Article

Science, Technology, & Human Values 2020, Vol. 45(2) 212-234 © The Author(s) 2019 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/0162243919835031 journals.sagepub.com/home/sth



Europeanizing the Danish School through DOI:1 National Testing: Standardized Assessment Scales and the Anticipation of Risky Populations

Helene Ratner¹0

Abstract

This paper explores "the peopling of Europe through data practices" in relation to standardized testing of students in Denmark. Programme for International Student Assessment (PISA) is a central component of Danish and European education infrastructures. In Denmark, mediocre PISA results spurred the introduction of national testing. With inspiration from Michel Foucault's notion of biopolitics, this paper analyzes how complementary Danish national test assessment scales make up population objects and student subjects and how these scales are aligned with European and transnational standards. A norm scale, standardized against the European Credit Transfer System (ECTS) grading scale, enacts a population whose performance can be tracked over time. A criteria scale introduces categories describing skills and enacts a moving student subject whose

¹Danish School of Education, Aarhus University, Aarhus, Denmark

Corresponding Author:

Helene Ratner, Danish School of Education, Aarhus University, Tuborgvej 164, 2400 Copenhagen, Denmark. Email: helr@edu.au.dk progression can be tracked. This paper argues that the three assessment scales enact the student population as bound to the nation and as simultaneously constituted in relation to transnational European categories and imaginaries of competition. As part of this, this paper discusses how the national test and PISA are used to single out students of non-European background, anticipated to be low PISA achievers and nonparticipants in a European knowledge economy.

Keywords

national testing, biopolitics, test scales, data, education, PISA, anticipation, ethnicity

Introduction

Education has long been a political priority for the European Union (EU). Considered a central asset in the "new knowledge economy of Europe," the EU vests considerable interest in how member states equip the population of tomorrow with the right entrepreneurial competences to survive in an increasingly globalized and competitive economy (Lawn and Segerholm 2011, 15; Hamre et al. 2015). For the EU and its member states, it is the very health and competitiveness of its future workforce that is at stake. Imagining EU member states' students as a European collective relies on an emerging "spatial infrastructure around education data" (Lawn and Segerholm 2011, 16). While global in scope, the Organization for Economic Cooperation and Development's (OECD) education flagship Programme for International Student Assessment (PISA)¹ is a central component of this European data infrastructure (Grek 2009; Lawn and Segerholm 2011).

It has been suggested that PISA produces "new scalar relations from the local through provincial to national and global in what emerges as a *global panopticism* (...) [resulting in] complementarity between international and national testing regimes" (Lingard, Martino, and Rezai-Rashti 2013, 545, original emphasis; see also Gorur 2016). This has certainly been the case in Denmark where standardized national testing was introduced to the Danish primary and lower secondary school in wake of disappointing PISA results in 2000 (Danish Parliament 2006). A statistical model has been developed to use national test results for anticipating future PISA results (Damvad 2014a), and Denmark moreover uses PISA to single out and monitor risky "population objects" (Ruppert 2011), "students with immigration

background" who statistically have lower PISA scores than "students without immigrant background" (National Agency for Education and Quality 2017, 3).

This paper is interested in how large-scale standardized testing is constitutive of "a people of Europe."² How do standardized test scales generate European population objects and student subjects? To engage with this question, this paper draws its conceptual inspiration from Foucault's (2009) notion of biopolitics, which characterizes population management as a fostering of its potentials, for example, of a healthy workforce, through various power/knowledge techniques. Dean (1996) warns against "the danger of missing the particularity [of techniques of government]" (p. 48), which are "so much more than a means or an instrument of government" (p. 58). Paying heed to this call, this paper explores the statistical techniques of ranking, comparing, categorizing, and anticipating students' national test performance. Indeed, as this paper will argue, assessment scales and statistical classifications in large-scale standardized testing have consequences for how population objects and student subjects come into being.

