantitative data. Quantitative data includes personnel resources (researchers, students, administrative personnel, technical staff, visiting researchers); funding (core and external); research output (number of types of scientific outputs, degrees earned); and national and international collaborations (Academy of Finland, 2008c, 2011). Qualitative information includes research strategies and activities; SWOT analysis (strengths, weaknesses, opportunities, threats); publications; self-assessment; infrastructure available; societal impact; and future prospects (Academy of Finland, 2008c, 2011). The Academy prepares a statistical report that summarizes and analyzes the quantitative data to aid the international expert panel in its judgments. These reports are complemented by on-site visits/interviews with research units, carried out over three or four days with each visit lasting about 1–1.5 hours (Academy of Finland, 2008c, 2011).

4.6 OBSERVATIONS AND LESSONS LEARNED

The Academy of Finland believes that although indicators are good support tools in evaluations, they cannot tell the whole truth on their own and must be used in combination with other methods such as expert judgement. In future evaluations, the aim is to exploit indicators even more systematically, and also to develop indicators that can measure research impact and trace linkages and dynamics between inputs, outputs, and impacts (Sirpa Nuotio, personal communication, March 18, 2011). In addition, the Academy of Finland acknowledges that future evaluations must further account for changing interaction between basic and applied research' the impact of funding for high-risk, breakthrough research; and the resource needs of emerging disciplines (Academy of Finland, 2008b).

As in many other European countries, the process of reform is ongoing in Finland. An international expert review panel , which assessed the state of the Finnish innovation system, including education, research, and the economy (Ministry of Employment and the Economy, 2009), identified the most important and pressing challenge as increasing the quality of Finnish research. The panel proposed to achieve this through basic institutional funding rules emphasizing quality. In light of the small size of Finland's research base, the panel stressed the need to specialize in areas of strength, even if it means closing down some activities. The panel supported bottom-up funding methods to build up these areas of strength, asserting that a top-down approach would be counterproductive. Once high-quality fields emerge, applied funding can be allocated to those fields that are promising in terms of commercial innovations.[13]

Finland's Ministry of Education has recently invited a working group to discuss institutional funding reform. The working group issued its own set of recommendations to the Ministry in Fall 2011. It remains too early in the process to draw any conclusions on the extent to which recommendations from the international review will be implemented. The working group was hesitant to give too much weight to quality because the Ministry must take into account political considerations and the issue of "fairness"; a balance needs to be struck between funding of high-quality research and provision of basic funding to universities in all regions. It is almost certain, however, that publication numbers will at least be adjusted based on journal quality, following the Norwegian model. The working group discussed a potential peer review component to the quality dimension, but cited concerns that this approach may be too complicated for Finland's relatively small university system (fewer than 10 universities conduct research in the NSE, and the Ministry is already aware of those capable of producing internationally competitive research, and those of importance at the national or regional levels). The U.K. model (which relies on peer review) may not be a relevant comparison, since U.K. polytechnics and universities compete for the same pool of funding, whereas in Finland the two are separated. Finland has studied other European institutional funding systems and a few outside of Europe, but has not found one model to be superior to any other.

Finally, there are both supporters and dissenters within universities for discipline-specific core funding allocations, and whether and how this would be implemented remains to be seen. A novel option debated by the working group is to provide a pool of around 10 indicators, and allow individual universities to choose which are best for their own strategic development. The goal of such a measure would be to diversify the system, creating different profiles for different institutions (whereas the current funding model steers all universities in the same direction). The working group is also discussing new types of indicators that would engage universities in increased collaboration and cooperation (Anita Lehikoinen, personal communication, June 15, 2011).

13 For a complete review please see *Evaluation of the Finnish National Innovation System – Full Report* from the Helsinki Ministry of Employment and the Economy, which can be found at http://www.tem.fi/files/24929/InnoEvalFi_FULL_Report_28_Oct_2009.pdf

4.7 REFERENCES

Academy of Finland (2008a). General Decision Criteria Retrieved May 2011, from http://www.aka.fi/en-GB/A/For-researchers/Decisions/General-decision-criteria/

Academy of Finland (2008b). Evaluations of disciplines and research fields Retrieved May 2012, from http://www.aka.fi/en-GB/A/Science-in-society/Evaluation-of-research/Evaluation-of-disciplines-and-research-fields-/

Academy of Finland (2008c). Mechanical Engineering Research in Finland 2000–2007: Evaluation Report. Helsinki, Finland: Academy of Finland.

Academy of Finland (2010a). The State and Quality of Scientific Research in Finland 2009. . Helsinki, Finland: Academy of Finland.

Academy of Finland (2011). Chemistry Research in Finland 2005–2009: Evaluation Report (pp. 137). Helskinki, Finland: Academy of Finland.

Häyrynen, M. (2007). Breakthrough Research: Funding for High-Risk Research at the Academy of Finland. Helsinki, Finland: Academy of Finland.

Ministry of Employment and the Economy for Finland (2009). Evaluation of the Finnish National Innovation System - Full Report. Helsinki, Finland: Ministry of Employment and the Economy Retrieved May 2012, from http://www.tem.fi/files/24929/InnoEvalFi_FULL_Report_28_Oct_2009.pdf

OECD (Organization for Economic Co-operation and Development) (2010a). Profile of Finland *Science, Technology and Industry Outlook 2010* (pp. 174–175). Paris, France: OECD.

OECD (Organization for Economic Co-operation and Development) (2010b). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, Paris, France.

OECD (Organization for Economic Co-operation and Development) (2010c). Web Annex: Additional Country Detail (Performance-based Funding for Public Research in Tertiary Education Institutions). Paris, France: OECD.

Statistics Finland (2009). Statistical Databases. Retrieved June 2011, from

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

5 Germany

Germany — Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP):	$72.2 billion
• GERD as a % of GDP:	2.54%
• Total Researchers (Full-Time Equivalent, 2007):	251,755
• # of Researchers per Million People (2007):	3,496
• # of Scientific Publications (2008):	76,368
• World Share of Exports in High-Technology Products (2007):	9.1%

Source: UNESCO Science Report, 2010

5.1 NATIONAL RESEARCH FUNDING CONTEXT

Germany's government consists of a federal parliament (the Bund) and 16 state governments (Länder), each with its own ministries of science, education, and economy. Government contributes roughly one-third of all spending on R&D in Germany (BMBF, 2011). There are three main performers of research: universities, government research organizations (the Bund and the Länder each have their own), and non-university public research organizations. The latter mainly consists of four associations each with a number of institutes: the Max Planck Society (focus is discovery research), the Fraunhofer Society (applied research), the Leibniz Association (problem-oriented research), and the Helmholtz Association (large infrastructure). These organizations carry out a large share of discovery research in Germany and are jointly financed by federal and state governments (European Commission, 2010). Around 70 per cent of national public funding to R&D performers is in the form of basic institutional funding (OECD, 2010a). The share of third-party competitive project funding for higher education institutions, however, rose from 18 per cent in 2000 to 21 per cent in 2006 (DFG, 2010).

The Länder, or state governments, are responsible for institutional funding of universities, which has resulted in a diversity of core funding procedures. The federal government contributes to university funding through large, long-term block grants to a limited number of universities based on competitive grants, as well as some thematic funding programs (European Commission, 2010). In 2006, 29 per cent of third-party funding for higher education institutions (HEIs) came from the German Research Foundation (DFG), an independent, self-governing organization funded jointly by the Bund and the Länder, with a total budget of €2.2 billion in 2009 (DFG, 2010). The DFG's main task is to fund discovery research in all fields, and its central form of funding is bottom-up through the Individual Grants Program (DFG, 2009). The Individual Grants Program appears comparable to NSERC's Discovery Grants Program, which also funds discovery research in all fields and receives about the same percentage of total NSERC funds. The DFG also funds coordinated (targeted) programs, a highly prestigious and influential Excellence Initiative, research infrastructure, prizes, and international cooperation.

5.2 NATIONAL RESEARCH PRIORITIES

Germany's first national strategy for S&T (its "High-Tech Strategy") was not developed until 2006. This strategy was necessary to coordinate R&D and innovation policy at the federal level, and its overall goal was to increase international competitiveness of Germany's knowledge-based economy (European Commission, 2010). The strategy has been recently updated with the High-Tech Strategy 2020, with an aim to make Germany a leader in solving global challenges (BMBF, 2010a). Five fields of action have been identified: health and nutrition, climate and energy, security, communication, and mobility. No programs at the DFG intentionally address the national priorities (Dr. Robert Paul Königs, personal communication, June 24, 2011); instead, these priorities are largely funded through the federal government's thematic funding programs (see examples in DFG, 2010). The strategy also aims to fund 11 cross-cutting "key technologies" relevant to progress in all fields, such as biotechnology, nanotechnology, and information technology. Individual states have their own research strategies as well. For example, the state of Berlin places an emphasis on three areas: health management; communication, media, and cultural industries; and mobility and transport (BMBF, 2010b).

5.3 RESEARCH FUNDING ALLOCATION

Most states include a formulaic performance-based component in their allocation of core funding to universities, but the specifics vary by state. The state of Berlin has the highest share (30 per cent) of core funding that is performance based, while for the other states it is at around 10 per cent (OECD, 2010a). Most states limit the maximum gain or loss possible from the performance-based portion by applying a safety net linked to the previous year's budget (OECD, 2010a). Germany's performance-based systems encompass teaching and research performance and often also include indicators of equality and/or internationalization (OECD, 2010b). Once funded, higher education institutions are free to distribute funds autonomously (OECD, 2010a).

The goal of the DFG's Individual Grants Program is to fund the highest quality proposals, irrespective of discipline or socio-economic impact. Proposals are first distributed among 203 subject areas, and reviewed by peers with expertise in a subject. Separate field-specific Review Boards then comparatively assess all proposals/reviews in several related subject areas (each of the 48 review boards is responsible for 1–28 subject areas). The Review Boards determine the structure of the review and evaluation of the proposals, which fall under their authority, and submit their funding recommendations to a Joint Committee, which makes the final funding decisions. Membership for the Joint Committee is drawn from the DFG's Senate as well as federal and state government representatives; the majority of the members (and votes) are active researchers drawn from the entire range of disciplines.

The Joint Committee, which meets six times per year, makes sure there are comparable levels of assessment among the Review Boards by discussing "borderline" proposals. These include proposals submitted by the Review Boards that do not meet quality standards, or high-quality proposals that Review Boards missed or underfunded, potentially because their value only becomes apparent in a DFG-wide context. In the few decades after the Second World War, the ratio of funding to proposals was higher, and all proposals competed against all other proposals. Since the late 1980s, due to the volume of proposals (the Individual Grants program currently receives about 10,000 proposals per year), the Review Boards are given preliminary budget limits. These preliminary allocations are based on a formula that gives equal weight to two factors: historical funding levels (over the past three years) and funding demand from the preceding year (amount of funding that was requested). The preliminary budget formula and figures are not published, but researchers are aware of their existence. There is an inherent flexibility in the budgeting process, including 5 to10 per cent "reserves" kept aside from the preliminary budgets. If there are significantly more good proposals that can be funded from the preliminary budget for a Review Board, additional funds can be allocated from other Review Boards or from the reserves.

To address the issue of emerging and interdisciplinary research, the DFG has a separate funding scheme, "Priority Programs," which generally nets eight per cent of the total DFG budget (about one-fifth the amount budgeted for the Individual Grants Program). The term "priority" in this context does not refer to national S&T priorities, but to scientific topics that show potential for novel research and interdisciplinary collaboration. Teams submit around 60–80 topic/project proposals each year (an example of a topic funded for 2011 is "Design and Generic Principles of Self-Healing Materials"), of which 15–16 are chosen by the DFG Senate based on the following criteria: novelty of the proposed project, clear objectives, collaboration potential, qualifications of the coordinator, measures to promote young researchers, international involvement, and placement within context of other funding activities. Again, no quantitative indicators are used in this peer review process. A nationwide call for individual proposals is then made, which funds about 20–30 teams of researchers to work on each of the 15–16 chosen topics for six years (Dr. Robert Paul Königs, personal communication, June 24, 2011).

5.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

The dominant form of assessment of research fields in Germany consists of subject-specific rankings and ratings designed to compare research performance of HEIs. Several organizations produce various types of ranking and ratings, including the Centre for Higher Education Development (CHE; an independent, non-profit think-tank committed to reforming the education system); the DFG; and, most recently, the German Council of Science and Humanities (WR; advises federal and state governments on structure and development of higher education and research, and also has some evaluation activities). The WR rankings "differ from evaluations in that they focus on the measuring and rating of outputs (rather than containing any recommendations for action or being process-oriented) and in that their purpose is to allow comparison…" (WR, 2004). The relevance of these rankings and ratings to the expert panels' charges may be limited in that there are no cross-subject comparisons, and the assessments do not inform funding decisions. Since the ratings and rankings measure research quality relying entirely or in part on quantitative indicators, several relevant lessons can be learned, especially as a result of studies undertaken by the WR (see Observations and Lessons Learned).

Two other field-specific evaluations are carried out in Germany: structural analyses (the WR) and research evaluations (Scientific Commission of Lower Saxony). In both cases indicators do not appear to play a major role. Since the 1990s foresight processes have been used increasingly in Germany (UNIDO, 2005) and influence budget priorities within the Federal Ministry of Education and Research (BMBF) (UNIDO, 2005). In the most recent foresight initiated by the federal government in 2007, experts came together to discuss emerging topics and future thematic developments for up to 10 years ahead (BMBF, 2010a).

The CHE and DFG rankings are based on quantitative indicators and/or reputation (surveys of opinion) (see Use of Indicators). The CHE conducts several different types of rankings targeting different audiences, including the CHE University Ranking, the CHE Research Ranking, and the CHE Excellence Ranking (CHE, 2011), some of which are published as a university guide in the magazine *Der Stern*

(WR, 2004). The CHE rankings are all subject-specific (no overarching ranking is calculated) and multidimensional (no overall value is derived from weighting individual indicators for a subject). The CHE Research Ranking compares German HEIs and their faculties on the basis of research performance, currently covering 16 subjects in the natural sciences, humanities, and social sciences. The DFG "Funding Ranking" is more a description than an assessment. It analyzes the distribution of public third-party funding to universities in Germany, and includes a broadly field-based analysis by the four DFG scientific disciplines. The goal of the report is to provide a basis for planning and management at HEIs (DFG, 2010); it does not influence funding allocation decisions at the DFG (Dr. Robert Paul Königs, personal communication, June 24, 2011).

The WR has recently initiated "Research Ratings," described as a "novel, discipline-specific, comparative appraisal process" (WR, 2011). According to the WR, ratings differ from rankings in that they are usually carried out by expert groups, and do not necessarily include aggregation in the form of ranking lists (WR, 2004). The purpose of the ratings is to compare and appraise the research performance of all German universities and non-university research institutions within a discipline, to inform decision-makers at the research institutions and various education ministries involved. The ratings are not yet meant to guide decisions related to basic funding (see Observations and Lessons Learned below) (WR, 2004). Following a 2004 report that reviewed best practices and issued recommendations on implementation of a comparative rating system, pilot ratings of chemistry and sociology were conducted and well received. A Research Rating for electrical engineering and information technology has recently been completed (Dr. Rainer Lange, personal communication, June 16, 2011), and one on English and American studies is in progress (WR, 2011). The WR aims to re-evaluate each subject every five to six years (WR, 2004); however, it is still open as to whether this rating will become standard procedure in Germany, at least in part due to the costly nature of peer review (Dr. Stefan Hornsbostel, personal communication, May 24, 2011).

The process for the Research Ratings is peer review (assessment boards of up to 15 researchers including 2 international experts and 2 experts from areas outside of publicly funded research institutions) informed by quantitative and qualitative data, including data gathered from profiles submitted by relevant research units and any external data that are available or can be obtained at a reasonable expense (WR, 2004). For example, the 2008 review of chemistry gathered data from 349 research units at 77 institutions as well as external data on awards, personnel, graduations, publications. and citations (WR, 2008a). The reviewers' judgments are not bound by the indicators, and also take into account contextual information such as innovative achievements or periods of fundamental change (WR, 2008a).

The research units/institutions are assessed based on six criteria under three dimensions: i) research, ii) promotion of young researchers, and iii) knowledge transfer (see Table 1). Each expert panel decides how to interpret the assessment criteria in the context of its discipline, which indicators to use to evaluate the criteria, and how to weigh the indicators. In the review of chemistry, the results were presented as overall results by individual criteria (including international comparisons), correlations between criteria, strengths and weaknesses of chemical research in Germany, and finally individual results by institution (WR, 2008a).

Besides its Research Ratings, the German Council of Science and Humanities produces reports on the structure and development of selected fields in German academia (WR, 2011). The most recent report covered theology and sciences concerned with religions (2010), while previous reports have covered communication and media studies (2007), agricultural sciences (2006), and the humanities (2006). The reports analyze the current situation of the field and provide structural recommendations aimed at the federal and state authorities as well as universities (WR, 2010). The 2010 report on the advancement of theologies and sciences concerned with religions was informed by a literature review and data on amount and distribution of external funding and numbers and distribution of students and professorial chairs (WR, 2010).

Finally, the Scientific Commission of Lower Saxony has carried out state-wide, subject-specific evaluations since 1999 (WR, 2004). The evaluations serve several purposes, including providing information to universities and local governments to help with strategic and structural planning, and to contribute to the development of criteria for performance-based funding from the state. Review groups assess the performance of research units within a discipline based on quality, relevance, effectiveness, and efficiency. The peer review process is informed by standardized self-reports submitted by the relevant research units and on-site visits by the review groups.

5.5 USE OF SCIENCE INDICATORS

Indicators are used systematically in the allocation of institutional funding, but vary by state and type of institution (e.g., public universities, universities of applied science, public colleges of art, university clinics, etc.) (OECD, 2010a). The most frequently used research-related indicators (used in all nine Länder surveyed by the OECD) are third-party funding and number of completed PhDs (OECD, 2010b). Other indicators used by certain states include participation in various research programs within the DFG or EU (states of Berlin and Hesse). Notably, bibliometrics play almost no role in institutional funding formulas (OECD, 2010a), although publications are taken into account for university clinics in Bavaria and universities of applied science in Berlin (OECD, 2010b). Five out of the nine states surveyed by the OECD apply subject-specific weightings (OECD, 2010a).

The DFG funding rankings use five groups of indicators: basic data (personnel, basic funds, third-party funds); funding for research projects (DFG grants, direct R&D project funding from the federal government, etc.); scientific expertise

Table 1

General guidelines for indicators used in Research Ratings by the WR

Dimension	Criterion	Indicators
Research	Research Quality	• Research outputs • Third-party funding • Scientific cooperation projects • *If applicable, relative citation indicators, proportion of highly cited publications*
	Impact/Effectiveness	• Quality-weighted publication figures • If *applicable, presentations at major international conferences* • *If applicable, absolute citation counts, impact-weighted publication counts*
	Efficiency	• Numerator: quality-weighted publication figures • Denominator: number of researchers • Resource input including third-party funding
Promotion of young researchers	Promotion of young researchers	• Structured doctoral programmes, median of duration of doctoral studies • Externally funded fellowships • Number of independent junior research groups • Subsequent career of doctoral students, postdoctoral students • Publications of young researchers • Subsequent career of doctoral students, postdoctoral students • Publications of young researchers
Knowledge transfer		• Research outputs • Cooperation projects • Funds from industry • Intellectual property rights, licences • Description of training and further training measures • Description of research-based consulting services and science communication activities

Source: Adapted from WR, 2004

(elected DFG review board members, DFG reviewers); international appeal (visiting researchers and scientists); and collaboration in research networks (participation in the DFG's Coordinated Research programs) (DFG, 2010). The indicators used in the CHE Research Rankings, all subject-specific, include level of third-party funding; number of doctorates; publications and citations (for certain subjects, publications are weighted by type and length)' and patent registrations or inventions (CHE, 2011). All of these indicators are quoted in absolute numbers and in proportion to the number of researchers (third-party funding) or professors (publications, doctorates) (WR, 2004). Professors are interviewed to gather information on reputation, but the results are not used to determine the top groups of universities. For the Research Ratings of the WR, the indicators used for each assessment criterion are determined by the subject-specific assessment boards. Table 1 shows the WR's general recommendations for indicators that could be used based on information provided in the institutional self-profiles. Table 2 displays the indicators chosen by the assessment board for the 2008 review of chemistry.

5.6 OBSERVATIONS AND LESSONS LEARNED

The inclusion of a performance-based formula in the core funding models of several states in Germany has had several benefits, including encouragement of the optimal use of state grants' increased transparency, and predictability of budget allocations; clear incentives for improvements in performance and quality; and stronger incentives for third-party funding and doctoral graduates (OECD, 2010a). Some concerns, however, have been raised around the implementation of the performance-based portion. The DFG rankings show that funding varies substantially by discipline. If third-party funding is used as an indicator of quality in performance-based research funding, which it is by most Länder, the DFG asserts that these discipline-specific differences need to be taken into account (DFG, 2010). Four out of nine states surveyed by the OECD, however, do not employ subject-specific weightings (OECD, 2010a). Research has suggested that the "discretionary/incrementalist" components of funding alongside performance-based components have the potential to undermine or neutralize the impact of indicator-based funding (see Orr & Jaeger, 2009). The WR notes that there is always a trade-off; relying on a small number of indicators may be unfair to some groups, but more complex funding formulas require data, which in turn require effort to report and gather on the part of researchers and administrators (Dr. Rainer Lange, personal communication, June 16, 2011).

