

Improving diversity, quality and efficiency in Dutch higher education using performance agreements

Ben Jongbloed, Frans Kaiser, Don Westerheijden

Center for Higher Education Policy Studies (CHEPS), University of Twente, the Netherlands

Corresponding author: b.w.a.jongbloed@utwente.nl

Paper presented at the 39th EAIR Forum 2017, Porto, Portugal

September 2017

Abstract

More and more governments have started to introduce elements of performance in the funding mechanisms for their higher education institutions. For instance, through Performance Agreements (PAs), which are contracts signed between funding authorities and individual universities or colleges. The key characteristics of the PAs in place in a number of OECD countries are summarized before turning to the Netherlands, where an experiment with PAs was recently (2016) concluded.

The question is whether this experiment actually improved performance in the higher education system, where 'performance' is understood in terms of students' study success, education quality and diversity in the provision of education and research. What has been achieved on these areas? And what can be learned from the Dutch performance agreements experiment in general?

Contents

1. Funding mechanisms in higher education	2
2. The move towards Performance Agreements.....	3
3. The Dutch experiment: intentions and outcomes.....	5
4. Encouraging diversity in Dutch higher education.....	10
5. Conclusion: lessons from the Dutch performance agreement experiment	16
References	20

1. Funding mechanisms in higher education

Many countries employ funding formulas to determine the core grant that a higher education institution (HEI) will receive for its day-to-day operations (Jongbloed & Vossensteyn, 2016). These formulas may be driven in part by measures of performance that are fed into the formula to determine the core funds. The measures then reflect the results achieved by an institution in the recent past (*ex post funding*). However, in case of a performance contract the funds are based on an agreement that reflects the performance that an institution promises to deliver in the (near) future. In this case, the HEI will receive its core budget based on a specification of the goals it intends to achieve in the coming period (*ex ante funding*). Every HEI is invited by the funding authorities to specify its ambitions, in line with its particular mission and given its context and its strengths. Surely, to qualify for public funds the HEI's goals will have to be in line with the overall national objectives for higher education. The agreement usually will include a financial penalty or sanction of some sort if the objectives are not achieved. Below we will present some examples of performance-based approaches.

The models for funding public HEIs vary enormously across countries. However, in most cases, the core grant that a HEI receives from its funding authority (say, a ministry or funding council) will primarily be based on the number of students it enrolls. In recent years, many countries have made at least part of the core grant dependent on measures of performance – either by including performance indicators in a funding formula or by introducing some sort of performance contract or performance agreement to reward HEIs for performance. Overlooking the countries and their performance-based funding arrangements, one can see many differences in the activities funded (depending on the broad institutional type), the performance indicators used, the weights attached to the indicators, or the proportion of funding that is dependent on performance. Obviously, what exactly is understood as being performance differs across different higher education systems as well as between different subsectors of the higher education system (i.e. research universities versus universities of applied sciences), depending on the challenges and ambitions of the particular country.

In line with the idea that funds should flow to institutions where performance is manifest, many countries have implemented formula-based funding where the institutions' core budget is based on a set of performance indicators or output criteria. Funding for research and teaching is sometimes awarded separately and managed by different funding schemes. The most frequently used performance indicators used in the funding formulas of OECD countries (see de Boer *et al.*, 2015) are:

- The number of Bachelor and Master degrees: Austria, Denmark, Finland, Netherlands, Germany, United States (e.g. Tennessee)
- Number of exams passed or credits earned by students: Austria, Denmark, Finland, US (e.g. Tennessee, Louisiana, South Carolina)
- Number of students from underrepresented groups: Australia, Ireland, Germany, US (e.g. Tennessee)
- Study duration: Austria, Denmark, the Netherlands, US (Tennessee)
- Number of doctoral degrees: Australia, Denmark, Finland, Germany, Netherlands
- Research output (e.g. research quality, impact, productivity): Australia, Denmark, Finland, Hong Kong, United Kingdom
- Research council grants won: Australia, Finland, Germany, Hong Kong, Ireland, Scotland, US (e.g. Tennessee)
- External income (i.e. non-core revenues): Australia, Denmark, Finland, Germany, Hong Kong

- Revenues from knowledge transfer: Australia, Austria, Scotland

Indicators like the relative success of graduates on the labour market and the number of graduates in jobs related to their training are used less frequently, most probably because there are doubts about whether these outcome measures can be controlled sufficiently by the HEIs. This points to one of the problems of performance-based funding (PBF), namely that the outcome measure has to be a fair representation of an institution's performance. And fairness should be seen in the light of an institution's context and mission. And performance measures so far have not been able to capture the multi-dimensional nature of an institution's output in terms of quality and above all, quality. When it comes to assessing the quality of research, the UK can point to its rather successful experience with its Research Assessment Exercises - in place since 1986, and recently (in 2014) mutated into the Research Excellence Framework. However, how to assess and measure the quality of teaching has proved to be more challenging. In Finland and some US states one can see some first experiences with using student satisfaction surveys to get closer to the quality of education (de Boer *et. al.*, 2015), but the measurement of students' learning outcomes is still in its infancy.

2. The move towards Performance Agreements

Funding formulas in many countries mostly tend to be a mix of input and output elements, whereby the most common criteria are the number of students enrolled and the number of degree completions. To ensure that a funding formula that financially rewards institutions on the basis of the number of degrees will not induce the institutions to lower their educational standards, countries operate a system of licensing and accreditation to guarantee that the quality of education meets national (and, increasingly, international) standards. When high-powered incentives are applied to imperfectly measurable teaching products, problems could arise with the unmeasured part of education production. High-powered incentives to produce graduates could lead to narrowly focused training programs. Non-measurable skills may be undervalued in such a system. Therefore, a hybrid funding structure combining low- and high-powered incentives may be the way out.

One such system that addresses some of the shortcomings of one-dimensional performance-driven formulas is a system of performance agreements. In such a system, there is more room for HEIs to see additional aspects of their performance reflected and connected to financial rewards. Performance agreements (see previous section) can handle situations where HEIs have multiple objectives (or performance dimensions) and – within some nationally-set boundaries – can set their own target levels, given their particular mission and strengths. The performance agreement sets out target levels, the institution's plans to achieve the targets and the budget that will be at stake. There is some form of oversight, either by the funding agency or through some independent commission, to guarantee that an institution does not over-focus on one of the objectives at the expense of another and to monitor how the performance agreements are carried out during the contract period.