Other researchers of standardized large-scale testing have explored it as a global "rescaling of the policy cycle" (Lingard, Martino, and Rezai-Rashti 2013) and the "perverse effects" of gaming the accountability regimes (Lingard and Sellar 2013). Sellar and Thompson (2016) further suggest that computer-adaptive testing marks a transition to Deleuze's sense of a "control society" that works by "the mechanism of prediction," arguing that this leads to a delimitation of learning (p. 494). Gorur (2011) has traced how objectivity is constructed "inside the PISA laboratory," a precondition for its global policy machine, through various negotiations and translations. This paper contributes to this corpus by investigating the specific assessment scales and probabilistic models that make it possible to know and govern the population-as a totality-and student-as an individual subject. Looking at how national test scales and anticipative models emerge over time further elicits how national testing as a technique of government is continually refined and made to relate with (trans)national data sets and standards. Eliciting these changes speaks to Gorur's (2011) suggestion that standardized tests are "ontologically variable" (p. 78). This paper demonstrates how these changing techniques of categorizing test results reorganize how population objects and student subjects come into being.

Empirically, this paper is based on documents and websites from the Ministry of Education, public debates in the media, and interviews with civil servants in the Ministry of Education. The aim was to generate knowledge about the statistical models and student classifications used in assessments scales, Danish PISA reports, and anticipative models (what are they and how do they "represent" the individual and the population) as well as local reasonings behind these models (why were these specific models chosen and how were they developed?).³

After this introduction, this paper presents a background section contouring the relationship between PISA, the EU, and the Danish national test. This is followed by a theory section that introduces the key concepts of population object and student subject as well as Foucault's notion of biopolitics. This paper then presents three analytical sections that bring to the fore the intersection between population objects, student subjects, and biopolitics in Danish large-scale testing. The first section explores how the national test's percentile scale and norm scale generate a population object that can be tracked over time, with the 2010 distribution of test results as a baseline. Aligned with the European Credit Transfer System (ECTS) grading scale, the norm scale's five categories imply an Europeanization in terms of how the lowest achievers are separated from those in the middle and the top of a population. These scales enact the student subject as a rank in a 2010 population, yet the student can neither be compared to its own population nor its own performance over time. The second section explores how a criteria scale, introduced in 2014, introduces categories describing skills, which both standardize the population object against PISA categories and enact it in relation to its distance or proximity to national achievement targets. Compared to the norm scale, this scale generates a moving student subject whose progression can be tracked. The third and final section explores how the Danish national test and PISA are used to enact students of ethnic minority as risky population objects through anticipatory techniques. Central to the biopolitics of Europeanizing "Danish" students, thus, are standardization and questions of singling out, knowing and acting on student populations anticipated to be nonparticipants in a European knowledge economy.

PISA and the Danish National Test

The OECD has famously promoted better skills in reading, mathematics, and science with reference to a "convergence" between the trajectory of global economic development and education (Tucker in Sellar and Lingard 2013, 718; see also Krejsler, Olsson, and Petersson 2014). Offering standardized measurements, PISA facilitates the generation, calculation, circulation, and comparison of data. This "[reconstitutes] the globe (...) as a single space of comparative and commensurate measurement of the

performance of school systems" (Lingard, Martino, and Rezai-Rashti 2013, 539), with the effect of "viewing of education as a global race with winners and losers" (Gorur 2016, 608). For example, in 2010, the Danish government launched the ambition that "Danish students are among the best in the world," operationalized as being among the "top five" in PISA (Danish Government 2010). International and national testing is thus a central component in the (global) government of student populations (Piattoeva 2015; Benavot and Tanner 2007; Gorur 2017). Large-scale testing, however, is not simply a technical and neutral measurement of students' proficiency. It shapes ideas about what students should learn and what counts as quality in teaching and student performance, and we might think of this as government "at a distance" (Krejsler, Olsson, and Petersson 2014; Miller and Rose 1990; Gorur 2016, Ratner and Gad 2018).

PISA is central in the EU's governance of education. Figuring as an indicator for "EU Targets for 2020 in Education," PISA feeds into policy reports benchmarking member states' ability to e.g. "reduce the share of 'low achievers'" (Directorate-General for Education and Culture 2016, 4). With significant funding of PISA and collaboration around data collection, Grek (2009) suggests that the EU actively uses PISA to build a "new European education space of competitiveness and cohesion" (p. 24). Testing student populations, thus, is ultimately about governing and improving a living population, to prepare it for an increasingly competitive and global knowledge economy.

