Is It Working? An Impact Evaluation of the German "Women Professors Program"

Andrea Löther

GESIS—Leibniz-Institute for the Social Sciences, 50667 Cologne, Germany; andrea.loether@gesis.org

Abstract: The Women Professors Program, which was initiated in Germany in 2008, aims to increase the proportion of women professors and to promote structural change in favour of gender equality at higher education institutions (HEIs). It is one of the central gender equality policies in higher education in Germany. The present study evaluates the impact of the program by estimating its causal effects on the proportion of women professors. By adopting a quasi-experimental approach and using a unique dataset—a long term census of German HEIs—the study proves that the proportion of women professors increased more than would have been expected in the absence of the program. Although the evaluation includes preliminary estimates of mechanisms driving the described impacts, the integration of context factors and mechanisms into the assessment of the impact of gender equality policies remains a desideratum. The study shows that the program is working, and it contributes to redressing the lack of impact studies on gender equality in science and research.

Keywords: impact evaluation; quantitative effects; quasi-experimental design; gender equality; higher education; program evaluation; program theory; proportion of women professors; Germany

1. Introduction

As in many other countries, women are still underrepresented in leadership positions at higher education institutions (HEIs) in Germany. In 2007, the proportion of women among the highest-ranking professors (Grade A according to the *She Figures*) in Germany was only 12%, which was considerably below the European average of 19% (European Commission 2009, p. 76). Ten years later, this proportion had increased to 20%. The Women Professors Program was initiated in Germany in 2008 with the aim of increasing the proportion of women professors and fostering structural change in favour of gender equality at HEIs (Zippel et al. 2016). One of the central gender equality policies in higher education in Germany, the program is financed jointly by the Federal Government and the Laender (federal states). An amount of 150 million euros was made available for the first and second phases of the program (2008–2012, 2013–2017). Funding was increased to 200 million euros for the third phase (2018–2022). The program combines structural changes—the implementation of a gender action plan—with financial incentives to appoint women to professorships. The first two program phases have been successfully evaluated (Löther and Glanz 2017; Zimmermann 2012).

The long duration of the program gives rise to the question of its impact on gender equality and on the representation of women in leadership positions. Is the program working? How can the impact of such a gender equality program be measured?

Focusing on the central objectives of the program, this study assesses its impact on the proportion of women professors at HEIs in Germany. In the next section, I elaborate the theoretical and methodological framework of the study by reviewing the literature on impact evaluation and gender equality policies. In Section 3, I outline the program theory and context. This is followed in Section 4 by a description of the research design and data basis of the study. After presenting the empirical

results in Section 5,[1] I will discuss them and draw some conclusions regarding the opportunities and challenges of impact evaluations of gender equality programs.

My main argument is that the program is working in that the proportion of women professors increased more than would have been expected in the absence of the program, but further investigation is needed to understand the mechanisms driving the impacts.

2. Gender Equality Policies and Impact Evaluation

Impact evaluations seek to understand the role of an intervention in producing measured and described changes (Fletcher 2015, p. 9). Estimating the counterfactual is the dominant principle to evaluate impact: "Evaluators assess the effects of social programs by comparing information about outcomes for program participants with estimates of what their outcomes would have been had they not participated". (Rossi et al. 2004, p. 234). Randomized experiments are still the gold standard for measuring the counterfactual, but for practical, ethical, or political reasons, quasi-experimental designs are often used to establish causal inference in evaluation studies (Khandker et al. 2012; Steiner et al. 2009). However, evaluation research has gone beyond this approach. Evaluation studies—and the social sciences in general—regard causation as "a probabilistic rather than a deterministic concept" (Steiner et al. 2009, p. 77), and aim to identify differing patterns before and after interventions rather than to perform definitive causal analysis (Stepan-Norris and Kerrissey 2016, p. 229). Instead of attributing outcomes to an intervention, evaluations investigate the way in which an intervention contributed to outcomes (Gates and Dyson 2017, p. 31).

Gates and Dyson 2017) also pointed to the growing acknowledgment "that there are multiple ways to think about causal relationships". They described five different ways of thinking about causality, including a generative logic, which "verifies a theory-based explanation of how causal processes happen by showing how mechanisms work within particular contexts to generate outcome patterns" (Gates and Dyson 2017, p. 36). This reference to mechanisms leads to another critique of impact evaluations. Müller and Albrecht (2016, p. 291) criticized the fact that, to date, impact evaluations had examined the effects of programs, but that there was a lack of knowledge about the mechanisms driving the effects. How the programs work thus remains a black box. Program model evaluation and theory-driven evaluation are approaches to studying the mechanisms that link the elements, context, and outcomes of a program (Astbury and Leeuw 2010; Funnell and Rogers 2011; Rogers 2000). Müller and Albrecht (2016, p. 291) suggest combining both approaches:

> Thus, from our point of view the future of impact evaluation has to be a mixture in the sense of combining approaches of black box and white box evaluation. Consequently, we suggest an integrative approach of impact evaluation that covers the estimation of overall intervention effects and the identification of the key drivers of effects by using adequate designs and methods of empirical research.

In the present study, I will follow this suggestion by analyzing the impact of the gender equality program using a quasi-experimental design and by exploring how quantitative data can be used to obtain knowledge about the mechanisms that drive the impact.

Evaluations of the impact of gender equality policies in higher education are still scarce (Kaplpazidou Schmidt et al. 2017, p. 27; Stepan-Norris and Kerrissey 2016, p. 228f.; Timmers et al. 2010, p. 720). As noted by Kalev et al. (2006, p. 590), "whereas there has been a great deal of research on the sources of inequality, there has been little on the efficacy of different programs for countering it". One decade later, Stepan-Norris and Kerrissey (2016, p. 228) made a similar observation, namely, that "far fewer studies empirically examine how interventions may matter". Even more striking is the lack of studies measuring quantitative effects (Timmers et al. 2010, p. 722f.). Whereas evaluations of gender equality interventions in

---

[1] The results will be published in German in *Zeitschrift für Evaluation*, issue 1/2019.

the development policy context often use quasi-experimental designs (Burde and Linden 2012; d'Agostino et al. 2016; Grillos 2018; World Bank n.d.), evaluation studies on gender equality programs in science and research rely mostly on qualitative methods, surveys without control groups and pre-test/post-test measurement, and descriptive and cross-sectional data (Höppel 2016; Munir et al. 2014; Williamson 2016).

The lack of studies on the impact of gender equality interventions in higher education is due partly to the complexity of the educational system. The combination of individual, social, and structural influence factors complicates the detection of causalities (Wroblewski et al. 2007, p. 19). It is difficult to attribute changes in gender equality to programs rather than to "wider contextual trends and factors" (Kaplpazidou Schmidt et al. 2017, p. 27). Data availability and financial and time restrictions, which reduce the possibility of using control groups and pre-post measurement, also contribute to the lack of impact studies. Furthermore, impact studies on gender equality have to deal with success indicators. Gender impact and gender equality cannot be limited to the increase in the proportion of women in leadership positions. Studies on women in senior management, for example, have shown that the increase in the proportion of women in these positions has not necessarily changed structural barriers (Neale and Özkanlı 2010; Peterson 2011). As Peterson (2011, p. 626) noted, "gender equality aspects of a qualitative character have to be taken into consideration to measure real change".

There are very few impact evaluations that endeavor to measure quantitative effects. With regard to effects on the level of individual participants, two evaluations of Austrian programs for the promotion of women scientists formed a control group comprising those applicants who were not funded (see Pohn-Weidinger and Grasenick 2011; Wroblewski et al. 2007, pp. 344ff.). A research design such as this requires a sufficiently large group of rejected applications, as may be the case in individual funding schemes (see Heidler 2016).