Table 1 below shows some of the characteristics of the various performance agreements in place in the OECD. In addition to the table it should be mentioned that in Germany and the US, as in other federal systems, states operate their own funding model. Furthermore, performance contracts are also part of the funding systems in France, Italy and Switzerland. Performance agreements are always part of a

broader set of policy instruments – some of these being funding arrangements like cost sharing, project-based subsidies and student financial support systems, while others are of a non-financial nature and relate to quality assurance regulations, information provision or organisational issues. In some of the countries covered above the agreements are linked to (performance-based) funding formulas. In others, the core funding and the performance-based funding are treated separately (e.g. Louisiana).

Performance agreements can handle situations where HEIs have multiple objectives (or performance dimensions) and – within some nationally-set boundaries – can set their own target levels, given their particular mission and strengths. A performance agreement sets out target levels, the institution's plans to achieve the targets and the budget that will be at stake. Performance agreements have the potential to address some of the shortcomings of one-dimensional performance-driven funding formulas, allowing more room for HEIs to see additional aspects of their performance reflected and connected to financial rewards.

The key characteristics of the performance agreements in place in a number of OECD countries are summarized in Table 1. The countries covered are: Australia, Austria, Canada, Denmark, Finland, Germany, Hong Kong, Ireland, Netherlands, Scotland, and the United States.

The share of the HEIs' public recurrent grant that is based on performance (see column on the right in Table 1) is difficult to determine in an exact way, because both the funding agreement and the formula are often a mix of input and output elements. For instance, in the Netherlands, the performance agreements constitute on average 7% of a university's teaching grant, whereas 20% of the (separate) formula-based teaching allocation is based on degrees and another 40% of the (separate) research allocation is also based on degrees. Thus, on average, a quarter (for universities) to a third (for universities of applied sciences) of funds is based on performance measures. Furthermore, the performance agreement in some countries is not always directly linked to (a separate portion of) the budget (see middle column). For example, in Australia, entering into a performance contract – a *compact* – is one of the quality and accountability requirements that a university must meet as a condition for receiving a grant.

The middle column of the table illustrates that performance agreements are not solely meant to strengthen performance but also have aims like encouraging HEIs to strategically position themselves (institutional differentiation), improving the strategic dialogue between the government and the HEIs and informing policy-makers and the public at large about the HEIs' performance, thus improving accountability and transparency.

The benefits of a diversified higher education system are well-recognised, and performance agreements are expected to help achieve this goal in countries like Austria, Ireland, Germany, Finland and the Netherlands. Performance agreements may prevent one of the risks of formula funding, namely that all HEIs will respond to the formula's indicators in the same way and produce more homogeneity instead of more diversity in the system (Codling and Meek, 2006). The broader set of objectives and indicators facilitated by the performance agreements therefore are expected to promote institutional differentiation.

Table 1: Performance agreements and funding across the world

	Performance agreement (year of introduction)	How is agreement linked to budget?	Period (in years)	% of budget for education and research based on performance (agreement and formula-based funds combined)
Australia	Mission-based compacts (2011)	Agreement is condition for receiving public funding and is accountability instrument	3	20% (includes performance-based formula funds)
Austria	<i>Leistungsvereinbarungen</i> (2007)	Budget linked indirectly (after negotiation) to agreed indicator targets	3	Close to 100%
Canada (Ontario)	Strategic Mandate Agreements (2014)	Budget not linked to agreement but meant to strengthen institutional differentiation and strategic dialogue	3	Around 2% of operating grant
Denmark	<i>Udviklingskontrakter/</i> Development Contracts (2000)	No direct link to funding. Is a Letter of Intent and an outcome of dialogue with ministry	3	60% (including funds based on students' credits and research performance)
Finland	Performance Contracts (1994)	Agreement is linked to 25% of public funding (the strategic component); remainder (75%) is indicator-driven	4	Between 75% and 100% (includes performance-based formula funds)
Germany (e.g. North-Rhine Westfalia))	<i>Ziel- und Leistungsvereinbarungen</i> (around 2002)	No link to budget, but meant as a negotiation and accountability device	2	Between 25% and 55% (includes performance-based formula funds)
Hong Kong	Academic Development Plans (2005)	Links the Performance and Role-related funding (10% of public funds) to an agreed Acad. Dev't Plan, based on institution's mission	3	Around 25% (includes performance-based formula funds)
Ireland	Institutional Performance Compacts (2012)	Compact determines 0.8% of funds, but mostly meant to strengthen strategic dialogue with funding authorities	3	0,8%
Netherlands	Performance Agreements (2012)	Determines 7% of an institution's education budget and meant to enhance institutional differentiation	4	27% - 32% (includes performance-based formula funds)
Scotland	Outcome Agreements (2012)	Non-compliance with agreement has various consequences (also financial)	3	85% (includes performance-based formula funds, e.g. through REF)
United States (e.g. Louisiana)	Performance Agreements (2011)	If targets and underlying story are judged as sufficient the institution qualifies for rewards (performance funds; more financial autonomy)	6	25% (includes performance-based formula funds for education)

Source: Jongbloed & Vossensteyn (2016)

3. The Dutch experiment: intentions and outcomes

The Netherlands has a binary system of higher education, which means there are two types of programmes: research oriented education, traditionally offered by research universities, and professional higher education, offered by universities of applied sciences (UASs). University and UAS programmes differ not only in focus, but also in access requirements, length and degree nomenclature.

There are 18 research universities in the Netherlands, including one Open University, and 38 UASs. The UASs have more of a regional function and focus in particular on their education mission, although in recent years the UASs also have started to strengthen the practice-based research they carry out, partly thanks to dedicated public funds for research and research-oriented staff positions.

The binary structure of the system is generally assessed favourably by the key stakeholders. A committee (the Committee on the Future Sustainability of the Higher Education System, also known as the Veerman committee), installed by the Education ministry in 2009 to look at system performance and

diversity in Dutch higher education, advised to strengthen the binary structure. The Committee regards the binary distinction as valuable and practical. It stated that eliminating binary divides would risk institutions competing with each other for the same students - making them more alike. Most importantly, in its advisory report (Veerman *et al.*, 2010) the committee called for a threefold differentiation in higher education:

1. a differentiation in institutional types (meaning: universities versus UASs)
2. a differentiation between institutions of the same type (i.e. institutions finding their own profile)
3. a differentiation in the range of programmes offered.

The third dimension of differentiation translates into the range of education programmes offered in response to the increased heterogeneity of the student population. It points at the need for a more tailored approach, or smooth transfers to other programmes in order to enable students to successfully complete their programmes. The UASs in particular have to contend with differences in terms of the educational preparation and the orientation (general – vocational education) of students embarking on UAS programmes. Student intake in research universities is relatively more homogenous: nearly two-thirds of all the students embark on an academic study programme on the basis of a pre-university education diploma.

