The impact of the performance agreements on the quality of education in Dutch higher education

Bachelor's Thesis

Hannah Schmidberger, s1832743
University of Twente
Faculty of Behavioral, Management and Social sciences
Center for Higher Education Policy Studies (CHEPS)

Supervisors

Dr. Don Westerheijden
Senior research associate
Centre for Higher Education Policy Studies (CHEPS)

Prof. Dr. Ariana Need
Full professor
Department of Public Administration

University of Twente, Enschede
July 4, 2018

Abstract

Performance agreements were introduced in Dutch higher education in 2012 in response to growing student numbers, high drop-out rates and unsatisfied needs of students and the labor market. This study investigated the impact of the implementation of the policy on quality of education at Dutch universities of applied sciences. The study applied an Interrupted time series research design using longitudinal data from pre-existing datasets. The data analysis consisted of two separate, repeated-measures ANOVA's and two separate Linear Mixed Models. The research found a statistically significant, slight increase in student satisfaction after introduction of the policy, whereas student retention rates have not increased. Universities of applied sciences with higher scores on student satisfaction showed relatively higher retention rates. Last, the research identified that two quality measures included in the performance agreements were not related to student satisfaction and retention rates. Weaknesses in the research forbid to make a definite statement on the effectiveness of the Dutch performance agreements to increase quality of education. It is recommended to focus future research on the underlying mechanisms of performance agreements and to identify circumstances under which performance agreements might produce their desired effects. Several policy recommendations are made for the design of the next round of the Dutch 'Quality' agreements that will be from 2019 to 2024.

Key words: Performance management, Performance agreements, Accountability, Higher education, Quality of education, Student satisfaction, Student retention, Principal-agent theory, Resource-dependence theory

Acknowledgments

I would like to give a big thank to Don, for your major support and your encouraging attitude throughout the last half year. Another major thank goes to Ariana, my second supervisor, for giving me so valuable and detailed feedback on my writing. Further, I want to sincerely thank Henk, for giving me your much-appreciated advice regarding statistics and lending me a listening ear for all my questions and doubts. Also, thank you a lot Frans, for making this research at all possible by providing the necessary data, being a contact person when all sort of related questions came up and your help with discussing the results. Also, thank you so much Llewellyn for all your help and support, especially on short-call, and making me believe in my goal to pursue a career in psychology. Thank you also to the girls from the Statistical Support Service for all your advices on my analyses and for letting me still in your office. An incredible thanks to Mama, Papa and Meike for all your love and support throughout my studies, and for bearing me also in difficult moments. Another major thank goes to Sina for your friendship for so such a long time and helping me out with your very much appreciated logical thinking skills. Last, a major thanks to my Enschede girls – Hengameh, Inês and Karen – for being my friends and encouraging me in everything I do.

Contents

Chapter One - Introduction

Performance agreements in Dutch Higher Education

Performance management has become increasingly prevalent throughout government. Originally drawn from the private sector, performance management has proliferated in the private sector since the 1980s in order to enhance performance, productivity, accountability and transparency of public services (Forrester, 2011). The adoption of management principles from the private sector happened with the emergence of New Public Management as the dominant approach to public administration. New Public Management came along with a shift from process accountability as in Traditional Public Administration towards accountability based on outcomes (Crosby, Bryson, & Bloomberg, 2014; Halachmi, 2005). One approach to performance management is performance-related pay (PRP) that links payment to performance measured against predetermined objectives or targets (Forrester, 2011). The pay-for-performance strategy is expected to increase outputs by holding agencies accountable for their results and increasing responsiveness to program stakeholders and constituencies (Heinrich, 2002).

One area where performance management and performance-related pay have become salient is higher education. Increasingly, public authorities connect the budget that higher education institutions receive to performance (De Boer et al., 2015). Researchers distinguish between different designs of performance policies (De Boer et al., 2015). Two main approaches are identified. The broader notion of performance-based funding is normally interpreted as an approach that allocates budgets based on actual performance in the past. In contrast, performance agreements refer to a type of funding that rewards institutions for expected performance. Whereas there are plenty of cases of performance-based funding in higher education across the globe, performance agreements occur less frequently.

One case is the Dutch higher education system. The landscape of Dutch higher education consists of 18 research universities (WO) and 36 universities of applied sciences (HBO) which employ respectively 276.713 (WO) and 453.354 (HBO) students (Vereniging Hogescholen, 2018b; VSNU, 2018). In the past, Dutch higher education faced a growth of student numbers, but also high dropout rates and an overall mismatch between the higher education system and the needs of students and the labor market (OCED, 2014).

As a result, the Association of the Universities in the Netherlands (VSNU) and the Netherlands Associations of the Universities of Applied Science (Vereniging hogescholen) signed collective strategic agreements in 2011 (Hoofdlijnenakkoord) with the Ministry of Education, Culture and Science in which the universities promised to sharpen their profiles as well as to conclude contracts on their teaching and learning performances (Hladchenko, 2014; VSNU, 2016). The goal of those performance agreements was, amongst others, to raise student success and educational quality, foster profiling of higher education institutions and to increase the accountability of higher education institutions to ensure quality of teaching (De Boer et al., 2015; OCED, 2014). In this way, it was desired that completion rates would eventually increase and relatively high drop-out rates of students decrease (De Boer et al., 2015).

The performance agreements were reached in the context of a management philosophy that allowed the individual institutions to make their own strategic choices (De Boer & Van Vught, n.d.; Higher Education and Research Review Committee, 2012). Dutch Higher education institutions were prompted to submit individual proposals with own suggestions for profiling and performance on specific indicators that they intended to accomplish until 2015 and the feasibility of their goals (OCW, 2011; Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c). The Dutch performance agreements explicitly included indicators on measures that were assumed to increase student outcomes (OCW & HBO-raad, 2011). Furthermore, the majority of universities stated to continue or implement interventions to reduce drop-out and switch rates, and hence to increase the retention rate (Reviewcommissie & Onderwijs, 2014). On that basis, every university and the ministry concluded individual agreements on education quality, study success, profiling and valorization (OCED, 2014). As such, higher education institutions committed themselves to incorporate alterations in the organization of their educational processes.

With the introduction of the performance agreements, changes in the university funding model came along. A new component has been introduced to the so far exclusively formula-based allocation of funds that made up for 7% of the education funding and that was determined on basis of a university's strategic plan for 'quality and profile' (OCED, 2014). The goal was to put less focus on student numbers in higher education funding and to reward instead quality and profile of higher education institutions (OCW, 2011). Consequently, 5% out of the 7% education funding was granted to the universities of applied sciences as conditional funding that required the agreement and signing of the performance agreements. The budget would continue after 2016 on the condition that the performance targets have been achieved. The remaining 2% was granted on a more competitive basis according to the ratings the universities had received for their proposals (De Boer et al., 2015).

Despite the wide adoption of performance-funding in higher education, the majority of previous research has found no significant effects on student outcomes (Hillman, Tandberg, & Gross, 2014; Li, 2018; Shin, 2010; Umbricht, Fernandez, & Ortagus, 2017). Nonetheless, the stakeholders involved in the policy face major efforts of planning, implementation, and evaluation of the policy. Hence, it can be questioned whether the performance-funding is at all worth its efforts. Researchers even warn of unintended consequences of performance-funding policies, such as weakened academic standards or more selective recruitment procedures to increase the likelihood that the students enrolled will graduate (Dougherty et al., 2016). Despite all findings of previous research (Hillman et al., 2014; Li, 2018; Shin, 2010; Umbricht et al., 2017), conclusions should be drawn with caution since findings cannot be easily generalized due to broad differences in the contexts of the higher education systems and implementation of performance funding policies.

Therefore, the goal of this bachelor thesis is to evaluate to what extent the introduction of the performance agreements has improved, as intended, the quality of education in Dutch higher education.

The research will be limited to Dutch universities of applied sciences which will be further justified in the context of the case selection method.

Policy context

Governance in Dutch Higher Education thoroughly changed in the 1980s with the introduction of a new concept of government steering that was labeled 'Steering at a distance' (Kickert, 1995). Classical government steering characterized by legislation, prohibitions, and regulations was dropped for a completely new concept of government control in which emphasis was put on the self-steering capacities of the system. The idea was to delegate responsibilities to the higher education institutions and to strengthen autonomy and self-responsibility of Dutch higher education institutions. The goal was to enable institutions to respond more effectively to the needs of society (De Boer & Van Vught, n.d.). Yet, the government did not withdraw entirely from intervening but has remained or even strengthened 'control mechanisms' and accountability requirements (De Boer & Van Vught, n.d.). During the last 25 years, researchers have interpreted this phenomenon as 'strange hybrid' between two different governance models – the state control model and the state supervising model (De Boer & Van Vught, n.d.). In the state control model, the dynamics of the higher education system are almost entirely controlled and regulated by the government. In the state supervising model, the state merely defines the 'broad terms' of regulation, whereas higher education institutions can express their self-regulating capabilities (De Boer & Van Vught, n.d.). Maassen and Van Vught (1988) chose the metaphor of the 'Janus-Head' to describe the contrasting tendencies. In recent years, a major change of perspective occurred in which researchers have criticised the dichotomy of governance models as too limited to capture reality (De Boer & Van Vught, n.d.). De Boer and Van Vught (n.d.) argue that there is a third governance model, the state contract model, and metaphorize the steering orientation of Dutch government as Trimurti, the Hindu triad of gods.

In the late 2000s grievances in the Dutch higher education system became public which caused growing societal resentment, mistrust and a crisis in the system (De Boer & Van Vught, n.d.). Higher education institutions were accused of displaying strategic behavior to increase graduate numbers and enrolments to maximize budgets which eventually happened at the expense of quality of education (De Boer & Van Vught, n.d.). Political responses criticized that higher education institutions had too much autonomy and called for stronger governmental steering to strengthen the needs and positions of the consumers of higher education, i.e. students (De Boer & Van Vught, n.d.). Collective agreements (2008 to 2011) on certain performance goals had been reached, yet were not successful in their realization (De Boer & Van Vught, n.d.). Against this background, the instrument of individual performance agreements was introduced in 2011 with the agenda 'Quality in Diversity' (De Boer & Van Vught, n.d.). The performance agreements display a contractual relationship between government and individual institutions. It is argued that this contractual relationship differs from the mere state supervising model since it constrains institutional autonomy and causes costs of administrative control. Yet, it also contrasts

a complete state control model since it allows for joint decision-making, flexibility, and controls via an incentive structure (De Boer & Van Vught, n.d.). By allowing higher education institutions to choose themselves their targeted indicator-based objectives, it was intended to convey a 'sense of ownership' to higher education institutions regarding their individual performance agreement (De Boer et al., 2015).

The awareness of the context in which the performance agreements have been introduced is essential for understanding the implications of the policy. The policy was driven by the aim to stronger recognize the needs and positions of the clients in higher education, that is particularly students (De Boer & Van Vught, n.d.). Grievances in Dutch higher education at the expense of quality should be corrected through a shift in the incentive structure of higher education institutions. The type of contractual relationship between ministry and individual institutions has been a new type of governance that should bring about promising results. If one is aware of those reasons, the importance of evaluating the policy becomes clear. Empirical evidence is necessary to evaluate whether the policy has eventually increased the quality of education and benefited the customers of higher education. The research can eventually provide some of that empirical evidence to advance the debate on the benefits and detriments of the new governance model in Dutch higher education.

The notion of quality

The notion of quality was inherent in the performance management debate. Literature highlights the relative nature of quality with widely differing conceptualizations of the term (Harvey & Green, 1993; Lindsay, 1992). Two main approaches can be distinguished (Lindsay, 1992). One approach takes a "production measurement" view and equals 'quality' with 'performance' (Lindsay, 1992). The production orientation revolves around the measurement of inputs, processes, and outputs and distinguishes between a "resources" view, an "outcomes" view and a "value-added" view (Lindsay, 1992). The second main approach to quality takes a "stakeholder-judgment" view and highlights the importance of stakeholders in judging the quality of any particular educational institution or program (Lindsay, 1992). Inherent in this conception is the concern that quantitative performance indicators are unable to capture intangible, but important dimensions of quality.

The Dutch performance agreements included indicators of both conceptualizations of quality. Most indicators revolved around the "production measurement" view and measured, among others, drop-out rates, switch, graduation rates ("outcomes" view) or the percentage of students in excellence programs ("value-added" view). Drop-out and switch rates provide information on the retention rates of students, i.e. "the maintenance of continued enrolment for two or more semesters" (Crawford, 1999; Wild & Ebbers, 2002). Yet, the performance agreements also incorporated an indicator according to the 'stakeholder-judgment' view of quality that showed the students' opinions on their programs.

Previous research has mainly focused on the impacts of performance-based funding on quality from a 'production measurement' view and has measured outputs in terms of graduation rates or retention rates (Hillman et al., 2014; Li, 2018; Sanford & Hunter, 2011; Shin, 2010). To the knowledge

of the author, no previous research has analyzed the outcomes of performance-based funding or performance agreements from a 'stakeholder-judgment' view. This seems surprising since there is increasing awareness among universities of the importance of student satisfaction for their survival and a growing research body has focused on the role and concept of student satisfaction (Browne, Kaldenberg, Browne, & Brown, 1998; de Lourdes Machado, Brites, Magalhães, & Sá, 2011; de Oliveria Santini, Ladeira, Hoffmann Sampaio, & da Silva Costa, 2017; Elliott, 2002; Elliott & Shin, 2002; Kotler & Fox, 1995; Mark, 2013; Schertzer & Schertzer, 2004). Against the background of rapidly expanding universities, demographic shifts in student populations and increasingly competitive marketplace dynamics, the target market of students gets more and more courted by higher education institutions (de Lourdes Machado et al., 2011; Elliott & Healy, 2001; Kotler & Fox, 1995). Furthermore, student evaluations of programs and institutions have become increasingly transparent which might impact the choices of future students at which university to enroll. Hence, higher education institutions need to continuously satisfy their target market needs to also retain these students at their institutions on the long-term (de Lourdes Machado et al., 2011; Elliott, 2002; Elliott & Healy, 2001).