With regard to impacts at the level of HEIs, Timmers et al. (2010, p. 732) investigated the efficacy of gender equality measures at all 14 universities in the Netherlands. They examined the correlations between implemented measures—which they classified according to individual, cultural, and structural perspectives—and changes in the Glass Ceiling Index and in the proportion of women among academic staff. In conclusion, they noted that "the relationships between policy measures and the reduction of the glass ceiling and between the policies in the cultural perspective and the increase in the proportion of women among professors reveal that the applied policy measures have been effective" (Timmers et al. 2010).

Stepan-Norris and Kerrissey (2016) evaluated the impact of an intervention—the ADVANCE Institutional Transformation Initiative—implemented at the University of California, Irvine (UCI). In order to estimate the counterfactual, they used as a control group seven other University of California campuses where no initiatives of that scale had been implemented. Using descriptive statistics, *t*-tests, and regression analysis of data on women's faculty representation, hiring, and retention, they could show that "ADVANCE awards have the potential to increase women's representation over and above what may occur otherwise. We find that although all the UC campuses increased women's representation, UCI under the ADVANCE program made the biggest gains" (Stepan-Norris and Kerrissey 2016).

Contrary to these studies attesting to the efficacy of gender equality policies, an econometric study on the impact of the Athena SWAN charter, an initiative aimed at advancing women's careers, in UK medical schools arrived at a negative conclusion, namely, that:

> "Despite a general increase in female employment and widespread adoption of the standards of Athena SWAN amongst UK medical schools there is no evidence yet to suggest that either the introduction of the Athena SWAN charter or the announcement of NIHR to tie future funding to Athena SWAN silver status has led to a measurable improvement in the careers of females employed in UK medical schools." (Gregory-Smith 2015, p. 21)

The aforementioned study compared medical schools at universities with and without an Athena SWAN award by using panel data on the proportion of women faculty and calculating fixed effects and

differences in difference. Gregory-Smith's findings contrast in part with those of Caffrey et al. (2016), who used a realistic evaluation approach to assess the effects of Athena SWAN in academic medicine. Based on in-depth interviews, focus groups, and participant observation, Caffrey et al.'s qualitative study at one HEI found that, although participants reported positive outcomes—for example, the creation of social space to address gender inequity—the program's implementation had reinforced gender inequalities because female staff had to carry out "a disproportionate amount of Athena SWAN work, with potential negative effects on individual women's career progression". Moreover, participants perceived the impact of the program to be "undermined by wider institutional practices, national policies, and societal norms", and early career researchers reported difficulties accessing program initiatives. The more differentiated results may stem from the different research designs and also from the fact that Gregory-Smith did not integrate his impact evaluation into a theoretical framework of gender equality. Without program theory, it remains unclear whether the Athena SWAN charter, which was implemented in the whole university, can increase the percentage of women faculty in the medical schools, as investigated by Gregory-Smith.

The literature review reveals that there is a lack of impact studies on gender equality in science and research, and especially of studies that calculate quantitative effects. With the aim of contributing to filling this gap in research, I will evaluate the quantitative impact of the Women Professors Program in Germany using a quasi-experimental design to estimate effects. Thus, my first research question is: Has the program had an impact on the proportion of women professors? Furthermore, literature about evaluation research and gender equality programs reflect on the limitations of impact studies. To understand whether and in what way a program works, we need to combine different approaches to assess its impact. Exploring the program theory holds the possibility of looking for mechanisms and context factors of the program. Taking these reflections into account, I seek to understand mechanisms and context factors, and to examine how these elements can be integrated into the quantitative research design. Hence, my second research question is: What are the mechanisms that contribute to the effects? My overall aim is to estimate the impact of a gender equality program and to understand the knowledge gains and the limitations of a quantitative impact evaluation of gender equality policies.

3. Program Theory and Context

Rossi et al. (2004, p. 64) define program theory as "the set of assumptions about the manner in which a program relates to the social benefits it is expected to produce and the strategy and tactics the program has adopted to achieve its goals and objectives". The program theory for the Women Professors Program (see Figure 1) shows that the program aims to produce structural changes and to improve gender equality at HEIs by providing start-up funding for professorships filled by newly recruited women. Funding is subject to the positive assessment of a gender action plan.[2] If an external evaluation committee assesses the gender action plan positively, the HEIs have the right to apply for start-up funding (for five years) for up to three tenured professorships filled by newly recruited women. The funding can be used for regular professorships, or for so-called "early appointments to a professorship" (*Vorgriffsprofessuren*), that is, appointments to professorships that have been created in addition to currently occupied chairs that are due to become vacant in the foreseeable future on the retirement of the incumbent. In the case of the funding for a regular professorship, the funds already allocated for this position in the HEI's budget—which are freed up by the program funding—must be spent on gender equality measures.

---

[2]   In the second program phase (2013–2017), HEIs that had participated in the first phase had to submit documentation of the implementation of the first gender action plan.
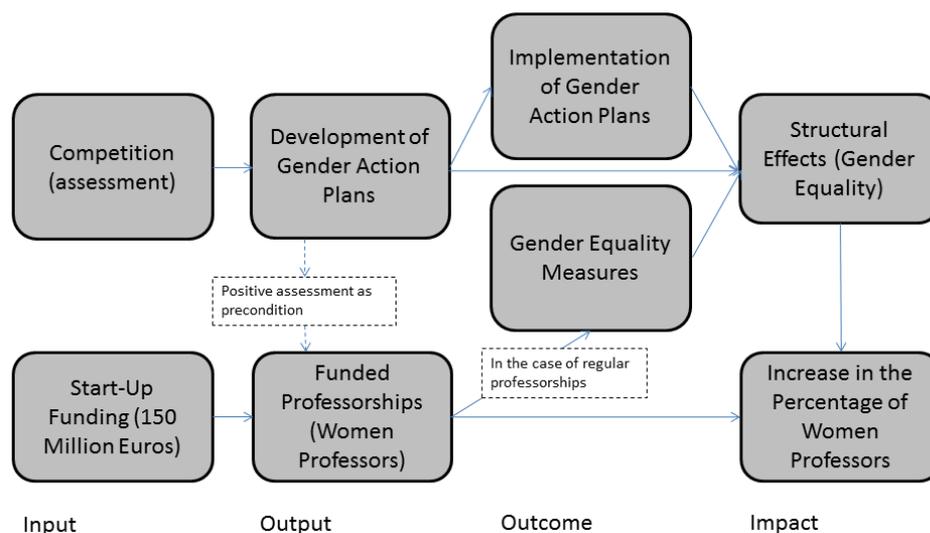
Figure 1. Program Theory.

The aim of the program is to improve gender equality in higher education and to promote structural change. Furthermore, the program seeks to increase the participation of women at all qualification levels, and especially in leadership positions and professorships. The present study focuses on the program objective to "increase the proportion of women professors". I will examine whether changes in women's representation in professorships can be causally attributed to the Women Professors Program. In doing so, I will focus on meso-level effects—that is, effects at the level of the individual universities—and on the macro-level impact, that is, the impact on the higher education system as a whole. Concerning the mechanisms, the program theory presupposes that an increase in the proportion of women professors can be achieved both through the professorships supported by the start-up funding and through the implementation of a gender action plan. The start-up funding may give HEIs an incentive to appoint women as professors. The implementation of the gender action plan may have a more indirect effect by raising awareness, increasing transparency in recruitment procedures, and other similar measures. Moreover, start-up funding also has an indirect effect: in the case of regular professorships, the resources freed up by the external funding of the professorship must be invested in gender equality measures and thus contribute to the implementation of the gender action plan. This possible mode of action is of great importance, as the funding of regular professorships is the dominant form of funding in the program, accounting for 60% of the funding in the first program phase (PP I) and 69% of the funding in PP II (Löther and Glanz 2017, p. 32). I will analyze whether the start-up funding has had direct and/or indirect effects on the proportion of women professors.