The Veerman committee stated that the generic quality of the education provided is in good order. However, at the same time, it perceived many weaknesses. It concluded that the way in which programmes are organised does not appeal to many students. Student drop-out was considered too high, students' talents were not properly addressed and there was too little flexibility in the system to serve the various needs of students and the labour market. In particular, the difference between the success rates of ethnic minority and native Dutch students is too great. The committee concluded that the system was not future-proof and that a long term strategy was needed to improve the quality and diversity of Dutch higher education.

Increasing diversity was regarded as an important part of the solution and, partly as a result of the recommendations of the Veerman Committee, performance agreements (PAs) were introduced in 2012. The agreements were signed between the Education Ministry and each individual HEI and formulated in terms of quantitative indicators and qualitative ambitions chosen by the institutions themselves. The agreements intended to encourage and reward performance in terms of the following goals:

- Improving the quality of education and the success rate of students in universities and universities of applied sciences;
- Enhancing the differentiation within and between HEIs, encouraging them to exhibit more clear education profiles and more focused research areas (including the creation of Centres of Expertise by UASs);
- Strengthening the focus of universities and UASs on their valorisation function (i.e. knowledge dissemination, commercialization, promoting entrepreneurship).

For the period 2012-2016, 7% of the core grant for education (some EUR 130 million for the research universities and EUR 170 for the UAS sector) was tied to performance agreements. The remainder of the core grant of HEIs continued to be based primarily on a funding formula that, already since the early

1990s, includes a significant performance orientation. An independent *Review Committee*, chaired by Frans van Vught, was installed by the Education ministry in 2011 to oversee the performance agreements, to develop criteria for assessing the agreements, to monitor each institution's progress in realizing its ambitions during the contract period, and, at the end of the period (i.e. in the year 2016), to make a recommendation to the minister about whether the goals in the agreement had been met or not. If a HEI did not achieve its agreed goals it risked losing part of the core grant for the years ahead. It should be mentioned that the performance agreements arrangements were set up as a policy experiment. Depending on an external evaluation, performance agreements could be continued (perhaps after some adaptations) and be included in the laws determining the funding of HEIs.

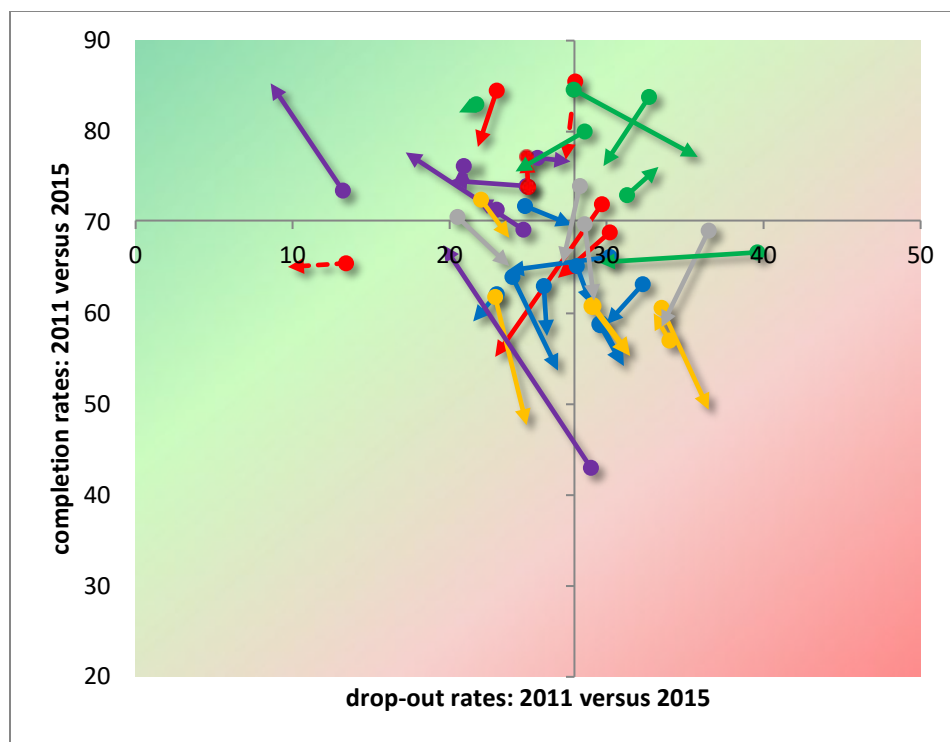
At the start of the performance agreements process the HEIs agreed to make use of seven mandatory indicators in order to state their ambitions with respect to student success and educational quality. Two of them, completion rates and drop-out rates received the most attention in the annual monitoring and, eventually, the final conclusion of the performance agreements. Ambitions with respect to differentiation and institutional profiling were stated in more qualitative terms relating to topics such as starting new degree programmes and phasing out old ones, introducing student mentoring programmes, setting up research centres, engaging in partnerships with local business et cetera.

In 2016 the agreements came to an end. The Review Committee made assessments of each institution's performance in light of the performance agreement and the information presented to the Review Committee through the institution's annual reports and a meeting with the committee.

The Review Committee published a summary of the results of the performance agreements in its 2016 Annual Monitoring Report (Review Committee, 2017a). The Committee concludes that many research universities have booked substantial results when it comes to reducing dropout and increasing completion rates. The average completion rates in universities have risen from 60% to 74%. Among the research universities, dropout rates declined from 17% to 15%. In professional higher education (i.e. the UASs), major efforts have been expended to achieve the targets set down in the performance agreements; nonetheless, some of the UASs have failed in their attempts. The average completion rate fell from some 70% to 67%. However, the dropout figures have been pushed back slightly, from 27% to 25.6%.

The disappointing results in the UAS sector with regard to student completion can in part be attributed to the trade-offs between access, quality, and completion. This trade-off is most strongly manifest in the large UASs that have a highly diverse student population. Handling the impact of the more stringent academic standards for academic performance that were placed upon the UASs by the accreditation agency has presented these institutions with more challenges than expected. The UASs argued that they did not wish to prioritise completion over quality and, in addition, felt they had an obligation to allow access also to new entrants that from an academic point of view are less well-prepared than others. To assess the quality of degree programmes, most UASs made use of the indicator "student satisfaction", based on the results of the national student experience questionnaires. During the time period covered by the performance agreements, this indicator saw an increase, both for universities and UASs.

Nevertheless, the Review Committee in its advise to the minister concluded that six UASs had not achieved their targets, despite of the efforts made to increase study success and to improve other areas of performance. The minister then decided to impose a financial penalty on these six institutions.