PISA's influence spans wider than simply assembling and comparing student populations worldwide. As the contemporary global "golden standard" in the assessment of education systems, it establishes OECD as a central policy broker with the authority to frame education issues and define policy problems (Breakspear 2012; Gorur 2016). Through PISA, OECD has shaped national education systems with a view to "increase productivity and sustain economic growth" (Sellar and Lingard 2013, 722), materializing both new accountability regimes and a narrowing of curriculum (Gorur 2016). Europeanization through PISA varies considerably, however, as the individual state take-up of PISA is adjusted to national programs and politics (Grek 2009). Moreover, member states' data practices, bringing European students into numerical existence as a population object, are neither "homogeneous" nor "symmetrical" (Lawn and Segerholm 2011, 16). How Europeanization plays out is mediated differently. This paper examines one such example, exploring how PISA has generated (numerical) population objects and student subjects in Denmark.

Researchers and politicians still refer to the "PISA-shock" following the first Danish PISA participation in 2000. Middle-range results raised questions about the quality of education, given the well-funded Danish school. Danish professor of education Niels Egelund, who has chaired several Danish PISA investigations, characterizes PISA as a necessary wake-up call: "PISA (...) woke us from the national beauty sleep that lulled us in self-satisfaction. We believed—like so many other countries—that we had the best education system in the world. But we didn't" (Egelund in Pedersen and Mehlsen 2010, 7). An OECD review of the Danish school system taking place in the wake of PISA recommended to introduce an "objective evaluation of students" and "proficiency standards" to improve student learning (Ekholm et al. 2004, 129). The introduction of mandatory standardized national testing in 2010 (decided by Parliament in 2006) was a response to this recommendation (Egelund 2008).

The development and implementation of the Danish national test were marked by several delays, two major challenges being the development of reading test items and the computational infrastructure for conducting the tests digitally in schools (REVIEW-panelet 2007). Since its implementation, the national test has been subject to critiques regarding both unintended student experiences, practices of teaching to the test and too high levels of statistical uncertainty (Bundsgaard and Puck 2016; Kousholt 2016). While these discussions are important, especially given how test data circuit still more accountability trails, the student experience and correctness of the national test are not my concerns. Instead, I explore how its assessment scales and anticipative models generate European population objects and student subjects, materializing a biopolitical logic of government. The next section will elucidate how I do this.

Biopolitical Governance: The Generation of Population Objects and Student Subjects

Up until this point, I have used the terms "population object" and "student subject" to characterize how large-scale testing enacts Danish students. Although psychometrics is about measuring innate abilities, the starting point of this paper is that the test scales are performative in terms of *enacting* population objects and student subjects (cf. Ruppert 2011). Historically, the development of a statistical apparatus is entwined with the government of populations, and statistics has effected an "ontological status of people as calculable objects" (Sellar and Thompson 2016, 493; see also Desrosiéres 1998; Hacking 1990). Offering to understand the population in

terms of "the action of the norm," individuals are organized through "the regular occurrence of certain events" (Ewald 1990, 138, 142). As Foucault (1991) wrote, "The perspective of population, the reality accorded to specific phenomena of population, render possible the final elimination of the model of the family and the recentering of the notion of economy (...) [Statistics reveal how] population has its own regularities [and makes possible quantifying] (...) specific phenomena of population" (p. 99).

It is in relation to this understanding of population objects that I explore the production of student subjects. With reference to biopolitics, Ruppert (2011) writes, "the embodied individual is of interest to governments insofar as the individual can be identified, categorized and recognized as a member of population" (p. 218). Student subjects, thus, can only be understood in relation to population objects. With Foucault (1982), government, however, is not about domination of the subject but "a total structure of actions brought to bear upon possible actions (...) [it is] a way of acting upon an acting subject or acting subjects by virtue of their acting or being capable of action" (p. 789). National test scales, for example, at once generate "student subjects" while also making possible students' participation in test practices, a precondition for producing the data upon which governance relies. With a Foucauldian framework, thus, one does not view techniques such as national testing in opposition to the subject. Rather, "power relations and forms of domination are conceived as operating through 'modes of subjectification' as much as forms of objectification. The subject is held to be fabricated in relation to domains of truth rather than falsified in its essence" (Dean 1996, 53).