Contextual factors may foster or hinder the impact of a program. The Women Professors Program is part of a bundle of innovative gender equality efforts initiated since 2006, which also includes the Research-Oriented Standards on Gender Equality of the German Research Foundation (DFG), Germany's leading research funder, and the gender equality standards of the German Excellence Initiative, which is aimed at making Germany a more attractive research location.[3] All three measures support a strategy shift toward more structural change in organisations rather than focusing on women scientists (see Weber 2017; Riegraf and Weber 2013; Simon 2013; Zippel et al. 2016). On the other hand, precarious working and qualification conditions, such as the high proportion of fixed-term contracts, the almost exclusive orientation of scientific careers toward obtaining a professorship, and

---

[3]   Between 2006 and 2012, the Excellence Initiative gave additional funding of 1.9 billion euros for three funding lines (graduate schools, clusters of excellence and institutional strategies to promote top-level university research). Only universities were eligible for funding under the Excellence Initiative. Forty-five universities (about 40% of all universities in Germany) received funding under the Excellence Initiative.

the intensified competition for professorships, are factors that may run counter to the objectives of the Women Professors Program. Women are more frequently affected by fixed-term and part-time contracts (see Beaufaÿs and Löther 2017; Courtois and O'Keefe 2015; Nikunen 2014). Furthermore, they have less occupational and non-occupational resources at their disposal with which to strengthen their position in the competition for academic advancement. Funken et al. (2013, p. 50f.) described this situation as a control paradox: for women early career researchers, increased competition together with greater efforts to increase gender equality in German science and research leads to an almost cynical situation where gender equality programs suggest that they have good academic career opportunities, but the intense competition acts as a structural brake on equal opportunities. When interpreting the results of the impact study, these context factors must be taken into consideration.

4. Research Design and Data Basis

The research approach that I have chosen for the evaluation of the impact of the Women Professors Program is a rigorous impact analysis based on the potential outcome model and the estimation of the counterfactual (Khandker et al. 2012; Meyer et al. 2008; Rossi et al. 2004; Rubin 1979; Wolbring 2014). In order to be able to causally attribute change (here: the increase in the proportion of women professors) to an intervention (here: the Women Professors Program), the results must be compared with the conditions that would be expected if the program had not existed: "The impact measurement must therefore take the counterfactual into account" (Meyer et al. 2008, p. 23f.). Because the counterfactual cannot be observed, evaluators use various methods to estimate it (Khandker et al. 2012; Legewie 2012; Morgan and Winship 2010). Experiments with random assignment to the treatment and control groups are viewed as the "gold standard" for assessing causal effects (Rossi et al. 2004, p. 237). However, in many evaluations, especially in the field of education, this research design is not possible (see Steiner et al. 2009 for experiments in evaluation).

The Women Professors Program was conceptualized as a competitive procedure. Participation of HEIs in the program is determined by self-selection (submission of a gender action plan) and external selection (assessment of the gender action plan by the evaluation committee). Due to these circumstances, the counterfactual was estimated using a quasi-experimental design: At the level of individual HEIs (1), the non-participating HEIs were the control group. The estimate was made by means of pre- and post-treatment measurements and the calculation of difference-in-differences. I calculated the differences between the proportion of women professors before and at the latest possible date during the intervention[4] for HEIs participating in the program (treatment group) and non-participating HEIs (control group). At the level of the higher education system (2), the counterfactual status was estimated by extrapolating the increase in the proportion of women professors during the years immediately preceding the introduction of the program to the duration of the program.

Information about the participation of HEIs in the program was based on administrative data provided by the DLR Project Management Agency,[5] which implements the program on behalf of the Federal Ministry of Education and Research (BMBF). The ministry published lists of positively evaluated universities after the decisions were made in 2008, 2009, 2013, and 2014. Data on academic staff were obtained from the Federal Statistical Office (Destatis), which provides long time series broken down by individual universities, categories of staff (e.g., professors), and gender.

---

[4]　The program was still ongoing when the evaluation took place.

[5]　Project management agencies implement projects in a professional and organizational capacity, for example, by providing advice, administrative handling, and technical support for projects. The contracting authorities are mainly ministries at federal and/or state level. The project management agencies are located at research institutions and other organizations. The DLR Project Management Agency, which is part of the German Aerospace Center (DLR), is an example of such an agency.

5. Empirical Results

5.1. Changes in the Proportion of Women Professors at Participating and Non-Participating HEIs

5.1.1. Establishment of Treatment and Control Groups

The treatment group comprised HEIs whose gender action plans had been positively assessed, and who were thus participating in the program. The control group consisted of those HEIs that did not submit a gender action plan or whose plan had not been positively assessed. They were assigned to the control group via a combination of self-selection (non-application) and external selection (non-positive assessment), although the data available[6] did not allow me to distinguish between self-selection and external selection. Taking the two program phases into account, there were four groups in total:

1. No application or not positively assessed
2. Positively assessed only in Program Phase I
3. Positively assessed only in Program Phase II
4. Positively assessed in both program phases

The analysis was carried out both by comparing all four groups, and in a binary comparison of the treatment group of participating HEIs (Groups 2–4 combined) and the control group of non-participating HEIs (Group 1). Unless otherwise stated, the treatment group thus included those HEIs whose gender action plans were positively assessed in Program Phase I and/or II.

The study population comprised all HEIs that could potentially have participated in the program. The 268 member HEIs of the German Rectors' Conference (HRK) were included in the sample as an approximation of the population, supplemented with five HEIs whose gender action plans were assessed positively but who are not members of the HRK (see Löther and Glanz 2017, p. 75f.). Thus, the evaluation covered almost 95% of academic staff and students at all German HEIs. The data for the population of HEIs were therefore available almost in the form of a census survey. Table 1. gives an overview of the population and the participation of HEIs in the Women Professors Program.

Table 1. Overview of All HEIs, Submissions, Positive Assessments, and Funding.

HEIs	Program Phase I		Program Phase II	
	Number	Percentage of All HEIs (%)	Number	Percentage of All HEIs (%)
All HEIs	268		273	
Applications	152	56.7	184	67.4
Positively assessed gender action plan	124	46.3	147	53.8
HEIs with at least one professorship funded under the program	111	41.4	120	44.0

Source: DLR Project Management Agency administrative data (see Löther and Glanz 2017, p. 23).

5.1.2. Comparison of the Treatment and Control Groups

Due to the long time series of higher education statistics, the data on the proportion of women professors at the individual HEIs were available in a panel design. The data were based on the data for the "professorships" category. This category includes the highest positions of tenured professorships as well as non-tenured junior professorships. As data showing the junior professorships separately were not always available for individual HEIs, this group could not be excluded from the calculation of the proportion of women professors. Using the difference-in-differences approach, I calculated

---

[6] Information on the number of HEIs that applied but were not positively assessed was available differentiated by federal state and HEI type but not by individual HEI.

the difference between the proportion of women professors at t_1 (2007) and t_2 (2015). t_1 was the year before the start of the program; t_2 was the last year for which data were available during the evaluation. A causal effect of the program may be assumed if the proportion of women professors at participating HEIs increased more than that at non-participating HEIs.

Table 2 shows that the proportion of women professors at participating HEIs increased by an average of 6.4 percentage points between 2007 and 2015, whereas at non-participating universities it rose by only 4.6 percentage points, and thus by 1.8 percentage points less than at participating HEIs. HEIs that were successful in both program phases were able to increase the proportion of women professors particularly significantly (by 6.9 percentage points). Thus, an effect of 1.8 percentage points between 2007 and 2015 and an annual average of 0.22 percentage points in increasing the proportion of women professors could be attributed to the Women Professors Program.

Table 2. Comparison of Means: Increase in the Proportion of Women Professors between 2007 and 2015 by Participation in the Women Professors Program.