The average completion rates in universities have risen considerably. The sharpest rise can be observed among the technical research universities (in green; see Fig 1a): from an average of 42% to 68%. At most of the research universities, the 2015 completion rates equalled or exceeded the ambition set for 2015. At two of the four universities that fell short, the 2015 completion rates were close to the target values.

In professional higher education, the average completion rate fell from approximately 70% to 67%. Large differences between the UASs can be observed in this respect (cf. Figure 1b). For example, among the large UASs in the western (and urbanized) part of the Netherlands (the blue arrows in Fig 1b), the average completion rates over 2015 (55%) were not only substantially lower vis-à-vis 2012 (63%), but also in many cases showed a persistent downward trend, that only recently had appeared to take a turn for the better. At thirteen UASs, completion rates over 2015 were significantly lower than over 2011, whereas ambitions for 2015 aimed at higher rates.

At the time of the conclusion of the performance agreements during the second half of 2016, most of the attention in the popular press and among the relevant stakeholders in the higher education sector was given to the indicators shown in figure 1. This was only natural because the financial sanctions attached to the performance agreements were very much tied to whether an institution had met its agreed ambition levels on seven indicators. In particular, indicators like completion and drop-out rates touch upon areas that are much less controllable than areas like the number of student contact hours (the ‘teaching intensity’ indicator), the share of academic staff holding a University Teaching Qualification (in Dutch: *Basiskwalificatie Onderwijs* – BKO) or the share of overhead (indirect) expenses. The Dutch higher education system is relatively open and UAS institutions are finding it difficult to influence the quality of the student cohort that starts their higher education career. Compared to universities, UAS institutions have much more challenging student population. They have relatively many students with non-Western backgrounds or students holding a vocational education and training (VET) degree. Whereas research universities have a much more homogenous population with many

entrants holding a pre-university education diploma. At the universities of applied sciences in the Randstad region (i.e. the urbanized Western part of the Netherlands) the proportion of incoming students from non-Western backgrounds is 26%.

The focus on indicators and quantitative targets is very much in line with a New Public Management (NPM) type of approach to governance in higher education, where financial consequences are tied to performance targets. NPM is a market-oriented approach to the provision of public services (Hood, 2007) and performance agreements laid down in contracts would qualify as NPM instruments. However, in NPM the targets would often be imposed from the top, with little room or acknowledgement for the professionals at the 'shop-floor level'. In the Dutch performance agreements, the target levels were set by the HEIs themselves, in light of their own strengths and weaknesses and respecting the national policy goals. In preparing the contents of their performance agreement, the HEIs had the opportunity to formulate ambitions that take into account the composition of their student populations. Looking back at the results of the performance agreement experiment, it turned out that some HEIs (in particular, UAS institutions) have set their ambitions too high and overestimated their power in influencing student success. The instruments that they put in place to raise student completion were either not effective enough or required more time to make their impact felt. In this respect, the performance agreements provided an important learning opportunity for the HEIs.

While the contents and results of the performance agreements in many ways stress measurable results, other valuable and often qualitative aspects of the HEIs' portfolio were also part of the agreements. In the Netherlands, performance agreements were placed in the policy debate of making Dutch higher education better aligned with the needs of society and in particular to create a higher education system that is 'fit for the future', offering increased quality and diversity. Recognising that a uniform policy tends to create uniform reactions, the performance agreements were seen as the way to create that diversity. It is to the issue of diversity that we now will turn.

4. Encouraging diversity in Dutch higher education

Looking back at the performance agreements in the Netherlands, the Review Committee, that was charged with overseeing the Dutch experiment, concluded that the results related to the objective of increasing differentiation and concentration were rather mixed (Review Committee, 2017b).

At the start of the experiment, part of the budget tied to performance agreements (on average: two-sevenths; the remainder was tied to quantitative indicators) had been awarded to HEIs in the form of competitive funds. This selective budget was awarded in proportion to the quality of a HEI's plans for programme differentiation and research concentration. HEIs that had submitted the best plans received relatively more than HEIs with a mediocre proposal.

In the performance agreements of the HEIs, the goals of encouraging differentiation and strengthening their focus on the basis of their strengths in education and research were very much tied to goals already included in the individual HEIs' strategic plans. The performance agreement goals often were linked explicitly to the strategic agendas of national/European policy-makers as well as regional policy-makers.

For the Review Committee, the concept of diversity proved difficult to grasp, because the phenomenon of diversity is naturally multidimensional and its nature is essentially qualitative – making it subjective and susceptible for criticism. In its annual monitoring reports and its assessment of the individual performance agreements, the Review Committee has operationalised diversity by looking at the range of programmes offered by HEIs (e.g. two-year Associate degrees, Bachelor’s degrees, Master’s degrees, broad-based bachelor programmes, two-year research master’s, selective honours programmes, professional Master’s programmes offered by UASs).

Diversity is a prominent theme in higher education (Birnbau, 1983; Marginson, 2017) and science and technology policy (Nowotny *et al.*, 2001). Diversity is held to be important as it is seen as a means to enhancing rigour and creativity, offering flexibility in the face of uncertain future progress, and promoting learning across programmes (Stirling, 2007). Diversity may act as a ‘resource pool’ in providing flexibility and resilience. More broadly, institutional and technological diversity are seen as stimuli for innovation and productivity. Diversity is a property of a *system* (e.g. the higher education system), rather than of its individual elements (e.g. universities). And it is an attribute of any system whose elements may be apportioned into *categories*. The concept of diversity however is quite unclear because it is a multi-faceted notion that combines many aspects. However, having a clear definition and measure of diversity can lead to an appropriate definition of the tools needed to improve diversity.

Stirling defines diversity as a combination of three basic properties – *variety*, *balance* and *disparity* (Stirling, 2007). These dimensions are not necessarily linked and do not evolve in the same way. Thus, it is impossible to interpret one of those dimensions without taking the other two into account.

Variety is the easiest dimension to understand and evaluate. It is the number of categories into which system elements is apportioned. It is the answer to the question: ‘how many types of a thing do we have?’ (Stirling, 2007, p. 709). All else being equal, the greater the variety, the greater the diversity. Obviously, a crucial issue here is to resolve the categories used to characterize variety. Variety is ‘simply’ the outcome of a category count. Distinguishing additional (sub-) categories however makes this a rather tricky issue.