Table 2

Indicators used in the WR Research Rating of chemistry (WR, 2008a)

Dimension	Criterion	Indicators	
		Quantitative	Qualitative
Research	Quality[14]	• Citations per publication normalized to average citation rate for the subject area (ZP/FCS_m) • Citations per publication normalized to average citation rate for the journal (ZP/JCS_m) • Citations per publication (ZP) • Number of publications (P) (additional information for assessing the citation indicators) • Ratio subject area-specific over journal-specific citation success (JCS_m/FCS_m) • Maximum number of citations of a single publication (Z_{max}) • Points score for support by the Chemical Industry Fund	• List of publications • Research output other than publications, e.g., databases and software, patents, etc. • List of third-party-funded projects • List of major research awards and prizes
	Impact/ Effectiveness[15]	• Number of publications (P) • Number of initial registrations, patents granted • Volume of third-party funding • Proportion of third-party-funded staff in total staff (full-time equivalent (FTE)) • Absolute number of citations (Z) (rating taking into account the maximum number of citations for a single publication (Z_{max}) and the number of publications never cited (P_{nz})) • Normalized number of citations: citations per publication normalized to the average number of citations for the journal (ZP/JCS_m) and citations per publication normalized to the average number of citations for the subject area (ZP/FCS_m) • Number of visiting scientists funded by DAAD and AvH • Citations from other subject areas	• Self report on interdisciplinarity • Elected/Appointed offices at other scientific institutions (list) • Plenary talks/Named lectures (list)

continued on next page

14 Assessment aspects include relative success of reception, output quality, and peer appreciation.
15 Assessment aspects include research productivity, research activity, visibility, interdisciplinarity, and reputation.

Table 2

Indicators used in the WR Research Rating of chemistry (WR, 2008a)

Dimension	Criterion	Indicators: Quantitative	Indicators: Qualitative
Research	Efficiency[16]	• Ratio number of publications / scient. staff (FTE total and FTE mainstream-funded, weighted) • Ratio number of citations (Z) / scient. staff (FTE total and FTE mainstream-funded, weighted) • Ratio third-party funding vol./ scient. staff (FTE total and FTE mainstream-funded, weighted) • Ratio patent registrations / scient. staff (FTE total and FTE mainstream-funded, weighted)	• Impact in proportion to total personnel input (FTE, weighed according to teaching duties, incl. doctoral students and post-docs), to mainstream-funded staff (FTE, incl. doctoral students and post-docs, weighted), to staff excl. doctoral students (FTE, weighted), and to technical staff (FTE number)
Promotion of young researchers[17]		• Number of postgraduate bursaries and fellowships, plus paid doctoral student posts • Number of PhD graduations • Proportion of female PhD graduates • Number of postdoc fellowships and junior group leader posts	• Structured postgraduate programs (list) • PhD prizes awarded (list) • Academic appointments for young scientists (list) • Prizes awarded to young scientists (list)
Knowledge transfer	Transfer to other areas of society	• Number of patents awarded • Number of licensed patents • License income • Third-party-funding from private-sector companies	• Spin-offs and shares in businesses (list) • Consulting functions outside the private sector (list)
	Promotion of the public understanding of science	• Number of vocational apprenticeships completed	• Advanced vocational training courses (list) • Description of exemplary measures to promote the understanding of science beyond the realm of science

The German Council of Science and Humanities has invested a considerable amount of effort in reviewing and analyzing best practices in assessment of research. A WR 2004 report concluded that to be effective and accurate, comparisons of research quality require research area-specific assessment in the form of peer review carried out on the basis of harmonized data and quantitative indicators ("informed peer review") on a predefined assessment scale (WR, 2004). The WR

16 Impact/effectiveness in relation to resources spent.

17 Assessment aspects include promotion of PhD students and of young postdocs.

noted an international trend towards increased usage of peer review, especially when the research assessment is meant to inform strategies and funding allocation decisions. It also noted, however, that use of data is important to prevent the review exercise from becoming an assessment of reputation, and, especially in cases of nationwide or international comparisons, to save time for the reviewers and help ensure reliability despite time constraints. Analyses of subsequent pilot studies of chemistry and sociology have supported these conclusions (WR, 2008b).

Germany's subject-specific research rankings and ratings purposely do not make comparisons between fields, nor are they used to determine funding allocations. In the CHE Rankings, each subject is analyzed differently, based on the assertion that preferred dissemination methods and available databases can vary highly in different disciplines (CHE, 2011). The WR recommended subject-specific ratings because assessment criteria differ for research areas, and declared that "a comparison across subject boundaries…will lead to bogus results" (WR, 2004). A sub-committee of the WR has determined appropriate assessment criteria for research in the individual disciplines (WR, 2011); unfortunately, the results are available only in German and scattered among various reports (Dr. Rainer Lange, personal communication, June 16, 2011). In terms of the link between assessments and resource allocation, the assessment board for chemistry noted that the allocation of public funds should not be directly linked to a "selective, one-off assessment," but rather should be based on trends that can only be identified after the rating has been repeated several years later (WR, 2008a).

Several lessons can be learned on use of indicators in Germany. According to the WR, the indicators used to assess a discipline should be determined by experts in that discipline (WR, 2004). In addition, indicators should not only be subject-specific, but also be given different weights by sub-field (Dr. Rainer Lange, personal communication, June 16, 2011). It is also important to distinguish between the *measurement* of activities and the *value-functions* that are applied to the activities; for instance, one can measure the number of PhDs, but it is not always a question of more is better. Since these non-linear value-functions can be complicated to capture mathematically, peer review groups are useful to discuss these issues. In the example of number of PhDs, peers may identify acceptable ranges that differ based on context (Dr. Rainer Lange, personal communication, June 16, 2011). The WR also cautions against using indicators based solely on accessibility of data.

The Institute for Research Information and Quality Assurance (iFQ), which is funded by the DFG to monitor and evaluate its funding activities, also conducts research on bibliometrics, and cautions that bibliometrics cannot be perfectly objective because they rely strongly on the methods adopted for their production.[18] For example, the quality of indicators depends on the robustness of the data on which they are based, which, in the WR's rating of chemistry, was found to be strong for the dimension of research but weak for the dimension of knowledge transfer (WR, 2008a). In terms of interdisciplinarity, the assessment board for chemistry also noted that more differentiation among disciplines leads to more frequent difficulties with the classification of research activities (WR, 2008b). Germany is conducting ongoing research in several of these topics, including performance-led governance, research ratings, and bibliometrics.

Through the GOMED project (*Intended and unintended effects of local incentive programmes using the example of performance-based funding in medical science*), the iFQ is researching the incentive effects of performance-based funding applied by university faculties (iFQ, 2011). The iFQ is also invested in improving quality of bibliometric analysis by providing information on reliability of bibliometric data and indicators, and is currently assessing use of bibliometric indicators in the field of electrical engineering, which is the subject of a current WR research rating (iFQ, 2011). This latter endeavour will assess the extent to which electrical engineering publications and patent registrations are covered and can be identified in special databases, as well as identify methodological problems arising when bibliometric analyses are used. The WR is currently examining the consequences of performance-led governance, and continues to monitor the results and application of its Research Ratings (WR, 2011).

5.7 REFERENCES

BMBF (German Federal Ministry of Education and Research) (2010a). Ideas, Innovation, Prosperity: High-Tech Strategy 2020 for Germany. Bonn, Germany: German Federal Ministry of Education and Research.

BMBF (German Federal Ministry of Education and Research) (2010b). Federal Report on Research and Innovation 2010: Abstract. Bonn, Germany: German Federal Ministry of Education and Research.

BMBF (German Federal Ministry of Education and Research) (2011). Research Funding System Retrieved May 2011, from http://www.research-in-germany.de/research-funding/59962/e2-research-funding-system.html

18 See a discussion of the h-Index by Bornmann (2006) on the iFQ's website for a preliminary discussion of bibliometric methods bias at http://www.research-information.de/iq/agora/H_Index/h_index.asp

CHE (Centre for Higher Education Development) (2011). CHE Ranking Retrieved May 2011, from http://www.che-ranking.de/cms/?getObject=614&getLang=en

DFG (German Research Foundation) (2009). Annual Report 2009: Projects in the Individual Grants Programme. Retrieved May 2011, from http://www.dfg.de/jahresbericht/projekte_0_index.html

DFG (German Research Foundation) (2010). Funding Ranking 2009: Institutions - Regions - Networks - Thematic Profiles of Higher Education Institutions and Non-University Research Institutions in Light of Publicly Funded Research. Bonn, Germany: DFG.

DFG (German Research Foundation) (2011). Priority Programmes Retrieved June 2011, from http://www.dfg.de/en/research_funding/programmes/coordinated_programmes/priority_programmes/index.html

European Commission (2010). ERAWATCH National Profiles: Germany. Retrieved March,, 2012, from http://erawatch.jrc.ec.europa.eu/erawatch/opencms/information/country_pages/de/country

iFQ (Institute for Research Information and Quality Assurance) (2011). Projects Retrieved May,, 2011, from http://www.research-information.de/Projekte/projekte_start.asp

OECD (Organization for Economic Co-operation and Development) (2010a). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, Paris, France.

OECD (Organization for Economic Co-operation and Development) (2010b). Web Annex: Additional Country Detail (Performance-based Funding for Public Research in Tertiary Education Institutions). Paris, France: OECD.

Orr, D. & M. Jaeger (2009). Governance in German higher education: Competition versus negotiation of performance. In J. Huisman (Ed.), *International Perspectives on the Governance of Higher Education: Alternative Frameworks for Coordination*. New York/London.

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

UNIDO (United Nations Industrial Development Organization) (2005). UNIDO Technology Foresight Manual (Vol. 1: Organization and Methods). Vienna, Austria: UNIDO.

WR (German Council of Science and Humanities) (2004). Recommendations for Rankings in the System of Higher Education and Research; Part 1: Research. (pp. 64). Hamburg, Germany: WR.

WR (German Council of Science and Humanities) (2008a). Research Performance of German Universities and Non-University Institutions in Chemistry. Cologne, Germany: German Council of Science and Humanities.

WR (German Council of Science and Humanities) (2008b). Steering Group Report on the Pilot Study Research Rating in Chemistry and Sociology (pp. 48). Cologne: German Council of Science and Humanities.

WR (German Council of Science and Humanities) (2010). Recommendations on the Advancement of Theologies and Sciences concerned with Religions at German Universities (pp. 158). Berlin, Germany: WR.

WR (German Council of Science and Humanities) (2011). Fields of Activity Retrieved May 2011, from http://www.wissenschaftsrat.de/1/home/

BMBF (German Federal Ministry of Education and Research) (2010a). *Federal Report on Research and Innovation 2010: Abstract*. Bonn: German Federal Ministry of Education and Research.

BMBF (German Federal Ministry of Education and Research) (2010b). *Ideas, Innovation, Prosperity: High-Tech Strategy 2020 for Germany*. Bonn: German Federal Ministry of Education and Research.

BMBF (German Federal Ministry of Education and Research) (2011). Research Funding System. Retrieved May 2011, from http://www.research-in-germany.de/research-funding/59962/e2-research-funding-system.html

CHE (Centre for Higher Education Development) (2011). CHE Ranking. Retrieved May 2011, from http://www.che-ranking.de/cms/?getObject=614&getLang=en

DFG (German Research Foundation) (2009). Annual Report 2009: Projects in the Individual Grants Programme. Retrieved May 2011, from http://www.dfg.de/jahresbericht/projekte_0_index.html

DFG (German Research Foundation) (2010). *Funding Ranking 2009: Thematic Profiles of Higher Education Institutions and Non-University Research Institutions in Light of Publicly Funded Research*. Bonn: German Research Foundation.

DFG (German Research Foundation) (2011). Priority Programmes. Retrieved June 2011, from http://www.dfg.de/en/research_funding/programmes/coordinated_programmes/priority_programmes/index.html

European Commission (2010). ERAWATCH National Profiles: Germany. Retrieved May 2011, from http://cordis.europa.eu/erawatch/index.cfm?fuseaction=ri.content&topicID=4&countryCode=DE

iFQ (Institute for Research Information and Quality Assurance) (2011). Projects. Retrieved May 2011, from http://www.research-information.de/Projekte/projekte_start.asp

OECD (Organization for Economic Co-operation and Development) (2010a). Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings. Paris: OECD

OECD (Organization for Economic Co-operation and Development) (2010b). Profile of Germany. *Science, Technology and Industry Outlook 2010*. Paris: OECD.

OECD (Organization for Economic Co-operation and Development) (2010c). *Web Annex: Additional Country Detail (Performance-based Funding for Public Research in Tertiary Education Institutions)*. Paris: OECD.

UNIDO (United Nations Industrial Development Organization) (2005). *UNIDO Technology Foresight Manual*. Vienna: UNIDO.

WR (German Council of Science and Humanities) (2004). *Recommendations for Rankings in the System of Higher Education and Research; Part 1: Research*. Hamburg: German Council of Science and Humanities.

WR (German Council of Science and Humanities) (2008a). *Research Performance of German Universities and Non-University Institutions in Chemistry*. Cologne: German Council of Science and Humanities.

WR (German Council of Science and Humanities) (2008b). *Steering Group Report on the Pilot Study Research Rating in Chemistry and Sociology*. Cologne: German Council of Science and Humanities.

WR (German Council of Science and Humanities) (2010). *Recommendations on the Advancement of Theologies and Sciences Concerned with Religions at German Universities*. Berlin: German Council of Science and Humanities.

WR (German Council of Science and Humanities) (2011). Fields of Activity. Retrieved May 2011, from http://www.wissenschaftsrat.de/1/fields-of-activity/#c2255

6 The Netherlands

The Netherlands — Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP):	$11.0 billion
• GERD as a % of GDP:	1.63%
• Total Researchers (Full-Time Equivalent, 2007):	51,052
• # of Researchers per Million People (2007):	3,089
• # of Scientific Publications (2008):	22,945

Source: UNESCO Science Report, 2010

6.1 NATIONAL RESEARCH FUNDING CONTEXT

The Netherlands has a dual-support research funding system, meaning that funding is provided both to institutions directly and to individual researchers. Institutional research funding for universities is provided in the form of block grants from the Ministry of Education, Culture, and Science (OCW) (this is referred to as "first-flow" funding in the Netherlands). These block grants were equal to approximately $2.6 million in 2010 (European Commission, 2010). Project-based funding ("second-flow" funding) is provided through grants to individual researchers from the Netherlands Organisation for Scientific Research (NWO). Annually, research funding through NWO exceeds $6.5 million (NWO, 2011). Recently, the government made a strategic decision to allocate more research funding through project-based granting programs, and has shifted approximately $130 million annually from first-flow funding through OCW block grants to second-flow funding through NWO granting programs for individual researchers and projects (European Commission, 2010).

6.2 NATIONAL RESEARCH FUNDING PRIORITIES

At the highest level, the government's science policy priorities and objectives are now defined in the Strategic Agenda for Higher Education Research and Science Policy (OCW, 2007). These consist of high-level policy objectives such as creating an ambitious research environment; increasing researcher autonomy; aligning research priorities with society needs; and enhancing collaboration of research institutes, businesses, and other organizations (OCW, 2007; European Commission, 2010). More focused research priorities, however, are developed by research funding organizations. For example, NWO also develops research themes to guide research funding decisions and priorities within its funding programs. The latest strategy, developed for the period 2011–2014, aims to solve "urgent problems in society" via the choice of six themes: Healthy Living; Water and Climate; Cultural and Societal Dynamics; Sustainable Energy; Materials: Solutions for Scarcity; and Connecting Sustainable Cities (NWO, 2011).

6.3 RESEARCH FUNDING ALLOCATION

"First-flow" or lump-sum research funding for universities is provided directly through transfers from the OCW. This funding is allocated based on a funding model using various parameters related to teaching and research activities (OCW, 2008b). Some of these parameters are performance based (OCW, 2008b). Grant-based funding from NWO is allocated primarily through peer review, in response to specific research programs or funding calls. Some NWO research support is allocated through personalized grants or "free competitions" where researchers propose their own research objectives (OCW, 2008b). The majority of NWO funding (56 per cent in 2006) supports research at universities, but NWO also funds research that takes place at NWO-administered research institutes.

6.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

The Netherlands conducts assessments of the national research system at several levels. While none of these evaluation exercises are directly tied to research funding allocation mechanisms or decisions, their outcomes may inform funding decisions through changes in research priorities or research management at the institutional level.

At the highest level of aggregation, the Netherlands Observatory of Science and Technology (NOWT) conducts biennial assessments of the Dutch national research system based on analysis of existing quantitative indicators. Eight of these evaluations have been conducted since NOWT's founding in 1992, with the latest completed in 2010 (see NOWT, 2010). These assessments are intended to provide comprehensive reviews of the national research system, including consideration of both research inputs and research outputs. The reports also contain some analysis of research strengths and weaknesses by field based on bibliometric data and other readily available quantitative evidence (there is no expert review or other qualitative evaluation element in NOWT's analysis of discipline strengths and weaknesses). Although these publications are not directly linked to research funding decisions, they are broadly intended to inform and support development of science and technology policy in the country and are therefore likely to influence development of research priorities.

With respect to more focused evaluations of university research, the Netherlands has developed a unique system of research assessment based on periodic, panel-based expert reviews of research programs or institutes within universities. In 1992, 13 universities and the Minister for Education agreed that the Association of Universities in the Netherlands (VSNU) should develop a system (Quality Assessment of Research) using external evaluation to complement internal quality controls (Geuna & Martin, 2003). This directive resulted in VSNU, NWO, and the Royal Netherlands Academy of the Arts and Sciences (KNAW) using a common protocol in use since 2003 known as the Standard Evaluation Protocol (the latest version of which — SEP 2009–2015 — was published in 2010) (ERiC, 2010). In this approach, research performing units (broadly defined) are evaluated once every several years (on a staggered basis) through a combination of self-evaluation and evaluation by an independent, external committee.

The current SEP approach can be characterized by both its units of evaluation and the focus and criteria of the evaluation process. The two levels of assessment within the framework are the research institute and the research program. Research institutes are broadly defined as "a group of researchers with an articulated shared mission, operating one or more research programmes under the same management" (KNAW, 2010). The exact nature of research institutes therefore varies among organizations, but includes defined research groups within universities (e.g., departments). The constituent research programs within an institute represent the second level of assessment. The current SEP framework for assessment emphasizes evaluation of a research institute's three main tasks (the production of results relevant to the scientific community, the production of results relevant to society, and the training of PhD students) and four general criteria (quality, productivity, societal relevance, and vitality and feasibility) (KNAW, 2010). More details on the current SEP assessment framework are provided in the following section.

While the SEP assessment process is not directly tied to specific research funding allocation decisions or mechanisms, one of its explicit goals is to provide accountability to research funding organizations. The framework document notes:

> With the external evaluation, the institute and its research groups account for their research activities to the board of the university, KNAW or NWO. In a broader sense, the external evaluations inform funding agencies, government and society at large of the quality and relevance of research activities, thus accounting for the public investments made in scientific research.
>
> (KNAW, 2010)

As a result, information resulting from the assessment process may have significant implications for internal management of research activities within institutes, as well as how those institutes articulate their objectives and account for progress to research funding organizations.

It is also worth highlighting that, in recent years, evaluation of the social relevance of research has taken on greater prominence in the SEP assessment process. As a consequence, KNAW, VSNU, the Netherlands Association of Universities of Applied Sciences (HBO-raad), and the Rathenau Institutes Science System Assessment department have partnered on an initiative aimed at improving research evaluation methods in this area, referred to as Evaluation of Research in Context (ERiC) (see ERiC, 2010). Pilot studies were conducted in various areas, and KNAW has now published a new guide on the subject (see ERiC, 2010).

6.5 USE OF SCIENCE INDICATORS

Use of science indicators and assessment methods in the Netherlands roughly corresponds to the various levels of assessment described in the preceding section.

First, the OCW reports that lump-sum funding for universities focuses on performance-based indicators in the funding allocation model. Second, the national S&T assessments from NOWT employ a variety of quantitative indicators. For example, the latest assessment includes analysis of R&D expenditures, human resources, scientific publications and patents, and various measures of R&D cooperation and knowledge transfer (NOWT, 2010). These indicators are presented in Table 1. While the overriding focus of NOWT reports is not on discipline-level assessments, the analysis of bibliometric indicators includes assessment of output and impact by research field. For example, the latest report finds that

Dutch research in the areas of information and communication science is both highly cited and represents a relatively high share of worldwide publication in this area. In contrast, Dutch research in mathematics has only an average level of citations and represents a smaller share of the Netherland's total research output (NOWT, 2010). Performance on these research output indicators is also analyzed by institution, highlighting differences in strengths and weaknesses of research across Dutch universities. NOWT assessments, however, remain focused on providing a comprehensive review of the national research system through existing quantitative indicators, rather than on a discipline by discipline evaluation for use in any funding allocation decisions.