HEIs	N	Mean	Median	Standard Deviation
No application or not positively assessed	99	4.594	4.088	7.265
Positively assessed (PP I and/or PP II)	162	6.379	6.268	4.735
Thereof				
Positively assessed only in PP I	22	4.761	4.312	4.433
Positively assessed only in PP II	43	6.050	5.450	6.501
Positively assessed in both program phases	97	6.891	6.939	3.729
Total	261	5.702	5.673	5.876

Source: Federal Statistical Office; DLR Project Management Agency administrative data; own calculations. *Note*: PP I = Program Phase I; PP II = Program Phase II. Significance test: Comparison of four groups: $p < 0.05$; comparison of participating and non-participating groups: $p < 0.05$. The differences were significant at the 5% level.

5.2. Selection Bias

Estimating the counterfactual by comparing treatment and control groups presupposes that there are no systematic differences between the groups that may influence the results. In contrast to random assignment to treatment and control groups, such differences in observed and unobserved characteristics—and thus selection biases—are to be expected in the case of self-selection and external selection. Hence, the possibility of selection bias must be borne in mind in the case of the Women Professors Program. In fact, as can be seen from Table 3, the treatment and control groups did indeed differ in terms of HEI type and size, as well as regional distribution.

The most significant difference between the treatment and control groups was the HEI type and, related thereto, the HEI size, operationalized as the number of students enrolled. Universities[7] were overrepresented among those HEIs whose gender action plans were positively evaluated, and universities are, on average, larger than other HEIs. In addition, HEIs from federal states in which the position of Equal Opportunities Officer is the incumbent's main job participate disproportionately frequently in the program. The reasons for these differences between participating and non-participating HEIs are the financial and human resources available for the preparation of a successful gender action plan and for the administrative implementation of the program (see Löther and Glanz 2017, pp. 23ff.). There are hardly any conspicuous features in the distribution between the western and eastern German states and Berlin. The Glass Ceiling Index (GCI) in 2007 was higher for participating HEIs than for non-participating HEIs. This contradicts the assumption that the HEIs that participate in the program are primarily those that were already pursuing a successful gender equality policy beforehand.

---

[7] Universities have the right to confer doctoral degrees. Universities of applied sciences focus on teaching professional skills. They specialize in specific fields—mainly engineering, computational sciences, business, and social work.

Table 3. Characteristics of Participating and Non-Participating HEIs (HEI Type, Regional Distribution, HEI Size, and Proportion of Women Professors).

HEIs	No Application or Not Positively Assessed		Positively Assessed (PP I and/or PP II)		Total HEIs	
	N	Percentage or Mean	N	Percentage or Mean	N	Percentage or Mean
Universities	17	16.3%	83	48.0%	100	36.1%
Universities of applied sciences	56	53.8%	76	43.9%	132	47.7%
Colleges of art and music	31	29.8%	14	8.1%	45	16.2%
Eastern German States	15	14.4%	32	18.5%	47	17.0%
Western German States	87	83.7%	127	73.4%	214	77.3%
Berlin	2	1.9%	14	8.1%	16	5.8%
Equal opportunities officer job model[8] Main-job model	24	23.1%	66	38.2%	90	32.5%
Time-off model	80	76.9%	107	61.8%	187	67.5%
Number of students	103	3270	166	12,614.28	269	9.036
Percentage of women professors in 2007	100	20.04	166	17.53	266	18.47
Glass Ceiling Index[9] 2007	96	2.90	164	3.20	260	3.09
Total	104	100%	173	100%	277	100.0%

Note. PP I = Program Phase I; PP II = Program Phase II.

The difference-in-differences approach used in this evaluation is a way of statistically controlling selection effects (Khandker et al. 2012, pp. 71–78; Caspari 2009; Legewie 2012, p. 135f.), assuming that the unobserved characteristics are constant over time and do not correlate with the intervention. According to Khandker et al. (2012, p. 76) however, there are doubts about these assumptions. Therefore, a subgroup analysis and a regression analysis were conducted to clarify whether the described differences between the treatment and control groups influenced the effect of the Women Professors Program on the increase in the proportion of women professors.

The subgroup analysis (see Table 4) revealed that the differences between the treatment and control groups did not distort the effect of the Women Professors Program. In almost all subgroups, the proportion of women professors rose more strongly at the participating HEIs (the treatment group) than in the control group. Only in the case of colleges of art and music did the treatment and control groups not differ. In Berlin, the higher increase among non-participating HEIs was probably due to the small number of cases.

The regression analysis (Table 5) confirmed that participation in the Women Professors Program had an independent effect on the increase in the proportion of women professors from 2007 to 2015, irrespective of the type of HEI or the regional distribution. In Model 1, participation in the Women Professors Program resulted in a 1.3 percentage point higher increase in the proportion of women professors (non-standardized regression coefficient). However, this influence was not statistically significant. At the same time, the regression analysis showed that the regional distribution and the proportion of women professors at the beginning of the program exerted a large and statistically significant influence. The proportion of women professors increased at the HEIs in eastern German

---

[8]   In most Laender, the equal opportunities officer is elected among members of the HEI, and the incumbent is fully or partially released from her other tasks and duties (time-off model). In some Laender (e.g., Berlin and Lower Saxony), the post of equal opportunities officer is externally advertised, and it is the incumbent's main job (main-job model).

[9]   The Glass Ceiling Index (GCI) is a relative index comparing the proportion of women in lower positions to the proportion of women in top academic positions see European Commission—DG Research (2016, p. 89f.). Here, I calculated the CGI by dividing the proportion of female students by the proportion of female professors: the higher the value, the stronger the glass ceiling effect.

states to a much lesser extent than at the HEIs in the western German states or Berlin. The variable "Eastern German HEI" led to a 2.1 percentage point lower increase (Model 1). HEIs that had a low proportion of women professors in 2007 were able to increase this proportion more strongly by 2015 than HEIs whose proportion of women professors was already high in 2007.

Table 4. Comparison of Means: Increase in the Proportion of Women Professors between 2007 and 2015 by participation in the Women Professors Program; Calculated Separately for Types of HEIs, Region, and Equal Opportunities Officer Job Model.

HEIs	No Application or Not Positively Assessed		Positively Assessed (PP I and/or PP II)		Total HEIs	
	N	Mean	N	Mean[10]	N	Mean
Universities	19	4.842	81	7.375 *	100	6.893
Universities of applied sciences	49	4.525	67	5.553	116	5.119
Colleges of art and music	31	4.551	14	4.566	45	4.556
Western German states	83	4.751	119	6.480	202	5.769
Eastern German states	14	1.690	31	4.957 *	45	3.941
Berlin	2	18.406	12	9.046	14	10.383
Equal opportunities officer job model: Main-job model	22	4.119	58	6.980	80	6.114
Time-off model	77	4.730	104	6.104	181	5.520

Note: * $p < 0.05$.

Table 5. Linear Regression: Changes in the Proportion of Women Professors between 2007 and 2015.

HEIs	Model 1		Model 2	
	Regression Coefficient and Standard Deviation	Standard Coefficient Beta	Regression Coefficient and Standard Deviation	Standard Coefficient Beta
(Constant)	6.963 *** (1.178)		7.002 *** (1.209)	
Participation in the program	1.268 (0.773)	0.105	1.273 (0.853)	0.105
Type (HEI)	0.998 (0.776)	0.083	0.940 (0.793)	0.078
Region (eastern Germany)	−2.108 * (0.941)	−0.136	−2.102 * (0.958)	−0.134
Equal opportunities officer job model (time-off model)	−0.531 (0.795)	−0.042	−0.526 (0.804)	−0.041
Proportion of women professors in 2007	−0.092 * (0.033)	−0.176	−0.093 * (0.033)	−0.178
Certified as a family-friendly HEI			−0.174 (0.794)	−0.015
TOTAL E-QUALITY (TEQ) award (received)			0.465 (1.029)	0.029
R^2	0.084		0.085	
R^2_{korr}	0.066		0.059	
N	260		259	

Note: The dependent variable was the change in the proportion of women professors between 2007 and 2015. * $p < 0.05$, *** $p < 0.001$.