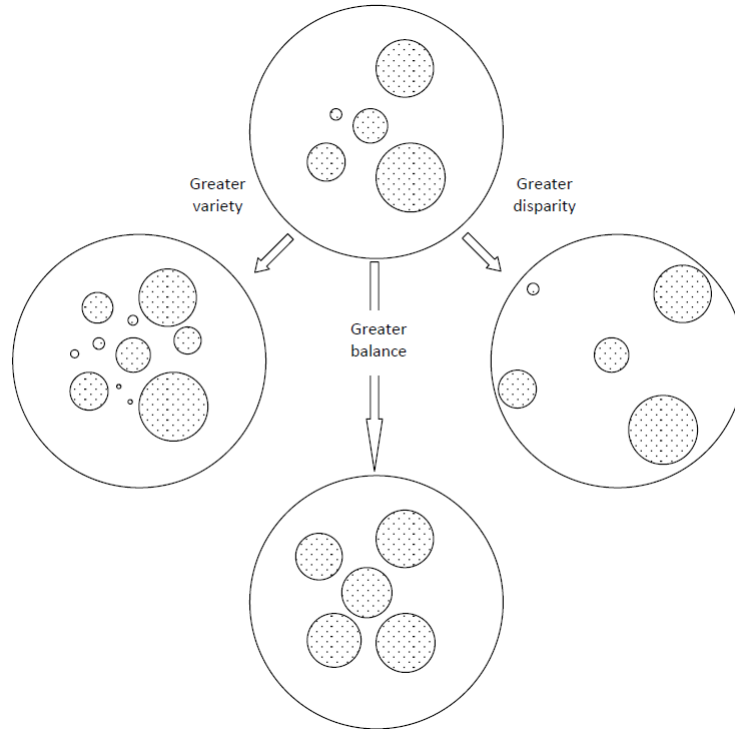
Balance refers to the pattern in the distribution of the quantity of a specific element across the relevant categories. Balance is the answer to the question: ‘how much of each type of thing do we have?’ (Stirling, 2007, p. 709) and is also known as evenness or concentration. Balance is perfect when each category is equally represented in the population. Balance refers to the extent to which different categories are equally well represented. It is considered that, all else being equal, the more even is the balance, the greater the diversity.

Disparity refers to the manner and degree in which the elements may be distinguished. It is the answer to the question: ‘how different from each other are the types of thing that we have?’ (Stirling, p. 709). Disparity goes beyond variety and balance by accounting for the nature of the categorization. All else being equal, the more disparate are the represented elements, the greater the diversity.

Diversity is a combination of these three basic properties. All else being equal, diversity increases when variety or balance or disparity increases (see Figure 2). However, one needs to recognise that each property constitutes the other two. Variety and balance, for instance, cannot be characterized without first considering disparity. The diversity of a system (e.g. the contents of degree programmes in higher

education) can only be assessed when its elements (i.e. programmes in this case) have been grouped into categories (e.g. disciplinary areas). Once this categorization has been done, variety corresponds to the number of categories; balance to the way the elements are spread among categories (e.g. the number of new students embarking on every category of programme); disparity to the level of difference between the categories (e.g. between every pair of them or between the two most distinct).

Figure 2: Increasing diversity along three dimensions



Source: Farchy and Ranaivoson (2011)

In order to judge different aspects of diversity, the Review Committee (RC) looked at a number of different diversity indicators too see how they developed over time (say, the period of the performance agreements compared to earlier periods). Some of the RC's diversity indicators are a combination of the three properties of diversity mentioned above. For instance, the combination of variety and balance is captured by the Herfindahl-Hirschmann Index (HHI; also known as the Simpson index). The HHI is often used to measure industrial concentration in a market. This indicator is defined as follows:

$$HHI = \sum_{i=1}^n MA_i * MA_i$$

where MA_i is the market share of institution i and n the number of institutions. The higher the value of the index, the weaker the diversity.

Another diversity indicator used by the RC is the Gini coefficient. This inequality index, often used in economics, is an indicator of balance (or evenness). A Gini coefficient of zero expresses perfect evenness, while a Gini coefficient of 1 (or 100%) expresses extreme inequality.

In its analysis of diversity, the RC is covering all three aspects of diversity. First of all it was confronted with the need to categorize HEIs and their degree programmes. For this, it felt that simply referring to research universities and universities of applied sciences (UASs) – the two categories of HEIs – was not enough. As already shown in the colour codes used for figure 1, a categorization in terms of the programme scope (say, comprehensive HEIs versus more specialised HEIs) and the size of the HEI (in terms of student enrolments) was employed to distinguish HEIs.¹ Of course, the question is whether such characteristics are the critical ones for categorizing HEIs. The RC choice seems to have been based more on convenience and tradition (i.e. referring to familiar categorisations) than on a truly rigorous analysis taking into account alternative characteristics. Critics might say that making an assessment of the evolution of disparity (are institutions belonging to one category becoming more distinct from institutions in another? Are there new categories appearing?) should be one of the *outcomes* of the RC's work and not be its starting point. After all, the RC essentially was installed to analyse differentiation and diversity.

Given that the objective of institutional differentiation ('profiling') was an important goal of the performance agreements, the RC paid a great deal of attention to the individual HEIs and was basing its analysis of diversity (a system level property, see above) on how individual HEIs have developed over time as a result of the performance agreements. The focus of the RC's diversity analysis was very much on the aspect of disparity and much less on balance or variety. As part of its analysis of institutional differentiation the RC analysed three partly overlapping features of a HEI's educational profile: (1) the range of programmes offered by a HEI, to see whether or not an institution is broadening the scope of its programmes by covering more disciplinary areas, (2) to what extent a HEI is focusing on particular programmes within that programme range, and (3) the market share of the programmes provided by the HEI.

In this paper we cannot possibly cover all aspects related to these profiling dimensions (see Review Committee 2017a for more on this topic) and we will focus now on the second and third aspect only: focus and market share. The first in particular is one aspect of the disparity in the system. It touches on the question of whether HEIs differ in terms of the emphasis they give to particular disciplinary areas – both in their education and in their research activity. To analyse this, the RC used information on student intakes in the institution's respective degree programmes, respectively the number of research publications published in various (sub-) disciplines.

¹ The research universities have been categorised into four groups: (1) the technical research universities together with Wageningen University & Research Centre ("technical"); (2) the general research universities with a relatively large intake ("large"); (3) the general research universities with a relatively small intake ("small"); and (4) the research universities set up on an ideological basis ("ideological"). The universities of applied sciences (UASs) have been categorised into seven groups. The first three groups comprise mono-sectoral universities: arts colleges ("arts"), agricultural UASs, and primary school teacher training colleges ("primary school teacher training"). The remaining UASs have been categorised into four groups based on the size of the entrance cohort (below 500 students; between 500 and 1000; between 1000 and 3000; larger than 3000).