Third, the SEP evaluation protocol used in research assessment in Dutch universities and research institutes is also based on a standardized evaluation method, which includes a set of defined assessment criteria. The SEP assessment cycle includes an internal, self-evaluation process and an external evaluation by a committee of experts including site visits. The external evaluation protocol outlines 4 evaluative criteria as well as 10 sub-criteria and various indicators or factors that may be considered in relation to these criteria (listed in Table 2). While this framework is intended to be suitable for all research disciplines and fields, the SEP recognizes the need for differentiation among fields and therefore allows for some variation in the indicators or evidence provided by research institutes and programs in different domains (KNAW, 2010).

At the end of the SEP external review assessment process, review panels are asked to provide summary ratings of the performance of the research institute or program on these criteria, based on a five-point scale. A rating of 5 denotes: "Research is world leading. Researchers are working at the forefront of their field internationally and their research has an important and substantial impact in the field;" whereas a 0 rating denotes: "Work is neither solid nor exciting, flawed in the scientific and or technical approach, repetitions of other work, etc." (KNAW, 2010).

The SEP also involves a periodic self-evaluation by the research unit. This self-evaluation process is governed by a standardized template, which requests the unit being evaluated to submit data on a variety of quantitative indicators including publication output (by type), staff and student numbers, data on research grants and funding, number of PhDs (completed and in progress), and measures of academic reputation and esteem and national and international positioning (see KNAW, 2010).

Table 1

NOWT S&T Scoreboard 2010 – R&D Performance in the Netherlands

Indicator	Value
R&D expenditures	
1. R&D intensity business sector (% of GDP)	1.03%
2. R&D intensity higher education sector (% of GDP)	0.45%
3. R&D intensity research institutes (% of GDP)	0.22%
4. Funding of public-sector R&D by business sector (% of total R&D)	10%
R&D human resources	
5. R&D personnel (% of labour force)	1.0%
6. Researchers (% of labour force)	0.5%
R&D outputs and impact	
7. Scientific output (research articles per 1,000 population)	1.8
8. Scientific impact (citation impact compared to world average)	+33%
9. Universities in the global top 100 (Shanghai ARWU-ranking)	2
10. Patents (triad patents per million population)	18
R&D cooperation and knowledge transfer	
11. International research co-publications (% of total publication output)	48%
12. Innovating firms collaborating with universities (% of all collaborating innovating firms)	29%
13. Innovating firms collaborating with public-sector research institutes (% of all collaborating innovating firms)	20%
14. Public-private research co-publications (% of total publication output)	6%

Source: NOWT, 2010

Table 2

Netherlands Standard Evaluation Protocol (SEP) Research Assessment Criteria

Criteria	Sub-Criteria	Aspects That May Be Considered
Quality	A1. Quality and scientific relevance of the research	Originality of the ideas and the research approach, including technological aspects; Significance of the contribution to the field; Coherence of the programme; Quality of the scientific publications; Quality of other output; Scientific and technological relevance
	A2. Leadership	Leadership of primary individuals; Mission and goals; Strategy and policy
	A3. Academic reputation	(Inter)national position and recognition; Prominence of the programme director and other research staff; Impact and significance of research results in the field
	A4. Resources	Human resources; Funding policies and earning capacity; Relevance of research facilities
	A5. PhD training	Objectives and institutional embedding; Structure of programmes; Supervision; Success rates; Educational resources
Productivity	B1. Productivity strategy	Productivity goals; Publication strategy; Rewards and sanctions
	B2. Productivity	Scientific publications and PhD-theses; Professional publications; Output for wider audiences; Use of research facilities by third parties
Relevance	C. Societal relevance	Societal quality; Societal impact; Valorisation
Vitality and feasibility	D1. Strategy	Strategic planning; Investments and collaboration; Research topics planned for the near future and their perspectives; Flexibility and anticipation of expected changes
	D2. SWOT-analysis	Analysis of the position of institute and programmes; Analysis of strengths and weaknesses
	D3. Robustness and stability	Research facilities; Financial resources; Staff competition; Mobility and attractiveness; Expertise within the institute

Source: SEP 2009-2015 (KNAW, 2010).

6.6 OBSERVATIONS AND LESSONS LEARNED

The SEP assessment process has now been through several iterations and was recommitted to for the period 2009–2015. A 2008 evaluation of the SEP concluded that the process was showing positive results and users emphasized the importance of continuity in the guidelines for research assessment (KNAW, 2010). The evaluation report highlighted concerns about the administrative burden created by the process, as well as the need for more emphasis on assessing societal relevance of research (KNAW, 2010). Both of these concerns were then considered in the development of the SEP 2009–2015, and KNAW has engaged in additional work on assessing social relevance of research through its ERiC initiative (see ERiC, 2010). In general, there is little evidence of widespread dissatisfaction with the system, or interest in moving to an assessment system with greater emphasis on quantitative metrics.

Use of thematic research priorities within NWO appears equally uncontroversial, as does the current allocation model for lump-sum funding for universities from the OCW, but it is unclear if the government is exploring alternative funding models. Perhaps the underlying implication of the Dutch experience with research evaluation is that the combination of a high-level national review of quantitative indicators (NOWT reports) with more detailed, periodic research assessments based on self-evaluation and external expert committees (the SEP assessment process) has provided researchers and policy-makers with a relatively robust base of knowledge to facilitate strategic planning and research priority setting. These conclusions, however, are based on relative paucity of critical research on the impacts of research evaluation in the Netherlands. Additional systematic analysis, as well as comparison with analogous systems in other countries, would likely be needed to provide more nuanced insights.

6.7 REFERENCES

ERiC (Evaluating Research in Context) (2010). Evaluating the Societal Relevance of Academic Research: A Guide. Amsterdam, the Netherlands: Royal Netherlands Academy of the Arts and Sciences.

European Commission (2010). ERAWATCH National Profiles: the Netherlands. Retrieved March,, 2012, from http://erawatch.jrc.ec.europa.eu/erawatch/opencms/information/country_pages/nl/country

Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva, 41*(4), 277–304.

KNAW (Royal Netherlands Academy of the Arts and Sciences) (2010). Standard Evaluation Protocol 2009–2015: Protocol for Research Assessment in the Netherlands. Amsterdam, The Netherlands: Royal Netherlands Academy of the Arts and Sciences.

NOWT (Netherlands Observatory of Science and Technology) (2010). Science and Technology Indicators 2010: A Summary Report on the Netherlands (pp. 44). Hague, The Netherlands: NOWT.

NWO (Netherlands Organization for Scientific Research) (2011). NWO Themes Retrieved May 13, 2011, from http://www.nwo.nl/nwohome.nsf/pages/NWOA_6XYDNE_Eng

OCW (Ministry of Education Culture and Science) (2007). Strategic Agenda for Higher Education Research and Science Policy. Hague, The Netherlands: Ministry of Education Culture and Science.

OCW (Ministry of Education Culture and Science) (2008b). The Science System in the Netherlands: An Organizational Overview (pp. 44). The Hague, the Netherlands: Ministry of Education, Culture and Science. Research and Science Policy Department.

OCW (Ministry of Education Culture and Science) (2011). Higher Education. Retrieved May 13, 2011, from http://english.minocw.nl/english/education/363/Higher-education.html

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

7 Norway

Norway — Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP):	$4.5 billion
• GERD as a % of GDP:	1.62%
• Total Researchers (Full-Time Equivalent, 2007):	26,062
• # of Researchers per Million People (2007):	5,468
• # of Scientific Publications (2008):	6,958

Source: UNESCO Science Report, 2010

7.1 NATIONAL RESEARCH FUNDING CONTEXT

The Norwegian research funding system is structured around a high proportion of core institutional funding of higher education institutions. The Ministry of Education and Research channels about 80 per cent of public funds to HEIs, a large proportion of which is distributed through the Research Council of Norway (RCN) directly to research institutes entitled to core funding from the government in the form of block funding (European Commission, 2010a). The RCN is the result of a 1993 merger of five research councils, and combines two funding roles that are discrete in other countries: the traditional research council funding role (focusing on research in universities and research institutes with an emphasis on scientific quality) is integrated with funding for user-oriented technology and innovation (relevance of user needs is very important) (Technopolis, 2001). The RCN also merges research funding and science policy advice, which are usually kept separate to avoid conflicts of interest (Technopolis, 2001).

7.2 NATIONAL RESEARCH PRIORITIES

As of 2008–2009, the government of Norway identified a number of thematic research priorities including global challenges (environment, climate change, oceans, food safety, energy research); health; social challenges; knowledge-based industry in all regions; and industry-oriented research (food, marine, maritime, tourism, biotech, ICT, new materials/nanotech) (European Commission, 2010b). In 2011 the RCN reorganized from three to four divisions including the Division for Science, Division of Innovation and two new departments under the Division for Strategic Priorities, the Department for Energy and Petroleum, and the Department for Climate and the Environment (European Commission, 2010b).

7.3 RESEARCH FUNDING ALLOCATION

A new funding structure was put in place in 2003 for core funding of universities, which consists of three components: 60 per cent block funding (without detailed specifications of its use);, a teaching component, in which funds are distributed based on reported student performance (25 per cent); and a research component, which amounts to about 15 per cent of institutional funding (European Commission, 2010a). The research component is then further sub-divided into a performance-based part, within which funds are redistributed among institutions on the basis of benchmarks for publications and competitive research; and a strategic component, within which earmarked funds are allocated to institutions for PhDs and equipment (European Commission, 2010a). There is no differentiation across scientific fields, and so far noticeable reallocations across fields have not been observed (OECD, 2010a).Once funded, universities allocate resources autonomously in line with their own research priorities (OECD, 2010a).

Much of the RCN's funding is linked to areas of research. Strategic or targeted research initiatives (top-down) receive about 50 per cent of the annual budget, while non-directed discovery research (independent research projects through the Independent Grants Program (FRIPRO) funding scheme) receive between 10 per cent and 15 per cent (RCN, 2011). Proposals submitted through FRIPRO are reviewed by field-specific referee panels comprising international experts, which meet to discuss and rank relevant proposals with scientific quality as the main criterion. A total of 37 panels and approximately 200 experts will be engaged for the 2011 call. Based on the panels' assessments and rankings, and consideration for the strategic guidelines set out in the call for proposals, such as gender equality and international cooperation, the RCN administration prepares funding recommendations for each of four Expert Committees. The committees, which have 7–10 members, all from Scandinavia, are broadly divided by field: Medicine, Health and Biology; Mathematics, Physical Sciences and Technology; Humanities; and Social Sciences. The recommendations form the basis for overall assessments and final decisions taken by the Expert Committees on grant allocations under FRIPRO. There is a predetermined allocation of funding among the four Expert Committees, which is historically determined. Any increases in the FRIPRO budget are distributed among the Expert Committees proportional to the total amount of funding received by afield at the national level, which is taken as an indicator of size and need. The FRIPRO success rate is about 10 per cent, so proposals funded by increases in the budget are already of very high quality (Asbjorn Mo, personal communication, June 30, 2011).

Within the FRIPRO scheme, it can be difficult for interdisciplinary research proposals to receive funding. In the past, this has been dealt with by having these proposals reviewed by two or three referees who are not members of the panels. In 2011 both the number of referee panels and Expert Committees will be reduced, which may help to broaden the knowledge base of the review process and allocation decisions. In general, the RCN funds interdisciplinary and emerging research through its more targeted discovery research funding schemes (Asbjorn Mo, personal communication, June 30, 2011).

7.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

For the past 5 to 10 years, the RCN's Division of Science has carried out comprehensive evaluations of scientific disciplines on a national level, similar to those done by the Academy of Finland. On average, two subjects are evaluated per year by international expert panels appointed by the RCN, and the evaluations cover five-year periods. The expert panels analyze information gathered through self-assessments by relevant research groups, meetings, site visits, and bibliometric analyses (RCN, 2009). The terms of reference for the most recent evaluation (human geography) are presented in Figure 1. As an example, the objectives of the evaluations for chemistry and physics are to review the strengths and weaknesses of research in the field, review the scientific quality of the discovery research in an international context, identify research groups that have achieved a high international level, identify areas of research that should be strengthened, and provide information to help the RCN in recruitment issues (RCN, 2009, 2010).

Each research group is also graded on a 5-point scale under three criteria: i) scientific quality and productivity; ii) relevance and societal impact; and iii) strategy, organization, and research cooperation, where a grade of 5 is "excellent," and described as follows: "Internationally leading position, undertaking original research and publishing in the best international journals. High productivity (including number of Ph.D. theses awarded). Clear and convincing strategy and future planning. Very positive overall impression of the research group and leadership" (RCN, 2009).

Following each evaluation, the RCN appoints a committee to develop a follow-up to the evaluation, including actions taken by universities, the RCN, and Norway's Ministry of Education and Research. Although the evaluations do not systematically dictate funding, many of these recommendations are directed towards funding decisions. For example, the evaluation of physics in 2010 included a recommendation to continue funding space research; the RCN increased funding in this area by 10 per cent. The eScience research program, a research and infrastructure program designed to address computing- and data-intensive challenges in science, technology, and medicine, was partly the result of 2002 evaluations of ICT and mathematics (Asbjorn Mo, personal communication, June 30, 2011).

7.5 USE OF SCIENCE INDICATORS

Bibliometric indicators are used in several capacities in Norway, including calculations of the core funding of higher education institutions, discipline evaluations, and a bi-annual national S&T report that serves as an information source on the status of Norwegian S&T for research policy-makers. This report is a collaboration of the RCN, the Norwegian Institute for Studies in Innovation, Research and Education (NIFU STEP), and Statistics Norway. It presents indicators in four key areas: R&D expenditures, human resources devoted to S&T, collaboration patterns, and outputs of R&D and innovation.

Currently, two input and two output quantitative indicators determine the performance-based portion of university core funding (OECD, 2010a). The output indicators are publication points and number of PhDs awarded. The publication points are calculated by adjusting publication numbers on three levels: quality of the journal, share of authorship, and publication form (e.g., book, article, etc.). Field differences in publication trends are not accounted for (OECD, 2010a). The input indicators, which are seen as indicators of quality, consist of the amount of funding from the EU Framework Programme for Research (which is seen as an incentive for institutions to compete on a European level) and the amount of RCN funding (OECD, 2010a).

The discipline evaluations rely primarily on international expert judgment, which is informed by qualitative and quantitative information provided by research group self-assessments, meetings, site visits, and commissioned bibliometric analyses. The qualitative information provided by research group self-assessments includes CV and publication lists for all staff in academic positions; SWOT (strengths, weaknesses, opportunities and threats) analyses; and information about organization, resources, collaborations, development, and future plans (RCN, 2009). Quantitative information gathered by the self-assessments includes numbers of personnel and graduates and R&D expenditures by source of funding. The bibliometric analyses are prepared by NIFU STEP and serve as background information for the evaluations and supplements to peer review in assessment of research performance (Aksnes, 2009). Bibliometric analyses are done at a variety of levels including national (with international comparisons), institutional, departmental, and division/research group. Indicators used include number of publications (including distributions of articles by sub-field, and number of publications by person, year, and research man-year); number of citations; relative citation indexes (including distributions by sub-field); co-authorship; and journal profiles such as impact factor.

As noted above, for assessment of research groups, different indicators are used to inform the 5-point rating scale for each of the three criteria (scientific quality and productivity; relevance and societal impact; strategy, organization and research cooperation). In determining the rating for "scientific quality and productivity," the following is considered: internationally applied standards for scientific quality based on bibliometric analysis, number of PhDs, Masters students and grades awarded, and participation in conferences (RCN, 2009).

7.6 OBSERVATIONS AND LESSONS LEARNED

In 2009 the Ministry of Education and Research evaluated the Norwegian funding model and saw no need for major changes, although some minor changes were made to take effect in 2012 (OECD, 2010a). Two issues, however, arising from use of quantitative indicators in the performance-based component of core funding to universities have been discussed. The first is the potential discrimination against fields of science where production of articles is not the norm, especially since no qualitative assessment explicitly feeds into the funding formula. The second has to do with incentive effects. The 2009 evaluation found that both publication counts and number of PhD students have increased in Norway, and thus the indicators may have resulted in positive incentive effects (although other factors undoubtedly contributed, such as use of publication count indicators for internal management at universities) (OECD, 2010a). Concern has been expressed that quality of publications would be sacrificed for number, but the Ministry of Education believes this risk is reduced by adjustment of the publication count based on quality of the journal (OECD, 2010a).

Dag Aksnes (2009) of NIFU STEP observed that use of bibliometrics in the RCN's discipline evaluations benefits the evaluation in several ways (in his presentation to the 14th Nordic Workshop on Bibliometrics and Research Policy in Stockholm 2009). He pointed out that although indicators cannot replace an assessment carried out by peers, they save time for the expert panel, and, more importantly, they give the panel increased credibility by providing an objective analysis that can counteract shortcomings or mistakes in the panel's inherent subjectivity, therefore contributing to fairness. Aksnes stated that although the application of bibliometrics is somewhat controversial in Norway, it does not arouse controversy when used to supplement peer review in discipline evaluations.

7.7 REFERENCES

Aksnes, D. W. (2009, September 29 - 30). *The Use of Bibliometric Indicators in Research Evaluations in Norway*. Paper presented at the 14th Nordic Workshop on Bibliometrics and Research Policy, Stockholm, Sweden Retrieved May 2012, from http://vr.se/download/18.227c330c123c73dc586800010980/12_Aksnes_Bibliometric_indicators_in_research_evaluations_in_Norway.pdf

European Commission (2010a). ERAWATCH National Profiles: Norway Retrieved March, 2012, from http://erawatch.jrc.ec.europa.eu/erawatch/opencms/information/country_pages/no/country

European Commission (2010b). Country Reports 2010: Norway Retrieved May 2012, from http://erawatch.jrc.ec.europa.eu/erawatch/export/sites/default/galleries/generic_files/X04-CR2010-NO-v2.pdfOECD (Organization for Economic Co-operation and Development) (2010a). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, Paris, France.

OECD (Organization for Economic Co-operation and Development) (2010b). Profile of Norway *Science, Technology and Industry Outlook 2010*. Paris, France: OECD.

OECD (Organization for Economic Co-operation and Development) (2010b). Web Annex: Additional Country Detail (Performance-based Funding for Public Research in Tertiary Education Institutions). Paris, France: OECD.

RCN (Research Council of Norway) (2009). Basic Chemistry Research in Norway (E. D. f. Science, Trans.). Oslo: RCN.

RCN (Research Council of Norway) (2010). Basic Physics Research in Norway: An Evaluation (E. D. f. Science, Trans.). Oslo: RCN.

RCN (Research Council of Norway) (2011). Funding Schemes Retrieved Apr. 2011, from http://www.forskningsradet.no/en/Funding_schemes/1138882212929

Technopolis (2001). A Singular Council - Evaluation of the Research Council of Norway. Norway: Technopolis.

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

8 Singapore

Singapore — Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP):	$3.0 billion
• GERD as a % of GDP:	2.52%
• Total Researchers (Full-Time Equivalent, 2007):	27,301
• # of Researchers per Million People (2007):	6,088
• # of Scientific Publications (2008):	6,813

Source: UNESCO Science Report, 2010

8.1 NATIONAL RESEARCH FUNDING CONTEXT

Singapore's public R&D funding is organized mainly along two tracks: (i) mission-oriented research at both public and private organizations funded largely by the Agency for Science, Technology and Research (A*STAR) and the Economic Development Board (EDB) under the Ministry of Trade and Industry (MTI); and (ii) academic, investigator-led discovery research at universities funded by the Academic Research Fund (AcRF) through the Ministry of Education (MTI, 2006). In addition, the National Research Foundation (NRF), founded in 2006 under the Prime Minister's Office, is designed to provide coherent strategic direction through coordination of research programs of different national agencies. The NRF also funds longer-term strategic research programs that align with national strategic initiatives (MIT, 2006). Singapore identified that, as a small country, its focus should be on creating excellence in key areas to be a competitor on a global scale. As such, a high percentage of the public R&D budget goes towards funding research in strategic areas. In 2009, 13.73 per cent of R&D expenditure from Foreign Governments and International Organizations was directed to the Private Sector (A*STAR, 2010), and the Ministry of Education received only 7 per cent of the R&D budget in 2006–2010 (MTI, 2006).