Given the relevance of student satisfaction for the survival of higher education institutions (Kotler & Fox, 1995) and the fact that the Dutch performance agreements even included such a 'stakeholder judgment view' indicator, a big interest arises in the impact of performance-based funding policies on student satisfaction.

Research questions

Based on the introduction to the research problem, the following overarching research question was formulated. *To what extent has the introduction of the performance agreements improved the quality of education at Dutch universities of applied sciences?* Related, the following sub-questions were posed.

RQ 1: To what extent has the introduction of the performance agreements improved student satisfaction at Dutch universities of applied sciences?

RQ 2: To what extent has the introduction of the performance agreements increased student retention rates[1] at Dutch universities of applied sciences?

RQ 3: To what extent has student satisfaction increased retention rates of students at Dutch universities of applied sciences?

RQ 4: To what extent have the quality measures included in the performance agreements increased student satisfaction and retention rates at Dutch universities of applied sciences?

[1] The choice of the concept 'Student retention' instead of 'Graduation rates' as dependent variable will be justified at a later stage in the Methodology chapter (Chapter Three).

Regarding the last research question, it shall be mentioned that the actual intention of the research was to investigate whether the actions taken by universities of applied sciences in response to the introduction of the performance agreements have increased student satisfaction and retention rates. These 'actions' would have involved both the performance on the quality measures included in the agreements (Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c), as well the individual interventions initiated by universities of applied sciences (Reviewcommissie & Onderwijs, 2014). Yet, only limited information was available on the different kinds of interventions implemented at Dutch universities and no information was available on the manner and extent of implementation. Furthermore, information was not available for all universities of applied sciences. Therefore, it was decided to focus instead on the effectiveness of the quality measures included in the performance agreements to increase student satisfaction and retention, for which enough information was available.

Social and scientific relevance

The research extends and refines existing empirical knowledge on the impacts of performance agreements. While previous research on the impact of performance funding on quality of education has mainly focused on the setting of the United States (Hillman et al., 2014; Li, 2018; Rabovsky, 2012; Sanford & Hunter, 2011; Shin, 2010), little is known about the impact of performance-related pay in European contexts. This is also the case for the Dutch higher education system where no research has been conducted yet on the actual impact of the performance agreements on quality of education. Therefore, the setting of the Netherlands contributes to a more differentiated view on the effects of performance funding. Furthermore, the case of the Dutch performance agreements is of high scientific relevance due to its particular policy design with bilateral performance agreements (OCW & HBO-raad, 2011). Previous research has mainly focused on the impact of performance-based funding on student outcomes (Dougherty et al., 2016; Hillman et al., 2014; Kelchen, 2018; Sanford & Hunter, 2011; Umbricht et al., 2017). Hence, the research extends empirical knowledge specifically on the impacts of performance agreements on student outcomes. Further, the research does not only focus on quality from a production measurement view as in most previous research (i.e. graduation rates or retention rates) but offers a new focus on the impact of performance agreements on quality from a 'stakeholder judgment' view, i.e. student satisfaction. Last, the research aligns with previous research concerning its practical relevance. Future students can benefit from the research in the long run since more evidence will be available on the effectiveness of performance agreements to improve quality of education and how institutions can contribute to higher student satisfaction and retention rates. Before eventually benefitting students in practice, the research provides a valuable policy evaluation of the Dutch performance agreements for the Dutch Ministry of Education, Culture and Science (and the remaining stakeholders) that can impact future policy decisions in Dutch higher education. On a more general level, more empirical evidence on the effectiveness of performance agreements will be collected that will

allow policy-makers to make better informed evaluations of the costs and benefits of performance agreements in higher education.

Outline of the thesis

The thesis is divided into 5 chapters. After this background chapter, Chapter Two provides an overview of previous research findings on the impact of performance-funding policies on student outcomes and the most relevant theories to understand their mechanisms. Furthermore, the reader will learn about the important role of the organization of the educational process for student satisfaction and student retention. The chapter ends with the formulation of hypotheses with the expected results for all research questions. In Chapter Three, the reader will be introduced to the research methodology of this study. The research design will be discussed, and the reader will be informed on the case selection and sampling procedure as well as the operationalization of concepts and the data collection procedure. Furthermore, it will be explained to the reader how to the study deals with missing data and the Descriptive statistics for the variables are presented. The chapter ends with an explanation of the statistical methods used in the data analysis to answer the research questions. Chapter Four presents the empirical results of the data analysis. The fifth and last chapter contains the conclusion and discussion of the results. Whereas the conclusion summarises the findings of the research, the discussion will interpret the results and integrate the findings into the existing body of research. Furthermore, limitations of the findings will be discussed. A discussion of the strengths and weaknesses of the overall research follows that eventually leads to suggestions for further research and policy recommendations.

Chapter Two - Literature review

There is mixed research evidence available on the impact of performance-based funding on student outcomes. Most studies have found no substantial impact on institutional performance that has been mainly conceptualized as graduation rates (Hillman et al., 2014; Sanford & Hunter, 2011; Shin, 2010; Umbricht et al., 2017). In contrast to those previous research findings, a recent article found that the introduction of performance funding incentives has improved STEM degree completions (Li, 2018). These contradictory results suggest an enormous complexity of the effects of performance funding policies. To gain a better understanding of the impacts of performance funding policies, the reader will be introduced in the following to the most relevant theories behind performance funding policies.

Principal-agent theory and resource-dependence theory

Several researchers (Hillman et al., 2014; Li, 2018) used principal-agent theory to explain why performance funding policies may or may not improve institutional performance. According to principal-agent theory, a principal enters into a contract with an agent who receives resources from the principal. The agent is expected to act on behalf of the principal and is delegated authority to make decisions (Hillman et al., 2014; Li, 2018). Both parties try to maximize their expected utility in this relationship (Hillman et al., 2014). Principals and agents pursue different interests and full monitoring of the agent's activities is impossible. Therefore, principals provide incentives to agents in order to align interests and ensure delivery of goals (Li, 2018; Sanford & Hunter, 2011). In theory, the "pay for performance" strategy should ensure that the agent does not evade their performance obligations (Hillman et al., 2014). Yet, challenges, such as the agent's own interest, information asymmetries, distrust or capacity constraints, can impede the agent from implementing the principal's agenda (Hillman et al., 2014). Information asymmetries can allow the agent to maximize his or her profit when the principal pays, and the agent does as little as he or she can.

Next to principal-agent theory, previous research used resource dependence theory to explain the effect of performance funding policies on institutional performance. According to resource dependence theory, universities respond to demands that impact their survival and growth (Sanford & Hunter, 2011; Shin, 2010). Hence, the incentive of a reward associated with a proposed policy determines whether institutions incorporate institutional changes (Shin, 2010).

In the case of the Dutch higher education system, the Ministry of Education, Culture, and Science (principal) concluded individual multi-year performance agreements with every individual university and university of applied sciences (agent) (OCED, 2014). Hence, the performance agreements can be conceptualized as the bilateral contract between principal and agent (De Boer et al., 2015). Resources by the principal, that is funding by government, is provided according to the achievement of the performance goals. This increases the pressure on the agent, who is dependent on the resources by the agent, to perform according to the principal's agenda. Following actions by the agent are expected.

Principal-agent theory and resource-dependence theory were included in the theoretical framework to help the reader to gain a basic understanding of the power relations between the contracting parties of the performance agreements, that is the Ministry of Education, Culture and Science and universities/ universities of applied sciences. Furthermore, the theories are considered valuable since they offer not only explanations for a 'positive' outcome (i.e. Performance agreements have increased the quality of teaching) but also for 'negative' outcomes (Performance agreements have not increased the quality of teaching). Because of their flexibility to account for different results the theories have been chosen to frame the discussion of the results at a later stage of this research.

Underlying mechanisms of the performance agreements

Whereas most previous research has focused on the direct causal chain between performance-based funding and institutional performance (Hillman et al., 2014; Sanford & Hunter, 2011; Shin, 2010; Umbricht et al., 2017), limited research is available on the intermediate links in the causal chain of performance funding policies (Kelchen & Stedrak, 2016; Rabovsky, 2012). Yet, knowledge on these intermediate links is relevant to give valid explanations for the ultimate outcomes observed.

Both Rabovsky (2012) and Kelchen and Stedrak (2016) were interested in the administrative responses by higher education institutions in response to performance policies. Rabovsky (2012) investigated whether performance funding policies have changed institutional spending patterns at higher education institutions. Similarly, Kelchen and Stedrak (2016) analyzed whether performance-funding has changed patterns and allocation of revenues, expenditures and financial aid at two-year and four-year colleges. As a theoretical framework, Rabovsky (2012) uses a conceptual model of the causal logic of performance funding policies that assumes that incentives will be restructured that results in an administrative response by institutions which will lead to improved outcomes (Rabovsky, 2012). The assumption of the causal logic of performance funding policies that financial incentives bring about results is basically the same as in principal-agent theory and resource-dependence theory. Still, the conceptual model was chosen since it makes explicit the intermediate links between performance policy and outcomes (Rabovsky, 2012).

Figure 1. Causal Logic of Performance Funding Policies (Rabovsky, 2012, p. 679)

In the case of the Dutch performance agreements, universities of applied sciences committed themselves to incorporate alterations in the organization of educational processes by signing of the contracts. The majority of universities stated to continue or implement interventions to reduce drop-out and switch rates, and hence to increase the retention rate (Reviewcommissie & Onderwijs, 2014). Many

interventions have started at the process of recruiting students with requesting motivation letters in the application, more selective admissions, intake conversations, or study choice checks (Reviewcommissie & Onderwijs, 2014). On the other hand, various interventions have focused on the actual process of studying and have included student accompaniment, mentoring, tutoring or coaching (Reviewcommissie & Onderwijs, 2014).

Furthermore, the Dutch performance agreements explicitly included indicators on measures that were assumed to increase student outcomes (OCW & HBO-raad, 2011). The measures included raising the quality of teaching staff, the intensity of education and to increase the proportion of teaching and research staff vs. support and management personnel (OCW & HBO-raad, 2011). Quality of teaching staff should be increased by employing more teachers with a master's or PhD degree to increase analytical and research competencies of students (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017). The intensity of education in the first year of bachelor's programs should be standardized to 12 contact hours per week (OCW & HBO-raad, 2011). Last, institutions should employ more teaching and research staff compared to support and management personnel to focus resources on the primary process of teaching that directly benefits students (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017)

Overall, there is evidence that universities of applied sciences have responded to the implementation of the performance agreements by incorporating changes in the organization of educational processes. Therefore, it is expected that quality of education has increased after the introduction of the agreements. The following paragraph gives more information on why it is assumed that these changes in the organization of educational processes have increased the quality of education.

Role of the organization of educational processes

The essential role of the organization of the educational process for increasing student satisfaction and student retention can be concluded from the conceptual model by Schertzer and Schertzer (2004). The authors argue for a 'relationship marketing' mindset in higher education to improve student retention. The customer focus in higher education draws on services marketing literature which assumes that customers are satisfied if the quality of service meets or surpasses their expectations (Mark, 2013). Hence, universities can contribute to student satisfaction if they meet students' needs and expectations (Elliott & Shin, 2002). Schertzer and Schertzer (2004) propose a model of student retention in which the role of student satisfaction is highlighted. The authors argue that student-institution values congruence and student-faculty values congruence determine the academic fit with the student. The right 'academic fit' is assumed to play an important role since it is positively related to student satisfaction and institutional commitment. It is argued that the more satisfied students are, the more likely they engage in institutional commitment which eventually increases the chance of retention (Schertzer & Schertzer, 2004).

Therefore, Schertzer and Schertzer (2004) argue that it is important for higher education institutions to target and attract the students with the 'right' fit. Against this background, the actions

taken of Dutch universities of applied sciences after the introduction of the performance agreements become relevant. It is assumed that recruitment interventions, such as intake conversations, have enhanced the 'academic fit' of students since students could experience themselves beforehand whether their priorities match the campus environment, that is whether their values are congruent with the institution. At the same time, students could interact with members of the faculty which contributes to student-faculty values congruence.

However, the choice of the 'right' students alone cannot ensure student satisfaction. Schertzer and Schertzer (2004) ask higher education institutions to continuously treat students as 'customers' to increase their satisfaction since student satisfaction is shaped continuously through repeated experiences in campus life (Elliott & Shin, 2002). In that context, interaction between the student and university personnel is important (Schertzer & Schertzer, 2004).

Research has shown that effective academic advising leads students to feel more positive about their institution in all (Schertzer & Schertzer, 2004). Overall, personal relationships with faculty and/or staff are often desired by many students and parents, and the chance of students recommending the university to others – understood as an indicator for student satisfaction - is strongly correlated with the extent of interaction between the student and faculty (Browne et al., 1998; Schertzer & Schertzer, 2004). There is empirical support for the hypothesis that those conditions are more likely to be found in higher education institutions with a lower student-staff ratio and it might be assumed that the same effects occur for small-scale universities with a smaller number of students enrolled (Bradley, Noonan, Nugent, & Scales, 2008; McDonald, 2013). In line with Schertzer and Schertzer (2004), Elliott (2002) found that the students' feeling of belonging and the provision of a quality education were key determinants of student satisfaction. De Oliveria Santini et al. (2017) found positive correlations between professor quality, academic service quality as well as teaching service quality and student satisfaction that were stronger than the correlations with the administrative service quality and the support service quality (de Oliveria Santini et al., 2017).

As a result, it is assumed that student satisfaction at Dutch universities of applied sciences must have increased after the introduction of the performance agreements, since various interventions have been implemented regarding the actual process of studying, such as student accompaniment, mentoring or coaching (Reviewcommissie & Onderwijs, 2014). Furthermore, it is assumed that the quality measures included in the performance agreements must have been effective in increasing student satisfaction and retention. By employing more teaching and research staff compared to support and management personnel (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017), more (personnel) resources are available for the primary process of teaching which might have increased academic service quality and teaching service quality perceptions among students (de Oliveria Santini et al., 2017). Furthermore, by employing more teachers with a master's or PhD degree with analytical and research competencies (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017), students might have perceived education to be of more quality (Elliott, 2002).