With respect to education activity, the distribution of new entrants across the programmes within a HEI is an indication of the institution's focus areas in the range of programmes on offer. The equality in this distribution can be quantified by means of the Gini coefficient: The more unbalanced the distribution, as reflected in a higher Gini coefficient, the more sharply the focus areas will stand out and the clearer the institutional profile. Focus areas do not by definition result solely from an institutional profiling strategy; they also evolve as a result of fluctuations in new students' interest in specific programmes – an aspect covered by means of the Herfindahl (market share) indicator.

The development of educational focus areas is not necessarily an indication of increasing diversity, because two different HEIs could choose to focus on the same areas or themes thus reducing diversity. Imitation strategies have been studied extensively in higher education (DiMaggio and Powell, 1983). This is an illustration of the contrast between the institutional level and the system level: what is optimal in one sense may not be optimal from the other: Isomorphistic behaviour risks reducing diversity (Van Vught, 2009). However, for students the presence of focus areas can make a HEI stand out more clearly in the crowd.

The RC (Review Committee, 2017a) noted that, from 2006 onwards, student intake in the majority of research universities was spread out increasingly more evenly across the Bachelor's programmes on offer and this trend largely continued during the period of the performance agreements. This was even more the case for student intake in the Master's programmes. The RC interprets this as a tendency towards fewer focus areas. In the UAS sector, most UASs show a more even spread across Bachelor's programmes, giving no indication of the strengthening of particular focus areas. However, when looking at the offer of Master's programmes in the UAS sector, there are more institutions that move to more clearly visible focus areas during the period of the performance agreements.

Another profile feature that can be related to the range of programmes offered by a HEI is the market share of an institution, as expressed through its HHI index (see above). In the particular version of the HHI used by the RC, the squared market shares for each programme offered by an individual HEI are summed to arrive at the index for the individual HEI. If a lot of the students of a HEI are enrolled in programmes where the institution only has a small market share, its aggregated market share is relatively low. If a lot of its students enrol in programmes where it has a large market share, its aggregated market share will be high. The aggregated market share can be interpreted as a measure of concentration and therefore as another profiling feature for a HEI.

In the period up to 2011, most research universities saw a decline in their market shares (thus: a decline in their concentration measure) for their Bachelor's programmes on offer. After the introduction of the performance agreements, however, the picture changed: nine out of 17 research universities saw a relative rise in intake in Bachelor's programmes with a large market share. For most research universities, the market share indicator for their Master's programmes has declined over the period up to the year 2015, especially during the period of the performance agreements. In professional higher education, the number of Bachelor's and Master's programmes with a large market share has increased up to 2011. During the period of the performance agreements, however, this trend subsided. There is no visible growth anymore in the institutions' market share for Bachelor's as well as Master's programmes.

Overlooking the time period of the performance agreements, the Review Committee presented an integrated assessment of whether the Dutch HEIs have worked on establishing a more clear educational profile for their institution by combining the results of its analyses of focus areas (by means of Gini

coefficients) and market shares (HHI index) in educational provision. This was done by placing the development of the Gini and HHI index for the three years 2006, 2011 and 2015 in a quadrant graph like Figure 3. If a HEI would see its market share (horizontal axis) rise while at the same time the inequality across the degree programmes (vertical axis) in its educational offer would grow then the HEI would distinguish itself more from other HEIs. In Figure 3 this would then be shown as a movement towards the top-right corner. A movement in the direction of a 'less prolific' HEI would be shown as an arrow that points towards the bottom-left corner.

Figure 3: Two profiling features in the provision of Master's programmes by research universities: Focus areas and Market shares from 2006, 2011 towards 2015

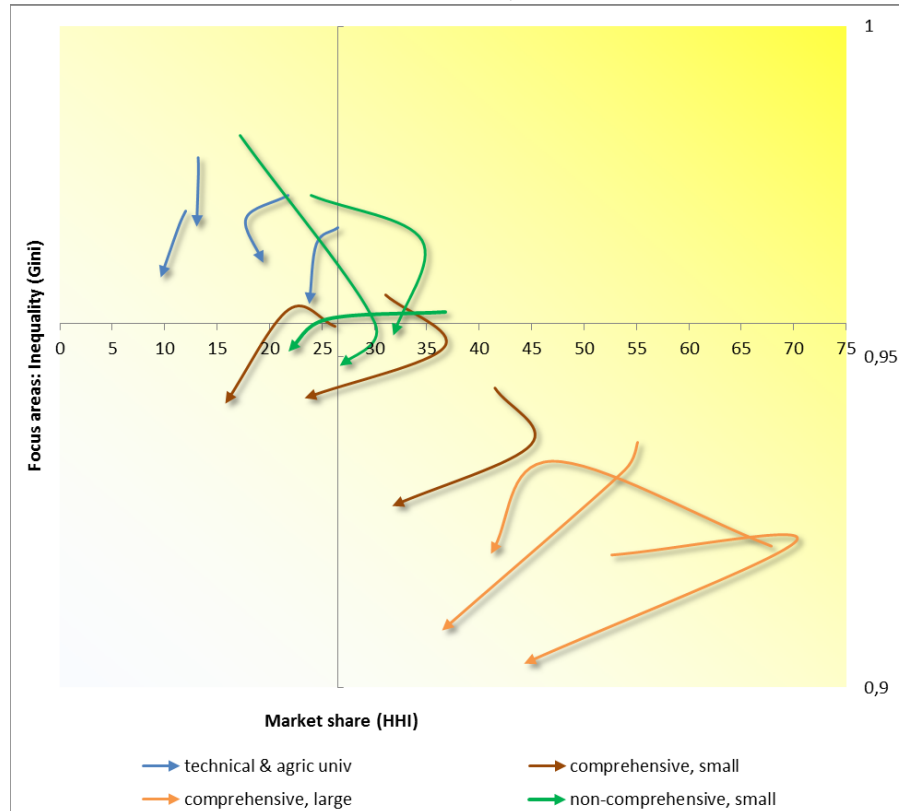


Figure 3 shows the result for the provision of Master's level programmes by research universities, where the changes over time are relatively larger and more clear than for areas like Bachelor's programmes. The arrows in Figure 3 clearly illustrate that research universities tend to get a less clear profile over time. For the UAS sector (that mostly offers bachelor's degree programmes) the picture shows no clear movement to one or the other quadrant – there is no indication of an increased profiling of institutions.