8.2 NATIONAL RESEARCH PRIORITIES

In 1991 the Ministry of Trade and Industry started publishing five-year S&T plans that lay out national priorities. The latest was developed by the Research, Innovation and Enterprise Council, a public and private council chaired by the prime minister that advises the Cabinet on national research and innovation policies (RIE, 2011). The 2011–2015 plan sets out five national research priorities:

electronics (specifically data storage and semiconductors), biomedical sciences, infocommunications and media, engineering (specifically precision and transport engineering), and clean technologies (water and solar). A*STAR funding decisions are based on the strategies laid out in the S&T plans (MTI, 2006). As well, the National Research Foundation holds calls for proposals in three of these priority areas (biomedical sciences, clean technologies, and interactive and digital media). Research funded by the Academic Research Fund is not tied to national strategic initiatives (personal communication, Wong Jee Pheng, August 25, 2011).

8.3 RESEARCH FUNDING ALLOCATION

The AcRF funds block grants to Singapore's four universities (Tier 1 funding, budget ~C$120 million[19] for the 2011–2015 budget year) as well as competitive project-based funding for discovery research (Tier 2, budget ~C$195 million for the 2011–2015 budget year and Tier 3, budget ~C$180 million for the 2011–2015 budget year). Budget allocation for block funding from Tier 1 among universities is currently based on historical funding amounts, although this process is currently under review and may in the future take into account indicators based on research publications and numbers of research-active faculty. Once universities receive Tier 1 funding, they have autonomy in how they distribute funding, and often administer their own competitive funding across faculties.

Tier 2 competitive funding is the main funding source in Singapore for academic, "blue sky" research. Each Tier 2 proposal must have a total project value between ~C$400,000 and C$800,000 over three years; otherwise, researchers are encouraged to compete in intra-university competitions for Tier 1 funding. To take into account differing costs of research across the five discipline clusters, however, the minimum funding request is lower for pure mathematics and statistics projects under EP2 (minimum ~C$200,000 over three years) and for projects under EP4 (minimum ~C$120,000 over three years).

Tier 3 funding was introduced in 2011, and funds high-impact, multidisciplinary research projects aiming to find integrative solutions to scientific issues. Tier 3 funding consists of very large grants (~C$20 million over five years, about 25 to 50 times the size of funding for Tier 2 projects) (personal communication, Wong Jee Pheng, August 25, 2011).

19 All C$ numbers on based on conversion from Singapore dollar to Canadian dollar on August 30, 2011.

For Tier 2 funding, proposal evaluation and budget allocation occur in four stages. In the first stage, each individual project proposal is reviewed by two members from the relevant expert panel (EP). There are currently five EPs based on discipline clusters: Chemistry and Chemical Engineering (EP1); Informatics and Mathematics (EP2); Biomedical Engineering and Life Sciences (EP3); Accountancy, Business, Humanities and Social Sciences (EP4); and Physics and Engineering (EP5). Each EP consists of 5–10 members, including a chairperson and local or international academics. When specific subject expertise is not available within the EP, input from international peer reviews is sought. If primary investigators consider a proposal to be interdisciplinary, they specify primary and secondary EPs. Each proposal is assessed on the scientific significance of the research, research approach, track record of the team, impact of the research environment on the project (research environment should be conducive to the success of the project), and project execution.

In the second stage, each EP meets to discuss the proposals and reviews, and to rank the proposals. In the third stage, the Ministry of Education allocates funds across the five EPs such that success rates across discipline clusters are similar.[20] It does so because it has not yet found an acceptable way to compare quality across disciplines. There is some flexibility in the system in that if an EP believes a set of proposals to be of exceptional quality, it can ask the ministry administration to adjust the success rate for that EP slightly upwards.

In the fourth stage, an Academic Research Council (ARC), which consists of 11 international, distinguished academics, endorses the funding recommendations, which are then approved by the Academic Research Board (the ARC chairman and the Ministry of Education's permanent secretary) (personal communication, Wong Jee Pheng, August 25, 2011).

8.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

Although there are no formal discipline evaluations in Singapore,, a number of national reviews and foresight exercises include trends at the field level. The national S&T plans are informed by months of discussions involving researchers from universities, research institutions, and hospital and industry representatives (MTI, 2006). In the 2006–2010 plan 17 expert panels were charged with different tasks. Of these, 13 panels performed foresight and technology scans according

20 The success rate is expected to be around 20 per cent. A budget increase of 60 per cent for the 2011-2015 budget term only resulted in a 5 per cent increase in success rate (from 15 per cent to 20 per cent) due to a large expansion in the university sector between 2006 and 2010.

to research area to identify new and emerging technology areas. A Collective Foresight Committee, made up of the chairpersons from the technology scan panels, was tasked with identifying overarching themes that emerged as a result of the individual scans. Other topics explored by expert panels included the role of private-sector research, ways in which public S&T infrastructure could be better shared by the research community, and ways in which government research institutes and universities could improve collaboration.

The National Survey of R&D in Singapore was conducted by the Singapore Science Council on a triennial basis from 1978 to 1987. Since 1990, it has been conducted and published annually by the Agency for Science, Technology and Research (formerly the National Science and Technology Board) (A*STAR, 2009). For the reviews, quantitative data are collected by surveying all organizations known to perform R&D, which in 2009 included 60 public institutions and 854 private-sector enterprises. There is some analysis at the level of fields, but only involving expenditures by field classified according to Singapore's priority areas.

The NRF participates in foresighting through its Competitive Research Program (CRP) International Evaluation Panel. One of the panel's roles is to provide insight on new emerging R&D areas, global R&D trends, and gaps in Singapore's R&D landscape (NRF, 2010). This may be tied to funding decisions, since the panel also advises the NRF on the overall funding scheme of the CRP.

8.5 USE OF SCIENCE INDICATORS

Quantitative indicators do not appear to be directly tied to funding allocation decisions in Singapore. Indicators are used, however, to measure progress and set targets on a national level and at funding agencies. Annual reviews of national R&D provide an overview of R&D data in the public and private spheres including expenditures, manpower, degrees granted, and patents (A*STAR, 2009). The 2006–2010 national S&T strategy collected data on number of researchers, R&D expenditures, and patents by field and by sector (private, government, higher education, public research institutes) (MTI, 2006). This strategy also set out national and agency-specific five–year S&T targets according to specific indicators. National targets were based on three indicators: GERD as a percentage of GDP, private-sector contributions to GERD, and research manpower (MTI, 2006). A*STAR target indicators included number of PhD students trained and graduated, number of government research institute staff spun out to locally based industry, number of primary patent applications, number of papers published, number of projects with industry, and industry funding levels (MTI, 2006).

Progress measurements at universities are informal and seem to rely on international rankings. The most recent S&T strategy cited the position of Singapore's universities relative to Thomson Reuters'- most cited institutions and *Nature's* Asia-Pacific Publishing Rankings (RIE, 2011). The government's response to a citizen's question on how universities measure the success of their research referenced the Times Higher Education Supplement (THES) and the Shanghai Rankings of World Universities — rankings based on indicators such as number of citations, number of publications, and number of patents (MOE, 2008).

8.6 OBSERVATIONS AND LESSONS LEARNED

The majority of research funding in Singapore is allocated according to national S&T priorities through A*STAR, EDB, or the NRF. There is some support for "blue sky" discovery research through the Academic Research Fund at the Ministry of Education. The ministry, however, has not found a satisfactory way to measure research quality across disciplines, and thus takes the egalitarian approach of ensuring equal success rates across discipline clusters. Since there are no formal discipline-specific evaluations of research performance, there may be little to learn from Singapore on use of indicators in research evaluation and budget allocation.

8.7 REFERENCES

A*STAR (Agency for Science Technology and Research) (2010). National Survey of R&D in Singapore 2009. Singapore: Agency for Science Technology and Research.

MOE (Ministry of Education for Singapore) (2008). Parliamentary Replies: Research in Universities Retrieved Apr. 2011, from http://www.moe.gov.sg/media/parliamentary-replies/2008/09/research-in-universities.php

MTI (Singapore Ministry of Trade and Industry) (2006). Science and Technology Plan 2010 (MTI, Trans.) (pp. 65). Singapore: MTI.

NRF(National Research Foundation of Singapore) (2010). Other Programmes and Initiatives: Competitive Research Program. Retrieved Apr. 2011, from http://www.nrf.gov.sg/NRF/otherprogrammes.aspx?id=178

RIE (Research Innovation and Enterprise Secretariat) (2011). Research, Innovation and Enterprise 2015: Singapore's Future (M. o. T. a. Industry, Trans.). Singapore: Ministry of Trade and Industry.

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

9 South Korea

South Korea — Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$41.3 billion
• GERD as a % of GDP:	3.21%
• Total Researchers (Full-Time Equivalent, 2007):	221,928
• # of Researchers per Million People (2007):	4,627
• # of Scientific Publications (2008):	32,781
• World Share of Exports in High-Technology Products (2007):	6.0%

Source: UNESCO Science Report, 2010

9.1 NATIONAL RESEARCH FUNDING CONTEXT

The public research funding system in South Korea (the Republic of Korea)[21] is undergoing transition. In the past, the majority of government-sponsored research in Korea was application-oriented and took place at Government Research Institutes (GRIs) rather than in universities (OECD, 2009). The government has increasingly focused on strengthening basic R&D capacity in the country to encourage more diverse project-based research in universities (OECD, 2009; Government of the Republic of Korea, 2008). In 2009 the government restructured several key agencies[22] that supported basic R&D, integrating them into a single agency, the National Research Foundation of Korea (NRF) (NRF, 2010). This leaves the NRF as the primary agency responsible for supporting discovery research in the natural sciences and engineering, through a range of funding programs and calls for proposals (see NRF, 2010). Other government departments and agencies continue to operate their own R&D funding programs, most of which are targeted towards technology development.

9.2 NATIONAL RESEARCH PRIORITIES

Korea has a tradition of establishing detailed government R&D priorities, particularly for applied research and technology development in major industrial sectors. Current research priorities are laid out in the government's latest S&T

21 "South Korea," "Korea," and "The Republic of Korea" are used interchangeably throughout.

22 These agencies included the Korea Science and Engineering Foundation, the Korea Research Foundation, and the Korea Foundation for International Cooperation of Science and Technology.

strategy, which aims to position Korea as one of the seven major S&T powers in the world (Government of the Republic of Korea, 2008). Established in 2008, this initiative identifies 50 critical technologies and 40 candidate technologies in seven major technology areas that the government will target for support. These technology and research priorities are highly specific compared to similar lists of national priorities developed by other governments. For example, within the category of "Key Industrial Technologies," "High-Precision Micro-machining and instrumentation control technology" is targeted for support. These priority technologies inform development of discovery research funding programs at the NRF and other departments.

Korea traditionally uses technology roadmaps to inform public R&D investments. In 2006 the government combined several such roadmap initiatives into a single, unifying national technology roadmap, the Total R&D Roadmap (OECD, 2009). This roadmap included medium- and long-term national R&D strategies, guidelines on strengthening discovery research capacity, and general guidance on coordination of government investments in R&D infrastructure and facilities (OECD, 2009). In addition to these technology development priorities and tools, the National Science and Technology Commission (NSTC) serves in an advisory role to the government in establishment of R&D priorities and science and technology policy in general (OECD, 2007).

9.3 RESEARCH FUNDING ALLOCATION

In 1999 the Ministry of Science and Technology (MOST) of Korea introduced a "pre-budget coordination" procedure, completed framing the "overall coordination" system of national R&D programs, and created the NSTC to support these activities (Oh & Kim, 2006). The NSTC draws up a broad spectrum of S&T policies and provides general planning for R&D programs funded by the government while the Korean Institute of Science and Technology (KISTEP) (also established in 1999) provides the actual coordination practice including impartial and objective evaluation and pre-budget coordination of those programs (Oh & Kim, 2006).

Since this evaluation and coordination process is program based, it also determines to a large extent internal allocation of resources within departments. The NRF, which receives its budget allocation directly from the Korean Ministry of Education, Science, and Technology (MEST), operates a wide range of funding programs that support discovery research. The majority of NRF funding programs are project based and rely primarily on peer review of grant applications. Others focus on providing institutional support to universities and other research centres based on established policy goals. In the past, university researchers believed that existing funding programs were biased against them in favour of researchers at GRIs, and, as a result, the government has expanded research funding opportunities for individual investigators and small groups (OECD, 2009).

9.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

At the national level, instead of administering systematic evaluations or assessments of research fields in the natural sciences and engineering, South Korea has a well-developed evaluation system for government R&D programs, which is directly implicated in R&D funding decisions and budget allocation. Government R&D programs have been systematically evaluated since the 1990s, and in 2006 a new evaluation system, the National Evaluation System of R&D, was introduced (OECD, 2009). This system is based on an internal evaluation completed by the responsible department, and a meta-evaluation and focused-evaluation completed by the NSTC (OECD, 2009).

The National Evaluation System evaluates government R&D programs by program type rather than by research field. The type corresponds to a program's objectives and orientation on the spectrum of discovery to applied research rather

Table 1

R&D Program Classification in South Korea

Major Classifications	15 Sub-Groups
R&D programs for basic, public and welfare technology	Mission oriented basic technology Public technology Welfare-related technology
R&D programs for industrial technology	Short-term industrial technology Middle and long-term industrial technology
R&D infrastructure	International co-operation, Development of human resources, Regional R&D centres of excellence R&D facilities and equipment
Support for public research institutes	National laboratories (3 sub-groups) Government supported research institutes for basic technology Government supported research institutes for industrial technology Government supported research institutes for public technology

Source: Oh & Kim, 2006

than to the field of research itself. There are four major classifications of R&D programs and 15 sub-groups (see Table 1). Each program category is evaluated by a committee of expert reviewers. This evaluation process is managed by KISTEP, and completed program evaluations are fed into the overall R&D priority-setting and budget planning process orchestrated by the NSTC.

In the past, R&D program evaluations occurred annually; however, the government has recognized the burden this creates for program managers and indicated in the *577 Initiative* that it would switch to a three-year evaluation cycle (OECD, 2007).

9.5 USE OF SCIENCE INDICATORS

There is no specific set of indicators or procedures to evaluate research fields or disciplines in Korea. The government, however, uses S&T-specific sets of indicators for research evaluation functions.

At the national level, the Korean government has identified a number of indicators to be used as national S&T performance targets. These indicators are listed in the *577 Initiative* with clear targets set for the period 2008–2012. Two general indices are included: an output index and an outcome index. The output index comprises four constituent indicators: average citations per paper in the Science Citation Index, international patent applications, the transfer ratio of publicly owned technology, and the ratio of technological balance of payments. The outcome index is based on three underlying indicators: R&D contribution to economic growth, HRST occupations as a percentage of total employment, and science and technology competitiveness scores from the International Institute for Management Development's (IMD) *World Competitiveness Yearbook*.

KISTEP uses a qualitative approach based on expert review to evaluate government R&D programs. Fifteen expert review committees are established, composed of non-government experts, to review programs of various types (see Table 1). Evaluation committees are then tasked to assess their respective programs on six general criteria: validity of program contents, efficiency of program management, effectiveness of program results, necessity of program, impact and utility of program, and appropriateness of budget size (evaluation committees have some discretion to modify these standard criteria) (OECD, 2007 from Oh & Kim, 2006). Table 2 provides a list of key questions considered by committee members in this process. The committees then score and complete a written evaluation of each program. Traditionally, programs were also given a final overall grade (OECD, 2007).

KISTEP representatives also report using a variety of indicators to identify emerging research areas and trends. According to a senior KISTEP official, the organization uses qualitative procedures (expert panels and surveys of experts) as well as quantitative indicators such as funding application trends by topic, publication trends by topic, and keyword analysis (D. Oh, personal communication, May 26, 2011). Indicators related to student or researcher population are not used.

Table 2

Criteria and questions used in reviews of government R&D programs in Korea

Criteria	Key Questions
Validity of program contents	• Are aim and scope of the program appropriate? • Does the program feature contemporary economical and social environments? • Are sub-projects in the program overlapping one another?
Efficiency of program management	• Is the detailed practice plan of the program set up systematically and strategically? • Is the conduct of the program efficient? • Does the program procedure concur with the practice plan? • Is the program carried out in cooperation with other stakeholders? • Is the budget for the program spent and distributed in an efficient way? • Are suggestions and recommendations presented in the previous year considered effectively?
Effectiveness in getting program results	• Does the program reach the main purpose (goal) for the appointed fiscal year? • Will the program be able to attain the final goal of the program in the future? • What are the scientific and technological achievements? • How effective the program is in nurturing human resources in R&D? • How effective is the program in building R&D infrastructure? • Does the program contribute to the strengthening of industrial competitiveness of the country? • Is the program conducive to the promotion of public welfare? • Are the research activities in the R&D institute suitable for its given mission?
Necessity of program	• Is the program better carried out by a private sector or does the program need the government's support? • Does the program concur with the government's strategic S&T policy?

continued on next page

Table 2

Criteria and questions used in reviews of government R&D programs in Korea

Criteria	Key Questions
Impact and utility of program	• What's the direct benefit from supporting the program? • What are the consequences that the technological development of the program incurs?
Appropriateness for budget size	• Is the amount of a budget requested for the program reasonable? • Is it necessary to reduce or raise a budget for the program?

Source: Oh & Kim, 2006

KISTEP also undertakes other activities related to S&T evaluation and planning, including technology foresighting studies and exercises, which may employ a range of S&T indicators. For example, KISTEP identifies 10 emerging technologies every year using a different set of methods and criteria — thus allowing an exploration of methodologies and emerging research trends (Lee, 2011).

9.6 OBSERVATIONS AND LESSONS LEARNED

South Korea does not regularly employ indicators in assessment of research fields at the national level. The government's model of R&D program evaluation based on expert review, however, appears to be regarded as generally successful, and has been recommended as a model for other countries (Oh & Kim, 2006). Historically, this model has contended with difficulties: finding competent and credible experts to serve on evaluation committees, frequency of evaluation, and absolute versus relative assessments of programs (OECD, 2007). In each of case, the model has been adjusted over the past seven years, and there is no indication of converting to a different assessment system.

The South Korean approach to science assessment and research funding can best be characterized as a combination of regular program evaluations based on expert review, and detailed technology development roadmaps and priority-setting, which are used to shape the development and direction of government research programs. This model has served Korea well and has led to an impressive record of technological and economic development. It remains to be seen whether the same assessment process and research priority-setting mechanisms will be equally suited to assessing the country's efforts in expanding discovery research capacity in the natural sciences and engineering.

9.7 REFERENCES

Government of the Republic of Korea (2008). Becoming a S&T Power Nation through the 577 Initiative: Science and Technology Basic Plan of the Lee Myung Bak Administration (S. a. T. Ministry of Education, Korea Institute of S&T Evaluation and Planning, Trans.). Seoul, South Korea: Ministry of Education, Science and Technology; Korea Institute of S&T Evaluation and Planning.

Lee, J. S. (2011). *Korea's S&T Knowledge-Sharing Activities with Developing Countries*. Paper presented at the AAAS, Washington (DC). Presentation retrieved from

NRF (National Research Foundation of Korea) (2010). About NRF Retrieved April 12, 2011, from http://www.nrf.re.kr/html/en/about/about_02_03.html

OECD (Organization for Economic Cooperation and Development) (2009). OECD Reviews of Innovation Policy: Korea. Paris, France: OECD.

Oh, D., & Kim, Y. (2006). Overall coordination of government-funded research and development programs in Korea. *Journal of MultiDisciplinary Evaluation, 3*(5), 153–170.

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

10 United Kingdom

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP):	$41.0 billion
• GERD as a % of GDP:	1.88%
• Total Researchers (Full-Time Equivalent, 2007):	261,406
• # of Researchers per Million People (2007):	4,269
• # of Scientific Publications (2008):	71,302
• World Share of Exports in High-Technology Products (2007):	3.6%

Source: UNESCO Science Report, 2010

10.1 NATIONAL RESEARCH FUNDING CONTEXT

The United Kingdom (U.K.) has a "dual-support" research funding system (HEFCE, 2011) providing for grants at both the institutional and project level. Institutional research funding is provided directly to universities and research-performing institutions by annual block grants from the Higher Education Funding Council of England (HEFCE) (or one of the equivalent regional funding councils). Total institutional funding for HEFCE now stands at approximately £7 billion annually. Roughly £1.6 billion of this is earmarked to support research, and this funding is allocated partially on the basis of quality assessments from the Research Assessment Exercise (RAE). Project-based funding is provided by seven research councils (the Research Councils United Kingdom, RCUK), with funding for the natural sciences and engineering coming from the Engineering and Physical Sciences Research Council (EPSRC).The research councils combined provide roughly £3 billion annually in research funding, approximately £850 million of which is dedicated to the natural sciences and engineering through EPSRC (EPSRC, 2011).

10.2 NATIONAL RESEARCH PRIORITIES

In general, U.K. research councils and universities appear to have a large amount of autonomy in setting funding priorities with no consistent, national set of research funding priorities. Research councils are responsible for determining their own funding programs and calls for proposals, and universities allocate institutional block funding from HEFCE however they deem most appropriate.

In April 2010 a House of Lords Science and Technology Committee report recommended that the new coalition government announce clear research funding priorities (Science and Technology Committee for the UK, 2010). The committee also recommended that the government improve the process by which science priorities are set. In response, the coalition government indicated that science and research funding priorities would not be announced independently, but would emerge from government fiscal planning in the Spending Review (Government of United Kingdom, 2010).