Although student satisfaction is assumed to predict retention rates of students (Schertzer & Schertzer, 2004), there might be other factors that limit the expected impacts. Open access institutions (as it is the case in the Netherlands) have reported major difficulties in responding to the funding formula regarding many at-risk students enrolled who face social and economic challenges and lack academic preparation (Dougherty et al., 2016). Therefore, one might assume that institutions with lower percentages of at-risk students have higher satisfaction and retention rates than an institution with a higher percentage of at-risk students. Since open-access institutions have no direct control over the number of at-risk students enrolled, but the student body is still assumed to have an impact on the retention rate, student characteristics are included as control variables in the empirical model. Furthermore, the analysis will control for the number of students enrolled and the student-staff-ratio at Dutch universities of applied sciences.

Hypotheses

Based on the theoretical framework, the following hypotheses were derived:

H1: *Student satisfaction at Dutch universities of applied sciences has increased over time after introduction of the performance agreements.*

H2: *Student retention rates at Dutch universities of applied sciences have increased over time after introduction of the performance agreements.*

H3: *The higher the student satisfaction, the higher the student retention rates at Dutch universities of applied sciences.*

H4: The lower the indirect costs and the higher the quality of teaching staff at Dutch universities of applied sciences, the higher their levels of student satisfaction and student retention rates. [2]

Conceptual models

The following figures present the conceptual models with the relationships to be tested in the analyses. The arrows are labeled according to the hypothesis that is tested. The signs indicate the predictions made for the relationships. Fig. 2 shows the conceptual model for the trend predictions made for student satisfaction and retention rates (H1 and H2). Fig. 3 presents the conceptual model with the hypotheses on the impact of the quality measures included in the performance agreements on student satisfaction and retention rates next to additional factors that are assumed to have an impact on student satisfaction and retention rates.

[2] The indicator 'Intensity of education' is not included in the analysis due to operationalization issues (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017) and since a majority of institutions had already reached the demanded threshold of 12 contact hours per week (1st year) before reaching the agreements.

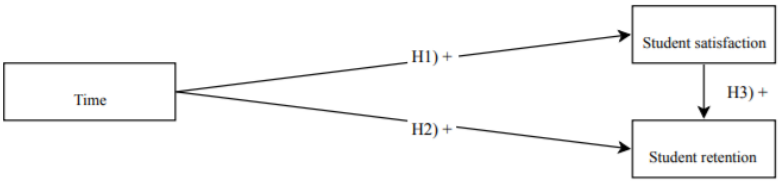

Figure 2. Conceptual model of the hypothesized development of student satisfaction (H1) and retention (H2) over time and the suggested impact of student satisfaction on student retention (H3).

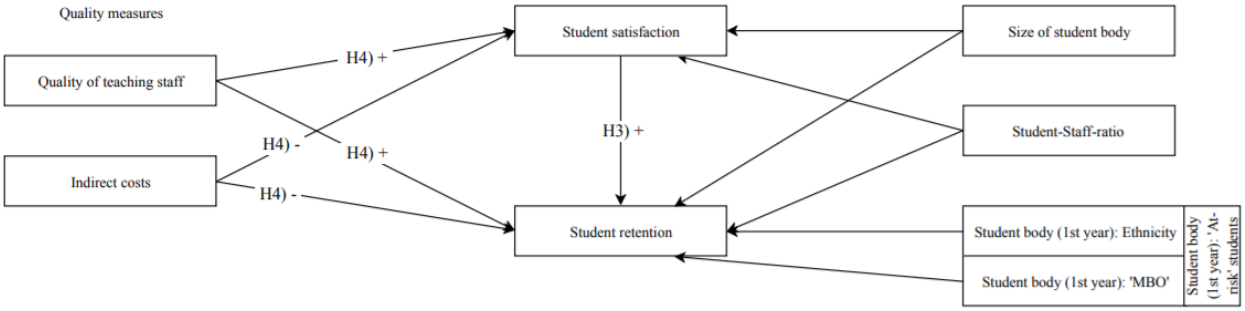

Figure 3. Conceptual model of the hypothesized impact of student satisfaction on student retention (H3) and the assumed impact of the quality measures on student satisfaction and student retention (H4).

Chapter 3 - Methodology

Research design

In the case of all research questions, the study makes use of an Interrupted time series research design. Interrupted time series is a common research design in many areas of study in order to consider the impact of large-scale interventions or public policy changes (Linden, 2015). It is suitable when observing an outcome variable over multiple, equally spaced time periods before and after the introduction of an intervention which is assumed to impact the trend (Bernal, Cummins, & Gasparrini, 2017; Linden, 2015). In the case of the Dutch performance agreements, it is assumed that the adoption of the agreements must have had an impact on the quality of education provided. The trend in the development of quality of education is investigated for the period 2011 to 2015 for student satisfaction, and for 2003 to 2014 for student retention. The research conducts a single-group analysis with the Dutch universities of applied sciences being the only group under study without comparison groups (Linden, 2015). Researchers argue that Interrupted time series require a clear differentiation of the pre-intervention period and the post-intervention period (Bernal et al., 2017). In the case of the Dutch performance agreements, the framework agreement ('Hoofdlijnenakkoord") was signed in 2011, followed by the individual performance agreements at the end of 2012. The clear cut between the pre-intervention period and the post-intervention period is considered the conclusion of agreements at the end of 2012. Next to a clear differentiation of the pre-intervention and post-intervention period, sequential measures of the outcome should be available before and after the intervention (Bernal et al., 2017). This applies for the performance agreements with yearly midterm reviews.

In interrupted time series, the change of the dependent variable in reaction to the intervention can take several forms varying from a gradual change in the gradient of a trend, a change in the level or both, an immediate change in the dependent variable or only after a lag period (Bernal et al., 2017). In the case of the Dutch performance agreements, data on the dependent variable (quality of education) are only available until 2015 and hence, can show trends until three years after the implementation. It might be that changes in educational quality require a longer implementation period so that interpretations on basis of data until 2015 might be misleading. The paper will take this into account as a limitation of the study, where further data collection is necessary.

Population of cases

The units of analysis are Dutch universities of applied sciences (n = 35)[3]. Although the Dutch performance agreements included both research universities and universities of applied sciences,

[3] The research excludes several universities of applied sciences that have been recently subject to merger with other universities of applied sciences. This concerns the Hogeschool Edith Stein, Hogeschool Helicon, Stoas Hogeschool, Christelijke Agrarische Hogeschool, CAH Vilentum University of Applied Sciences as well as Aeres and Thomas More University of applied sciences.

universities of applied sciences were chosen for units of analysis as (almost all) universities of applied sciences included the student satisfaction indicator in the performance agreements, while most universities did not (Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-a, n.d.-b). The teaching focus makes universities of applied sciences valuable units of research, as major influences through research or knowledge transfer activities can be excluded. Apart from that, the choice of universities of applied sciences as units of analysis ensure a large enough sample size for running the statistical analysis.

Operationalization of concepts and data collection methods

The variable 'Student retention'[4] is conceptualized as the student body at a higher education institution that is not affected by drop-out or switch (Student retention = 1 – (% drop-out + % switch). Data on student retention are derived from the 'Dienst Uitvoering Onderwijs' (DUO). The dataset contains longitudinal data covering the period from academic year 2003/2004 to 2014/2015. 'Drop-out' is operationalized as the percentage of the total number of full-time bachelor's students (first-year higher education) who are no longer enrolled in the same higher education institution after one year. "Switch" is operationalized in terms of the percentage of the total number of full-time students (first-year higher education) who are enrolled after one year of study in another study at the same institution (Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c).

Data on student satisfaction are derived from National Student Survey (NSE - De Nationale Studenten Enquête) conducted by the Dutch independent foundation Studiekeuze123. Student satisfaction is operationalized as the proportion of respondents (full-time students) that is satisfied (score in category 4) or very satisfied (score in category 5) with the program in general compared to the total number of respondents (full-time students at the institution) (Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c). Data at the program level were aggregated at the institutional level. Data are available for the period academic year 2011/2012 to 2015/2016.

Data on the measures 'Teacher quality' and 'Indirect costs' are derived from the 'Reviewcommissie Hoger Onderwijs en Onderzoek". The data are originally retrieved from the annual reports of the universities of applied sciences and cover the years 2012 to 2015. Teacher quality is operationalized as the share of teaching staff with a master's or PhD (regardless of the nature of their employment) in the total number of teaching staff. The variable 'Indirect costs' is operationalized as the ratio between the number of teaching and research staff vs. the number of support and management personnel (Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c).

[4] Student retention instead of student graduation rates has been chosen as dependent variable since trends in graduation rates after introduction of a policy are conclusive at the earliest at the graduation of the first year of students enrolled after the introduction of the policy, and hence in 2016. In contrast, trends in 'Switch' and 'Uitval' are already observable after 1 year since the introduction of the policy, which is why the concept of retention rate is considered the best early-warning indicator for study success in the available data.

The variable 'student-staff ratio' is operationalized as the number of students at a university of applied sciences divided by the number of staff at a certain university of applied sciences at a certain moment in time. A student is counted as enrolled if he or she is active on October 1 of a certain academic year (Vereniging Hogescholen, 2018a). The denominator 'staff' is operationalized as the number of personnel employed on full-time or another hours-base at the moment of October 1 of each year (Vereniging Hogescholen, n.d.-a). The data for the 'student-staff ratio' are publicly available and can be found on the website of the 'Vereniging Hogescholen' (Vereniging Hogescholen, n.d.-b). The data are available for the academic year 2008/2009 to 2017/2018.

The variable 'Size of student body' is operationalized as the total number of students enrolled at a certain university of applied sciences in October of the respective academic year (Vereniging Hogescholen, 2018a). A student is counted as enrolled if he or she is active on October 1 of a certain academic year (Vereniging Hogescholen, 2018a). Data on the number of enrolled students are available for the academic year 2008/2009 to 2017/2018 and have been retrieved from the 'Inschrijvingen' dataset provided by the 'Vereniging Hogescholen' (Vereniging Hogescholen, n.d.-b).

'At-risk' students are conceptualized as students who are either 'Non-Western' or with an 'MBO' background (Dougherty et al., 2016; Zijlstra et al., 2013). Both 'Non-Western' students and students with an 'MBO' background are assumed to be 'at risk'. Data on the ethnic background of students are derived from 'Vereniging Hogescholen" and cover the academic years 2008/2009 to 2017/2018 (Vereniging Hogescholen, 2018a). 'Non-western' is operationalized in terms of students of which at least one parent comes from abroad, that is from Turkey, African countries, Latin-America or Asia, except former Dutch-India, Indonesia, and Japan. Data on the educational background are also derived from the "Vereniging Hogescholen" and cover the academic years 2008/2009 to 2017/2018 (Vereniging Hogescholen, 2018a). The facet 'MBO' applies if the MBO degree is the highest preliminary training that a student has successfully completed before the current degree program. The predictor 'Share of at-risk students' is operationalized as the share of first-year at-risk students out of the total number of first-year students enrolled. A first-year student is operationalized as 'student who has maximal one enrolment of this type per type of higher education' (HBO) (Vereniging Hogescholen, 2018a).[5]

[5] The control variable 'Share of at-risk students' is included in the analysis of the extent to which the quality measures have impacted student retention rates at Dutch universities of applied sciences (RQ 2). It is not included in the analysis of the impact of the quality measures on student satisfaction (RQ 1). The reason therefore is that information on the share of 'Non-Western' and 'MBO' students are only available for the first-year student body (Vereniging Hogescholen, n.d.-b). This matches the variable 'Student retention' that also refers to first-year bachelor's students. Yet, the variable 'Student satisfaction' refers to the total number of full-time students at an institution, so that first-year student characteristics as predictor would not be meaningful.

Data analyses

Two single-factor repeated measures Analyses of Variance (ANOVA) are used to analyze the impact of the introduction of the performance agreements on student satisfaction and student retention at Dutch universities of applied sciences (H1 and H2). In both analyses, the independent variable is the categorical variable 'Time' (within-subjects-factor) and the dependent variables are the continuous variables 'Student satisfaction' and 'Student retention' respectively. The trend for student satisfaction is analyzed for the period AJ 2011/12 to 2015/16, whereas the trend in student retention is analyzed for the period AJ 2003/4 to 2014/15. To test the hypotheses 3 and 4, two separate Linear Mixed Models were conducted. The first linear mixed model included 'Student satisfaction' as the dependent variable, with 'Time', 'Quality of teaching staff', 'Indirect costs', 'Student-staff-ratio' and 'Size of student body' as predictors. In the second linear mixed model, 'Student retention' was used as the dependent variable, with 'Time', 'Student satisfaction', 'Quality of teaching staff', 'Indirect costs', 'Student-staff ratio', 'Size of student-staff-ratio' and 'Share of first-year at-risk students' as explanatory variables. Both linear mixed models included panel data for the period AY 2011/2012 to AY 2014/15.