---

[10] The significances refer to the comparison of participating and non-participating HEIs for the respective subgroups.

The regression analysis thus showed, on the one hand, that the Women Professors Program had an independent effect on changes in the proportion of women professors. On the other hand, the program did not have the strongest influence compared to other variables included in the model. Overall, the variables examined explained only a small proportion of the variance in the changes in the proportion of women professors that existed between the HEIs.

In this study, the regression analysis did not aim to explain changes in the proportion of women professors, but rather to estimate the impact of the Women Professors Program on these changes. The independent influence of the Women Professors Program was confirmed by the regression analysis. Both the regression analysis and the subgroup analysis (Table 4) revealed the particular significance of the program for the eastern German HEIs that participated in the Women Professors Program: the proportion of women professors increased strongly compared to the non-participating HEIs, and the effect of the Women Professors Program was particularly high in this region.

5.2.1. Contextual Factors

Contextual factors that foster or hinder gender equality could not be operationalized in such a way that they could be integrated into the analysis. Data on employment conditions differentiated by individual HEIs were not available. Moreover, as German HEIs typically do not recruit professors from within their own institutions, the employment conditions at a particular HEI—for example, the proportion of temporary or part-time academic staff—cannot be linked to the proportion of women professors at the institution.

The evaluation of reports on the implementation of the German Research Foundation's (DFG) Research-Oriented Standards on Gender Equality (German Research Foundation (Deutsche Forschungsgemeinschaft DFG)) could be considered to be an operationalization of beneficial contextual factors—at least for HEIs with the right to confer doctorates (i.e., universities). In fact, there were correlations between these assessments and participation in the Women Professors Program: in both program phases, universities participating in the Women Professors Program were rated higher on average in the DFG assessment.[11] However, as the reports assess the implementation of the Standards on Gender Equality, these assessments constitute an outcome variable and cannot be included in the regression analysis as a factor influencing the increase in the proportion of women professors.

The TOTAL E-QUALITY (TEQ) award and certification as a "family-friendly higher education institution" (familiengerechte Hochschule) after successfully undergoing the "Audit familiengerechte Hochschule" offer limited possibilities to assess the influence of other gender equality policies.[12] These two certificates were included in a second model in the regression analysis (see Table 5). The analysis showed that receipt of the TEQ award or certification as a family-friendly HEI had only a minor influence on the increase in the proportion of women professors.

5.2.2. Mechanisms

With the aim of understanding the mechanisms driving the effects of the program, I examined which program elements produced the described effects. According to the program theory (Figure 1) an increase in the proportion of women professors can be achieved both through the professorships supported by the start-up financing and through the implementation of a gender action plan. A first approach to understanding the mechanisms of the program is to look at the start-up funding. A high

---

[11] The DFG evaluated the universities according to four stages (Stage 1: first steps of implementation; Stage 4: successfully implemented concept). Comparing the treatment and control groups, the means of the stages were as follows: PP I: 3.17 vs. 2.64; PP II: 3.10 vs. 2.67. Note: Only six universities for which DFG assessment results were available did not participate in the Women Professors Program.

[12] For more information, see the English-language web pages of the TEQ award (https://www.total-e-quality.de) and the family-friendly HEI audit (https://berufundfamilie.de/english-info/audit-familiengerechte-hochschule).

influence of this part of the program may be assumed if the proportion of women professors at a HEI increases with the number of funded professorships.

For both program phases, there was a weak to medium correlation ($r = 0.285$) between the number of funded professorships and the increase in the proportion of women professors. However, the effect was apparent only for regular professorships ($r = 0.263$); no correlation, and thus no effect ($r = 0.079$) was observed for so-called "early appointments to a professorship" (*Vorgriffsprofessuren*). The lack of correlation in the case of these professorships may be due to the lower number of HEIs using this type of appointment (79 HEI vs. 129 HEIs for regular professorships). Furthermore, more than half of the HEIs that appointed women to regular professorships used solely this type of funding, whereas this was the case for only a quarter of the HEIs that appointed women to additional professorships. My main argument is, that this finding suggests that the promotion of appointments has both direct incentive and indirect effects via the financing of gender equality measures and the implementation of the gender action plan. With the available data, it is not possible to weight the two effects. Although the above correlations do not establish causation, they suggest an effect of the start-up funding on the increase in the proportion of women professors.

In the present quantitative part of the evaluation, it was not possible to analyze how exactly the funding works in the HEIs. However, the qualitative part of the evaluation provided some starting points for future research (Löther and Glanz 2017): the appointment procedures followed the usual rules. Only after completion of the appointment procedure and only where the recruitment conformed to the program rules, did the HEIs decide to apply for funding under the Women Professors Program. The funding does not make the HEIs more competitive to recruit women, because it cannot be used to provide higher remuneration or better equipment to the recruited professor. Nevertheless, the recruitment committees might feel some pressure or incentive to hire women in order to obtain program funding. Concerning the indirect effects, the evaluations reveal discursive effects on gender awareness. Moreover, the program strengthens the university management's responsibility for equality. Additionally, many gender actions plans provide for measures for transparent and gender-oriented recruitment procedures. These are ways in which the program might have affected the recruitment of women professors. To obtain an in-depth understanding of how the program works in the HEIs, qualitative data on recruitment procedures and other decision-making processes are needed.

5.3. Real and Expected Proportion of Women Professors

The impact of the program at the level of the higher education system was estimated by means of a time series analysis (see Rossi et al. 2004, pp. 291ff. on time series analysis in evaluations), comparing the expected and the actual proportion of women professors. The expected proportion of women professors is the counterfactual situation that would have existed in the absence of the Women Professors Program.

The time series analysis was based on data for the "professorships" category in higher education statistics provided by the Federal Statistical Office (Destatis).[13] As the first appointment of women to tenured professorships is funded under the Women Professors Program, the proportion of women professors was calculated without non-tenured junior professorships.[14] Whereas the analysis at the level of individual HEIs included only the members of the German Rectors' Conference, data for the time series analysis included the total number of professors at all German state, private and church-run HEIs (universities, universities of applied sciences, colleges of art and music). The proportion of women professors in the two samples did not differ.

---

[13] Available online: https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PersonalHochschulen.html. The "professors" category includes the salary groups C2–C4 and W1–W3-professors as well as full-time visiting professors.

[14] Junior professorships have existed in Germany since 2002.