Biopolitics seeks to optimize the subject, not as part of a unified and homogeneous population but as "a series of [unfolding] elements" "that will have to be regulated within a multivalent and transformable framework" (Foucault 2009, 35). I am particularly interested in the biopolitical techniques of "discipline" and "security." They both take the multiplicity of a population as their starting point but engage with this differently. Discipline is concerned with the classification and disciplining of individuals:

Discipline ... [is a] particular way of managing and organizing a multiplicity, of fixing its points of implantation, its lateral or horizontal, vertical and pyramidal trajectories, its hierarchy, and so on. (...) Discipline is a mode of individualization of multiplicities rather than something that constructs an edifice of multiple elements on the basis of individuals who are worked on as, first of all, individuals. (Foucault 2009, 29)

In this view, standardized testing is a disciplinary technique for comparing and classifying students, materializing them as "bodies capable of performances" (Foucault 2009, 37, 51). "Security," in turn, is concerned with probability and temporality and works on the multiplicity as a "series of events or possible elements" (p. 35). Whereas discipline works on the population with an eye for productivity, security "foregrounds 'the population'...as a nonunified, complex and indeterminate object of government" (Villadsen and Wahlberg 2015, 7). However, rather than analyzing discipline and security as two separate mechanisms, Foucault (2009) emphasizes their entanglement:

With the establishment of these mechanisms of security there is a considerable activation and propagation of the disciplinary corpus. For in order actually to guarantee this security one has to appeal, to take just one example, to a whole series of techniques for the surveillance of individuals, the diagnosis of what they are, the classification of their mental structure, of their specific pathology, and so on; in short one has to appeal to a whole disciplinary series that proliferates under mechanisms of security and is necessary to make them work. (p. 29)

In that way, a biopolitical analytics conceives of data practices as "bound up with a particular ontology of the subject and governing logic" that are not only generative of certainty and knowledge but also of "uncertainty and instability in both how the subject is known and governed" (Ruppert 2012, 119–20).

While techniques of security and discipline differ in how they render populations knowable, this paper is concerned with statistical techniques of standardized testing. These statistically enact normal performances and their deviations, prescribe optimal performances, predict future outcomes, and the "bandwidth of the acceptable that must not be exceeded" (Foucault 2009, 21). Norms are important in that they are at once "detachable from the practices that are its support and able to be rendered into particular formulas or diagrams of rule which ramify through a range of different settings" (Dean 1996, 58). Through test scales (e.g., a rank), psychometric measurements orient administrative and pedagogical practices "toward conduct that takes the form of a strategic rationality concerned with the optimization of performance, aptitudes and states" (Dean 1996, 48).

This paper analyses three national test scales and anticipatory uses of PISA. These techniques reassemble the population by relating data elements to one another and categorizing them as "normal," "deviant," or

"risky" objects of government. Yet, as the analysis will also show, this takes place in different ways in the rather short Danish history of national testing, materializing different intersections of discipline and security.

Ranking and Tracking the Population through Percentiles and Norms

When the Danish national test was first implemented in 2010, the Ministry of Education chose a percentile scale and a norm scale for reporting students' test results. The scales, I argue, enact a population object with the purpose of tracking its progression over time. A former civil servant responsible for implementing the national test characterizes the choice of these scales in terms of their potential for making the abstract Rasch scale, used as a base in national testing, meaningful for teachers:

The task was to translate [the Rasch-scale] to something meaningful for teachers. Reporting the test result -0.5 logits on Rasch—a scale ranging from negative infinity to positive infinity where zero is defined as the difficulty of the median item in the item bank, which, by the way, is confidential—that won't make sense to anyone (...). However, if you report that -0.5 is equivalent to the 40-percentile, and that the 40-percentile equals ability just below the average, then it's meaningful, right. So we needed, like, to transform the results [in Rasch logits] to something we can communicate (...) and there we had the norm, the average. (Interview, former civil servant, January 2018)

A test result is reported as 67 percentiles in the norm category of "above average." The percentile scale ranks a student in a population between 1 and 100, and the norm scale simplifies the percentile scale by dividing the 100 values into five broad segments: "Well below average" (lowest ten), "Below average" (next 25), "Average" (middle 30), "Above average" (next 25), and "Well above average" (highest ten; Ministry of Education 2018c).⁴ These cut scores match the five passing grades (A, the highest 10 percent, to E, the lowest 10 percent) in the European Transfer and Accumulation System (ECTS), a grading scale developed by the European Commission to have a common measure across different EU countries. The very choice of norm scale and its cut scores thus indicates a Europeanization in terms of both the number of categories and how distinctions are drawn between, for example, "Below Average" (from 11-35 on the percentile, similar to D in the ECTS scale).