In the case of both Linear Mixed Models, time was included as a covariate to treat time as a linear trend (Field, 2013, p. 854). Time was specified as a fixed effect to add the potential growth curves in the model (Field, 2013, p. 854). Furthermore, it was expected that the relationship between time and both dependent variables 'Satisfaction' and 'Retention' have random intercepts and random slopes so that these parameters had been specified in the model (Field, 2013, p. 855). The random intercept assumes that at the beginning of the measurement, x = 0, there was variability in student satisfaction and retention rates at Dutch universities of applied sciences. The random slope assumes that there is a lot of variability in student satisfaction and student retention rates at Dutch universities of applied sciences also after that. The remaining covariates were specified as fixed effects. The following model was formulated:

$$
\begin{aligned}
Satisfaction_{ij} = {} & b_{0j} + b_{1j}Time_{ij} + b_2 TeacherQuality_{ij} + b_3 IndirectCosts_{ij} \\
& + b_4 StudentStaffRatio_{ij} + b_5 SizeStudentbody_{ij} + \varepsilon_{ij} \\
b_{0j} = {} & b_0 + u_{0j} \\
b_{1j} = {} & b_1 + u_{1j} \\
Retention_{ij} = {} & b_{0j} + b_{1j}Time_{ij} + b_2 Satisfaction_{ij} + b_3 TeacherQuality_{ij} + b_4 IndirectCosts_{ij} \\
& + b_5 StudentStaffRatio_{ij} + b_6 SizeStudentbody_{ij} + b_7 FirstYearEthn_{ij} \\
& + b_8 FirstYearMBO_{ij} + \varepsilon_{ij} \\
b_{0j} = {} & b_0 + u_{0j} \\
b_{1j} = {} & b_1 + u_{1j}
\end{aligned}
$$

The analysis will start with a bivariate analysis to show the value of Pearson's correlation coefficient (r) between every pair of variables as well as the one-tailed significance of each correlation

(Field, 2013). The correlation matrix will provide a preliminary look for multicollinearity (p) (Field, 2013). To account for more subtle form of multicollinearity, the variance inflation factor (VIF) will be considered. The output of the main analysis will be first analyzed regarding the statistical significance of the contributions of the predictors (p-value). The b-values will indicate the effect size, i.e. the individual contribution of each predictor to the model if the effects of all other predictors are held constant (Field, 2013). Beta-coefficients will be analyzed to compare the strengths of relationships between a certain predictor of many and the outcome (Field, 2013).

Missing data

Several variables are affected by missing data over the time periods included in the analyses. This concerns the variables 'number of personnel employed', 'Teaching quality', 'Indirect costs', and 'Student satisfaction'. In total, 66 out of 1894 values are missing which make out 3.367% of all values included in the dataset.[6] Since the dataset provides for every variable longitudinal data, missing data are imputed manually via linear imputation. Hence, a missing value is imputed through calculation of the series mean with the two measurement points around the measurement gap – both for enclosed measurement gaps or missing start- or end- measurements. If a case is missing data for the majority of measurement moments, it is excluded from analysis.

For the variable 'Number of personnel employed', 5 values are missing for the measurement moment 2013 (14.3%). Simple linear imputation was used.

The variable 'Teacher quality' showed 16 missing values for the time JVS 2012 (45.7%), 2 missing values for 2013 (5.7%) and 1 missing value both for JVS 2014 and 15 (2.9% respectively). The series mean is imputed for JVS 2013, 2014 and 2015. Since almost half of the values are missing for the JVS 2012, it was decided to exclude the measurement moment from the analysis. Instead, the variable is replaced through the measurement moment 'Nulmeting' with full data being collected at the moment the proposal was handed in, i.e. the end of 2011/start 2012. For the measurement moment JVS 2013,

The variable 'Indirect costs' shows 19 missing values for the time JVS 2012 (54.3%), 4 missing values for JVS 2013 (11.4%), 2 missing values for JVS 2014 (5.7%) and 1 missing value for JVS 2015 (2.9%). The series mean is imputed for JVS 2013, 2014 and 2015. One case (H37) was removed from the dataset since a majority of values were missing for the particular case. Since more than half of the values are missing for JVS 2012, also here, the measurement moment JVS 2012 was excluded from analysis. Instead, the variable is replaced through the measurement moment 'Nulmeting' with full data being collected at the time when the proposal was handed in, i.e. End of 2011/Start of 2012.

[6] The analysis of missing data was conducted only for the data included in the analysis. Available data for measurement moments not included in the analysis were not considered for the analysis of missing values. Hence, 'Switch' and 'Uitval' were scrutinized for the period AY 2003 to 2014, 'Student satisfaction' for AY 2011 to 2015, and all remaining variables included for the period AY 2011 to 2014.

For the variable 'Student Satisfaction', 6 values are missing for AY 2011/12 (17.1%), and 3 values each for AY 2012/13, 2013/14, 2014/15, respectively (8.6% respectively). The missing data are imputed via linear trend imputation as described above. 3 cases (H05, H09, and H36) were excluded from the analysis since a majority or all values were missing. The remaining missing values are filled up through simple linear imputation. The following analysis will exclude all cases with missing values that have not been imputed due to the reasons mentioned earlier.

Descriptive statistics

The following table reports the descriptive statistics of the variables included in the analyses. Since longitudinal data are included in the analysis, the table will provide information on the time periods for which data are included. Furthermore, the tables report the number of values included in the analysis (N), the max and min value, the mean value as well as the standard deviation, i.e. an estimate of the average variability (spread) of the data (Field, 2013). The variables 'Satisfaction' and 'Retention' are included twice in the overview of Descriptive statistics since they are used in different analyses for different time periods. To make it clear for the reader which duplicated variables belong to which analysis, indications are given in the table.

Table 1. Summary Descriptive statistics table with all variables included in the analyses. Assignment of variables to statistical analyses according to time periods.

Descriptive Statistics						
	Period of time	N	Min.	Max.	Mean	Std. Dev.
Satisfaction	2011 – 2015 (ANOVA)	160	54.60	93.80	73.411	8.399
Retention	2003 – 2014 (ANOVA)	420	46.65	92.63	69.978	8.483
Satisfaction	2011 – 2014 (LMM)	128	54.60	91.90	72.422	8.471
Retention	2011 – 2014 (LMM)	128	52.37	92.63	67.642	8.551
Teacher quality	2011 – 2014	128	27.00	91.00	66.477	14.466
Indirect costs	2011 – 2014	128	1.00	3.30	1.672	.375
Student-staff-ratio	2011 - 2014	128	2.972	13.478	8.903	2.611
Size student body	2011 - 2014	128	431	48220	13293.520	13944.976
1st year students 'Non-Western'	2011 - 2014	128	.003	.315	.106	.080
1st year students 'MBO'	2011 - 2014	128	.056	.522	.280	.087
Time	2011 - 2014	128	1	4	2.50	1.122

Chapter Four - Empirical results

The effect of the introduction of the performance agreements on student satisfaction

To test the hypothesis that the introduction of the Dutch performance agreements has increased student satisfaction, a single-factor repeated measures Analysis of Variance (ANOVA) was used with time as predictor (within-subjects factor) and student satisfaction as dependent variable. It was expected that student satisfaction did not increase significantly before introduction of the performance agreements, but only after. Figure 4 displays the trend in student satisfaction over the observed time span from AY 2011/12 to AY 2015/16.

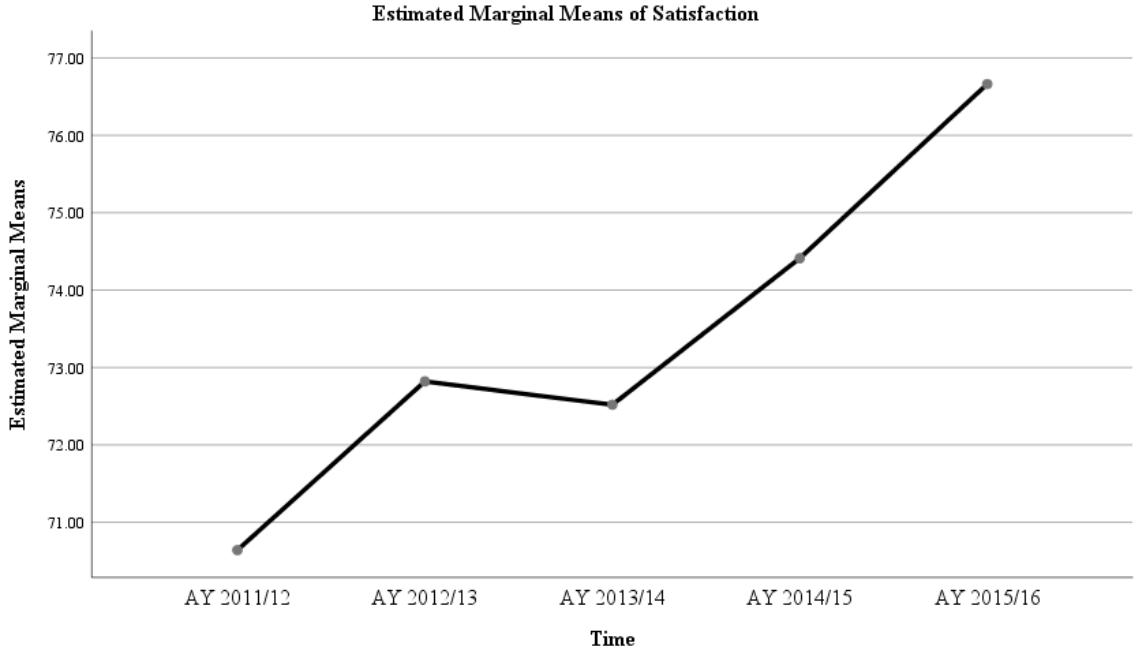

Figure 4. Mean Student satisfaction rates at Dutch universities of applied sciences from 2011 to 2015. N = 160.

The output has produced pairwise comparisons of the categories of the between-subjects factor, time. The pairwise comparisons were used to determine whether differences between the categories of the between-subjects factor, time, were statistically significant. The SPSS output can be found in Table 2. It was checked whether the contrast between the year of introduction (T2) and the year(s) before (T1) and after (T3,4,5) was significant. It was expected that the contrast of T1 (before introduction) and T2 (introduction) was not significant, whereas the contrasts T2 and T3, T2 and T4 and T2 and T5 should have been significant.

The graph shows that overall, satisfaction has increased over the years. When considering the pairwise comparisons to assess whether the contrasts between the years were significant, the following results show up: The contrast between T2 (AY 2012/2013) and T1 (AY 2011/2012) was statistically significant ($p = .027$). Hence, student satisfaction significantly increased (AY 2011/2012). A non-significant impact was found for the contrast between T2 (AY 2012/2013) and T3 (AY 2013/2014), p =

1.000 as well as between T2 and T4 (AY 2014/2015), p = .528. The contrast between T2 and T5 (AY 2015/2016) was found significant, p = .002. Hence, one can conclude that the performance agreements had a positive effect on the trend in student satisfaction, but only for the AY 2015/2016, and hence 3 years after the introduction of the policy. When interpreting the mean difference of the contrast, that is -3.842, it is useful to take the results of the descriptive statistics (Table 1) in the last chapter into account. The actual range of the variable 'Student satisfaction' was 55 (Min. Value) to 94 (Max Value). If using that scale, the results indicate only a pretty small increase in student satisfaction after the introduction of the performance agreements. Nonetheless, the results confirm hypothesis H1 that predicted that the introduction of the performance agreements has increased student satisfaction at Dutch universities of applied sciences. However, the results must be questioned critically, since time had a significant effect on student satisfaction already before the introduction of the performance agreements (contrast T2 and T1, p = .027). This issue will be discussed in the next chapter in the discussion. Given the slight increase in student satisfaction, the author has become interested in whether the trend differed across institutions. Not being directly part of this research, an extension of the analysis can be found in the appendix.

Table 2. Results of the repeated-measures ANOVA for the time trend of student satisfaction at Dutch universities of applied sciences in 2011 to 2015 (N = 160). Display of pairwise comparisons to show contrasts between groups (Differences in mean student satisfaction rates in universities of applied sciences at a certain time point x).

Pairwise Comparisons						
Measure: Satisfaction						
(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
2	1	2.180*	.668	.027	.160	4.200
	3	.302	.505	1.000	-1.223	1.826
	4	-1.592	.791	.528	-3.982	.798
	5	-3.842*	.904	.002	-6.574	-1.111

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

The effect of the introduction of the performance agreements on student retention

To test the hypothesis that the introduction of the performance agreements has increased student retention rates at Dutch universities of applied sciences, the same analysis was conducted as for the trend in student satisfaction. As in the previous analysis, a single-factor repeated measures Analysis of Variance (ANOVA) was used with time as predictor (within-subjects factor) and student retention as dependent variable. Figure 5 displays the trend of student retention over the observed time span from AY 2003/04 to AY 2014/15.

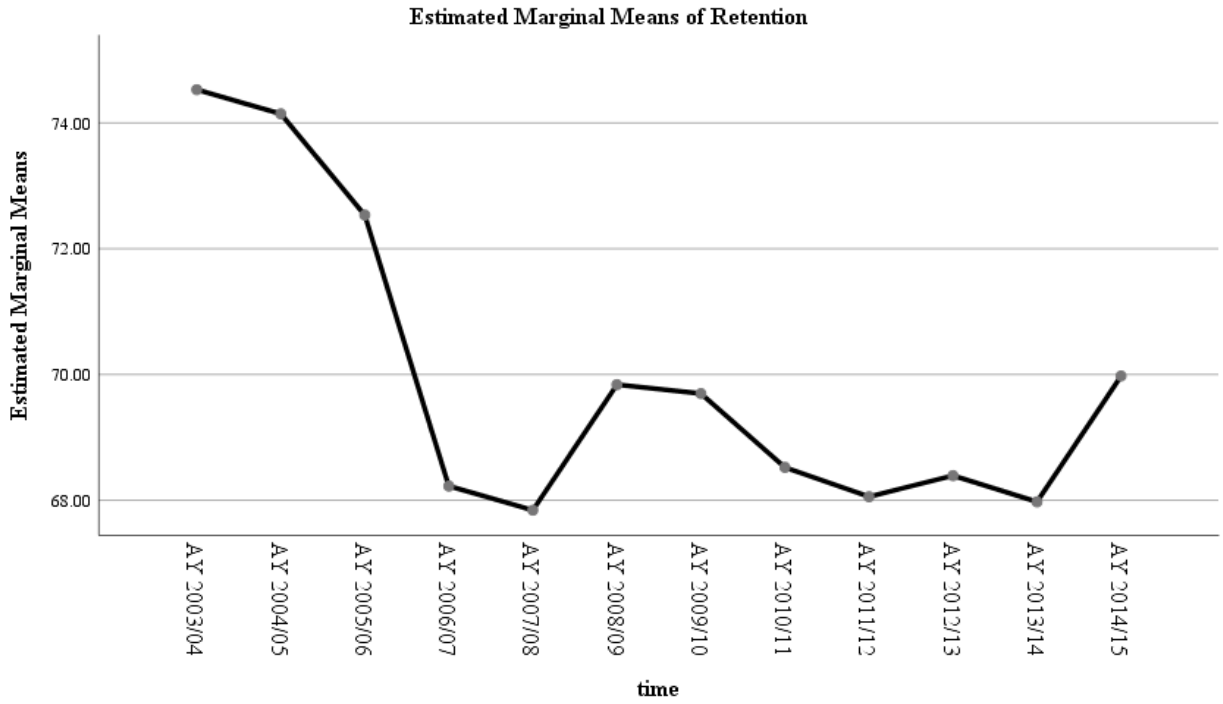

Figure 5. Mean Retention rates at Dutch universities of applied sciences from 2003 to 2014. N = 420.