In order to determine the estimator for the expected proportion of women professors, the average annual increase before the start of the program was extrapolated to the program duration. The year 1985 was chosen as the starting point for the time series because, in that year, an equality mandate ("to eliminate the disadvantages existing for women scientists") was included in the German Higher Education Framework Act. Since then, the promotion of gender equality has been a task of HEIs. Furthermore, 1985 was the year in which the first women's representatives were appointed at German HEIs, namely at the Universities of Hamburg, Kassel, and Oldenburg (Blome et al. 2014, p. 97). The proportion of women in permanent professorships rose from 5.1% in 1985 to 16.0% in 2007 (see Figure 2). The average annual increase was 0.47 percentage points. To calculate the expected proportion of women professors, this value was multiplied by the number of years since the start of the program and added to the initial value in 2007. This resulted in an expected proportion of women professors of 19.7% for the year 2015. Thus, the actual proportion of women professors—22.1%—was 2.3 percentage points[15] higher than expected; the average annual was 0.29 percentage points.
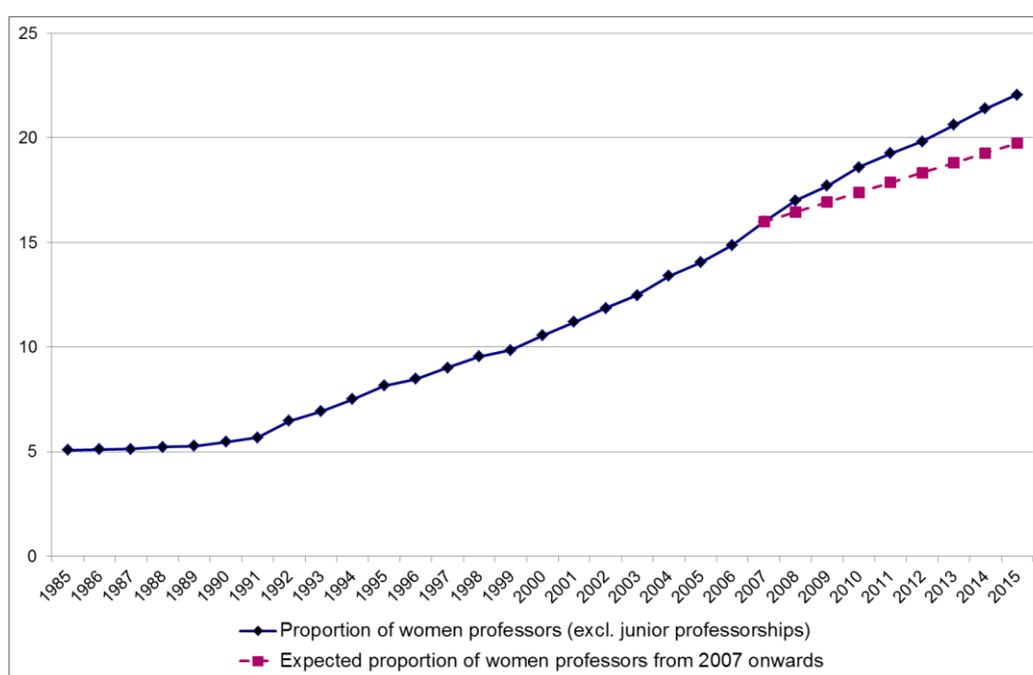
Figure 2. Proportion of Women Professors (Excl. Junior Professorships) 1985–2015; Expected Proportion of Women Professors from 2007 Onwards.

Retirements could be an interfering factor in these calculations. On average, women professors are younger than men professors and thus, the proportion of men professors is higher among retirees. However, this pattern did not affect the time series analysis: retirements of professors peaked in 2005/2006 and dropped dramatically after 2009. As a consequence, one would rather expect a higher increase in the proportion of women professors in the years before the program—with high rates of retirements. In fact, retirements did not disturb the time series analysis in such a way that the impact of the program was lower.

The effects estimated by means of time series analysis are not attributable solely to the Women Professors program. Simultaneous gender equality policies, in particular the DFG's Research-Oriented Gender Equality Standards, may also have contributed to this change. Because the various initiatives

---

15   The fact that the figure was 2.3% rather than 2.4% is due to rounding differences.

and programs were implemented during the same time period, their effects cannot be differentiated in the time series analysis.

Contextual factors that may run counter to the objectives of the Women Professors Program could not be integrated into this analysis either. During the period under review, the proportion of temporary and part-time scientific staff below the level of professor increased, and the ratio of early career researchers to professorships deteriorated (Konsortium Bundesbericht Wissenschaftlicher Nachwuchs 2017). Employment conditions and increasing competition are among the reasons why more women than men end their scientific careers after obtaining a doctorate (Funken et al. 2015). It may be assumed that the proportion of women professors would have increased more had these conditions not existed. However, these contextual factors could not be quantified for the time series analysis.

6. Discussion and Conclusions

By using long time series data on academic staff differentiated by individual HEIs, it was possible to carry out an impact evaluation of the German Women Professors Program employing a quasi-experimental research design. At the level of individual HEIs, I calculated changes in the proportion of women professors at participating and non-participating HEIs before and during the second program phase, thus using pre- and post-treatment measurements and difference in differences to estimate the counterfactual. The analysis showed that participating HEIs were able to increase the proportion of women professors by an average of 6.4 percentage points between 2007 and 2015, whereas the figure for non-participating HEIs was 4.6 percentage points. Thus, an effect of 1.8 percentage points between 2007 and 2015 and an annual average of 0.22 percentage points in increasing the proportion of women professors can be attributed to the Women Professors Program. Observable selection effects such as type of institution and the region did not influence this result. At the level of the German higher education system as a whole, I estimated the counterfactual by means of time series analysis. This analysis showed that the proportion of women professors rose more strongly than would have been expected based on the annual growth rate during the years preceding the introduction of the program. Concerning my first research question, the impact evaluation demonstrated a causal effect of the Women Professors Program on the increase in the proportion of women professors.

One could argue that an effect of 0.22 percentage points (comparison of participating and non-participating HEIs) and 0.28 percentage points (time series analysis) as an annual average is small compared to the amount of money spent on the program. However, in relation to the average annual increase of 0.76 percentage points since the start of the program, the effect of the program is about one third of the regular increase and thus quite high. Furthermore, the program aims not only to increase the percentage of women professors but also to foster structural changes in favour of gender equality. This study focused on the quantitative effect of the program, but qualitative aspects should be taken into consideration when measuring change towards gender equality (Peterson 2011; Neale and Özkanlı 2010).

Facilitating and impeding contextual factors could not be satisfactorily included in the analysis. Their influence thus remains an open research question and the integration of such factors into the quantitative analysis is a research desideratum. Nevertheless, interactions can be described between the Women Professors Program, the DFG's Research-Oriented Standards on Gender Equality, and the gender equality standards of the German "Excellence Initiative" (see Heidler 2017; Schacherl et al. 2015). Combining the results of the impact evaluation of the Women Professors Program with those of other studies reveals distinct mechanisms of the different initiatives. First, the way in which the initiatives foster gender equality depends on the context, in particular, the type of HEI. University managers attribute the greatest influence on their gender equality policies to the DFG's Standards on Gender Equality, followed by the Women Professors Program and the gender equality standards of the Excellence Initiative. Universities of applied sciences are neither members of the DFG nor are they eligible for funding under the Excellence Initiative. This is why managers of this type of HEIs consider the Women Professors Program, in particular, to be a strong driving force (Schacherl et al. 2015, p. 177).

Furthermore, the program theories of the aforementioned initiatives differ. The Research-Oriented Standards on Gender Equality link gender equality to the award of research funding by the DFG—Germany's leading funding organization. The members of the DFG, the universities, comply with the standards on a voluntary basis. However, the required reporting on gender equality increases the pressure on HEIs that apply for funding. As an effect of the standards, "university administrators and other key players regard successful gender equality policy as being linked to the prestige and competitiveness of the university" and "gender equality has increasingly come to be regarded as a prerequisite for innovative and excellent research" (see Heidler et al. 2017; Heidler and Reichwein 2018, p. 8). The Women Professors Program, on the other hand, exerts an effect through financial incentives and (in the case of regular professorships) the freeing up of financial resources for gender equality measures. HEIs have described the (indirect) provision of financial resources for gender equality and awareness-raising and reflection on gender equality as the most frequent positive changes brought about by the program (Löther and Glanz 2017, p. 64). The present impact analysis shows that start-up funding for the appointment of new women professors has had both direct effects and indirect effects (via the funding of gender equality measures for regular professorships) on increasing the proportion of women professors.

Concerning the second research question, it must be stated that the integration of facilitating and impeding contextual factors into the quantitative analysis remains a matter for further research. Nevertheless, combining the impact evaluation with other studies and research designs revealed the way in which different gender equality initiatives contribute to the increase in the proportion of women professors. By basing the impact evaluation on the program theory, it was possible to assess some mechanisms that link the start-up funding for professorships, gender equality measures, and the increasing proportion of women professors. Future research is needed to deepen our knowledge of the mechanisms that drive the effects of gender equality programs in higher education. In particular, we want to know how exactly these programs work within the HEIs. This research should bear in mind that mechanisms might work differently under specific conditions, characterized, for example, by the institution type, region, and previous experiences in gender equality. Case studies, document analysis, and a survey have provided insights into the different ways in which the Women Professors Program has been adopted at HEIs: depending on the status of the gender equality policy at the institution, the program has been used flexibly to test new instruments and expand activities or to implement measures for the first time and establish gender equality structures (Löther and Glanz 2017).