10.3 RESOURCE ALLOCATION PROCESS

Research council funds are filtered through funding programs and allocated on the basis of calls for proposals and peer review. EPSRC divides its programs into groups that focus on supporting investigator-led research (Research Base Programs) and those that focus on supporting research themes and "maximizing economic and social impacts" (Business Innovation Programs) (EPSRC, 2011). Institutional research funding from HEFCE is allocated using several formulas. The largest portion is targeted at supporting teachers and allocated primarily on the basis of number of students at receiving institutions. A substantial portion of HEFCE's research funding, however, is performance based or Quality-Related (QR), and allocated on the basis of the outcomes from the RAE — a large-scale national research assessment exercise based on informed peer review. This QR funding is first allocated across disciplines, based on assessment of the volume of research (as captured by number of research-active staff) that meets a certain research quality threshold (as measured by past RAE ratings), and the relative costs of research in different fields as well as research quality (HEFCE, 2010a). QR funding for each discipline is then allocated across institutions on the basis of RAE assessments for departments or research groups. (Final funding allocations for each field are combined in a lump-sum transfer to institutions, and university administrators have full discretion as to how that funding is allocated across units.)

10.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

Two types of national research assessments undertaken in the U.K. are relevant to evaluation of research performance at the field level.

First, U.K. research councils periodically undertake in-depth studies of research capacity and performance in specific fields. For example, in recent years EPSRC has undertaken a number of international review reports, field-based analyses using an international expert review model (EPSRC, 2011). These studies are based on models of informed, expert panel review and include analysis of data on research performance and research capacity in specific fields, as well as site visits and interviews with key informants. Some bibliometric data were frequently included for consideration in these reviews, but analysis was not taken to a high level of detail (Arnold *et al.*, 2005). For nearly all reviews, all participating reviewers were from countries other than the U.K., presumably to curtail any bias that might arise from direct affiliations with U.K. research institutions. There is no direct linkage between report findings and funding outcomes from these international reviews; however, findings may be used by EPSRC (or others) to inform policy priorities and research funding decisions in the future.

Second, the U.K. undertakes a large-scale national research evaluation, the RAE. The RAE was first carried out in 1986 and has since been repeated in 1989, 1992, 1996, 2001, and 2008. Its primary purpose is to inform allocation of institutional research funding through HEFCE, and incentivize and reward research quality by providing increased funding for high-quality (i.e., world-leading or internationally recognized) research. The assessment is based on evaluation of research fields at specific universities or institutions, and employs a peer review-based model. Research activities are categorized into approximately 60 to 70 research fields (referred to as Units of Assessment or UoA), and a panel of 10 to 15 experts is asked to evaluate the quality of research for each institution in a field.[23] Participating institutions are given a standard template to organize a submission to the panel in each area, and the Panel is then asked to provide a final quality rating for each submission. In the past, this rating was based on a seven-point scale; however, the most recent RAE in 2008 used only a five-point scale. RAE assessments correspond to evaluation at the field (as defined by the UoA) and the institutional level. In some cases, however, RAE scores may be aggregated across a range of institutions to assess the U.K.'s overall level of research quality in an area.

The RAE has evolved over time, and will in the future be replaced with a similar, though modified exercise, the Research Excellence Framework (REF). The first REF is scheduled to take place in 2014. While the REF will remain based on a peer/expert review model of assessment, it is expected to make additional use of bibliometric data at the discretion of individual review panels. The REF will also include consideration of research impact (to be evaluated through case studies submitted by participating institutions) and the vitality of the research environment.

23 RAE Units of Assessments generally correspond to "cost centres" within institutions, which generally equate to departments or schools, rather than research fields that may have researchers from many disciplines.

10.5 USE OF SCIENCE INDICATORS

Expert reviews of research fields undertaken by EPSRC and other U.K. research councils typically include consideration of all available data on a field, though review panels may have difficulty fully analyzing all available data (Arnold *et al.*, 2005). Common sources of data include U.K. policy documents, human resource data for a field such as numbers of students and researchers, existing field-specific reports, bibliometric data, and technology foresight reports where available (Arnold *et al.*, 2005). Review panels may make use of past RAE scores and reports relevant to the field. In addition, several panels have conducted surveys of researchers in the field to supplement existing information (Arnold *et al.*, 2005).[24]

The RAE is based on informed, expert review. Panels evaluate research quality in particular fields (UoAs) based on submissions from participating institutions. For the RAE 2008, universities made submissions in a standard format containing information about current staffing, details of publications and other forms of assessable outputs produced between 2001 and 2007, and data about research students and research income with a textual commentary relating to the six-year assessment period. Submissions were not required to be directly related to administrative units, but had to be related to a body of research activity within a unit, research, or department. Based on information in these submissions, the review panel then provided an overall ranking of research quality for that institution in that unit of assessment. In general, overall quality rankings are geared towards identifying areas in which U.K research achieves a high level of international and national recognition of research excellence, with the highest rating (four stars) implying research that is "world leading in terms of originality, significance, and rigor." The quality bands and definitions used in the 2008 assessment are reproduced in Table 1.

As discussed previously, RAE assessments for institutions then feed directly into the allocation of annual institutional research funding from HEFCE. In addition to RAE scores, HEFCE allocations also take into account research volume (as captured by the number of research-active staff) and variations in the cost of research by field (HEFCE, 2010a). According to HEFCE, analysis of the cost of research is based on differences in the expenditure data reported by institutions to the U.K. Higher Education Statistics Agency.

24 Examples of these reports, along with the data and indicators used to support them, can be found on EPSRC's website at http://www.epsrc.ac.uk/pubs/reports/Pages/internationalreviews.aspx

The overall methodology of the REF — to be introduced in 2014 — will be broadly similar to the RAE, though it remains unclear what indicators or data will be used to assess environmental factors. It will remain based on informed peer/expert review; however, review panels may choose to incorporate additional bibliometric analysis into their review process, and HEFCE has been undertaking pilot studies to explore several models of bibliometric analysis, with various procedures for data collection and indicator construction (see HEFCE, 2011). In addition, the REF will employ a case study-based methodology to provide assessments of research impact (HEFCE, 2010b), where institutions will be asked to identify instances where research they have carried out has resulted in broader, non-academic impacts. Case study submissions will then be assessed *within* research fields (i.e., units of assessment) based on a standardized qualitative rating scale.

Table 1

RAE 2008 Research Quality Assessment Scale

Rating	Definition
4*	Quality that is world-leading in terms of originality, significance and rigour
3*	Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence
2*	Quality that is recognized internationally in terms of originality, significance and rigour
1*	Quality that is recognised nationally in terms of originality, significance and rigour
Unclassified Quality	that falls below the standard of nationally recognized work. Or work which does not meet the published definition of research for the purposes of this assessment

Source: RAE (2011).

10.6 OBSERVATIONS AND LESSONS LEARNED

As the most widely studied and analyzed example of a large-scale, performance-based research funding regime, the RAE is a valuable source of observations and lessons learned.

First, some initial observations can be made about the international reviews conducted by EPSRC and other research councils. A 2005 review of the first cycle of EPSRC's international reviews concluded that these field-based evaluations had yielded useful insights into the structural characteristics of the fields surveyed, insights of a type that do not emerge in the RAE (Arnold *et al.*, 2005). The review also found, however, that the U.K. scientific community had expressed concern that the role of the reviews was not clear and that they had few visible consequences, even if it is said they do influence debate and science policy more generally. Other areas of possible improvement included additional consideration of panel composition, particularly with respect to coverage of research fields and sub-fields; more fully incorporating other types of data into the analysis; and providing additional analytical and policy support to the panel.

The RAE has been subject to a range of internal, government-sponsored reviews and external scrutiny from academics and other stakeholders. A prominent government-sponsored review by an independent commission led by Sir Gareth Roberts (2003) T validated the overall approach and concluded that, in its current form, the RAE was a qualified success. The review made a number of substantive recommendations resulting in significant changes between the 2001 and 2008 RAEs. Roberts strongly emphasized that any future evaluations should remain founded on a peer review process, stating that "we [the panel] are now convinced that the only system which will enjoy both the confidence and the consent of the academic community is one based ultimately upon expert review. We are also convinced that only a system based ultimately upon expert judgement is sufficiently resistant to unintended behavioural consequences to prevent distorting the very nature of research activity" (Roberts, 2003).

The panel review also recommended a change from a seven-point grading scale for overall quality assessments to a quality profile where the percentage of research-active staff qualifying for each quality rating would be identified. This recommendation was later implemented in the 2008 RAE, and resulted in more blended assessments of quality at each institution, which also had a significant impact on funding (OECD 2010). The panel recommended more clarity and transparency in recruitment of panellists and in definition of the criteria by which panellists adjudicated submissions. Several recommendations were not acted upon, including not assessing non-research intensive universities, using bibliometrics to analyze middle-tier universities, and reserving intensive peer review only for top research universities that receive the bulk of research funding.

Many other studies have examined the impacts of the RAE on the U.K. research system. A recent OECD (2010) review provides a useful overview of the literature. Other relevant analyses include McNay (1998), Evaluation Associates Ltd. (1999), and Corbyn (2009). Some key findings from these studies include:

- Impact on funding: Throughout most of its history, the RAE has resulted in an increasingly concentrated distribution of research funding, where a select group of institutions (particularly the research-intensive Russell Group) has received a greater share of HEFCE's QR block grants. This pattern changed, however, in the 2008 RAE due to the introduction of graduated quality profiles versus the former single rating. See Corbyn (2009) and Adams & Gurney (2010) for discussions of RAE impact on funding selectivity.
- Impacts on researchers: The introduction of the RAE has led to significant changes in overall patterns of human resource management and contracting in U.K. universities. In general, it resulted in the creation of a "transfer market" in which universities compete to attract high-quality researchers in advance of upcoming RAE assessments. The development of this market led to an increasing variety of positions, job descriptions, and contracting options for academic researchers. The RAE may have had significant impacts on staff promotion and movement within institutions as well as between them (HEFCE, 1997). Researcher morale and collegiality were also found to be adversely affected in some cases due to sensitivities around which personnel within a department were selected as "research-active" staff in RAE submissions (HEFCE, 1997). Institutions do not always respond to pressures from the RAE in the same way. Evidence suggests that the RAE led some universities to focus on hiring younger staff with research potential, while others have taken a more conservative approach and focused on hiring well-established researchers (OECD, 2010; HEFCE, 1997).
- Impacts on research productivity: Moed (2008) reports evidence suggesting that the RAE has had substantial impacts on research output and productivity in the U.K. — impacts that have varied through time depending on different emphases of successive assessment rounds. In the RAE's early years, total output of research publications in the U.K. increased. In response to greater emphasis on quality over quantity in the 1996 RAE, the U.K.'s share of world publications declined, but its share of articles in highly cited journals increased. Finally, in recent years the overall share of research publications appears to be increasing again, possibly in response to attempts to increase number of staff included as research active (OECD, 2010; Moed, 2008).
- Impacts on research quality: There is a general consensus within the U.K that the RAE has resulted in increased research quality (OECD, 2010; Government of United Kingdom, 2006). Bibliometric evidence appears to confirm this in

so far as the overall impact of U.K. research, as measured by citation patterns, has increased since the introduction of the RAE (see Adams & Gurney, 2010; Moed, 2008). There is no definitive proof of a causal relationship, but the correlation is suggestive. The RAE has also likely changed author behaviour and research publication patterns, with an increasing emphasis on publishing in well-known, highly cited scientific journals over other outlets (OECD, 2010; RIN, 2009; McNay, 1998).

- Impacts on research focus: A persistent concern in the U.K. is that the RAE is biased against multidisciplinary and interdisciplinary research (OECD, 2010; Government of the United Kingdom, 2006; Roberts, 2003). Two studies examined the issue and found no evidence to suggest this was the case. The first, a survey of U.K. researchers undertaken in 1998, found that few respondents reported moving away from interdisciplinary work in response to the RAE (McNay, 1998) despite the fact that many researchers reported being concerned about the issue. The second study, conducted by Evaluation Associates Ltd. (1999), found that departments with a high proportion of interdisciplinary research did not appear to be penalized in their RAE ratings. A more significant issue, however, may be that the RAE discourages "blue sky" research — especially in cases where researchers worry about not being able to generate research outputs in advance of the next round of assessment (see OECD, 2010; McNay, 1998; Evaluation Associates Ltd. 1999).
- Impacts on researcher autonomy and collaboration: Many researchers in the U.K. have reported their autonomy and ability to determine their own research direction has been eroded by the RAE. Institutional preoccupation with RAE scores and their funding implications has led university administrators and managers to increasingly control the overall research directions of their staff and departments (OECD, 2010; McNay 1998). Researchers., however, have also expressed concerns that the RAE does not do enough to encourage research collaboration, particularly with researchers outside higher education institutions (OECD, 2010; Evaluation Associates Ltd., 1999).
- Impacts on research departments: The RAE has clearly had substantial impacts on restructuring of research departments in the U.K., with several universities closing departments in response to poor ratings (OECD, 2010).

In response to many of the issues and concerns identified above, research assessment practices in the U.K. continue to evolve. In advance of implementation of the REF, the U.K. has undertaken additional pilot studies on various aspects of research assessment methodology. Two of these are worth highlighting. First, HEFCE conducted a series of pilot studies on use of bibliometrics in research evaluation. These studies explored three models of bibliometric analysis of research quality at participating institutions and concluded: "Bibliometrics are not sufficiently robust at this stage to be used formulaically or to replace expert review in the REF. However there is considerable scope for citation information to be used to inform expert review" (HEFCE, 2009).

Second, HEFCE completed a pilot study on use of a case study methodology to assess research impact (i.e., non-academic impacts from research). The results were deemed encouraging, and in response HEFCE and the other funding councils decided to include evaluation of research impacts in the REF based on this approach (HEFCE, 2010b). Additional lessons and observations may emerge in the future from ongoing research at HEFCE in preparation for the first round of the REF in 2014.

10.7 REFERENCES

Adams, J., & Gurney, K. (2010). Funding Selectivity, Concentration and Excellence - How Good is the UK's Research? Oxford, England: Higher Education Policy Institute.

Arnold, E., Muscio, A., & Zaman, R. (2005). Reviews Reviewed: Lessons from the First Six International Panel Reviews, 1999–2004. London, England: EPSRC.

Corbyn, Z. (2009). Reversal of fortunes. *Times Higher Education, 5 March, 2009*.

EPSRC (Engineering and Physical Sciences Research Council) (2011). International Review Reports Retrieved March 7, 2012, from http://www.epsrc.ac.uk/newsevents/pubs/reports/Pages/internationalreviews.aspx

Evaluation Associates Ltd. (1999). Interdisciplinary Research and the Research Assessment Exercise. Bristol, UK: HEFCE.

Gareth, R. (2003). Review of Research Assessment: Report by Sir Gareth Roberts to the UK Funding Bodies. London, UK: RA Review; Prepared for UK Funding Bodies.

Government of the United Kingdom (2010). Government Response to the House of Lords Science and Technology Select Committee Report: "Setting Priorities for Publicly Funded Research". London, UK: Coalition Government.

Government of United Kingdom (2006). Science and Innovation Investment Framework 2004 - 2014: Next Steps. London, UK: Department for Business, Innovation and Skills.

HEFCE (Higher Education Funding Council of England) (1997). The Impact of the 1992 Research Assessment Exercise on Higher Education Institutions in England. London, UK: HEFCE.

HEFCE (Higher Education Funding Council of England) (2009b). Report on the Pilot Exercise to Develop Bibliometric Indicators for the Research Excellence Framework. London, UK: HEFCE.

HEFCE (Higher Education Funding Council of England) (2010a). Guide to Funding: How HEFCE Allocates its Funds (pp. 56). London, UK: HEFCE.

HEFCE (Higher Education Funding Council of England) (2010b). Research Excellence Framework Impact Pilot Exercise: Findings of the Expert Panels. London, UK: HEFCE.

HEFCE (Higher Education Funding Council of England) (2011). REF2014: Research Excellence Framework. Retrieved April 13, 2011, from http://www.hefce.ac.uk/research/ref/

McNay, I. (1998). The research assessment exercise (RAE) and after: 'You never know how it will all turn out'. *Perspectives: Policy and Practice in Higher Education, 2*(1), 19–22.

Moed, H. F. (2008). UK research assessment exercises: Informed judgments on research quality or quantity? *Scientometrics, 74*(1), 153–161.

OECD (Organization for Economic Co-operation and Development) (2010). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, Paris, France.

RIN (Research Information Network) (2009). Communicating Knowledge: How and Why Researchers Publish and Disseminate Their Findings. London, UK: RIN.

Science and Technology Committee for the UK (2010). Setting Priorities for Publicly Funded Research. London, UK: House of Lords.

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

11 United States

Key Statistics

- Gross Expenditures on R&D (GERD), 2007 (PPP): $398.1 billion
- GERD as a % of GDP: 2.82%
- Total Researchers (Full-Time Equivalent, 2007): 1,425,550
- # of Researchers per Million People (2007): 4,663
- # of Scientific Publications (2008): 272,879
- World Share of Exports in High-Technology Products (2007): 13.2%

Source: UNESCO Science Report, 2010

11.1 NATIONAL RESEARCH FUNDING CONTEXT

The U.S. research funding system is extensive, complex, and generally decentralized. Public funding for research is provided by both federal and state governments, with the federal government providing nearly 60 per cent of research funding in universities (European Commission, 2010). The majority of federal funding is provided through project-based grants awarded by mission-oriented departments or agencies in response to calls for proposals. Of these departments, the National Science Foundation (NSF) is the most significant for funding discovery research in the natural sciences and engineering, and now provides roughly US$5.5 billion in funding for research and related activities annually (NSF, 2011). Many other federal departments and agencies also have substantial research expenditures, including the Department of Defense, the Office of Science within the Department of Energy, the National Institutes of Health (NIH), the National Aeronautics and Space Administration (NASA), the National Institute of Standards and Technology (NIST), etc.

Budgets for federal departments and agencies are established annually by budget appropriations from Congress, and are guided by the President's budget request (European Commission, 2010). Departments are then relatively autonomous in establishing research funding priorities and programs.

11.2 NATIONAL RESEARCH PRIORITIES

In 2009 the Obama administration published a national strategy for innovation, which was updated in 2011 (The White House, 2011). This strategy focuses on national priorities such as clean energy; biotechnology, nanotechnology, and advanced manufacturing; health care technologies; and education technologies (The White House, 2011). While this strategy does not have the force of law, these priorities can be expected to guide funding program in federal departments and agencies in the near future. For example, administrators at the NSF include funding programs (e.g., calls for proposals) connected to these priorities in their 2012 budget request (NSF, 2011). There has also been a commitment to double the budgets for the three key discovery research funding organizations (NSF, NIST, and the Office of Science at the Department of Energy) over the next decade (see USOMB, 2011).

11.3 RESOURCE ALLOCATION PROCESS

Most federal research funding in the United States is project based and distributed through granting programs that evaluate grants on the basis of peer review. The NSF receives approximately 55,000 grant applications every year, of which approximately 13,000 receive funding (NSF, 2011). In particular, the NSF approach to peer review is based on assessing all proposals relative to two general criteria: intellectual merit and broader impacts. The latter criterion specifically includes consideration of integration of research and education, and cultivation of diversity in NSF programs, projects, and activities (NSF, 2010). NSF program officers are also encouraged to fund high-risk research (NSF, 2010). Other federal departments and agencies such as the NIH, the Department of Defense, and the Department of Energy have their own models of peer review used to assess grant applications.[25]

11.4 ASSESSMENT & EVALUATION OF RESEARCH FIELDS

There is no large-scale systematic assessment of research performance by field in the United States that is directly tied to research funding decisions or allocation. The federal government does, however, directly or indirectly, support several science assessment initiatives worth noting.

25 Note that NIH funding includes substantial support for the behavioural and social sciences as well as the health and life sciences.

First, federal departments and agencies are subject to reporting and evaluation requirements created by passage of the *Government Performance and Results Act* (GPRA) in 1993 (updated in 2010), which requires federal departments to explicitly identify measurable performance indicators related to research funding (see Cozzens, 1997). Guidelines have since been developed to assist departments and agencies that fund discovery research in this regard, including development of federal R&D investment criteria (USOMB, 2003) and a Program Assessment Rating Tool (PART) (USOMB, 2006). Science-funding agencies like the NSF were also given authorization by the United States Office of Management and Budget (USOMB) to use alternative performance metrics under GPRA based on difficulties in assessing the outcomes of R&D investment in short timeframes. While these federal reporting and assessment requirements are not directly tied to field-level assessments of research performance, they directly influence approaches taken by federal departments and agencies to monitor performance of research investments.