The analysis produced pairwise comparisons of the categories of the between-subjects factor, time. The pairwise comparisons were used to determine whether differences between the categories of the between-subjects factor, time, were statistically significant. The SPSS output can be found in Table 3. As for the ANOVA for student satisfaction, it was expected that student retention did not increase before the introduction of the performance agreements, but only after. Hence, it was checked whether the contrasts between the year of introduction (10) and the years before and after were significant. It was expected that the contrasts between T10 and the years before are not significant, whereas the contrasts between T10 and the years after (T11 and T12) were assumed to be significant. The graph shows the trend of student retention rates over time. When considering the pairwise comparisons to assess whether the contrasts between the years were significant, the following results show up: The contrasts between T10 (introduction of the policy) and T1, T2 and T3 were significant with $p = .001$, $p = .001$ and $p = .029$, respectively. No significant effects could be found for the remaining contrasts between T10 and T4 to T12. The results indicate that the introduction of the performance agreements in

2012/13 (T10) had no significant effect on retention rates. In contrast, time had a significant effect on retention rates for T1(AY 2003/4) to T3(AY 2005/6), but not anymore after.

Table 3. Results of the repeated-measures ANOVA for the time trend of student retention at Dutch universities of applied sciences in 2003 to 2014 (N = 420). Display of pairwise comparisons to show contrasts between groups (Differences between mean student retention rates in universities of applied sciences at a certain time point x).

Pairwise Comparisons						
Measure:	Retention					
(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
10	1	-6.137*	1.187	.001	-10.529	-1.744
	2	-5.753*	1.147	.001	-9.998	-1.507
	3	-4.143*	1.065	.029	-8.083	-.203
	4	.166	1.063	1.000	-3.768	4.100
	5	.550	.820	1.000	-2.486	3.585
	6	-1.446	.654	1.000	-3.866	.975
	7	-1.304	.968	1.000	-4.886	2.278
	8	-.131	.688	1.000	-2.675	2.414
	9	.335	.757	1.000	-2.467	3.138
	11	.417	.605	1.000	-1.822	2.655
	12	-1.584	.657	1.000	-4.014	.846

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

The impact of the quality measures on student satisfaction

To analyze to what extent the quality measures 'Quality of teaching staff' and 'Indirect costs' included in the Dutch performance agreements have increased student satisfaction, a Linear Mixed Model was used. Before the main analysis, bivariate correlations were examined to take a preliminary look for multicollinearity (p) (Field, 2013). No very high (>.80/>.90) correlations were observed that might have indicated problems of multicollinearity (Field, 2013). The correlation matrix can be found in the appendix. To account for more subtle form of multicollinearity, the collinearity statistics were considered. No Variance Inflation Factor (VIF) was greater than 10. The average VIF was 1.552. Bowerman and O'Connell (1990) argue that the analysis might be biased if the average VIF is substantially greater than 1 (In Field, 2013). Hence, no risk of bias was assumed.

The output of the Linear Mixed Model analysis can be found in Tables 4 and 5. First, the random effects included in the model as random intercept and random slope will be reported (Table 4). The variance of random intercepts was Var(u_{0j}) = 83.896, p = .000. That implies that student satisfaction rates at baseline differed significantly between universities of applied sciences. Next, the change in student satisfaction over time differed significantly across universities of applied sciences, Var(u_{1j}) = 3.563, p = .001. Last, the significant covariance between the slopes and intercepts (-.753) implies that slopes decreased with an increase in intercepts (Field, 2013). The meaning of that will be explained in the following. The analysis results for the fixed effects (Table 5) found that student satisfaction has increased significantly over time with b = 1.003, t(2.373) and p = .022, provided that the effects of all other predictors are held constant. Per year, student satisfaction has increased by 1.003%. Against this background, the covariance between slopes and intercepts (-.753) implies that universities of applied sciences with a higher student satisfaction at baseline had weaker increases in student satisfaction over time than universities of applied sciences with lower scores on student satisfaction at baseline. The discovered increase in student satisfaction is consistent with the findings of the earlier ANOVA analysis for the trend in student satisfaction (Table 2). Likewise, the increase in student satisfaction by 1.003% per year (Table 5) is interpreted as rather small regarding the actual range of the variable 'Student satisfaction' that goes from 55 (Min. value) to 94 (Max. Value).

Table 4. Results of random parameters in the Linear Mixed Model analysis of the impact of quality measures on student satisfaction (H4).

Estimates of Covariance Parameters					
Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		3.939	.700	5.631	.000
Intercept + Time [subject = Hogescholen]	Var: Intercept	83.896	22.555	3.720	.000
	Var: Time	3.563	1.104	3.228	.001
	ARH1 rho	-.753	.089	-8.491	.000

Dependent variable: Satisfaction

Regarding the remaining fixed effects, the following results were found. Not all independent variables had a statistically significant relationship with student satisfaction. No significant relations were found for the quality measures 'Quality of teaching staff' and 'Indirect costs' (p = .625 and p = .216). The analysis showed that the predictor 'Size of student body' was associated with a significant decrease in student satisfaction rates with b = -.000261, t(-3.020), p = .005, provided that the effects of all other predictors were held constant. This means that an increase of the student body by one student was related to a decrease by .000261% in the proportion of full-time students who are satisfied (score 4) or very satisfied (score 5) with their program. For every 100 students, this was a decrease of .0261% and for every thousand students a decrease of .261%.

An objective estimation of the strength of the effect could not be made since SPSS does not produce estimates of effect sizes for Linear Mixed Models. Further calculations would have allowed to determine the estimate of the effect size, yet this was not possible within the given time. More information on the procedure of calculating the estimates of effect sizes can be found in Taylor (2014).

Last, no significant relation was found between the control variable 'Student-staff-ratio' and student satisfaction (p = .515).

Table 5. Results of fixed effects in the Linear Mixed Model analysis of the impact of quality measures on student satisfaction (H4).

Estimates of Fixed Effects[a]

	Unstandardized Coefficients		Standardized Coefficients			
Parameter	Estimate	Std. Error	Beta	df	t	Sig.
Intercept	70.827	4.613	.000	82.410	15.355	.000
Time	1.003	.423	.133	48.604	2.373	.022
Teacher quality	.028	.058	.048	108.629	.490	.625
Indirect costs	1.832	1.471	.081	106.707	1.245	.216
Size student body	-.000261	.000	-.431	36.949	-3.020	.005
Student-staff-ratio	-.267	.408	-.082	90.547	-.654	.515

a. Dependent variable: Satisfaction

The impact of the quality measures on student retention

In the same way as for the previous analysis, a Linear Mixed Model was used to analyze to what extent the quality measures 'Quality of teaching staff' and 'Indirect costs' in the performance agreements have increased student retention at Dutch universities of applied sciences. Before the main analysis, bivariate correlations were examined to take a preliminary look for multicollinearity (p) (Field, 2013). No very high (>.80/>.90) correlations were observed that might have indicated problems of multicollinearity (Field, 2013). The correlation matrix can be found in the appendix. To account for more subtle form of multicollinearity, the variance inflation factor (VIF) was considered (Field, 2013). No Variance Inflation Factor (VIF) was greater than 10. The average VIF was 1.81. Bowerman and O'Connell (1990) argue that the analysis might be biased if the average VIF is substantially greater than 1 (In Field, 2013). Hence, no risk of bias was assumed.

The output of the Linear Mixed Model analysis can be found in Tables 6 and 7. First, the random effects included in the model as random intercept and random slope will be reported (Table 6). The variance of random intercepts was Var(u_{0j}) = 21.072, p = .017. Hence, student retention rates at baseline differed significantly between universities of applied sciences. In contrast, the developments in student retention rates over time did not differ significantly between universities of applied sciences, Var(u_{1j}) = .025 p = .767.

Table 6. Results of random parameters in the Linear Mixed Model analysis for the impact of quality measures on student retention (H4).

Estimates of Covariance Parameters[a]					
Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		8.274	1.247	6.638	.000
Intercept + Time [subject =	Var: Intercept	21.072	8.838	2.384	.017
Hogescholen]	Var: Time	.025	.083	.296	.767
	ARH1 rho	-1.000[b]	.000	.	.

a. Dependent Variable: Retention.
b. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

Regarding the fixed effects included in the analyses, the following results occurred. The output of the Linear Mixed Model analysis in Table 7 showed that student retention rates have not significantly increased over time (p = .331). This finding is consistent with the results of the earlier ANOVA analysis for the trend in student retention (Table 3).

However, the results demonstrated that student satisfaction scores were positively related to student retention rates, b = .226, t(2.890) and p = .005, provided that all other predictors were held constant. One percentage point increase in the proportion of respondents (full-time students) that is satisfied (score 4) or very satisfied (score 5) with their program was associated with an increase in retention rates by .226%.

The results showed no significant relationship between performance on the quality measures 'Quality of teaching staff' and 'Indirect costs' and student retention rates (p = .526 and p = .279).

Regarding the control variables, the analysis found a negative relation between the student-staff-ratio and student retention rates, b = -1.050, t(-2.873) and p = .005, assuming that the effects of all other predictors were held constant. One unit increase in the student-staff-ratio, that is every additional student per teacher, was associated with a decrease in student retention rates by 1.05%.

Furthermore, the results indicated a significant, negative relation between the predictor 'Size of the student body' and student retention, b = -.00021, t(-2.804) and p = .006, assuming that the effects of all other predictors were held constant. This means that an increase in the student body by one student is associated with a decrease in student retention rates of .00021%. For every 100 students this is a decrease of .021% and for every thousand students a decrease of .21%.

Last, a significant relation was found between the predictor 'Share of first-year 'MBO' students' and student retention with b = -22.355, t(-2.775) and p = .006. This is only true if the effects of all other predictors were held constant. The results imply that a 1% increase in first-year 'MBO' students was associated with a decrease in student retention rates by .224%. Last, no significant relation was found between the predictor 'Share of first-year, non-Western students' and student retention rates (p = .624).

As in the case of the previous analysis on student satisfaction, an objective estimation of the strengths of the effects could not be made since SPSS does not produce estimates of effect sizes for Linear Mixed Models. It is again referred to Taylor (2014) for further instructions on how to calculate the estimates of effect sizes in Linear Mixed Models.

To provide at least some comparison of the strengths of the relationships between the predictors and student retention, the standardized beta coefficients were analyzed. Compared with the other predictors, the size of the student body was strongest related to student retention rates (Beta = -.342), followed by the predictor 'Student-staff-ratio' (Beta = -.320). The relationship between the 'Share of first-year 'MBO' students' and student retention rates was relatively weaker with Beta = -.226. Almost the same size of standardized coefficient was found for the predictor student satisfaction with Beta = .224.

Table 7. Results of fixed effects in the Linear Mixed Model for the impact of student satisfaction on student retention (H3) and the impact of quality measures on student retention (H4).

Estimates of Fixed effects[a]						
	Unstandardized Coefficients		Standardized Coefficients			
Parameter	B	Std. Error	Beta	df	t	Sig.
Intercept	68.770	7.285	.000	119	9.440	.000
Time	.310	.318	.041	119	.977	.331
Satisfaction	.226	.078	.224	119	2.890	.005
Teacher quality	.035	.055	.059	119	.636	.526

Indirect costs	-1.697	1.559	-.074	119	-1.088	.279
Student-staff-ratio	-1.050	.365	-.320	119	-2.873	.005
Size student body	-.00021	.000	-.342	119	-2.804	.006
1st year students 'Non-Western'	6.000	12.216	.056	119	.491	.624
1st year students 'MBO'	-22.355	8.055	-.226	119	-2.775	.006

a. Dependent variable: Retention

Conceptual model revisited

To visualize the results from the Linear Mixed Model analyses, Figure 6 shows the revised conceptual model that was established at an earlier stage. The original conceptual model (Figure 3) displayed the hypothesized impacts of student satisfaction on student retention (H3), the impact of the quality measures 'Teacher quality' and 'Indirect costs' on student satisfaction and retention (H4), as well as the control variables included. The revised conceptual model indicates only those relationships that were found to be significant in the analyses ($p \leq .05$). Attached to the arrows are the standardized beta coefficients (β) that allow for a comparison of the strengths of the relationships.

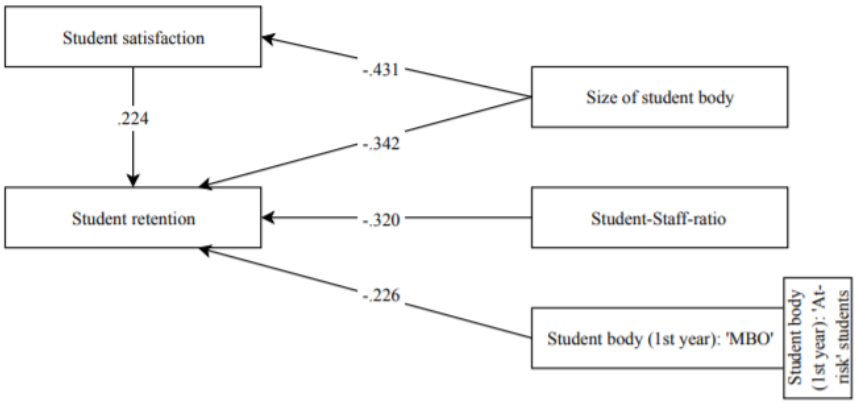

Figure 6. Conceptual model (Figure 3) revisited. Outcome of the analyses of the impact of student satisfaction on student retention (H3) and the impact of the quality measures on student satisfaction and student retention (H4).