Another challenge of the impact evaluation was the question as to how the two levels of impact—the increase in the proportion of women professors and structural changes—interact. To further investigate quantitative effects, structural changes, and mechanisms, one must start with an in-depth analysis of the program theory. This analysis enables a better understanding of how, and under what circumstances the program produces specific results in the HEIs (see Gates and Dyson 2017, p. 36; Haunberger and Baumgartner 2017). The impact evaluation presented here offers a basis for these investigations. The analysis shows that it is possible to estimate causal effects of a gender equality policy program using a quasi-experimental research design, thus realizing an evaluation approach that is only rarely applied in the policy field of gender equality in science and research.

Funding: The evaluation of the Women Professors Program was funded by the German Federal Ministry of Education and Research, grant number 01FP1501 (2015–2017).

Acknowledgments: I would like thank Angela Wroblewski and Christof E. Müller for their valuable feedback.

Conflicts of Interest: The author declares no conflict of interest. The funders described the design of the overall program evaluation in the terms of reference but had no role in the design of the impact study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript, or in the decision to publish the results.

References

Astbury, Brad, and Frans L. Leeuw. 2010. Unpacking Black Boxes. Mechanisms and Theory Building in Evaluation. *American Journal of Evaluation* 31: 363–81. [CrossRef]

Beaufaÿs, Sandra, and Andrea Löther. 2017. Exzellente Hasardeurinnen? Beschäftigungsbedingungen auf dem wissenschaftlichen Arbeitsmarkt, Geschlechterungleichheit und Exzellenzförderung. *WSI-Mitteilungen*, 346–55. Available online: https://www.boeckler.de/wsi-mitteilungen_110014_110024.htm (accessed on 12 November 2018).

Blome, Eva, Alexandra Erfmeier, Nina Gülcher, and Sandra Smykalla. 2014. *Handbuch zur Gleichstellungspolitik an Hochschulen. Von der Frauenförderung zum Diversity Management?* Wiesbaden: Springer VS.

Burde, Dana, and Leigh Linden. 2012. *The Effect of Village-Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan*. Cambridge: National Bureau of Economic Research.

Caffrey, Louise, David Wyatt, Nina Fudge, Helena Mattingley, Catherine Williamson, and Christopher McKevitt. 2016. Gender equity programmes in academic medicine. A realist evaluation approach to Athena SWAN processes. *BMJ Open* 6: 1. [CrossRef]

Caspari, Alexandra. 2009. 'Rigorose' Wirkungsevaluation-methodische und konzeptionelle Ansätze der Wirkungsmessung in der Entwicklungszusammenarbeit. *Zeitschrift für Evaluation* 8: 183–213.

Courtois, Aline, and Theresa O'Keefe. 2015. Precarity in the Ivory Cage: Neoliberalism and Casualisation of Work in the Irish Higher Education Sector. *Journal for Critical Education Policy Studies* 13: 43–66. Available online: http://www.jceps.com/wp-content/uploads/2015/06/13-1-3.pdf (accessed on 21 February 2019).

d'Agostino, Giorgio, Margherita Scarlato, and Francesca Capparucci. 2016. Evaluating CCTs from a Gender Perspective: The Impact of Chile Solidario on Women's Employment Prospect. *Journal of International Development* 28: 177–97.

Deutsche Forschungsgemeinschaft (DFG). 2013. Forschungsorientierte Gleichstellungsstandards der DFG. Bewertung der Abschlussberichte 2013-Gesamtübersicht. Available online: http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/chancengleichheit/abschlussberichte/gesamtuebersicht_abschlussberichte_2013.pdf (accessed on 22 February 2018).

European Commission. 2009. She Figures 2009. Statistics and Indicators on Gender Equality in Science. Luxembourg: Office for Official Publications of the European Communities. Available online: http://ec.europa.eu/research/science-society/document_library/pdf_06/she_figures_2009_en.pdf (accessed on 19 February 2019).

European Commission—DG Research. 2016. She Figures Handbook 2015. Available online: https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015_Handbook_final.pdf (accessed on 19 February 2019).

Fletcher, Gillian. 2015. *Addressing Gender in Impact Evaluation. What Should Be Considered?* London: Overseas Development Institute.

Funken, Christiane, Sinje Hörlin, and Jan-Christoph Rogge. 2013. Generation35plus. Aufstieg oder Ausstieg? Hochqualifizierte und Führungskräfte in Wirtschaft und Wissenschaft. Available online: http://www.mgs.tu-berlin.de/fileadmin/i62/mgs/Generation35plus_ebook.pdf (accessed on 22 February 2018).

Funken, Christiane, Sinje Hörlin, and Jan-Christoph Rogge. 2015. *Vertrackte Karrieren. Zum Wandel der Arbeitswelten in Wirtschaft und Wissenschaft*. Frankfurt: Campus.

Funnell, Sue C., and Patricia J. Rogers. 2011. *Purposeful Program Theory. Effective Use of Theories of Change and Logic Models*. San Francisco: Jossey-Bass.

Gates, Emily, and Lisa Dyson. 2017. Implications of the Changing Conversation About Causality for Evaluators. *American Journal of Evaluation* 38: 29–46. [CrossRef]

Gregory-Smith, Ian. 2015. The Impact of Athena SWAN in UK Medical Schools. Available online: https://www.sheffield.ac.uk/polopoly_fs/1.449704!/file/paper_2015010.pdf (accessed on 14 January 2019).

Grillos, Tara. 2018. Women's participation in environmental decision-making. Quasi-experimental evidence from northern Kenya. *World Development* 108: 115–30. [CrossRef]

Haunberger, Sigrid, and Edgar Baumgartner. 2017. Wirkungsevaluationen in der Sozialen Arbeit mittels Realistic Evaluation. empirische Anwendungen und methodische Herausforderungen. *Zeitschrift für Evaluation* 16: 121–47.

Heidler, Richard. 2016. Karrierewege im Emmy Noether-Programm und beim Heisenberg-Stipendium. Available online: http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/infobriefe/ib02_2016.pdf (accessed on 22 February 2018).

Heidler, Richard. 2017. Forschungsförderung und Karrierewege. Vergleichende Studie zu den DFG-Programmen zur Förderung der wissenschaftlichen Karriere. Available online: http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/studien/studie_karrierewege.pdf (accessed on 25 October 2017).

Heidler, Julia, and Eva Reichwein. 2018. The DFG Research-Oriented Standards on Gender Equality. Selected Findings from a Study on Implementation and Impact. Available online: http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/infobriefe/ib01_2018_en.pdf (accessed on 22 February 2018).

Heidler, Richard, Eva Reichwein, Nike Alkema, and Ursula von Gliscynski. 2017. Die Forschungsorientierten Gleichstellungsstandards der DFG. Umsetzung und Wirkungsweisen. Available online: http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/studien/studie_gleichstellungsstandards.pdf (accessed on 22 February 2018).

Höppel, Dagmar, ed. 2016. *Aufwind mit Mentoring. Wirksamkeit von Mentoring-Projekten zur Karriereförderungen von Frauen in der Wissenschaft*. Baden-Baden: Nomos.