On the *research* side of higher education, the promotion of focus areas (research concentration is the English term; 'focus & massa' is the Dutch term) was another goal of the performance agreements. Since research in the UAS sector is still a rather small part of the institutions' activity robust judgements can

only be made for the research universities. To achieve the goal of research concentration, the performance agreements of universities set out many initiatives related to (inter-institutional and within-institutional) collaboration, making professorial appointments and internal reallocations of research capacity and resources. It obviously takes time for such initiatives to have an effect in terms of outcomes like research publications or other research outputs. Nevertheless, the Review Committee has sought to monitor the development of focus areas in research by analysing the spread of the scientific publications per university across the various (in this case: 250) sub-disciplines that can be distinguished in the world of science. An unbalanced spread across disciplines within a university (a high Gini coefficient) then would be an indication of research concentration in focus areas.

The Gini index is showing a decline over time, indicating that research publications are spread more equally across the disciplines. This proved to be the case in particular for the technical universities (of which there are four in the Netherlands). This trend started to become manifest already more than 15 years ago, but it has levelled off in recent years. The RC interpreted this trend as a decline in the number of focus areas in universities, implying a decline in the diversity of the entire Dutch university research system. The committee also witnessed an increase in the number of subdisciplines covered by each university's research publications. The differences between the various categories of universities (e.g. comprehensive versus technical universities) are becoming smaller. All of this suggests a decrease in the differences between the university research profiles. However, the RC noted that these tendencies are difficult to attribute to the performance agreements alone, since they already had manifested themselves much earlier. The debate on the effectiveness of the performance agreements will be continued in the next section.

Concluding this section on the diversity impact of performance agreements, it has to be noted that the initiatives undertaken by HEIs to enhance differentiation and diversity were not only understood and evaluated in terms of programme offer or research publications. The performance agreements not only included a set of quantitative targets but also specified the initiatives that the HEIs were to put in place for achieving those targets. These were often stated in qualitative terms and related to institutional strategies with respect to issues such as improving programme choice by potential students, enhancing student counselling services, introducing new didactical approaches in teaching and learning, or collaborating with vocational institutions to facilitate student transfer and study success.

5. Conclusion: lessons from the Dutch performance agreement experiment

Overlooking the outcomes of the performance agreements experiment in the Netherlands one can state that the results in terms of performance are mixed. In the previous sections we showed that in terms of increasing quality and completion in education the research universities showed significant progress while the UAS sector experienced problems in achieving the wished for completion rates. In terms of the second key objective of the performance agreements, increasing diversity by means of a differentiation in degree programmes and a strengthening of the HEIs' education and research profiles, results again are rather mixed. In terms of degree programme diversity, there is no clear sign of institutional differentiation: most institutions exhibit a more equal spread of activity over their programmes. On the research side, however, we can witness a clear tendency towards a decrease in differences between

universities. When looking at the spread of a university's scientific publications across the various disciplines we see no indication of the appearance of clear focus areas in research.

This may leave the reader with the impression that performance agreements have not achieved a lot. However, the policy evaluations that were made by three different committees were much more positive. The Review Committee itself produced an evaluation report, looking back at the experiment (Review Committee, 2017b). The association of UAS institutions also ordered an evaluation (Slob, 2017). Finally, the minister of Education had ordered an independent committee (the 'Commissie Van de Donk') to evaluate the experiment and make recommendations for a future system of (what was to be called) Quality Agreements (Evaluatiecommissie Prestatiebesteding Hoger Onderwijs, 2017). The three committees agreed on many issues. On the positive side, they concluded that the performance agreements (PAs) had contributed to the following outcomes:

- Putting the improvement of students' study success more prominently on the institutions' agendas
- Intensification of the debate about the drivers of study success (both among HEIs and within HEIs)
- More attention for the profiling (differentiation, focus areas) of HEIs
- Improvement of the dialogue between stakeholders in higher education (executive boards of HEIs, Ministry, department heads, associations of HEIs, review committee, representatives of business and community)
- Increased transparency and accountability, thanks to the setting of targets and the use of indicators
- Appreciation of the possibility for HEIs to share their 'story behind the numbers' with the Review Committee

Less positive were HEIs and students about:

- The decline of the HEIs' autonomy, due to the setting of national targets and uniform indicators
- The additional bureaucracy and administrative cost due to the emphasis on indicators
- The financial penalty associated with the non-achievement of goals
- The choice and definition of indicators, that in some cases contributed to unintended effects
- The lack of time available for a well-considered construction of the rules surrounding the experiment
- The fact that the experiment was managed largely by stakeholders that are quite distant from the 'shop floor level' (executive boards, managers, ministry, national committees and organisations), with a small role only for students in this process

In the evaluations of the performance agreements by the three evaluation commissions and the follow up discussion the need was reaffirmed for incorporating a performance-oriented component in the funding mechanism for HEIs. Already in an earlier stage, the Minister of Education expressed her intention to continue with some form of performance agreements, but was keen to stress that that the agreements should ultimately be about the quality of higher education and that quantitative targets should not receive priority over qualitative ones. The future agreements will therefore have the label Quality Agreements. On the topic of potential financial sanctions tied to quality agreements there still is no clarity (also due to the still ongoing negotiations about a new coalition government). The associations

of HEIs (for the research universities, respectively the UAS institutions) showed little enthusiasm for financial sanctions. However, the Review Committee in its evaluation (Review Committee, 2017b) concluded that attaching financial consequences to agreements fosters effectiveness. It argued that both the international literature and the Dutch experiment have shown that agreements are taken more seriously by all the parties, and have greater impact if financial consequences are attached.

Several options are under consideration for the next round of Quality Agreements, with the UAS sector expressing a wish to abolish any potential financial consequences. They feel that HEIs are autonomous and responsible institutions that should receive the freedom to decide on their ambitions in dialogue with its internal and external stakeholders. In their view, HEIs need to have the (financial) room to set their own targets (and indicators) and primarily report to their own stakeholders about the process towards achieving those goals. The term used here is horizontal accountability; the UASs prefer this over a vertical type of steering and see the government mostly playing a facilitating role.