Chapter 5 – Conclusion and Discussion

Conclusion

Research question 1 asked to what extent the introduction of the performance agreements has improved student satisfaction at Dutch universities of applied sciences. It was expected that student satisfaction rates have increased after the introduction of the performance agreements (H1). The analysis showed that student satisfaction has increased significantly 3 years after the introduction of the performance agreements. Consequently, hypothesis 1 was accepted.

Research question 2 asked to what extent the introduction of the performance agreements has increased student retention rates at Dutch universities of applied sciences. It was expected that student retention rates have increased after the introduction of the policy. The research showed that there were no significant changes in retention rates after the introduction of the agreements. Hence, the answer to the research question is that student retention rates have not increased after the introduction of the performance agreements.

Research question 3 asked to what extent student satisfaction has increased student retention rates at Dutch universities of applied sciences. It was expected that higher scores on student satisfaction would lead to higher student retention rates at Dutch universities of applied sciences. The analysis found that student satisfaction scores were associated with a significant increase in student retention rates. Hence, the answer to the research question is that higher student satisfaction was related to higher scores on student retention at Dutch universities of applied sciences.

Yet, no definite statement could be made on the individual strength of the effect. In comparison with the other predictors for student retention, student satisfaction was not the predictor with the strongest effect. Instead, the predictor ‘Size of the student body’ was related strongest with scores on student retention. It was found that universities of applied sciences with larger numbers of enrolled students scored relatively lower on student satisfaction and student retention rates compared to universities of applied sciences with smaller numbers of enrolled students. Furthermore, the student-staff-ratio at Dutch universities of applied sciences was stronger related to student retention than were scores on student satisfaction. Institutions with relatively higher student-staff ratios were related to relatively lower scores on student retention. The relation between the share of first-year ‘MBO’ students and student retention had an equal strength as the relation to student satisfaction. Universities of applied sciences with relatively larger shares of first-year ‘MBO’ students were associated with relatively lower student retention rates.

Research question 4 asked to what extent the quality measures included in the performance agreements have increased student satisfaction and retention rates at Dutch universities of applied sciences. It was expected that the lower the indirect costs and the higher the quality of teaching staff at Dutch universities of applied sciences, the higher their levels of student satisfaction and retention rates. The analysis found that the quality measures ‘Quality of teaching staff’ and ‘Indirect costs’ were not

significantly related to scores on student satisfaction and retention rates at Dutch universities of applied sciences.

Reflection on theory and hypotheses

In line with the causal logic of performance funding policies (Rabovsky, 2012), the slight increase in student satisfaction might suggest that the introduction of the performance agreements has restructured incentives of universities of applied sciences. Consequent administrative responses might have eventually resulted in the outcomes observed. It is known that the majority of universities of applied sciences stated to continue or implement interventions to reduce drop-out and switch rates (Reviewcommissie & Onderwijs, 2014). Furthermore, the Dutch performance agreements explicitly included indicators on measures that were assumed to increase student outcomes (OCW & HBO-raad, 2011). Obviously, actions taken by universities of applied sciences were effective in increasing student satisfaction.

Nonetheless, the results must be questioned critically since student satisfaction had increased significantly already before the introduction of the performance agreements. Consequently, rises in student satisfaction might be not solely a result of the introduction of the performance agreements, but due to additional causes such as earlier policies in universities of applied sciences that aimed to increase student outcomes (De Boer & Van Vught, n.d.). Furthermore, nation-wide economic conditions or political situations might have influenced the trend in student satisfaction. In that regard, it is acknowledged that the analysis did not incorporate any control variables since panel data were not available for all measurement moments. Furthermore, limitations in the trend analysis are acknowledged since the trend is only observed for the period AY 2011/2012 to 2015/2016. Given that the policy was introduced in the AY 2012/2013, this gives merely one pre-intervention measurement moment (AY 2011/2012), and three post-intervention measurement moments (AY 2013/2014 to AY 2015/2016). Hence, limitations of the findings must be acknowledged since long-term trends of student satisfaction after the introduction of the policy could not be considered. Furthermore, it was only considered in the analysis whether the intervention (performance agreements) was present or not, but not the extent to which it was implemented.

The finding that retention rates have not increased after the introduction of the performance agreements is in line with the majority of previous empirical evidence on the impact of performance-based funding policies on graduation rates (Hillman et al., 2014; Sanford & Hunter, 2011; Shin, 2010; Umbricht et al., 2017). Nonetheless, it seems surprising that student satisfaction has increased after the introduction of the performance agreements whereas student retention rates have not, given that the study found a positive relationship between satisfaction scores and retention rates at Dutch universities of applied sciences.

Since there is evidence that universities of applied sciences have taken actions after the introduction of the performance agreements (OCW & HBO-raad, 2011; Reviewcommissie &

Onderwijs, 2014), it seems unlikely that financial incentives for implementing changes were not large enough or that threats to the principal-agent relationship, such as an agent's own interest, information asymmetries, distrust or capacity constraints, have impeded universities of applied sciences from implementing the principal's agenda, i.e. the performance agreements (Hillman et al., 2014; Rabovsky, 2012; Shin, 2010).

Instead, one can argue that the trend of retention rates was only analyzed for the period AY 2003/2004 to AY 2014/2015. This gives ample pre-intervention measurement points (AY 2003/2004 to AY 2011/2012), but only limited post-intervention data (AY 2013/2014 and AY 2014/2015), so that long-term trends of student retention rates after the introduction of the policy could not be considered. Since the increase in student satisfaction became only significant for the academic year 2015/2016, which was not considered anymore in the trend analysis of student retention, it might be that student retention would have shown similar results, if data had been available.

Yet, one might also argue that changes in the organization of the educational process at Dutch universities of applied sciences in response to the introduction of the performance agreements were possibly effective in increasing student satisfaction, whereas obviously not effective in increasing retention rates. This might be due to additional predictors that were associated with stronger effects on student retention rates than student satisfaction. Stronger relations were found for the control variables 'Student-staff-ratio' and 'Size of the student body'. The, by comparison, weaker relationship between student satisfaction and student retention might be explained on basis of theory that suggested an indirect effect of student satisfaction on student retention via the mediator 'Institutional commitment' (Schertzer & Schertzer, 2004). Since no data were available, this variable was not included in the analysis. Hence, one might conclude that actions taken by universities of applied sciences were effective in increasing student satisfaction, but in the end, other factors had bigger impacts on student retention rates.

Given that the performance agreements explicitly included the indicators 'Teacher quality' and 'Indirect costs' as measures to improve student outcomes, it was surprising that both predictors were not significantly related to student satisfaction and retention rates at Dutch universities of applied sciences. One might conclude that the measures were ineffective as tools to increase quality of education which points to weaknesses in the program design of the Dutch performance agreements (Hillman et al., 2014). At first glance, it seems contradictory that teacher quality was not significantly related to student satisfaction and retention rates given previous research findings that highlighted the importance of professor quality, academic service quality and teaching service quality perceptions of students for student satisfaction (de Oliveria Santini et al., 2017). These unexpected results might be a hint for an inappropriate conceptualization of 'Quality of teaching staff' in the performance agreements. The variable 'Quality of teaching staff' was conceptualized as the share of teaching staff with a master's or PhD (regardless of the nature of their employment) in the total number of teaching staff (Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c). Quality of education was intended to be improved by increasing the share of teachers with analytical and research competencies

(Reviewcommissie Hoger Onderwijs en Onderzoek, 2017). Yet, these initiatives were highly criticized by the biggest national student organization, 'Interstedelijk Studenten Overleg' (ISO), that argued that teacher quality is not only predicted by the educational attainment of a teacher (Interstedelijk Studenten Overleg, 2016). The perception of teacher quality was found to be the second most important indicator for the satisfaction of students with their studies (van den Broek, Wartenbergh, Bendig-Jacobs, Braam, & Nooij, 2016). This might explain why the indicator 'Teacher quality' showed no significant effect on student satisfaction and retention rates. Regarding the results for the predictor 'Indirect costs', the Review Committee has acknowledged in an own evaluation operationalization issues due to diffuse mechanisms of assigning personnel either to teaching and research staff or support and management personnel. In the current indicator, the primary process does not only refer to teaching, but also to research and valorization, which might have distorted results (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017).

It is furthermore surprising that higher student-staff-ratios were related to relatively lower retention rates, whereas no significant relationship was found for student satisfaction. This finding is not in line with theory that assumes that personal relationships with faculty and/or staff and the extent of interaction between student and faculty predict student satisfaction – environments that are assumed to occur more likely in institutions with lower student-staff-ratios (Bradley et al., 2008; Browne et al., 1998; McDonald, 2013; Schertzer & Schertzer, 2004). Yet, operationalization issues might again be an explanation for the findings. The variable 'Student-staff-ratio' was operationalized as the number of enrolled students per employee on full-time or another hours-base. The operationalization was chosen since universities of applied sciences employ relatively high numbers of professionals with minor teaching assignments next to their professional careers. It is questionable then, whether minor teaching assignments can ensure the extent of interaction and personal relationships that is desired by many students (Browne et al., 1998; Schertzer & Schertzer, 2004). Therefore, a high number of teachers with minor teaching assignments might satisfy the demands of teaching, yet the extent of interaction that satisfies students might be not achieved. The number of full-time employees may be in many cases a better indicator for the extent of interaction with students due to fixed office hours. Yet, it may be not very predictive for the case of Dutch universities of applied sciences due to the reasons mentioned above.

In line with previous research that found lacking academic preparation of students as an impediment to institutional performance of colleges (Dougherty et al., 2016), the research has shown that universities of applied sciences with larger shares of 'MBO' students scored relatively lower on student retention. Other than theory suggested (Zijlstra et al., 2013), the share of non-Western, first-year students was not significantly related to student retention rates. Yet, these interpretations demand great caution since findings might be distorted through the problem of 'fallacy at the wrong level' (Dansereau, Cho, & Yammarino, 2006). Group-level effects might be wrongly attributed to individuals.

In the end, the research findings are too limited in their effect size (Student satisfaction) and significance (Student retention) to make a definite statement against or in favor of the effectiveness of

the Dutch performance agreements in increasing quality of education. Furthermore, it seems quite bold to make a final statement on the reasons for a rise in student satisfaction, since only few information was available on the actions taken by universities of applied sciences after the introduction of the performance agreements. Yet, what can be said is that there were obviously weaknesses in the design of the performance agreements. Performance on the quality measures included in the agreements was not significantly related to student satisfaction and retention rates. In the following, the overall strengths and weaknesses of the research will be discussed, followed by suggestions for further research and policy recommendations.

Strengths and weaknesses

The research has contributed to the body of empirical knowledge on performance-related pay by testing the results of previous research (Hillman et al., 2014; Sanford & Hunter, 2011; Shin, 2010; Umbricht et al., 2017). The research can be regarded as a pioneer study because of several reasons. Against the background of a major research focus on performance-based funding, the research extended empirical evidence on the impacts of performance agreements as a less common form of performance-related pay. Second, it is not known to the author that the impact of the Dutch performance agreements on quality of education has been researched in any way before this study. Third, the research analyzed the impact of the introduction of the performance agreements on the satisfaction of consumers of higher education, that is students. This is considered a particularity since previous research has consistently analyzed student outcomes in terms of graduation or retention rates (Hillman et al., 2014; Li, 2018; Sanford & Hunter, 2011; Shin, 2010).

Despite its strengths, the author is also aware of several weaknesses of the research. First, the research considered only one small part of the changes in the organization of educational processes after the introduction of the performance agreements. These were two of the three quality measures included in the performance agreements (OCW & HBO-raad, 2011; Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c). What was not considered in the research were the interventions that universities of applied sciences stated to continue or implement after the introduction of the agreements (and the extent of their implementation). This also limits the evidence to eventually explain why satisfaction has increased 3 years after the introduction of the policy, whereas student retention has not. Consequently, the research has to acknowledge that the underlying mechanisms of the Dutch performance agreements remain, metaphorically spoken, largely hidden in a ‘Black Box’ (Dougherty et al., 2016).

Second, the research could not observe long-term trends in student satisfaction and retention since only few post-intervention (and partly pre-intervention) data were available. Since the policy had been introduced already in 2012 and ended in 2015, the author had to rely on pre-existing datasets for the research. Consequently, the research could only use those data that were already available so that the trends in student satisfaction and retention could not be observed for longer periods.

Third, all variables included in the analysis were measured or aggregated at the institutional level. Hence, the research did not consider variations within institutions themselves. The author is aware that graduation rates, scores on student satisfaction, student-staff-ratios, the share of 'at-risk' students or the share of teachers with a master's or PhD-degree can widely differ across different study programs or faculties within an institution. Furthermore, the author is aware of the risk of 'fallacy of the wrong level' (Dansereau et al., 2006). Great caution is demanded to not wrongly attribute group-level effects to the individual level. For example, an increase in the share of 'MBO' students was related to a decrease in student retention rates. Yet, from that result, one cannot conclude that 'MBO' students were more likely to switch or drop out of their studies. Therefore, no definite statements could be made on the causality between the predictors and scores on student satisfaction and student retention rates. Instead, it was merely speculated on the reasons for the identified relationships between the variables.

Recommendations for further research

The previously discussed weaknesses of the study point to the need for further research. Whereas a major research interest has focused on the direct outcomes of performance policies, future research should attempt to 'bring light into the dark' of the underlying mechanisms of performance agreements to give more evidence-based explanations for the observed outcomes. Detailed documentation is required of the actions taken by higher education institutions after the introduction of performance agreements. For example, what type of institution has adopted what type of interventions? To what extent differs the implementation of interventions across institutions?