Kalev, Alexandra, Frank Dobbin, and Erin L. Kelly. 2006. Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies. *American Sociological Review* 71: 589–617. [CrossRef]

Kaplpazidou Schmidt, Evanthia, Susanne Bührer, Martina Schraudner, Sybille Reidl, Jörg Müller, Rachel Palmen, Sigrid Haase, Ebbe K. Graversen, Florian Holzinger, Clemens Striebing, and et al. 2017. A Conceptual Evaluation Framework for Promoting Gender Equality in Research and Innovation. Toolbox I—A Synthesis Report. EFFORTI—Deliverable 3.3. Available online: https://www.efforti.eu/sites/default/files/2018-03/EFFORTI%20D3.3%20FINAL%20report%2027032018.pdf (accessed on 5 April 2018).

Khandker, Shahidur R., Gayatri B. Koolwal, and Hussain A. Samad. 2012. *Handbook on Impact Evaluation. Quantitative Methods and Practices*. Washington, DC: World Bank.

Konsortium Bundesbericht Wissenschaftlicher Nachwuchs. 2017. *Bundesbericht Wissenschaftlicher Nachwuchs 2017. Statistische Daten und Forschungsbefunde zu Promovierenden und Promovierten in Deutschland*. Bielefeld: Bertelsmann, Available online: https://www.bmbf.de/files/buwin_2017.pdf (accessed on 22 February 2018).

Legewie, Joscha. 2012. Die Schätzung von kausalen Effekten. Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64: 123–53. [CrossRef]

Löther, Andrea, and Sabrina Glanz. 2017. Evaluation des Professorinnenprogramms des Bundes und der Länder: Zweite Programmphase und Gesamtevaluation: Erweiterter Bericht. cews.publik 22. Available online: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-54112-9 (accessed on 4 April 2019).

Meyer, Wolfgang, Nicolá Reade, and Reinhard Stockmann. 2008. Rigorous Impact Evaluation. *Ökologisches Wirtschaften* 23: 23–24. Available online: http://www.oekologisches-wirtschaften.de/index.php/oew/article/download/601/601 (accessed on 4 April 2019). [CrossRef]

Morgan, Stephen Lawrence, and Christopher Winship. 2010. *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

Müller, Christoph Emanuel, and Maria Albrecht. 2016. The Future of Impact Evaluation is Rigorous and Theory-Driven. In *The Future of Evaluation. Global Trends, New Challenges, Shared Perspectives*. Edited by Reinhard Stockmann and Wolfgang Meyer. New York: Basingstoke, Hampshire: Palgrave Macmillan, pp. 283–93.

Munir, Fehmidah, Carolynne Mason, Hilary McDermott, John Morris, Barbara Bagilhole, and Mary Nevill. 2014. Advancing Women's Careers in Science, Technology, Engineering, Mathematics and Medicine. Evaluating the Effectiveness and Impact of the Athena SWAN Charter. Available online: https://www.ecu.ac.uk/wp-content/uploads/2014/07/advancing-women%E2%80%99s-careers-in-STEMM-evaluating-the-effectiveness-and-impact-of-the-Athena-SWAN-Charter-FINAL.pdf (accessed on 14 January 2019).

Neale, Jenny, and Ozlem Özkanlı. 2010. Organisational barriers for women in senior management. A comparison of Turkish and New Zealand universities AU—Neale, Jenny. *Gender and Education* 22: 547–63. [CrossRef]

Nikunen, Minna. 2014. The 'entrepreneurial university', family and gender. Changes and demands faced by fixed-term workers. *Gender and Education* 26: 119–34. [CrossRef]

Peterson, Helen. 2011. The gender mix policy—Addressing gender inequality in higher education management. *Journal of Higher Education Policy and Management* 33: 619–28. [CrossRef]

Pohn-Weidinger, Simon, and Karin Grasenick. 2011. Elita-Evaluierung der FWF Progamme Elise Richter und Hertha Firnberg. Endbericht. Available online: https://www.fwf.ac.at/fileadmin/files/Dokumente/Ueber_den_FWF/Publikationen/FWF-relevante_Publikationen/elita.pdf (accessed on 22 February 2018).

Riegraf, Birgit, and Lena Weber. 2013. Exzellenz und Geschlecht in der unternehmerischen Hochschule. In *Die unternehmerische Hochschule aus der Perspektive der Geschlechterforschung. Zwischen Aufbruch und Beharrung*. Edited by Kristina Binner, Bettina Kubicek, Anja Rozwandowicz and Lena Weber. Münster: Westfälisches Dampfboot, pp. 67–85.

Rogers, Patricia J. 2000. Program Theory. Not Whether Programs Work, But How They Work. In *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*. Edited by Daniel L. Stufflebeam, George F. Madaus and Thomas Kellaghan. Boston: Springer Science + Business Media, pp. 209–32.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. 2004. *Evaluation: A Systematic Approach*. Thousand Oaks, London and New Dehli: Sage Publications.

Rubin, Donald B. 1979. Discussion of "conditional independence in statistical theory", by A.P. Dawid. *Journal of the Royal Statistical Society* 41: 27–28.

Schacherl, Ingrid, Melanie Roski, Maresa Feldmann, and Birgit Erbe. 2015. *Hochschule verändern. Gleichstellungspolitische Innovationen im Hochschulreformprozess*. Leverkusen: Barbara Budrich.

Simon, Dagmar. 2013. Organisation und Evaluation: Gleichstellungspolitische Optionen durch die "neue" Hochschule? In *Die unternehmerische Hochschule aus der Perspektive der Geschlechterforschung. Zwischen Aufbruch und Beharrung*. Edited by Kristina Binner, Bettina Kubicek, Anja Rozwandowicz and Lena Weber. Münster: Westfälisches Dampfboot, pp. 51–66.

Steiner, Peter M., Angela Wroblewski, and Thomas D. Cook. 2009. Randomized Experiments and Quasi-Experimental Designs in Educational Research. In *The SAGE International Handbook of Educational Evaluation*. Edited by Katherine E. Ryan and J. Bradley Cousins. Los Angeles: Sage, pp. 75–95.

Stepan-Norris, Judith, and Jasmine Kerrissey. 2016. Enhancing Gender Equity in Academia. Lessons from the ADVANCE Program. *Sociological Perspectives* 59: 225–45. [CrossRef]

Timmers, Tanya M., Tineke M. Willemsen, and Kea G. Tijdens. 2010. Gender diversity policies in universities: A multi-perspective framework of policy measures. *Higher Education* 59: 719–35. [CrossRef]

Weber, Lena. 2017. *Die unternehmerische Universität. Chancen und Risiken für Gleichstellungspolitiken in Deutschland, Großbritannien und Schweden*. Weinheim: Beltz Juventa.

Williamson, Charmaine. 2016. 'Views from the nano edge'. Women on doctoral preparation programmes in selected African contexts. *Studies in Higher Education* 41: 859–73. [CrossRef]

Wolbring, Tobias. 2014. Kausalanalyse und Wirkungsevaluation. Potential Outcomes, Graphenmethodologie und ihre Anwendung am Beispiel der Bologna-Reform. *Zeitschrift für Evaluation* 13: 243–270.

World Bank. n.d. enGENDER IMPACT. A Gateway to Gender-Related Impact Evaluations. Available online: http://www.worldbank.org/content/dam/Worldbank/document/Gender/enGENDER%20IMPACT,%20finding,%20including,%20extracting%20data%20from%20IEs.pdf (accessed on 19 February 2019).

Wroblewski, Angela, Michaela Gindl, Andrea Leitner, Ada Pellert, and Birgit Woitech. 2007. *Wirkungsanalyse frauenfördernder Maßnahmen im bm: Bwk*. Wien: Verlag Österreich.

Zimmermann, Karin. 2012. Bericht zur Evaluation des "Professorinnenprogramm des Bundes und der Länder". Available online: http://www.hof.uni-halle.de/dateien/ab_6_2012.pdf (accessed on 6 March 2018).

Zippel, Kathrin S., Myra Marx Ferree, and Karin Zimmermann. 2016. Gender equality in German universities: Vernacularising the battle for the best brains. *Gender and Education* 28: 1–19. [CrossRef]