Second, federal departments and agencies have undertaken a range of science assessment initiatives not directly tied to funding allocation. At the NSF, data on research inputs and outputs in the natural sciences and engineering (e.g., funding, students, faculty, publications) are collected and published in the biennial *Science and Engineering Indicators* series (e.g., National Science Board, 2010). Perhaps more relevant, the National Academies of Science and the National Research Council (NRC) periodically undertake a variety of assessments of scientific performance by field, including a series of international benchmarking studies of U.S. research capacity and performance by field compared to other countries (see NRC, 2000). These studies typically consist of an expert panel using informed judgment to evaluate research performance and capacity of the field based on in-depth analysis of all available data for that field (typically including consideration of inputs, such as research funding and infrastructure, and outputs, such as publications).

The NRC also periodically undertakes "decadal surveys" of research fields, which are comprehensive reviews of the state of research in a field or sub-field, generally aimed at informing development of future research priorities (see NRC, 2007a, 2007b, 2010). These studies are also founded on expert judgment rather than quantitative indicators, and generally include a comprehensive engagement/ consultation process with the relevant research community. While these studies are not directly tied to specific funding decisions, they will likely have significant impact on funding decisions and priorities within federal departments and agencies.

Third, the U.S. government is actively promoting new developments in science assessment and evaluation through a variety of means. In 2005 President George W. Bush's Science Advisor and Director of the Office of Science and Technology (OSTP) Policy, John Marburger III, called for a new "science of science policy," which would include creation of the data sets, tools, and methodologies needed to assist policy-makers in decisions about scientific research and funding. Since that time OSTP has remained engaged in this effort, and in 2008 commissioned a *Science of Science Policy: Federal Research Roadmap*, which outlines a federal research agenda for improving the science of science policy (see NSTC, 2008). OSTP maintains a website to disseminate information (http://scienceofsciencepolicy.net/), and continues to post new publications and resources on this site.

Finally, increased funding for scientific research included in the U.S. economic stimulus package (i.e., the *American Recovery and Reinvestment Act*) led to development of the STAR METRICS program to assess the impact of this additional research investment. STAR METRICS is an interdepartmental initiative (as well as a partnership between the federal government and participating universities), led by the NIH and NSF under auspices of OSTP, and focuses on developing "uniform, standardized, and auditable" science indicators related to socio-economic impacts of research funding (see Lane, 2010; NIH, 2010).

11.5 USE OF SCIENCE INDICATORS

Since the federal government has no single, systematic assessment of research fields, it does not use a standardized set of indicators. Different sets of indicators are used for different initiatives, depending on the methodology and intended objective.

The NSF *Science and Engineering Indicators* report data on a range of research inputs and outputs by field, including research funding (public and private), students, faculty, student/faculty migration, and publications. Publication trends in the sciences and engineering are reported by field and sub-field, but no attempt is made to link them to a rigorous assessment of either research quality or research impact (e.g., socio-economic impacts). This is consistent with the purpose of this series as presenting a broad range of data on U.S. science and engineering trends rather than assessing or evaluating those trends.

NRC international benchmarking studies use a variety of indicators. In particular, the NRC notes that the ability to tailor selection of indicators for specific fields is one of the strengths of the benchmarking methodology (NRC, 2000). The dominant evaluative methodology is expert judgment based on informed opinion of the expert panel. Any quantitative indicators used in these studies may inform the judgment of the panel, but final evaluations or conclusions need not be directly based on the included indicators. General methods used by these panels include a virtual congress, citation analysis, journal-publication analysis, analysis of other quantitative data (such as numbers of graduate students and degrees, employment status, etc.), analysis of scientific prizes and awards, and analysis of international congress speakers.

Regarding evaluations of research impact, the STAR METRICS program is divided into two phases. The first phase focuses entirely on developing measures of research funding impact on job creation. The second phase focuses on developing indicators in four main areas: economic growth (e.g., patents, spin-offs, etc.); workforce outcomes (student mobility and employment); scientific knowledge (e.g., publications and citations); and social outcomes (e.g., health outcomes and environmental impact measures).

Finally, government departments and agencies identify individual performance metrics and measures associated with departmental strategic outcomes under GPRA, using various tools, indicators, and guidelines. The NSF received authorization from the OMB to use an alternative format for assessing the agencies' performance under GRPA, based on challenges with evaluating R&D investments over short timescales. The NSF does, however, use a range of quantitative output indicators in assessing its activities, including published research results, journal publications, student participation and demographics, new tools and technologies, etc. (NSF, 2011).

11.6 OBSERVATIONS AND LESSONS LEARNED

The lack of a unified assessment system in the United States makes it difficult to extract a single observation or lesson learned. The NRC has reported initial experimental efforts with international benchmarking of disciplines to be generally regarded as successful and informative by participating panel members — in contrast to some panel members' initial expectations (NRC, 2000). Other NRC committees have continued to grapple with issues involved in research funding in the interim. One relevant example is a 2007 report on strategies for assessing science, which highlights the challenge of predicting future research performance based on past results, and emphasizes the usefulness of methods that combine the virtues of analytical and deliberative methods (NRC, 2006). Efforts underway at OSTP related to the science of science policy may yield useful insights in the future, but it appears too early to draw conclusions — especially for recent initiatives such as the STAR METRICS program. Some participants in these efforts are generally optimistic about the ability to significantly improve on existing approaches to measuring impacts of research (Lane, 2010).

The development of the science of science policy website and blog provides a clearing house for the science policy community to monitor (and comment on) ongoing work in this area. Overall, science assessment in the federal government has generally remained consistent with the conclusions of the Committee on Science, Engineering, and Public Policy (COSEPUP), which reviewed the issue of science assessment with respect to GPRA and concluded that informed, expert review and judgment was the most reliable existing method of evaluating federal investments in science (NRC, 1999).

11.7 REFERENCES

Cozzens, S. E. (1997). The knowledge pool: Measurement challenges in evaluation fundamental research programs *Evaluation and Program Planning, 20*(1), 77–89.

European Commission (2011). ERAWATCH National Profiles: United States Retrieved March, 2012, from http://erawatch.jrc.ec.europa.eu/erawatch/opencms/information/country_pages/us/country

Lane, J. (2010). Let's make science metrics more scientific. *Nature, 464*(7300), 488–489.

National Science Board (2010). Science and Engineering Indicators 2010 (pp. 566). Arlington (VA): NSF.

NIH (National Institutes of Health) (2010). STAR METRICS: A Federal Collaboration with Research Institutions. Retrieved April 5, 2011, from https://www.starmetrics.nih.gov/

NRC (National Research Council) (1999). Evaluating Federal Research Programs: Research and the Government Performance and Results Act. In E. Committee on Science, and Public Policy (Ed.), (pp. 96). Washington (DC): NRC.

NRC (National Research Council) (2000). Experiments in International Benchmarking of U.S. Research Fields. Washington (DC): NRC.

NRC (National Research Council) (2006). A Strategy for Assessing Science: Behavioral and Social Research on Aging. In I. Feller & P. C. Stern (Eds.), (pp. 176). Washington (DC): NRC.

NRC (National Research Council) (2007a). Decadal Science Strategy Surveys: Report of a Workshop. Washington (DC): NRC.

NRC (National Research Council) (2007b). Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond (pp. 456). Washington (DC): NRC.

NRC (National Research Council) (2010). New Worlds, New Horizons in Astronomy and Astrophysics. Washington (DC): NRC.

NSF (National Science Foundation) (2010). Merit Review Retrieved April 5, 2011, from http://www.nsf.gov/bfa/dias/policy/meritreview/

NSF (National Science Foundation) (2011). NSF FY 2012 Budget Request to Congress. Arlington (VA): National Science Foundation.

NSTC (National Science and Technology Council) (2008). The Science of Science Policy: A Federal Research Roadmap. Washington DC: US Office of Science and Technology Policy.

The White House (2011). A Strategy for American Innovation: Driving Towards Sustainable Growth and Quality Jobs Retrieved April 30, 2012 from http://www.whitehouse.gov/administration/eop/nec/StrategyforAmericanInnovation/

UNESCO (United Nations Educational Scientific and Cultural Organization) (2010). UNESCO Science Report 2010: The Current Status of Science around the World. Paris, France: UNESCO.

USOMB (United States Office of Management and Budget) (2003). Budget Procedures Memorandum No. 861: Completing the Program Assessment Rating Tool (PART) for the FY 2005 Review Process. Washington (DC): USOMB.

USOMB (United States Office of Management and Budget) (2006). Program Assessment Rating Tool (PART). Washington, (DC): USOMB.

USOMB (United States Office of Management and Budget) (2011). Winning the Future through Innovation: The Federal Budget Fiscal Year 2012. Washington (DC): USOMB.

Appendix C
Review of Quantitative Indicators of Research Quality & Trends

1 About this Appendix

In the course of this assessment the Panel undertook a systematic review of major types of science indicators relevant to assessing research quality and research trends in the natural sciences and engineering (NSE). The Panel evaluated each indicator type against a set of general criteria, and considered the merits and limitations of each indicator with respect to each criterion. This appendix presents the results of that assessment process.

The evaluation criteria used by the Panel are described and defined in Section 1.1. Section 2 reports the results of assessing indicators relevant to measuring research quality, and Section 3 reports on research trend indicators. The Panel also undertook a basic analysis of the three major data sources for bibliometric indicators: Thomson Reuters' Web of Science, Elsevier's Scopus, and Google Scholar. The results are presented in Section 4.

1.1 CRITERIA FOR ASSESSING QUANTITATIVE INDICATORS

To evaluate commonly used indicators, the Panel focused on criteria defined in Table C1. Validity was the overarching criterion. If an indicator was found to be valid for the purpose of science assessment at the field level, the other criteria were also considered. If an indicator was *not* deemed valid, the Panel did not evaluate it against the other criteria.

The assessment process focused explicitly on use of indicators at the level of nationally aggregated research fields. Therefore judgments about indicator validity reflect whether an indicator is appropriate for use at this level of aggregation. The Panel also assessed the relevance of each indicator type to the Natural Sciences and Engineering Research Council's (NSERC) allocation or reallocation of Discovery Grants Program (DGP) funding. The intention behind this criterion was not to be prescriptive, but to highlight which indicators the Panel deemed most relevant to NSERC.

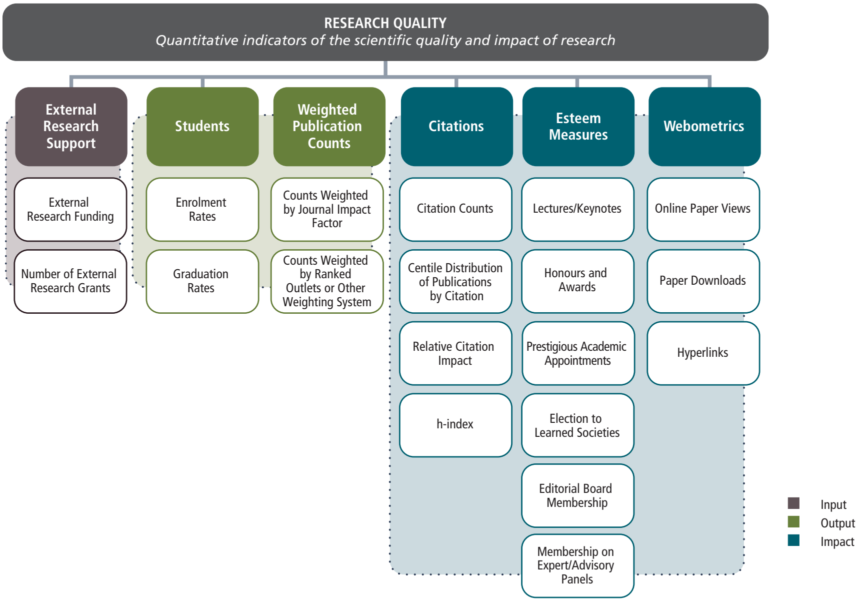
Table C1

Criteria considered in evaluating quantitative indicators

Rating	Definition
Validity	To be valid for assessing science performance at the field level in the NSE, indicators must be well researched, internationally recognized, validated by existing research and past experience, and able to support cross-field comparisons.
Timeliness	The indicator must relate to recent activities. Data related to research undertaken many years previously do not reflect current dynamics of the research environment and may lead to inappropriate funding decisions.
Behavioural impact	The indicator should not present a high risk of resulting in unintended and negative behavioural responses in the research community.
Level of aggregation	The indicator should be relevant and valid in assessments at the field level. Appropriate levels of aggregation for each indicator are denoted by the following abbreviations: N (national), F (field), I (institutional), G (group), and R (researcher).
Transparency	The indicator should be transparent and based on publicly available methodologies and data.
Relevance to NSERC	For valid indicators types, the Panel considered to what extent the indicator is of relevance to NSERC, particularly in the context of informing allocation or reallocation of DGP research funding across research fields.

2 Indicators of Research Quality

The most commonly used quantitative indicators of research quality fall into six broad categories: external research support, student population, weighted publication counts, number of citations, esteem measures, and webometrics. These indicator types, and selected sub-types, are presented in the taxonomy below.



2.1 EXTERNAL RESEARCH SUPPORT

2.1.1 Funding

External research funding is used as a measure of research quality based on the rationale that higher quality (or more promising research) is more likely to attract substantial funding from other sources. This indicator can also be looked at as a derivative of peer review processes, to the degree that the other funding sources included are allocated based on peer review.

These measures are calculated as the amount of external (i.e., from other sources) research funding received by individuals, research groups, institutions, etc. They can be calculated annually or based on rolling multi-year averages. Metrics based on external research funding can be limited to particular sources (e.g., specific public granting programs, industries, international sources, peer-reviewed granting mechanisms, etc.), and calculated for various units of assessment (e.g., individuals, research groups, institutions, fields of research, etc.). Indicators of this type may also be normalized by number of research-active staff to compensate for differences in research capacity across different groups or institutions.

Criterion	Rating (Y/N)	Rationale/Comments
Validity	N	The indicator is not generally valid for assessments at the field level due to varying resource requirements (i.e., costs of research) across fields. As a result, total amounts of external funding received are not necessarily informative for comparing quality across fields.
Data sources	Custom databases, with data drawn from various national and international research funding sources; will vary by country.	
References	REPP, 2005; OECD, 2010; Tognolini *et al.*, 1994; Gillet, 1991; Hornbostel, 2001; Laudel, 2006.	

2.1.2 Numbers of Grants

This indicator is analogous to those based on external research funding amounts (see Section 2.1.1.), but is based on number of grants received rather than dollar amounts awarded. (One rationale for this is that it may compensate for differences in the cost of research across fields and therefore be more suitable for comparisons.)

There are no known examples of use of this indicator in major national research evaluation exercises; however, information of this type is often used informally in assessments of individual researchers and research groups. Similar to funding amount indicators (Section 2.1.1.), it can be calculated for specific funding sources or grants as well as at different levels of aggregation. It can also be calculated annually or over a pooled set of years.

Criterion	Rating (Y/N)	Rationale/Comments
Validity	N	This indicator is not generally valid for assessments at the field level because different research fields can depend on different sources of funding, have different levels of diversity in core funding sources, and have different average grant sizes, which can result in differences in average number of grants received.
Data sources	Custom databases, with data drawn from various national and international research funding sources; will vary by country.	
References	REPP, 2005; OECD, 2010; Tognolini et al.,1994; Gillet, 1991; Hornbostel, 2001; Laudel, 2006.	

2.2 STUDENT POPULATION

2.2.1 Student Enrolment Rates

Data on student population are commonly used in research assessment exercises. For example, student-based metrics are used in Australia's Excellence in Research Australia (ERA), the United Kingdom's Research Assessment Exercise (RAE), and numerous other performance-based research funding mechanisms in OECD countries (see OECD, 2010). In most cases, these measures are not used as a direct measure of research quality, but rather as a measure of student output. Student enrolment rates, however, may be used as an indicator of research quality because they capture information about the competitive ability of different research groups or institutions to attract new students (OECD, 2010). Enrolment rates are typically calculated annually by program level (e.g., BA, MA, PhD) and field of study based on data collected from institutions or national statistical agencies.

Criterion	Rating (Y/N)	Rationale/Comments
Validity	N	Student enrolment rates are not valid as a core metric of research quality. Although they may reflect underlying competitiveness of research units or institutions in attracting students, they are also influenced by many other factors including geographic location, labour market trends, teaching quality as opposed to research quality, etc. Enrolment rates are also not useful as a measure of quality at the field level because enrolment trends by field of research are more likely affected by factors other than student perceptions of research quality. Student data are a valid measure of other aspects of research such as research trends and capacity, and therefore may complement other measures of research quality.
Data sources	Institutional data systems, national statistical agencies, international organizations.	
References	REPP, 2005; OECD, 2010; Phillimore, 1989.	

2.2.2 Student Graduation Rates

Student graduation rates may be used in conjunction with enrolment rates to analyze the productivity (*vis à vis* training) of different research groups or institutions.

Student graduation rates are typically calculated annually by program level (e.g., BA, MA, PhD). These metrics are not subject to significant variation (but program levels and fields may be classified differently based on different statistical and educational practices in countries or regions).

Criterion	Rating (Y/N)	Rationale/Comments
Validity	N	Student graduation rates are not valid as a core metric for assessing research quality. They are more likely a reflection of other factors than of the underlying quality of research in a field. Graduation rates may be useful in some contexts in assessing the effectiveness of particular research groups, faculty, or institutions in training and matriculating new students. Student data are a valid measure of other aspects of research such as research trends and capacity, and therefore may complement other measures of research quality.
Data sources	Institutional data systems, national statistical agencies, international organizations.	
References	REP, 2005; OECD, 2010; Phillimore, 1989.	

2.3 WEIGHTED PUBLICATION COUNTS

2.3.1 Weighted Publication Counts

Indicators based on weighted publication counts assign different weights to different types of publications (e.g., books or refereed journal articles may be weighted more heavily than other types of publications), or to journal articles based on the "quality" of the journal in which a paper is published. A common approach is to weight publications by the journal impact factor of the journals in which they appear. Alternatively, journal articles may be weighted according to a separate ranking of journals such as that formerly used in Australia's ERA initiative.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Weighted publication counts (e.g., publications weighted by journal impact factor or other journal ranking system), based on the prestige of outlets in which publications appear, have been shown to correlate strongly with quality assessments if field-specific (and year-specific for citation based) weights are applied. They can be used to analyze whether publication output is appearing in disproportionately more or less prestigious outlets, which in turn can provide an indication of research quality. This provides, however, only an indication of the potential impact of publications, and actual impact can vary markedly.
Timeliness	Y	One major advantage of metrics based on weighted publication counts is that they are more current than indicators based on citations.
Behavioural impact		The application of weighted publication counts can produce undesirable behaviours. Research managers' reaction to use of this measure in Australia's ERA initiative provides a stark illustration of how the data can be abused. When the ERA's ranked outlets measure was introduced, university mangers began specifying targets for individual researchers that did not take into account relevancy. As a result, the measure was subsequently dropped from the next ERA round.
Level of aggregation	R,G,I,F	This measure can be used at all levels of aggregation. Some analysts favour these metrics in analysis of output from early career researchers who have had little time to achieve actual impact.
Transparency	Y	Publication output data are typically transparent, and weighting methods are usually publicly available and easily verified.
Relevance to NSERC	Y	Publication counts by field are relevant to NSERC. Knowing whether Canadian research appears in high-impact journals can be an informative indicator. To ensure comparability across fields, the ranking of journals or outlets (either by citation data or peer assessments) must take field differences into account.
Data sources	Web of Science, Scopus, Google Scholar, institutional databases.	
References	REPP, 2005; Moed, 2005; Van Leeuwen *et al.*, 2003.	

2.3.2 Journal Impact Factors

Various bibliometric indicators are constructed based on journal impact factors (JIFs). JIFs, which were developed for the Web of Science database, are calculated based on average number of citations received by articles in a journal in the previous two years. As such, it depends on two components: (i) total number of "citable" items published in that journal over the period (typically excludes editorials and letters to the editor), and (ii) number of citations to those items over the period. Alternative journal impact measures have been developed, the most notable being the source normalized impact by paper (SNIP) and MJR measures based on Scopus data.

As a measure of quality, JIFs are interpreted as indicating whether or not a particular unit of analysis is publishing articles in highly cited (i.e., prestigious, competitive) scientific journals.