At this point, the author wants to refer to Meredith I. Honig (2006) who wrote an inspiring chapter on the complexities of policy implementation in education nowadays. New approaches to implementation research aim to discover how and why interactions among policy, people, and places shape implementation (Honig, 2006). It is aimed to find out the conditions under which education policies get implemented successfully (Honig, 2006).

Future research on the impact of performance agreements or performance-based funding may certainly benefit from these insights. A small attempt was made as an extension of this research to investigate the conditions under which the increase in student satisfaction at Dutch universities of applied sciences occurred. The extended analysis can be found in the appendix. This attempt may inspire future researchers to extend their own analyses on the effectiveness of performance-related-pay. In sum, future researchers must gain a deeper understanding of the underlying mechanisms of performance agreements and performance-based funding. Eventually, it is important to find out under which circumstances performance agreements are effective.

Hereby, future research might also benefit from more qualitative research designs to unveil more nuanced, and 'soft' mechanisms that play a role in the implementation of a certain policy. For example, to what extent are higher education institutions more willing to incorporate changes in the organization of educational processes when they are given autonomy in their decisions? This research question might

be addressed in the context of the second round of the 'Quality agreements' in Dutch higher education after 2018 in which more autonomy shall be granted to higher education institutions (Interstedelijk Studenten Overleg, Landelijke studentenvakbond, Ministerie van Onderwijs Cultuur en Wetenschap, Vereniging Hogescholen, & Vereniging van universiteiten, 2018).

It is furthermore strongly recommended to collect longitudinal data over long pre- and post-intervention periods, since only this information allows researchers to make explicit statements on the long-term efficiency of an intervention. Also, future research is encouraged to extend the body of empirical knowledge on the impact of performance policies on student satisfaction, next to common outcome indicators such as retention or graduation rates.

Against the background of a largely US-dominated debate on the effectiveness of performance-based funding (Dougherty et al., 2016; Hillman et al., 2014; Kelchen & Stedrak, 2016; Li, 2018; Rabovsky, 2012; Shin, 2010; Umbricht et al., 2017), future research might add a new perspective by extending research on performance agreements in different contexts. Settings for those research projects might be the higher education systems of Australia, Austria, Denmark, Finland, North-Rhine-Westphalia, Thuringia, Hong-Kong, Ireland or Scotland, where performance agreements are currently implemented (De Boer et al., 2015).

Policy recommendations

Since the findings of this research are too limited to make a definite statement on the effectiveness of the performance agreements in increasing quality of education at Dutch universities of applied sciences, the author will address instead several policy recommendations on how the latest design of the performance agreements can be improved in the future. The recommendations are of high actuality since new 'Quality agreements' ('Kwaliteitsafspraken') for the period 2019 to 2024 are currently being arranged (Interstedelijk Studenten Overleg et al., 2018). As such, the policy recommendations can offer valuable inputs to the Ministry of Education, Culture and Science, The Netherlands Association of Universities of Applied Sciences, the Association of Universities in the Netherlands, the Dutch Student Union and the Intercity Student Consultation[7] on how to shape certain features in the design of the future 'Quality agreements' (Interstedelijk Studenten Overleg et al., 2018).

The research findings underpin a certain relevance of student satisfaction for universities of applied sciences in retaining students at their institution. In the past, higher education institutions were accused of displaying strategic behavior to increase graduate numbers and enrolments for maximizing budgets, which went at the expense of students' impressions of the quality of education (De Boer & Van Vught, n.d.; Reviewcommissie Hoger Onderwijs en Onderzoek, 2017). This highlights the relevance of holding Dutch universities of applied sciences accountable for satisfying the needs of students.

[7] The policy recommendations do not address the Higher Education and Research Review Committee that was abolished in 2016. Monitoring of the Quality agreements will be done by the official accreditation agency NVAO (Interstedelijk Studenten Overleg et al., 2018)

Therefore, it is recommended to the contracting parties to include indicators with stakeholder judgments also in the future 'Quality agreements'.

Since the research found no significant relationship between the quality measures included in the performance agreements and student satisfaction and retention rates, it is recommended to the contracting parties to reconsider the inclusion of these indicators in future performance agreements. In terms of the indicator 'Indirect costs', this recommendation matches own evaluations by the Review Committee itself who acknowledged operationalization issues of the indicator (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017). Regarding the indicator 'Teacher quality', it is undisputed that students should acquire analytical and research competencies during their studies (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017). Yet, there might be better suited indicators for 'Teacher quality' to include in future performance agreements. In the case of Dutch research universities, teacher quality was measured in terms of the share of teachers with a 'University Teaching Qualification' (In Dutch: BKO) out of the total number of teachers. It is an institution-overarching qualification for didactic competencies of teaching staff (VSNU, n.d.) that has become increasingly implemented at Dutch research universities (Reviewcommissie Hoger Onderwijs en Onderzoek, 2017). In the case of Dutch universities of applied sciences, the Netherlands Association of Universities of Applied Sciences agreed in 2013 to implement a similar 'Basic Qualification of Didactical Competence' (in Dutch: 'Basiskwalificatie Didactische Bekwaamheid; 'BDB') at Dutch universities of applied sciences (Zestor, n.d.). Since critics have highlighted that teacher quality is not only ensured by the educational background of a teacher (Interstedelijk Studenten Overleg, 2016), policy-makers might consider including the share of teachers with a 'BDB' qualification as a better indicator for teacher quality in future performance agreements.

Given the negative relationship between student-staff-ratios and student retention rates, policy efforts should aim at decreasing student-staff ratios at Dutch universities of applied sciences. In the performance agreements from 2012 to 2015, the student-staff-ratio (teaching staff) was included as an optional indicator that universities of applied sciences could choose to include in the agreements (Reviewcommissie Hoger Onderwijs en Onderzoek, n.d.-c). Given its effect on student retention rates, policy makers should consider including the ratio as a prescribed component in the future 'Quality agreements'.

Regarding the findings that larger shares of 'MBO' students were associated with lower retention rates, the contracting parties of the future 'Quality agreements' might consider adjusted budget allocations for higher education institutions. As Honig (2006) suggested, not every policy may work in every setting and implementation also depends on the people. Hence, the budget allocation might not suit every higher education institution due to different student compositions. Universities of applied sciences with higher shares of 'MBO' students might require proportionally more financial resources to implement adequate support interventions for these students. This recommendation is given as a new impulse for discussion rather than as a concrete recommendation. This is because the study could not

draw causal relationships between student characteristics and outcomes. Nonetheless, it can give an interesting input for the discussion on how to customize the design of the ‘Quality agreements’ to the individual needs of institutions.

Finally, the author awaits with eagerness further developments in the new round of the Dutch ‘Quality agreements’ starting 2019 to 2024. With certainty, the next round of quality agreements will offer valuable, new insights into the mechanisms of the policy tool ‘Performance agreements’. The author hopes that the suggestions made for further research and policy recommendations will find at least some attention on the path until then.

References

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, *46*(1), 348–355. https://doi.org/10.1093/ije/dyw098

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed.). Belmont, CA: Duxbury.

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian Higher Education. Final Report.*

Browne, B. A., Kaldenberg, D. O., Browne, W. G., & Brown, D. J. (1998). Student as Customer: Factors Affecting Satisfaction and Assessments of Institutional Quality. *Journal of Marketing for Higher Education*, *8*(3), 1–14. https://doi.org/10.1300/J050v08n03_01

Crawford, L. M. (1999). Extended Opportunity Programs and Services For Community College Retention. In *The Eighth Annual California Community Colleges Chancellor's Office Statewide Conference* (pp. 1–32). Monterey, CA. Retrieved from https://files.eric.ed.gov/fulltext/ED429642.pdf

Crosby, B. C., Bryson, J. M., & Bloomberg, L. (2014). Public Value Governance: Moving Beyond Traditional Public Administration and the New Public Management. *Public Administration Review*, *74*(4), 445–456. https://doi.org/10.1111/puar.12238

Dansereau, F., Cho, J., & Yammarino, F. J. (2006). Avoiding the "Fallacy of the Wrong Level." *Group & Organization Management*, *31*(5), 536–577. https://doi.org/10.1177/1059601106291131

De Boer, H., Jongbloed, B., Benneworth, P., Cremonini, L., Kolster, R., Kottmann, A., … Vossensteyn, H. (2015). *Performance-based funding and performance agreements in fourteen higher education systems: Report for the Ministry of Education, Culture and Science*. Retrieved from www.utwente.nl/cheps

De Boer, H., & Van Vught, F. (n.d.). Higher Education Governance in the Netherlands: from a Janus-head to a Trimurti. Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/5596904

de Lourdes Machado, M., Brites, R., Magalhães, A., & Sá, M. J. (2011). Satisfaction with higher education: Critical data for student development. *European Journal of Education*, *46*(3), 415–432. https://doi.org/10.1111/j.1465-3435.2011.01489.x

de Oliveria Santini, F., Ladeira, W. J., Hoffmann Sampaio, C., & da Silva Costa, G. (2017). Student satisfaction in higher education: a meta-analytic study. *Journal of Marketing for Higher Education*, *27*(1), 1–18. https://doi.org/10.1080/08841241.2017.1311980

Dougherty, K. J., Sosanya, M. J., Lahr, H., Natow, R. S., Pheatt, L., & Reddy, V. (2016). Looking Inside the Black Box of Performance Funding for Higher Education: Policy Instruments, Organizational Obstacles, and Intended and Unintended Impacts. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(1), 147–173. Retrieved from https://www.rsfjournal.org/doi/pdf/10.7758/RSF.2016.2.1.07

Elliott, K. M. (2002). Key determinants of student satisfaction. *Journal of College Student Retention*, *4*(3), 271–279. Retrieved from http://journals.sagepub.com.ezproxy2.utwente.nl/doi/pdf/10.2190/B2V7-R91M-6WXR-KCCR

Elliott, K. M., & Healy, M. A. (2001). Key factors influencing student satisfaction related to recruitment and retention. *Journal of Marketing for Higher Education*, *10*(4), 1–11. Retrieved from https://www-tandfonline-com.ezproxy2.utwente.nl/doi/pdf/10.1300/J050v10n04_01?needAccess=true

Elliott, K. M., & Shin, D. (2002). Student Satisfaction: an alternative approach to assessing this important concept. *Journal of Higher Education Policy and Management*, *24*(2). https://doi.org/10.1080/1360080022000013518

Field, A. (2012). Repeated Measures ANOVA. Issues with Repeated Measures Designs. Retrieved from http://health.uottawa.ca/biomech/courses/apa6101/Repeated Measures ANOVA.pdf

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. (M. Carmichael, Ed.) (4th ed.). SAGE Publications.

Forrester, G. (2011). Performance management in education: milestone or millstone? *Management in Education*, *25*(1), 5–9. https://doi.org/10.1177/0892020610383902

Halachmi, A. (2005). Performance measurement: test the water before you dive in. *International Review of Administrative Sciences*, *71*(2), 255–266. https://doi.org/10.1177/0020852305053884

Harvey, L., & Green, D. (1993). Defining Quality. *Assessment & Evaluation in Higher Education*, *18*(1), 9–34.

Heinrich, C. J. (2002). Outcome-Based Performance Management in the Public Sector : Implications for Government Accountability and Efectiveness. *Public Administration Review*, *62*(6), 712–725. https://doi.org/10.1111/1540-6210.00253

Higher Education and Research Review Committee. (2012). *Work Method and Assessment Framework*. Retrieved from http://www.rcho.nl/images/NL PerformAgr Assessment Framework 2012.pdf

Hillman, N. W., Tandberg, D. A., & Gross, J. P. K. (2014). Performance Funding in Higher Education: Do Financial Incentives Impact College Completions? *The Journal of Higher Education*, *85*(6), 826–857. https://doi.org/10.1080/00221546.2014.11777349

Hladchenko, M. (2014). Performance agreements in Dutch higher education. *Euromentor Journal*, *5*(3), 34–47. Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/18868866

Honig, M. I. (2006). New Directions in Education Policy Implementation: Confronting Complexity. In M. I. Honig (Ed.), *New Directions in Education Policy Implementation: Confronting Complexity* (pp. 1–23). Albany, New York: State University of New York Press. Retrieved from https://www.sunypress.edu/pdf/61303.pdf

Interstedelijk Studenten Overleg. (2016). *Goed onderwijs begint bij de docent*. Retrieved from https://www.iso.nl/website/wp-content/uploads/2016/06/Docentkwaliteit-Goed-onderwijs-begint-bij-de-docent.pdf

Interstedelijk Studenten Overleg, Landelijke studentenvakbond, Ministerie van Onderwijs Cultuur en Wetenschap, Vereniging Hogescholen, & Vereniging van universiteiten. (2018). *Investeren in Onderwijskwaliteit. Kwaliteitsafspraken 2019-2024*.

Kelchen, R. (2018). Do performance-based funding policies affect underrepresented student enrollment? *The Journal of Higher Education*, *3*, 1–26. https://doi.org/10.1080/00221546.2018.1434282

Kelchen, R., & Stedrak, L. J. (2016). Does Performance-Based Funding Affect Colleges' Financial Priorities? *Journal of Education Finance*, *41*(3), 302–321. https://doi.org/10.1353/jef.2016.0006

Kickert, W. (1995). Steering at a Distance: A New Paradigm of Public Governance in Dutch Higher Education. *Governance*, *8*(1), 135–157. https://doi.org/10.1111/j.1468-0491.1995.tb00202.x

Kotler, P., & Fox, K. F. A. (1995). *Strategic marketing for educational institutions* (2nd ed.). Englewood Cliffs; Upper Saddle River, N.J.: Prentice-Hall.

Li, A. Y. (2018). Performance Funding Policy Impacts on STEM Degree Attainment. *Educational Policy*, *00*(0), 1–38. https://doi.org/10.1177/0895904818755455

Linden, A. (2015). Conducting interrupted time series analysis for single and multiple group comparisons, *15*(2), 480–500.