Table 2: Performance agreements: good practices from OECD countries

<p>Evidence from several OECD countries has shown that performance agreements:</p> <ul style="list-style-type: none"> • are not solely meant to strengthen performance but also have aims such as encouraging HEIs to strategically position themselves, given their particular mission and strengths • can handle situations where HEIs have multiple objectives (education, research, innovation, entrepreneurship) and—within some nationally-set boundaries—can set their own targets; • improve the strategic dialogue between the government and the HEIs • help informing policy-makers and the public at large about the HEIs’ performance, thus improving accountability and transparency • can be used to promote horizontal collaboration between different actors.
<p>The evidence also points at the following lessons for an effective design of this type of agreements:</p> <ul style="list-style-type: none"> • Performance agreements are taken more seriously by all the parties and have greater impact if financial consequences are attached. They should include a mechanism to reward “overachievement” and not just be focused on budget cuts as a result of failure to meet indicator-based targets. • The nature of financial incentives must be carefully chosen. The budget linked to the agreements must be sufficiently large to have an impact, yet not so sizeable to the extent that the incentive becomes a goal in itself, or could lead to perverse effects. • Agreements must primarily pertain to goals and results. The indicators related to the targets should meet the requirements of validity, relevance, and reliability. Organisation-specific performance indicators can sometimes limit the scope for horizontal collaboration, with HEIs focusing solely on meeting the performance targets assigned to them.

Source: OECD (2017)

What the future Quality Agreements will look like is still unclear. However, despite the concerns expressed by associations of universities and students, there is support for connecting budgets to these

agreements. Most likely we will see some form of rewarding (and not: punishing) HEIS for meeting self-stated ambitions in terms of quality and differentiation. The fact that those ambitions are very much agreed in close dialogue with the HEIs' relevant stakeholders implies that the agreements will lose more of their New Public Management character and develop more into a steering instrument that fits the Public Value Management paradigm (Stoker, 2006; Vossensteyn & Westerheijden, 2016).

Whether performance agreements, or indeed performance-based funding formulas matter for the performance of higher education is a question that cannot be answered on the basis of the Dutch experiment with performance agreements alone. Although the Review Committee claims that the agreements were indeed effective, causality is difficult to prove. First of all, because the experiment was integrated in a larger policy framework with connections to other policy instruments and policy domains. Second, one needs to be aware of the fact that national policies and their associated incentives have to trickle down from the ministry (i.e. system level) to the HEI and, then, the level of the student or the academic (say teacher, researcher) to have an effect. The individual HEI and its specific characteristics is an important intermediate layer, where many intervening (either obstructing or facilitating) factors are located (Jongbloed & Vossensteyn, 2016). And, thirdly, one needs to understand the concept of performance much better (de Boer *et al.*, 2015); it is a multi-dimensional concept and highly subjective for that matter.

As described in section 2, many countries do employ performance-based funding mechanisms. However, they mostly continue to do so without necessarily having evaluated how effective the approach actually is (Jongbloed & Vossensteyn, 2016). Policy evaluations on the impact of PBF are rare, but there is some scattered evidence (see Table 2, based on OECD, 2017) that points to the benefits of performance agreements (see also: Claeys-Kulik & Estermann, 2015; De Boer *et al.*, 2015). In fact, the Dutch performance agreements are an experiment on the way towards finding an effective design for funding mechanisms and their associated incentive structure. While we have not answered the million dollar question "Do Performance Agreements Matter For Performance?", we have summed up some of the evidence that may be used to inform the design of future funding mechanisms.

References

- Birnbaum, R. (1983), *Maintaining Diversity in Higher Education*. San Francisco: Jossey-Bass.
- Claeys-Kulik, A. L., and Estermann, T. (2015), *DEFINE Thematic Report: Performance-based Funding of Universities in Europe*, Brussels. European University Association.
- Codling, A., & Meek, L. V. (2006), Twelve propositions on diversity in higher education. *Higher Education Management and Policy*, Vol. 18, No. 3, pp. 1-24.
- De Boer, H., Jongbloed, B. and other CHEPS colleagues (2015), *Performance-based funding and performance agreements in fourteen higher education systems*. Report for the Ministry of Education, Culture and Science. The Hague: Ministry of Education, Culture and Science.
- DiMaggio, P.J. and Powell, W.W. (1983), The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, Vol. 48, No. 2, pp. 147-160.
- Evaluatiecommissie Prestatiebesteding Hoger Onderwijs (2017), *Van Afvinken naar Aanvonken*. Den Haag: Ministerie van OCW.
- Farchy J. and H. Ranaivoson (2011), An international comparison of the ability of television channels to provide diverse programming: Testing the Stirling Model in France, Turkey and the United Kingdom. In UNESCO-UIS, *Measuring the diversity of cultural expressions*, pp. 77-138. Montreal: UNESCO.
- Hood, C. (2007), Intellectual obsolescence and intellectual makeovers: Reflections on the tools of government after two decades. *Governance*, Vol. 20, No.1, pp. 127-144.
- Jongbloed, B.W.A. and J.J. Vossensteyn (2016), University funding and student funding: international comparisons, *Oxford Review of Economic Policy*, Vol. 32, No.4, pp. 576–595.
- Marginson, S. (2017), *Horizontal diversity in higher education systems. Does the growth of participation enhance or diminish it?* Paper for the CGHE seminar 6 July 2017, London: UCL institute of Education / University College London
- Nowotny, H., Scott, P. & Gibbons, M. (2001), *Re-thinking science: knowledge and the public in an age of uncertainty*. London, UK: Polity Press.
- OECD (2017, to be published), *Supporting entrepreneurship and innovation in higher education in the Netherlands*. Paris: OECD
- Review Committee / Higher Education and Research Review Committee (2017a), *System report 2016. Fourth annual monitoring report on the progress of the profiling and quality improvement process in higher education and research*. The Hague: Review Committee.
- Review Committee / Higher Education and Research Review Committee (2017b), *Prestatieafspraken: Het Vervolgproces na 2016. Advies en Zelfevaluatie*, The Hague: Review Committee.
- Slob, A., Jeene, B.G., Rouwhorst, Y.T.M., Theisens, H.C., van Welie, E.A.A.M. (2017), *Kwaliteit door Dialoog. Eindrapport van de commissie prestatieafspraken hbo*. Den Haag: Vereniging Hogescholen.
- Stirling, A. (2007), A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, Vol. 4, pp. 707-719.
- Stoker, G. (2006), Public value management: a new narrative for networked governance? *The American review of public administration*, Vol. 36, No. 1, pp. 41-57.
- Van Vught, F. (2009), Diversity and Differentiation in Higher Education. In: F.A. van Vught (ed.), *Mapping the Higher Education Landscape*, pp. 1-16. Dordrecht: Springer.
- Veerman, C. et al. (2010), *Threefold Differentiation. For the sake of quality and diversity in higher education*. The Hague: Ministry Of Education.
- Vossensteyn, H., & Westerheijden, D.F. (2016), Performance Orientation for Public Value. In Pritchard, R.M.O., Pausits, A., and Williams, J. (eds.), *Positioning Higher Education Institutions* (pp. 227-245). Rotterdam: Sense Publishers.