Although JIFs themselves pertain only to journals, average JIFs can be calculated for specific groups of publications or researchers. In addition, the average *relative* journal impact factor can be calculated by comparing the average for a particular unit to an appropriate benchmark (e.g., the average world journal impact factor for publications in that field of research), which roughly corresponds to the expected level of citations for an article based on its journal.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	JIFs alone are not useful as a measure of research quality of a paper because they relate specifically to journals. Indicators based on average JIFs, however, can be used in assessing research quality — though these types of indicators are not valid for comparisons at the field level unless they are field normalized. JIF-based indicators are also a measure of expected research quality, rather than actual research quality.
Timeliness	Y	These indicators can be calculated based on recent publication histories. One of the potential advantages of measures such as average relative impact factors is that they can be calculated for papers as soon as they are published rather than having to wait for a citation history to develop.
Behavioural impact		JIFs have resulted in negative unintended consequences as applied to journals because journal editorial boards may tailor content to deliberately increase their impact factor (e.g., by publishing numerous review articles). The most perverse occurrences of such behaviour, however, are detectable, and journals have been removed from Web of Science for this reason. Researchers may be asked to publish in journals covered in Web of Science and Scopus simply because they have a JIF.

continued on next page

Criteria	Rating (Y/N)	Rationale/Comments
Level of aggregation	(Journals/ All)	Measures based on average JIFs can be calculated at multiple levels including researchers, research groups, institutions, and fields. But only journals included in Web of Science or Scopus can be attributed a JIF, which may affect fields less represented in those databases.
Transparency	Y	The methodology for calculating JIFs is generally transparent. There may be questions, however, about how "citable" items are identified in journals (PLoS Medicine Editors, 2006).
Relevance to NSERC	Y	Field-normalized metrics based on journal impact factors calculated at the field level are relevant to NSERC. They can provide useful information complementary to other measures because they provide evidence of Canada's presence (or absence) in high-impact journals in a field.
Data sources	Scopus, Web of Science, Google Scholar.	
References	REPP, 2005; Garfield, 2006; Moed, 2005; Van Leeuwen et al., 2003; PLoS Medicine Editors, 2006.	

2.4 CITATION COUNTS

2.4.1 Citation Counts

Simple citation counts are one of the most basic bibliometric measures. They are commonly used on their own and to construct more complex bibliometric indicators. The rationale for using citations is that they capture information on the extent of research impact i.e., the influence of cited research on subsequent work. There are many possible variations. They may be reported at various levels of analysis and for various time periods. Some bibliometricians exclude self-citations, though there is considerable debate about the validity of this methodology.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	N	Because the number of citations a publication attracts depends not only on the intrinsic quality of its content, but also on the field it relates to and the year it was published, comparing the total citation counts of any two units of assessments is rarely valid. Comparability is further reduced if the units are significantly different in size (i.e., number of publications produced).
Data sources	Web of Science, Scopus, Google Scholar.	
References	REPP, 2005; Moed, 2005; Van Leeuwen *et al.*, 2003; Van Leeuwen & Moed, 2005.	

2.4.2 Average Citations and Average Relative Citations

The average or mean number of citations received by a unit (e.g., researcher, group, institution, field, etc.) is also frequently used as a basis for bibliometric indicators. Although simple averages may be used, average relative citations (ARC), which compare average citation levels for a unit to a benchmark level (typically the world average for that field of research), are more common. For example, the actual average citations achieved for the unit assessed is compared to the average citation rate in the field to which it refers for the year(s) in question. Benchmarks for each field are calculated using world data for a specified set of journals. The ARC for all publications in a unit's *oeuvre* can then be summed and an ARC index calculated. An ARC of 1.0 denotes that a unit's publications are achieving the average rate of citations for the fields in which they publish.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Indicators based on average relative citations are considered the most adequate measure of research performance (Van Raan, 2000) where an acceptable set of journals can be developed for a field. As the reference value is 1.0 for all fields, cross-field comparisons can be made.
Timeliness	Y	With the exception of articles published in the last 12 to 18 months, citations can be calculated for fairly recent publications. Provided the analysis is undertaken on robust publication numbers (at least 100), data can be relatively recent.
Behavioural impact		The measure, which encourages researchers to improve impact of their research rather than simply rewarding them for productivity, focuses on the highest-impact publications. Using field-specific benchmarks with the same reference value enables units with output in diverse fields to calculate an aggregate performance value. Unlike total citation calculations, units are not penalized for publication in low-impact fields.
Level of aggregation	I,F	Due to the skewed nature of the citation distribution, this indicator needs to be made on a robust number of publications (100 or more); hence it is rarely applicable to individuals and can be problematic for small groups.
Transparency	Y	The indicator uses a straightforward methodology that is readily understood and easy to interpret.
Relevance to NSERC	Y	This is a highly relevant indicator for NSERC as it allows for direct cross-field comparisons of citation impact.
Data sources	Web of Science, Scopus.	
References	Moed, 2005; Van Leeuwen et al., 2003.	

2.4.3 Distributions of Publications by Citations

These indicators are based on distribution of publications by number of citations received. Since average citation scores can mask quite different performance profiles, it is important to determine whether an average is driven by a small number of outstanding (and highly cited) publications or a steady stream of publications that receive only average citations.

Distributions of publications by citations can be calculated in many ways, the most common being by centile. For example, a country's share of the top one per cent cited publications in a field has been used as an indicator of research quality (King, 1987). Distribution of citations across publications is very skewed, with few high-impact articles. For this reason, many commentators criticize use of citation means, and prefer to highlight performance of units relative to a median citation value, or to identify the proportion of very highly cited articles coming from the unit. Calculations of the number of citations required to place a publication into a particular centile (typically top 1 per cent, top 10 per cent, top 50 percent — above the median) are made based on world data and are field and year specific.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Studies have shown that it is valid to assume that a unit with outstanding (highly cited) publications performs better than a unit producing a steady stream of publications that receive only average citations(Van Leeuwen *et al.*, 2003); though there are of course always exceptions.
Timeliness	Y	With the exception of articles published in the last 12 to 18 months, citations can be calculated for fairly recent publications. Provided the analysis is undertaken on robust publication numbers (at least 100), data can be relatively recent.
Behavioural impact		The measure, which encourages researchers to improve impact of their research rather than simply rewarding them for productivity, focuses on the highest-impact publications. One concern is that researchers may focus on review articles or methodology papers, which on average tend to attract more citations. Once again journal editors would play a gate-keeping role in this regard. Additionally, thresholds can be calculated separately for different types of publications, and hence negate these strategies.
Level of aggregation	All	While care needs to be taken in the use of this indicator at the individual level, it can nevertheless provide useful insight into the existence of high-impact articles in a researcher's publication *oeuvre*.

continued on next page

Criteria	Rating (Y/N)	Rationale/Comments
Transparency	Y	The methodology to calculate citation thresholds, and the distribution of a unit's publications across centiles, is straightforward, readily understood, and easy to interpret.
Relevance to NSERC	Y	Data based on citation centiles provide additional information to those based on simple averages. They highlight the extent to which Canadian output in a field achieves the highest citation impact and can also identify areas with a higher than expected proportion of low/uncited articles. This is a highly relevant indicator for NSERC as it allows for cross-field comparisons.
Data sources	Web of Science, Scopus, Google Scholar.	
References	Moed, 2005; Van Leeuwen, *et al.*, 2003.	

2.4.4 h-index

The h-index was developed by Jorge E. Hirsch to combine productivity and impact of a scientist into a single measure. A scientist has an index of h if his or her total publication list includes h publications with at least h citations. For example, a scientist has an h-index of 15 when 15 publications have each been cited 15 times or more. The level of the h-index is therefore strongly correlated with number of papers published and thus on the number of years an author has been active and the field to which the publications relate (Hirsch, 2005). Many variations have been proposed to overcome the shortcomings of the original h-index — including the g-index and the m-index — but none of these are generally accepted as adequate.

Criterion	Rating (Y/N)	Rationale/Comments
Validity	N	This indicator is not valid for assessing research quality for many reasons. It combines quantity and quality and is thus not a homogeneous measure of quality. Also, h-indices are highly correlated with overall publication output. Its maximum value is bounded by the total number of papers. As a result, the metric is greatly influenced by the quantity of publications — making it more a measure of volume output than quality since it also eliminates the few papers which may have a very large number of citations.
Data sources	Web of Science, Scopus, Google Scholar.	
References	Hirsch, 2005; Van Leeuwen, 2008.	

2.5 ESTEEM MEASURES

2.5.1 Lectures and Keynote Speeches

These indicators are based on the number of keynote addresses or lectures given to major national and international conferences by a researcher or group of researchers. The assumption is that a keynote address is more prestigious than a standard presentation. Since there are no standard data on these activities, standard indicators cannot be built. Peer evaluation based on CVs, however, usually takes this information into account.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	N	This is not a valid metric for assessments at the field level due to the technical challenges involved in constructing such counts at the national level, and applying them in cross-field comparisons. Without a unified national CV database, this indicator would be almost impossible to develop. Even if data could be generated at this level, it would be extremely difficult to develop robust field-sensitive benchmarks (what is "average" in a field?), and impossible to find international comparisons to determine if Canadians received more or less invitations than the norm. It may, however, be a valid measure of research quality at lower levels of aggregation. An invitation to present a paper at an international conference implies that a scientist is held in high regard by peers in the international scientific community (King, 1987). But this may be open to questions of bias because of the existence of personal networks (Wood, 1989).
Data sources	Custom databases; generated from public/institutional records and CVs.	
References	Donovan & Butler, 2007; King, 1987; Wood,1989.	

2.5.2 Honours/Awards

These indicators are constructed from simple counts of awards based on an established set of awards or honours that qualify for inclusion.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	N	Metrics based on counts of honours or awards are not generally feasible at the field level. They may, however, be valid measures of quality at lower levels of aggregation. Awards and honours given for research achievements are generally regarded as an acceptable measure of esteem and accomplishment, though they may relate more to past achievements than current activity (Donovan & Butler, 2007). The timeliness of this type of metric is often problematic as honours or awards may be given in recognition of research accomplishments from many years, or even decades, before.
Data sources	Custom databases; based on public records of awards and prizes.	
References	Donovan & Butler, 2007; Luwel *et al.*, 1999.	

2.5.3 Prestigious Appointments

Quantitative esteem measures can also be based on counts of prestigious academic appointments such as endowed chairs or other academic positions typically reserved for accomplished or promising researchers.

Criterion	Rating (Y/N)	Rationale/Comments
Validity	N	Since metrics based on counts of prestigious appointments are not standardized or based on an agreed source, they are not applicable to assessments at the field level. Again, they may be valid measures of quality at lower levels of aggregation. As with other esteem measures, there is little evidence or experience on which such measures can be judged. Most applications are used informally or at the institutional level. Donovan & Butler (2007) suggest that esteem measures based on such appointments may be acceptable. As with most other esteem measures, the danger is that they will not relate adequately to current strengths if appointments were made in the distant past or in recognition of research accomplishments from many years before.
Data sources	Custom databases; based on records of such appointments.	
References	Donovan & Butler, 2007; Luwel *et al.*, 1999.	

2.5.4 Election to Learned Societies

These metrics are similar to those based on prestigious appointments (see Section 2.5.3). They are based on counts of researchers or scientists elected to national or internationally recognized learned societies (e.g., Fellows of the Royal Society in the United Kingdom, Fellows of the National Academy of Science in the United States).

Criteria	Rating (Y/N)	Rationale/Comments
Validity	N	Metrics based on election to learned societies are not generally applicable to assessments at the field level. Since election to such societies are often recognized as a valid measure of accomplishment, they may be used at lower levels of aggregation. As with other esteem measures, the key limitation is timeliness. Election to societies may have occurred many years before, or may be in response to past achievements. As a result, they generally are not a reliable indicator of current research strengths.
Data sources	Custom databases; based on records of such appointments.	
References	Donovan & Butler, 2007; Luwel *et al.*,1999.	

2.5.5 Editorial Board Memberships

Service to scientific or academic journals may also be used as the basis for quantitative, esteem-based indicators of research quality. Such metrics could be based, for example, on editorial roles at journals or on participation as peer reviewers of submitted publications.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	N	These measures have not been conclusively established as useful measures of research quality. One analysis suggests that indicators based on service to scientific journals could be regarded as a valid measure of quality, but only as a "low" measure of quality (i.e., not an indicator of major quality or accomplishment) (Donovan & Butler, 2007). This assertion is debatable as others have argued that such activity is a standard part of scientific work and does not indicate any level of accomplishment above what is expected.
Data sources	Custom databases, with data drawn from various national and international research funding sources; will vary by country.	
References	Donovan & Butler, 2007; King, 1987; Wood, 1989.	

2.5.6 Participation in Expert Advisory Panels

Participation in expert advisory panels can also be used as the basis for quantitative esteem measures. Such participation may be awarded in recognition of particular expertise or accomplishment in a field of research. The issues associated with these metrics are largely similar to those based on prestigious appointments and election to learned societies.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	N	These metrics are not valid for field-level assessments due to difficulties in ensuring the equivalence of different panel processes across different fields of research, or of opportunities associated with different fields (i.e., some fields may be more likely to be called upon to serve on panels). These metrics could be valid measures of quality, but validity would depend on the nature of the advisory panel and the selection process for panel members. In general, panel selection in such cases is likely to reflect significant past accomplishment in a field, but the timeliness of that accomplishment may vary.
Data sources	Custom databases, with data drawn from various national and international research funding sources; will vary by country.	
References	REPP, 2005; OECD, 2010; Tognolini *et al.*, 1994; Gillett , 1991; Hornbostel, 2001; Laudel, 2006.	

2.6 WEBOMETRICS

2.6.1 Online Paper Views

Webometrics are science indicators based on quantitative analysis of online activity. They are similar to traditional bibliometrics in that they capture data on usage and influence of publications or sets of publications.

One common indicator is based on online paper views. For example, for papers published in online science journal *PLoS ONE*, the total number of times each paper has been viewed is presented for each article published.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Unclear	The validity of using online paper views as a measure of research quality is unclear. In many respects it is similar to the use of citations. Although Brody, *et al.* (2006) show the correlation between the number of online paper views and the citations they obtain, paper views may represent a lower threshold of impact (viewing a paper does not indicate the same degree of influence as citing a paper in subsequent work). Since papers may be viewed by many people outside the scientific community who may not publish scientific articles, the meaning of such "views" is not clear. They may easily be manipulated or give rise to a popularity contest as anybody can view a paper in open access.
Timeliness	Y	These types of metrics, compared to citations, make it possible to capture very recent (even immediate) information on research uptake.
Behavioural impact		Mechanically incorporating online paper views into a research assessment exercise would be dangerous because of the potential for artificially inflating such counts.
Level of aggregation	R (G,I,F)	To date, these metrics are typically only reported at the level of individual researchers. But there is no theoretical or practical reason why they could not also be aggregated to higher levels.
Transparency	Y	Metrics based on counts of online paper views are generally transparent. They are less transparent than citations, however, because the identity of individual viewers is typically unknown and one person can view the same paper many times.
Relevance to NSERC	N	Metrics based on paper views are always likely to be vulnerable to gaming and therefore unlikely to be sufficiently robust for use in support of funding allocation.
Data sources	Selected online journals such as those hosted by the Public Library of Science (PLoS).	
References	Thelwall, 2009, 2008; Van Noorden, 2010; Brody *et al.*, 2006.	

2.6.2 Paper Downloads

These web-based metrics, similar to those based on paper views, are triggered by actual downloading of a paper. Aside from this possibly constituting a slightly higher threshold of impact (in so far as downloading a paper, possibly for future reference or use, may indicate more serious consideration as opposed to viewing a paper online and scanning its content instead of really reading it), there is no substantive methodological difference between the two indicator types.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Unclear	The validity of using online paper downloads as a measure of research quality is unclear. Paper downloads may constitute a higher threshold of impact than online views, though it may often be technically difficult to distinguish between paper views and paper downloads depending on the nature of the online platform.
Timeliness	Y	One of the strengths of these types of metrics is their ability to capture very recent (even immediate) information on research uptake.
Behavioural impact		Mechanically incorporating paper downloads into a research assessment exercise would be dangerous because of the potential for artificially inflating such counts.
Level of aggregation	R (G,I,F)	To date, these metrics are typically only reported at the level of individual publications. But there is no theoretical or practical reason why they could not also be aggregated to higher levels.
Transparency	Y	Metrics based on counts of paper downloads are generally transparent, though little information is available about the identities of individuals downloading papers.
Relevance to NSERC	N	Metrics based on paper downloads could at some point be developed to be relevant to field-level assessments. Currently, they are not sufficiently developed for such use, and it may take time to overcome serious concerns about the potential for artificially inflating counts.
Data sources	Selected online journals such as those hosted by PLoS.	
References	Thelwall, 2008, 2009; Van Noorden, 2010.	

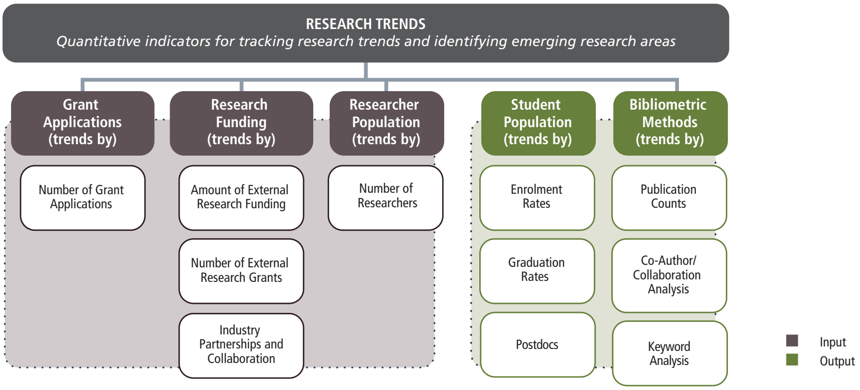
2.6.3 Hyperlinks

This indicator is based on counts of hyperlinks. When performing an online search, Google's PageRank algorithm essentially weights each page by the number of other webpages linking to it, thereby identifying those pages most frequently referenced by other sources on the web. Likewise, counts of hyperlinks can also be used to identify websites (or web-based resources) that are highly referred to by other websites. For example, since 2004 a research group in Spain has been publishing a webometric ranking of world universities based partially on hyperlinks (see http://www.webometrics.info/). Such links can potentially be interpreted as a measure of the influence of a researcher or research institution.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Unclear	The validity of hyperlink analysis as a measure of research quality is largely untested, and thus unclear. Given the importance of online material for researchers and research institutions (e.g., researcher, lab, or institution websites; online CVs or publications, etc.), hyperlinks may be an important measure of the extent of online presence or influence. Although hyperlinks may capture the extent or range of influence of an entity (as reflected online), that influence may be determined by factors other than research quality (e.g., connection to controversial or otherwise topical issues, scale of relevant online community, extent of online activity among relevant research users, investment in a technical infrastructure, etc.).
Timeliness	Y (N)	Web-based analysis of hyperlinks is current. One challenge with hyperlink analysis, however, is that link numbers change relatively quickly as the internet evolves, making it difficult to maintain stable comparisons over time.
Behavioural impact		There is a significant risk of gaming with these metrics by artificial inflation through otherwise superfluous links (this is analogous to processes currently used now to enhance rankings on search engine returns through "link farms"). It has, however, been shown that it is possible to detect the most perverse attempts at such manipulation and to remove those institutions from the rankings.
Level of aggregation	R,G,I	This metric is limited to units of analysis with a defined online presence e.g., website of a researcher, lab, or institution. Results could possibly be aggregated to the level of research fields; however, this would also require normalization by field as different fields can reasonably be expected to have different extents of online presence.
Transparency	N	The methodologies behind hyperlink analysis are relatively transparent. But actual analytical results are *not* entirely transparent because online search engines scan a sample of all webpages and do not specify actual contents of that sample (Thelwall, 2009).
Relevance to NSERC	N	This metric is not relevant to NSERC because it is not sufficiently robust to support cross-field comparisons.
Data sources	Hyperlink analysis based on existing search engine coverage of the web.	
References	Thelwall, 2008, 2009; Van Noorden, 2010.	

3 Indicators for Assessing Research Trends

Much of science measurement and evaluation addresses research quality. Funding agencies, however, need information beyond assessments of quality. Foremost among these are *research trends*, which illustrate changing direction of scientific research and therefore provide useful insights into field-level funding allocation decisions.



The figure above presents a taxonomy of indicator types used to capture changes in key aspects of research activity over time.

3.1 GRANT APPLICATION TRENDS BY TOPIC

This metrics includes indicators based on numbers of grant applications received for research funding programs, as analyzed over time and by research field or topic.

According to the Panel, research funding councils frequently track funding application trends — both formally and informally — to assess research trends. The applicability and usefulness of this indicator, however, are largely dependent on the nature of the funding programs. For example, thematic calls for proposals are of limited use in estimating trends over time. Since funding application trends by topic may be determined in part by changes in funding levels across fields or granting programs, they may not be an accurate measure of trends that would be occurring in the absence of changes occurring in the different fields.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	When applied to stable research funding programs open to proposals in a wide range of fields, grant application trends can be a reliable and useful measure of research trends. Interpretation of these indicators must include consideration of other factors that could potentially influence these trends including changes in funding levels across fields and in other funding mechanisms, which may result in an increase or decrease in grant applications for specific programs, etc.
Timeliness	Y	These indicators can be current up to the year in question; however, their ability to capture trends may be limited by any significant changes in the funding programs being analyzed (or related programs).
Behavioural impact		Behavioural impacts would be generally minimal, though there is the possibility of researchers preferentially applying for funding in research areas with growing numbers of applicants.
Level of aggregation	R,G,I,F	Grant application trends are potentially of use at all levels of aggregation; however, trend analysis is likely more useful and reliable at higher levels of aggregation such as the field level.
Transparency	Y	In most cases, this indicator is transparent. Full transparency requires that data on relevant grant application trends be publicly available.
Relevance to NSERC	Y	Grant application trends by research field or topic are relevant to NSERC and could be used for the DGP.
Data sources	Research granting council databases; public records of past funding awards.	