Lindsay, A. (1992). Concepts of Quality in Higher Education. *Journal of Tertiary Education Administration*, *14*(2), 153–163. https://doi.org/10.1080/1036970920140203

Maassen, P. A. M., & Van Vught, F. A. (1988). An Intriguing Janus-Head: The Two Faces of the New Governmental Strategy for Higher Education in the Netherlands. *Source: European Journal of Education European Jfournal of Education*, *23*(1/2), 65–76. Retrieved from http://www.jstor.org/stable/1502965

Mark, E. (2013). Student satisfaction and the customer focus in higher education. *Journal of Higher Education Policy and Management*, *35*(1), 2–10. Retrieved from https://www-tandfonline-com.ezproxy2.utwente.nl/doi/pdf/10.1080/1360080X.2012.727703?needAccess=true

McDonald, G. (2013). Does size matter? The impact of student-staff ratios. *Journal of Higher Education Policy and Management*, *35*(6), 652–667. https://doi.org/10.1080/1360080X.2013.844668

McLellan, D. L., Cabán-Martinez, A. J., Nelson, C. C., Pronk, N. P., Katz, J. N., Allen, J. D., … Sorensen, G. (2015). Organizational characteristics influence implementation of worksite health protection and promotion programs: Evidence from smaller businesses HHS Public Access. *Journal of Occupational & Environmental Medicine*, *57*(9), 1009–1016. https://doi.org/10.1097/JOM.0000000000000517

OCED. (2014). *OECD Reviews of Innovation Policy: Netherlands 2014*. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264213159-en

OCW. (2011). *Quality in diversity: Strategic Agenda for Higher Education, Research and Science*. Den Haag.

OCW, & HBO-raad. (2011). *Hoofdlijnenakkoord OSW - HBO-RAAD*.

Rabovsky, T. M. (2012). Accountability in Higher Education: Exploring Impacts on State Budgets and

Institutional Spending Patterns. *Journal of Public Administration Research and Theory*, *22*(4), 675–700. https://doi.org/10.1093/jopart/mur069

Reviewcommissie, A., & Onderwijs, H. (2014). Interventies uitval en rendement, (november), 1–25.

Reviewcommissie Hoger Onderwijs en Onderzoek. (n.d.-a). Eindadviezen hogescholen. Retrieved June 30, 2018, from http://www.rcho.nl/eindadviezen/eindadviezen-hogescholen

Reviewcommissie Hoger Onderwijs en Onderzoek. (n.d.-b). Eindadviezen universiteiten. Retrieved June 30, 2018, from http://www.rcho.nl/eindadviezen/eindadviezen-universiteiten

Reviewcommissie Hoger Onderwijs en Onderzoek. (n.d.-c). *Overzicht van indicatoren (verplicht en facultatief)*. Retrieved from http://www.rcho.nl/images/RC overzicht indicatoren.pdf

Reviewcommissie Hoger Onderwijs en Onderzoek. (2017). *Prestatieafspraken: het vervolgproces na 2016. Advies en zelfevaluatie*.

Sanford, T., & Hunter, J. M. (2011). Impact of performance-funding on retention and graduation rates. *Education Policy Analysis Archives*, *19*(33), 1–30.

Schertzer, C. B., & Schertzer, S. M. B. (2004). Student Satisfaction and Retention: A Conceptual Model. *Journal of Marketing for Higher Education*, *14*(1), 79–91. Retrieved from https://www.tandfonline.com/doi/pdf/10.1300/J050v14n01_05?needAccess=true

Shin, J. C. (2010). Impacts of performance-based accountability on institutional performance in the. *Higher Education*, *60*(1), 47–68. https://doi.org/10.1007/s10734-009-9285-y

Taylor, A. (2014). Standardised Effect Size in a Mixed/Multilevel Model.

Umbricht, M. R., Fernandez, F., & Ortagus, J. C. (2017). An Examination of the (Un)Intended Consequences of Performance Funding in Higher Education. *Educational Policy*, *31*(5), 643–673. https://doi.org/10.1177/0895904815614398

van den Broek, A., Wartenbergh, F., Bendig-Jacobs, J., Braam, C., & Nooij, J. (2016). *Monitor Beleidsmaatregelen 2015*. Nijmegen. Retrieved from https://www.mboraad.nl/sites/default/files/monitor_beleidsmaatregelen_2015.pdf

Vereniging Hogescholen. (n.d.-a). *Definities personeel kengetallen*. Retrieved from https://www.verenigingh ogescholen.nl/system/knowledge_base/attachments/files/000/000/172/original/Definities_personeel_kengetallen.pdf?1438936767

Vereniging Hogescholen. (n.d.-b). Feiten en Cijfers. Retrieved May 12, 2018, from http://www.verenigingh ogescholen.nl/kennisbank/feiten-en-cijfers

Vereniging Hogescholen. (n.d.-c). Vereniging Hogescholen | Sectoren. Retrieved June 26, 2018, from https://www.verenigingh ogescholen.nl/sectoren

Vereniging Hogescholen. (2018a). *Bron en definities studentenaantallen Vereniging Hogescholen* (Vol. 1). Retrieved from http://www.verenigingh ogescholen.nl/system/knowledge_base/attachments/files/000/000/171/original/Bron_en_definities_studentenaantallen.pdf

Vereniging Hogescholen. (2018b). *Factsheet Feiten & Cijfers*. Retrieved from

http://www.vereniginghogescholen.nl/system/knowledge_base/attachments/files/000/000/841/original/factsheet_studentenaantallen_2017_definitief.pdf?1518013848

VSNU. (n.d.). Basiskwalificatie Onderwijs. Retrieved June 14, 2018, from https://www.vsnu.nl/bko.html

VSNU. (2016). *Prestaties in perspectief. Eindrapport Hoofdlijnenakkoord en prestatieafspraken*. *VSNU Prestaties in perspectief*. Retrieved from https://vsnu.nl/files/documenten/Nieuwsberichten/VSNU_Prestaties_in_perspectief_%28web%29.pdf

VSNU. (2018). Aantal ingeschreven studenten. Retrieved May 1, 2018, from https://www.vsnu.nl/f_c_ingeschreven_studenten.html

Wild, L., & Ebbers, L. (2002). Rethinking student retention in community colleges. *Community College Journal of Research and Practice*, *26*(6), 503–519. https://doi.org/10.1080/2776770290041864

Zestor. (n.d.). Inventarisatie trajecten Basiskwalificatie Didactische Bekwaamheid (BDB). Retrieved June 25, 2018, from http://www.zestor.nl/inventarisatie-trajecten-basiskwalificatie-didactische-bekwaamheid-bdb

Zijlstra, W., Asper, H., Amrani, A., Tupan-Wenno, M., Van Stapele, M. C., & Van Stapele, N. (2013). *Generiek is divers. Sturen op studiesucces in een grootstedelijke context*. Utrecht.

Figures and tables

Appendix. Extended analysis

The output of the repeated-measures ANOVA for the development in student satisfaction rates over time showed a significant contrast in student satisfaction scores between the year of policy implementation (T2) and 3 years after (T5). This means that student satisfaction scores have significantly increased after the introduction of the performance agreements. A spontaneous interest aroused in whether the trend in student satisfaction differed according to institutional factors. This extension of the analysis aimed to obtain more detailed information about the conditions under which the trend in satisfaction occurred.

The first step of this extended analysis was to define institutional factors according to which universities of applied sciences would be categorized. The author oriented herself toward previous research to define institutional factors (McLellan et al., 2015). McLellan et al. (2015) investigated whether organizational factors in terms of size, sector, leadership support and organizational capacity were associated with the implementation of worksite health protection and promotion programs in smaller businesses. The authors defined size as the number of employees in each company (operationalized as a dummy variable) and categorized businesses across four broad sectors on a nominal variable.

Since information on size and profile could be easily retrieved in the case of Dutch universities of applied sciences, the author chose to investigate whether the trend of student satisfaction differed across institutions regarding their size and profile.

The idea behind including the factor ‘Size’ was that universities of applied sciences with more students and thus more financial resources might have had more capacities to respond to the introduction of the performance agreements so that the trend of student satisfaction might have been stronger at these universities. Universities of applied sciences were rated from ‘very small’ (‘1’), ‘small’ (‘2’), ‘large’ (‘3’), and ‘very large’ (‘4’). Similar to the procedure in McLellan et al. (2015), the categories had been determined on basis of the max. and min. value and the median of the number of enrolled students in academic year 2011/2012, i.e. the first measurement moment included in the ANOVA. Table 8 provides the Frequency distribution for the variable ‘Number of enrolled students’ in AY 2011/2012. The median was chosen to create categories due to the uneven distribution of universities of applied sciences regarding their size. The boxplot can be found in the appendix. Universities of applied sciences were assigned categories on basis of the following intervals: Category 1 [473; 1797], Category 2 [1798; 4067], Category 3 [4068; 20548] and Category 4 [20549; 45165]. The frequency distribution of the newly created variable can be seen in Table 9.

Table 8. Frequency distribution of the Number of enrolled students in academic year 2011/2012

Statistics		
Inschrijvingen_2011		
N	Valid	35

Missing	0
Mean	12038.51
Median	4068.00
Std. Deviation	13557.058
Minimum	473
Maximum	45165

Table 9. Frequency distribution of the created variable 'Size' of universities of applied sciences, N = 35

Size					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very small	12	34.3	34.3	34.3
	small	5	14.3	14.3	48.6
	big	7	20.0	20.0	68.6
	very big	11	31.4	31.4	100.0
	Total	35	100.0	100.0	

Second, universities of applied sciences were categorized regarding the sector(s) they serve. Given the small sample size of 32 institutions, it was decided against categories for every single sector (Vereniging Hogescholen, n.d.-c), since this would have led to even lower sample sizes per category. To obtain more cases per category, the sectors were not categorized regarding their specific profile but according to the extent of specialization. Thereafter three categories were created: 'Not specialized' ('1'), 'Specialized' ('2') and 'Highly Specialized' ('3'). 'Not specialized' was attributed in case that a university of applied sciences offered education for several sectors, whereas 'specialized' was attributed institutions that covered specific sector clusters (more than one sector, but related sectors). Last, 'specialized' was attributed if an institution was specialized in only one sector. The frequency distribution of the newly created variable can be found in Table 10. It was speculated that more specialized universities of applied sciences would probably serve a more homogenous student body so that it might be easier for those institutions to implement policies aimed to benefit students.

Table 10. Frequency distribution of the created variable 'Profile' of universities of applied sciences, N = 35

Profile					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not specialized	18	51.4	51.4	51.4
	Specialized	3	8.6	8.6	60.0
	Highly Specialized	14	40.0	40.0	100.0

Total	35	100.0	100.0	

Hence, the model of the ANOVA was extended by including separately the factors 'Size' and 'Profile' that may have conditioned the observed increase in student satisfaction. Both variables were specified as ordinal variables. They were included as between-subjects factors respectively. The output of the analyses was analyzed according to the test of within-subjects effects, that is time. It was tested whether there was a significant interaction between the within-subjects effect (time) and the between-subjects effects (size/profile).

The output for the interaction with the factor 'Size' can be found in Table 11. The analysis showed that the interaction between the factors 'Time' and 'Size' was not significant ($p = .467$)[8]. This means that the trend in student satisfaction over time did not vary significantly across universities regarding their size.

Table 11. Non-significant interaction effect between time and the size of universities of applied sciences

Tests of Within-Subjects Effects

Measure: Satisfaction

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	502.570	4	125.643	11.984	.000	.300
	Greenhouse-Geisser	502.570	1.812	277.322	11.984	.000	.300
	Huynh-Feldt	502.570	2.138	235.058	11.984	.000	.300
	Lower-bound	502.570	1.000	502.570	11.984	.002	.300
time * Size	Sphericity Assumed	118.554	12	9.879	.942	.508	.092
	Greenhouse-Geisser	118.554	5.437	21.806	.942	.467	.092
	Huynh-Feldt	118.554	6.414	18.483	.942	.476	.092
	Lower-bound	118.554	3.000	39.518	.942	.433	.092
Error(time)	Sphericity Assumed	1174.220	112	10.484			
	Greenhouse-Geisser	1174.220	50.742	23.141			
	Huynh-Feldt	1174.220	59.866	19.614			
	Lower-bound	1174.220	28.000	41.936			

Dependent variable: Satisfaction

[8] Greenhouse-Geisser corrected values were used for interpretation. Further explanations of the output interpretation can be found in (Field, 2012).

Table 12 provides the tests of the within-subjects effect (Time) with the factor 'Profile' included. The analysis showed that the interaction between time and the degree of profile specialization at Dutch universities of applied sciences was not significant (p = .819)[9]. Hence, the trend of student satisfaction over time did not vary significantly across institutions regarding their degree of profile specialization.

Table 12. Non-significant interaction effect between time and the profile of universities of applied sciences

Tests of Within-Subjects Effects

Measure: Satisfaction Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	323.784	4	80.946	7.443	.000	.204
	Greenhouse-Geisser	323.784	1.817	178.201	7.443	.002	.204
	Huynh-Feldt	323.784	2.065	156.769	7.443	.001	.204
	Lower-bound	323.784	1.000	323.784	7.443	.011	.204
time * Profile	Sphericity Assumed	31.285	8	3.911	.360	.940	.024
	Greenhouse-Geisser	31.285	3.634	8.609	.360	.819	.024
	Huynh-Feldt	31.285	4.131	7.574	.360	.842	.024
	Lower-bound	31.285	2.000	15.642	.360	.701	.024
Error(time)	Sphericity Assumed	1261.489	116	10.875			
	Greenhouse-Geisser	1261.489	52.692	23.941			
	Huynh-Feldt	1261.489	59.895	21.062			
	Lower-bound	1261.489	29.000	43.500			

Dependent variable: Satisfaction

[9] Greenhouse-Geisser corrected values were used for interpretation. Further explanations of the output interpretation can be found in (Field, 2012).