How do the two scales (norm and percentile) generate population objects and student subjects? Both scales are self-referential scales, scoring individuals according to their position within a group, which severs any relationship to "external defining factor[s]" (Ewald 1990, 146). However, rather than updating the two scales with each new national testing, as a conventional norm scale would do (to reflect the distribution of a contemporary population or the rank of a student within its own group), the first year of mandatory testing (2010) has become a baseline for tracking progression of the population over time (Danish Ministry of Education 2018c). This means that if a student scores thirty-three in the percentile scale, his or her performance is as good as the lowest performing 33 percent of students taking the test in 2010 and not in his or her own group.⁵ NordicMetrics, a company offering consultancy services in the Danish national test, explains the rationale this way: "it is an advantage that all results are comparable over time: A result in '43' in for example text comprehension represents the same skill in 2010 and 2017" (Nordic-Metrics 2017, 2). The student results of 2010 have thus become a scale of their own, against which subsequent populations are compared (Ministry of Education 2018b). This means that populations (second-grade students, fourth-grade students, etc.) can be compared over time. When generating population objects, the Ministry of Education has prioritized the possibility of tracking populations over time over updating the percentile and norm scales to new test results. The student subject differentiates itself and gains its particular value as a rank, either through the percentile number or the broader norm category that summarizes that percentile in terms of its proximity or distance from the middle values. The student's performance over time, however, cannot be compared as the student is ranked against different test populations (year 2010 students in second grade, third grade, etc.). This means that the test results cannot be used for longitudinal measurements of individual student, that is, for comparisons of the student's achievements over time.⁶

In conclusion, the percentile scale and norm scale enact a student subject as a position among students taking the test in 2010. Standardized against the ECTS scale's five cut scores (A–E), and maintaining the 2010 percentiles as baseline, in turn, generates a population object whose progression from earlier years is monitored. The norm scale, then, prioritizes a longitudinal measurement of the population, articulating its overall qualities in comparison to past performances.

Tracking the Individual Student through Criteria

In 2014, the Ministry of Education added a criteria scale to the national test as part of a new school reform, which introduced national achievement targets (Regeringen and Dansk Folkeparti 2013). The target that "at least 80 percent of all students are good in the national test" required the Ministry of Education to set criteria for the "good" performance, and a criteria scale with six categories resulted from this work. Currently, all scales (percentile, norm, and criteria) are being reported together.

The criteria scale introduced six hierarchical categories of proficiency ranging from "insufficient" to "excellent" (Danish Ministry of Education 2018c). Each category is specified through a description of the attainment level a student is expected to have at the time of testing. While the norm scale allows tracking the population over time, using the 2010 distribution of test results as a baseline, the criteria scale, with its definition of the "good" performance, allows tracking whether the population reaches the national target of 80 percent being "good" in reading and mathematics. This implies an overall change from monitoring the population in relation to its performance in 2010 toward monitoring the population, and holding institutions accountable, as more or less "good." The 2014 national targets further differentiate the population into two categories: "the share of best students," which is to increase, and "the share of students with poor results," which is to decrease (Regeringen and Dansk Folkeparti 2013). The latter target is similar to the EU target "to reduce the share of 'low achievers'," mentioned earlier. Thus, there is an overlap between EU and Denmark, in terms of the targets set for schools, operationalized through national test. This is one sense in which the criteria scale contributes to the making of a European student population, by replicating EU targets as national targets.