3.2 TRENDS IN R&D FUNDING BY TOPIC

Trends in research investment levels can be analyzed to assess the evolving direction of research. Indicators based on these trends can be constructed from (i) funding levels from national research councils or other public research funding mechanisms; (ii) external grant applications (i.e., analysis of grant application numbers from other sources); and (iii) industry R&D investment.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Trends in R&D investment by field of research are extremely important for characterizing research capacity and changes in capacity across fields. R&D funding trends represent an important metric on the flow of resources to support different area of research, and can therefore be valid and useful at the field level. Care should be taken, however, in interpretation of this metric as changes in funding levels may or may not correspond to new scientific developments or emerging research areas.
Timeliness	Y	These metrics are generally current, but can be variable depending on the underlying data.
Behavioural impact		There is a low risk of a bandwagon effect when these measures are formulaically included in research funding allocation mechanisms.
Level of aggregation	All	Funding levels and changes in funding levels by research topic can be analyzed at all levels of aggregation.
Transparency	_	Variable depending on the source of funding; full transparency requires that data be publicly disclosed and accessible.
Relevance to NSERC	Y	Various types of indicators based on analysis of R&D investment trends may be of interest to NSERC. These types of measures, however, should generally be considered as more pertinent to general examination of research capacity across fields, rather than as measures of trends or developments independent of funding levels.
Data sources	University financial records and databases; research council funding databases; relevant data on industry investment collected by national statistical agencies.	

3.2.1 Industry Partnerships & Collaborations by Topic

Related to measures of investment trends by research topic, metrics based on absolute numbers of partnerships, collaborations, or contractual agreements between researchers and private industry are also sometimes used in research assessment. Such measures capture information about the level of private-sector investment and interest in research activity in particular fields, areas, institutions, etc. These measures can be calculated for individual researchers, research groups, or institutions, and could possibly be analyzed at the level of nationally aggregated research fields.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Changes in numbers of research partnerships or collaborations may be indicative of new developments in an area of research. As such, these indicators may be worth tracking, though findings from these sources would need to be verified or validated with other methods.
Timeliness	Y	Current; though again analysis over time is subject to challenges arising from changes in other, unrelated environmental factors e.g., climate for business investment, nature of industry/university relations, etc.
Behavioural impact		Any formulaic use in research funding allocation would be subject to considerable behavioural risks as it could incentivize development of superficial or otherwise vacuous agreements with industry solely for the purposes of registering changes.
Level of aggregation	R,G,I (F)	These indicators are generally applied to the researcher, group, or institution level. Theoretically, data could be aggregated to the field level, though this is rarely done in practice.
Transparency	_	Variable; full transparency requires that underlying data be fully public and accessible. This may or may not be the case with university financial data and records of external contractual arrangements.
Relevance to NSERC	N	Due to challenges in aggregating these measures to the field level, and their greater applicability to funding of applied rather than discovery-oriented research, they are of limited relevance to NSERC in the current context.
Data sources	University financial records and contractual agreements; possibly industry surveys undertaken by public statistical agencies.	

3.3 TRENDS IN RESEARCHER POPULATION BY TOPIC

Metrics based on researcher population are commonly used in research assessment practices. They may require identifying numbers of "research-active staff" or "full-time-equivalent or FTE" staff. A key challenge is accounting for differences between part-time and full-time staff. The rationale for using these metrics as a measure of research trends is that active and growing research topics might be expected to show growing numbers of researchers over time, whereas research topics or fields of study that are stagnating or inactive would show declining levels of researchers.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Metrics based on research population are a valid measure of research trends, but only over sufficiently long time horizons. Researchers may take many years to shift activity into new research topics or areas, and, in many cases, these trends may be tied closely to overall levels of researcher recruitment and retirement. Analysis of such trends is also challenging as it requires accurately differentiating among research sub-fields, which may not be captured in standard disciplinary categorizations at the university department or program level.
Timeliness	-	There are significant challenges associated with the timeliness of this indicator. In the long term, changes in research population by topic of research may be an accurate indicator of research trends. These changes, however, likely occur over many years. In the short term, any dramatic changes may be determined more by other factors (e.g., university financial constraints or context).
Behavioural impact		Generally, behavioural impacts related to use of trend indicators are likely to be relatively minor. In the context of assessing research quality, however, these metrics have been shown to have significant behavioural consequences. In the United Kingdom, discrimination between research-active and non-research-active staff resulted in tensions in researcher morale at some institutions (OECD, 2010). Linking funding decisions directly to metrics related to researcher population risks incentivizing the concentration of researchers in larger fields/topics.
Level of aggregation	I,F	These indicators can be aggregated and analyzed at the group, institution, and field level.
Transparency	Y	Such metrics are typically based on public data on faculty appointments and positions.
Relevance to NSERC	Y	Metrics on researcher population are relevant to NSERC as measures of research trends and research capacity.
Data sources	University faculty databases (in Canada these data are available through Statistics Canada's University and College Academic Staff System, UCASS).	

3.4 TRENDS IN STUDENT POPULATION

3.4.1 Student Graduation and Enrolment Rates by Topic

Student graduation and enrolment rates can be tracked and analyzed by program/topic of research to measure research trends. The rationale is that students will gravitate towards growing, active, or emerging fields of research, while stagnating or declining research areas are more likely to see declining enrolment. Such metrics can be calculated for various levels (BA, MA, PhD), though graduate level enrolments may be most relevant to actual trends in research performance.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Student graduation and enrolments at upper program levels (PhD, MA) can be plausibly expected to shift in relation to student expectations about overall viability and promise of research activity in a field. Care must be taken to differentiate between other possible factors when interpreting these data. At lower levels of aggregation, student enrolments may be more a function of perceived teaching quality than research activity. Student expectations about labour market outcomes must also be factored in as these may drive changes in enrolment rates across programs. Finally, it may also be important to consider application rates in conjunction with enrolments because actual enrolments may be capped due to constraints in capacity and therefore not fully reflect changes in numbers of students applying to programs in specific areas.
Timeliness	Y	Student enrolment and graduation data are current and up to date. Student enrolments may also be reasonably responsive to research trends in the shorter term than is the case for the researcher population.
Behavioural impact		Linking assessments of research trends to student enrolment metrics is not likely to have significant, negative behavioural repercussions. Student decisions may be affected by perceptions of funding availability, but there is a low risk of major consequences given that any allocation of funding would presumably be equitable on a per student basis, and institutions would bear additional costs associated with training additional students. Such a system, however, may favour fields of research that can readily increase their capacity to train new students over those where the cost of training is higher or training is limited by other constraints related to infrastructure, facilities, equipment, etc.
Level of aggregation	I,F (R)	These metrics can be analyzed at the institution or field level. They may also be analyzed at the level of individual researchers or labs at the graduate level.
Transparency	Y	Student data are typically readily available and transparent.
Relevance to NSERC	Y	Measures of student enrolment and graduation rates by field of research are relevant to NSERC as a way of assessing field-level research trends in Canada.
Data sources	University databases; public education statistics collected by national statistical agencies.	

3.5 BIBLIOMETRIC METHODS

3.5.1 Publication Counts by Topic

A variety of bibliometric indicators can be used to analyze research trends and identify emerging research areas at the national level. The most straightforward of these are simple publication counts by field or topic of research. Changes in total output of research publications in research areas can be analyzed over time to differentiate between areas with a growing amount of activity and work, and those with declining activity and output. Such measures can be computed for various types of publications (e.g., books, book chapters, etc.), with indexed, peer-reviewed journal articles the most common. Indicators based on publication counts can be aggregated at multiple levels (e.g., researcher, institution, field, etc.) and analyzed annually or over pools of years.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Trends in publication output by field or topic of research are a valid measure of research trends. Information of this type can identify areas of research with increasing activity and publication output. The primary limitation of such indicators relates to their dependence on previously existing definitions of fields and sub-fields. Traditional field categories used in bibliometrics may not adequately reflect truly novel research areas, which may in some cases cut across fields. (This limitation may be mitigated by using publication counts in conjunction with keyword analysis).
Timeliness	Y	Publication counts are current and available for recent years.
Behavioural impact		Any direct application of publication-based indicators to funding decisions can result in behavioural consequences related to increasing publication output. These can have negative, unintended consequences. Such risks can be mitigated, however, by avoiding any formulaic connection of funding levels with publication outputs across fields and by using a suite of complementary indicators.
Level of aggregation	R,G,I,F	These types of indicators can be used at all levels of aggregation.
Transparency	Y	Publication counts are transparent when conducted on standard, bibliometric databases such as Web of Science and Scopus. Counts based on Google Scholar are not fully transparent because Google does not disclose the exact contents of its database.
Relevance to NSERC	Y	Publication counts and trends in publication counts by topic or field of research are relevant to NSERC. They can be used to assess research trends at the field level and potentially identify new or emerging research areas.
Data sources	Web of Science, Scopus, Google Scholar.	

3.5.2 Co-author/Collaboration Analysis by Topic

Bibliometric techniques can also be used to study patterns in research collaboration through analyzing co-authorship among researchers and co-citation patterns. Data on paper co-authorship can be used to assess trends in levels of collaboration between researchers in different institutions or fields. Patterns in co-citation can be used to assess the degree of impact of research in one area on another field. Although patterns in scientific collaboration are not directly relevant to identifying emerging research areas, significant changes in collaborative activity between research fields, or development of new collaborations or patterns of cross-citation, may signify new research developments. As such, these indicators may also be useful and informative in tracking research trends over time.

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Bibliometric indicators based on co-authorship and collaboration are reliable indicators that can be used to assess research trends and possibly identify emerging research areas.
Timeliness	Y	These indicators are current and can capture information on recent trends in publication activity.
Behavioural impact		These indicators are unlikely to have significant negative or unintended behavioural impacts. If other bibliometric indicators are formulaically included in research funding allocation, however, there could potentially be consequences if researchers artificially inflate co-citation or co-authorship levels in order to boost counts.
Level of aggregation	R,G,I,F	Patterns in co-authorship and collaboration can be analyzed at all levels, including nationally aggregated research fields.
Transparency	Y	Bibliometric indicators based on co-authorship and co-citation patterns are transparent and based on well-understood methodologies and data.
Relevance to NSERC	Y	These indicators are relevant to NSERC in analyzing and tracking research trends at the level of nationally aggregated research fields in the NSE.
Data sources	Web of Science, Scopus.	

3.5.3 Keyword Analysis

A number of bibliometric techniques and indicators can now be constructed based on analysis of key words within publications and of other key terms that may occur within a journal article or abstract. Key words can be analyzed to identify and track emergence of research areas or topics of study. Simple counts of the occurrence of key words can be analyzed to identify topics of growing or declining interest. Various methodologies can also be used to identify clusters of related key words and therefore to study and assess relationships between related areas of research. These kinds of analysis can also be coupled with different types of visualization tools to produce various "science maps."

Criteria	Rating (Y/N)	Rationale/Comments
Validity	Y	Indicators and techniques based on keyword analysis (and other important terms) are a valid approach to tracking research trends.
Timeliness	Y	Keyword analysis is current, and useful in analyzing up-to-date trends in research publication.
Behavioural impact		Given the typical pattern of use for these types of indicators and approaches in general assessments of research trends, there is limited potential for negative unintended behavioural consequences.
Level of aggregation	All	Keyword analysis can be potentially applied to any level of aggregation; however, in most cases it is used at a national or international level to identify broad research trends, rather than in the context of a specific assessment of a research group, institution, or field.
Transparency	Y	These indicators and methods are based on transparent, well-understood methodologies and data. These methodologies, however, may be complex and not easily understood by a layperson.
Relevance to NSERC	Y	These indicators are relevant to NSERC and to tracking research trends and identifying emerging research areas within Canada and internationally.
Data sources	Web of Science, Scopus, Google Scholar.	

4 Assessment of Bibliometric Data Sources

The Panel reviewed and assessed three primary sources of bibliometric data analogous to the preceding assessment of indicators: Thomson Reuters' Web of Science, Elsevier's Scopus, and Google Scholar. These sources were evaluated based on the criteria in the table below:

Criteria	Description
Timeliness	Data must be collected and available for the specified period of time required for analysis and comparisons. Old data do not reflect the current dynamics of the research environment and may lead to inappropriate funding decisions.
Accuracy and reliability	The contents of the database are accurate. Queries of the data must produce the same result regardless of when and by whom they are run.
Transparency	Construction of the database should be grounded in transparent and publicly available methodologies.
Cost	Cost of accessing the data in terms of money and time.
Coverage	The database should cover a large proportion of the activity being assessed.

4.1 WEB OF SCIENCE

For many years, the Institute for Scientific Information's (ISI) Web of Science (WoS) (now owned by Thomson Reuters) was the only available platform for undertaking large-scale, rigorous bibliometric analysis. WoS consists of a family of related citation indices, the three most prominent of which are the *Science Citation Index*, the *Arts and Humanities Citation Index*, and the *Social Science Index*. Together these indices currently include coverage of over 12,000 scientific journals.[1] Journals are selected for inclusion based on a policy of covering the large majority of referenced articles (i.e., coverage prioritizes more highly cited journals).

1 http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/#tab2

Criteria	Y/N	Rationale/Comments
Timeliness	Y	Bibliometric data from WoS citation indices are kept current and up to date. The database also has retrospective coverage dating back to 1900, making it flexible with respect to periods of analysis.
Accuracy and reliability	Y	In general, WoS bibliographic data meet acceptable standards of accuracy and reliability. In the past, WoS citation indices have been criticized due to significant potential for errors related to improper referencing, author name ambiguity, variable referencing conventions, etc., which may result in loss or misappropriation of up to seven per cent of total references (Moed, 2005). These sources of error, however, are common to all bibliometric databases. When data are cleaned and validated through independent publication records (e.g., author CVs), or used at a high enough level of aggregation, the potential for error is sufficiently mitigated to yield useful, informative data in many contexts.
Transparency	Y	Coverage in WoS is sufficiently transparent; a master list of all included journals is maintained online and freely accessible. Raw data behind the online search engine can be accessed to ensure accuracy and undertake detailed, complex analyses.
Cost	-	Access to WoS bibliometric data is by license and can be costly; however, the data are available through various licensing agreements (e.g., multi-site licenses negotiated on behalf of national university peak bodies), giving most researchers ready access to the online version of WoS. Bibliometricians can purchase custom data runs to undertake specific analyses as well as evaluate data coverage and reliability.
Coverage	Y	Most past analyses of WoS have concluded that coverage is sufficient in most NSE fields to support robust analyses. The WoS coverage policy explicitly targets incorporating the majority of highly cited scientific literature.

4.2 SCOPUS

Introduced in 2004, Elsevier's Scopus database is now a major competitor to WoS, and provides an alternative source of data for bibliometric analysis. Scopus includes coverage of over 19,000 scientific journals.[2] Scopus journal coverage is recognized to be generally more extensive than WoS, particularly for non-English language journals and the social sciences and humanities (Hicks, 2004; HEFCE, 2009). In general, however, bibliometric analyses based on Scopus and WoS typically yield similar results at higher levels of aggregation despite differences in coverage (Archambault *et al.*, 2010).

2 http://www.info.sciverse.com/scopus/scopus-in-detail/facts

Criteria	Y/N	Rationale/Comments
Timeliness	Y	Scopus bibliometric data are generally timely and up to date, and can be used for analysis of various time periods. Scopus has comprehensive coverage of journal output dating back to 1996, but currently does not have the extensive historical coverage of WoS.
Accuracy and Reliability	Y	Scopus, like WoS, meets basic standards of accuracy and reliability. The data suffer from many of the same potential sources of error as WoS citation indices; however, these sources can either be corrected at lower levels of aggregation, or are unlikely to significantly skew results at higher levels of aggregation (e.g., national/field level).
Transparency	Y	The contents of the Scopus database are transparent, and a current master list of all included journals is freely available through Elsevier's website. The raw data behind the online search engine can be accessed to ensure accuracy and undertake detailed, complex analyses.
Cost	-	Like WoS, although access to Scopus data is subject to license agreements and can be costly, data are available for license and bibliometricians have accessed data to explore issues of coverage and accuracy.
Coverage	Y	Scopus coverage is sufficient to allow robust analyses of research performance at the NSE field level. Scopus journal coverage is recognized to be more extensive than WoS, though this difference in coverage is unlikely to be significant in national/field level analyses in the NSE.

4.3 GOOGLE SCHOLAR

Due to ease of use and free data, Google Scholar, Google's online database of scholarly literature, is increasingly used in amateur bibliometric analyses in addition to WoS and Scopus. But given the lack of knowledge on what exactly is contained in the database, it is impossible to calculate any world benchmarks against which to assess research performance at the field level. This makes Google Scholar largely unsuitable for rigorous bibliometric analysis. The database, however, may have some value for comparative assessments of journals in fields not well covered by Scopus and/or WoS. Ann-Wil Harzing has developed her *Publish or Perish* tool for this purpose.

Criteria	Y/N	Rationale/Comments
Timeliness	Y	The database is continuously updated, thus is very current.
Accuracy and Reliability	N	It is not possible to specify a citation window. Because the database is continually updated, the same query, run on consecutive days, may return different citation counts.
Transparency	N	Unlike Scopus and WoS, Google Scholar does not supply the raw data behind the search engine under license and does not provide a list of sources for citation counts. This makes it impossible for bibliometricians to interrogate the data to identify problems and determine the extent to which citations are "cleaned" and aggregated, etc.

continued on next page

Criteria	Y/N	Rationale/Comments
Cost	-	Free
Coverage	N	Due to the lack of transparency of the database's contents, and its variability over time, it is not possible to accurately characterize its coverage of publication types (e.g., journal articles) or fields or domains (e.g., NSE).

REFERENCES

Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Association for Information Science and Technology (JASIST), 57*(8), 1060–1072.

Donovan, C., & Butler, L. (2007). Testing novel quantitative indicators of research 'quality', esteem and 'user engagement': an economics pilot study. *Research Evaluation, 16*(4), 231–242.

Garfield, E. (2006). The history and meaning of the Journal Impact Factor. *Journal of the American Medical Association, 295*(1), 90–93.

Gillett, R. (1991). Pitfalls in assessing research performance by grant income. *Scientometrics, 22*, 253–263.

HEFCE (Higher Education Funding Council of England) (2009). *Report on the Pilot Exercise to Develop Bibliometric Indicators for the Research Excellence Framework*. London, UK: Higher Education Funding Council of England.

Hicks, D. (2004). The Four Literatures of Social Science. In H. Moed (Ed.), *Handbook of Quantitative Science and Technology Studies*. Netherlands: Kluewer Academic.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569–16572.

Hornbostel, S. (2001). Third party funding of German universities: an indicator of research activity? *Scientometrics, 50*, 523–537.

King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science, 13*, 261–276.

Laudel, G. (2006). The quality myth: promoting and hindering conditions for acquiring research funds. *Higher Education, 52*, 375–403.

Luwel., M., Moed, H., Nederhof, A., De Samblanx, V., Verbrugghen, K., & Van der Wurff, L. (1999). *Towards Indicators of Research Performance in the Social Sciences and Humanities*. Leiden, Netherlands: Centre for Science and Technology Studies.

Moed, H. (2005). *Citation Analysis in Research Evaluation*. Dordrecht, Netherlands: Springer.

OECD (Organisation for Economic Co-operation and Development) (2010). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*. Paris, France: OECD

Phillimore, A. (1989). University research performance indicators in practice: the university grant committee's evaluation of British universities. *Research Policy, 18*, 255–271.

PLoS Medicine Editors (2006). The impact factor game. *PLoS Medicine, 3*(6), e291.

REPP (Research Evaluation and Policy Project) (2005). *Quantitative Indicators for Research Assessment - A Literature Review*. Canberra, Australia: Australian National University.

Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science, 34*(4), 605–621.

Thelwall, M. (2009). *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*. San Rafael (CA): Morgan and Claypool (Synthesis Lectures on Information Concepts, Retrieval, and Services, 2009, Vol. 1, No. 1).

Tognolini, J., Adams, K., & Hattie, J. (1994). A methodology to choose performance indicators of research attainment in universities. *Australian Journal of Education, 38*, 105–117.

Van Leeuwen, T. (2003). The holy grail of science policy: exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics, 57*(2(2003)), 257–280.

Van Leeuwen, T. (2008). Testing the validity of the Hirsch-index for research assessment purposes. *Research Evaluation, 17*, 157–160.

Van Leeuwen, T., & Moed, H. (2005). Characteristics of journal impact factors: the effects of uncitedness and citation distributions on the understanding of journal impact factors. *Scientometrics, 63*, 357–371.

Van Noorden, R. (2010). Metrics: A profusion of measures. *Nature, 465*(7300), 864.

Wood, F. (1989). Assessing research performance of university academic staff: measures and problems. *Higher Education Research and Development, 8*, 237–250.