The criteria scale also materializes a new student subject. As mentioned, the norm scale enacts a student in relation to her position in the 2010 population with the implication that a student's test results (e.g., reading in second, fourth, sixth, and eighth grade) cannot be compared as the percentile scales refer to different populations. The criteria categories, in turn, act as constants across different tests and allow tracking a student's progression through comparison of test results. A statistician in the Ministry of Education describes it as a matter of enabling teachers to follow progression: "We would like teachers to use national tests to understand how much the student or the class progresses. And [the norm-based visualization] is not very good for that" (interview, January 2018). A feature in the Danish teachers' union's magazine *Folkeskolen* articulated the same concern using a distinction between rank and progression:

In the worst case, the norm scale is used to celebrate the highest achieving students at the cost of the lowest achieving students. (...) if you want numbers on progression, you need to test twice and look at the difference in the test results. However, this requires reporting test results on the same scale. (...) The challenge is that it is not possible to compare results from the different national reading tests as they are calculated on different [percentile] scales. (Wandall and Wåhlin 2014)

The possibility of comparing a student's performance over time generates a student subject that can be tracked across tests. Paraphrasing Sellar and Thompson (2016, 493), this "disaggregates the individual into patterns of behavior or performance" over time. Compared to the percentile and norm scales, which compare student populations, the criteria scale folds comparability into the individual student. To use the words of the feature referenced above, this implies a shift in focus from "high achieving schools" to "those who move the students the most, even if their achievements are low" (Wandall and Wåhlin 2014). The criteria scale prompts a new individuality where, in addition to high achievement, there is a focus on progression, a continuous improvement toward a better performance, regardless of one's rank. We might think of this as an "intensified individualization" as it is to focus teachers' attention on individual students' progression between tests (Ruppert 2012, 125). Here, the criteria scale becomes technique of subjectivation, using test results "to create and maintain a sense of progress" (Finn 2016, see also Ratner, Andersen, and Madsen 2019).

This enactment of the Danish student subject operates through a rearticulation of disciplinary techniques. Enacting the student subject as *in progression*, the criteria scale resembles Foucault's (2009) depiction of disciplinary normalization: "Disciplinary normalization consists first of all in positing a model, an optimal model that is constructed in terms of a certain result, and the operation of disciplinary normalization consists in trying to get people, movements, and actions to conform to this model" (p. 85). That progression is now measureable, and desirable, for *all* students, further speaks to the biopolitical logic where, rather than "eliminating the undesired... [we find] the more limited ambition of facilitating and optimizing the processes already inherent in this reality" (Villadsen and Wahlberg 2015, 9).

Europeanizing the Criteria Scale: National and Transnational Data Relationality

Whereas the norm scale was standardized against ECTS thresholds, the Ministry used PISA's criteria scale to standardize the new national test criteria. To develop the criteria scale, the Ministry of Education commissioned "item commissions," members with expertise in the test subjects, to propose a "translation model" between the test items and the six criteria categories. They chose thirty representative test items from each item database and criteria-categorized them with reference to Common Objectives (Fælles Mål), national targets outlining in text what students should learn in the different subjects (Ministry of Education 2018a). The development of the criteria-categories also made use of the existing distribution of test results as indicated by the percentile scale. The statistician characterized the division of labor between item commissions and statisticians as follows:

It was important to have a scale with a solid professional and proficiencybased anchoring as opposed to one that is purely data-driven. (...) we accepted many deviations from the existing distribution—then we would simply go "aha"—but we did use the percentile scale to check if the bar was too high or too low, if say, 80% of a population would be good in the translation model. (Interview, September 9, 2017)

Common Objectives, with qualitative descriptions of skills, offered a proficiency basis for criteria-relating test items. The percentile scale, in turn, was used to tame the qualitative estimations of test items, in order to place the bar neither "too high" nor "too low."

PISA was also used to standardize the criteria scale. As already noted, PISA holds great authority among Danish politicians and civil servants. A report (Rosdahl 2014) documenting "significant correlation between (...) [students' PISA] results and their chances for completing secondary education, obtaining employment and level of income" (Greve and Krassel 2017, 7) emerged as a point of reference in several interviews and documents. A Ministry of Education civil servant, for example, explained the wish to standardize the criteria scale with PISA as follows:

PISA definitely made a political agenda and (...) defined criteria for the insufficient and the excellent performance (...) we see a clear correlation between PISA performance and one's success in further life, in the education system and in terms of employment.... Of course, we were interested in

knowing: are the PISA categories random or are they compatible with our [criteria] categories? Do we share the same view, more or less, of scientific competencies? Obviously, we would be annoyed to have PISA-results telling us that 20% cannot read if our own system estimated that to be 10%.... We want... to work systematically with what constitutes "good" and "poor," when a skill is sufficient and so on. It is no good if... there are no standards for what "good" or "poor" really is. (Interview, June 5, 2015).

The wish to align the national tests with PISA indicates that calibrating European students, here in the case of Denmark, marks an early ambition to anticipate students' future education and employment by having similar categories of "good" and "poor." For this reason, the Ministry of Education commissioned the consultancy house Damvad (2014b) to clarify the correlation between the categories in national test and PISA. This was used to decide on the number and thresholds of criteria categories:

We had discussions of whether to have five, six or seven [criteria] categories. We used the probabilistic model to check the correlation between our six categories and those of PISA. If (\ldots) [the Damvad report] had given us a cloudy picture, say, that two of our categories were contained by one PISA category, we would have changed our scale to contain five categories. (Interview, September 17, 2017)

Through statistical analysis of two data sets (same students taking the national test and PISA), Damvad (2014a, 2014b) explored (and established) a correspondence between the national test and PISA scales, leading to the conclusion that the Ministry of Education could keep the six categories they had already related to the test items in cooperation with the item commissions. The criteria scale required relationality with both Danish qualitative targets for teaching, assisted by the percentile scale, *and* making the thresholds between categories correspond to those in PISA.

Students with Immigration Background as Risky Population Objects

In this final analytical section, I explore how the Ministry of Education uses PISA to single out ethnic minority students as a matter of concern. The ministry has oversampled students with a different ethnic background than Danish in PISA 2009, 2012, and 2015 (National Agency for Education and Quality 2017, 1). With PISA used as an indicator of future education and

employment status, and this part of the population scoring lower in PISA than its Danish peers, the rationale is to achieve "a more precise picture of the results for this group of students" (National Agency for Education and Quality 2017, 1). The resulting "PISA Ethnic" reports, as they are called, look for differentiated patterns in the performance of students with non-Danish ancestry such as the significance of whether they are immigrants or descendants of immigrants, their age of immigration, their country of origin, the (ethnic) composition of the school, and the school's overall performance (Greve and Krassel 2017). Here, anticipative techniques of security rely on disciplinary techniques of classification. Disaggregating performance "on the basis of gender, migration status (...) and other dimensions (...) [isolate] specific areas for intervention" and are thus performative in how populations are seen and governed (Gorur 2016, 164).

The Damvad (2014a, 2014b) report, used to PISA standardize national test's criteria categories, specifies this anticipatory mode of testing further through a probabilistic model, developed to generate relationality between national test data and PISA test data. Using test results from the *same* students taking both tests, *dislocated* in time with the national test taken three years before PISA, Damvad concluded that national test results are predictive of PISA. This probabilistic configuration of population objects, such as the PISA Ethnic reports, also generates racialized distinctions where non-Europeans are singled out as risky population elements.

The probabilistic model makes use of fictional examples to illustrate how it can be applied. We meet thirteen-year-old student "Aisha" of "non-Danish ancestry" who started school a year late:

In the national test in mathematics, Aisha achieved the following results in the criteria-based categories: *mediocre* in "numbers and algebra," *faulty* in "geometry," and *faulty* in "mathematics in use." (...) We can expect Aisha in two years' time will score around 290 in the PISA examination in mathematics. This is equivalent of the level "under 1" in PISA's system of categories. (Damvad 2014a, 18)

Accompanying this account is a table showing the digits to add or subtract to Aisha's predicted PISA results if she had different national test results or different "background factors" such as gender or ancestry. The choice of PISA level and ancestry is not coincidental. "Under 1" is the lowest category in PISA's scale of assessment, with level 2 being defined as "the lowest acceptable level of competencies... students with skills under level 2 are estimated to have insufficient... skills in relation to what they are expected to exercise in secondary education or in a job" (National Agency for Education and Quality 2017, 3). With this rationale, the national test predicts not only a student's PISA results but also anticipates his or her possibility of further education and employment.

In that sense, Damvad's choice of fictional example is somewhat stereotypical. Rather than being a question of inclusion/exclusion, however, the PISA Ethnic is an example of "biopolitical regulation . . . [that] is prepared to discover and operate upon diverse forms of life" (Villadsen and Wahlberg 2015, 112). PISA Ethnic discerns students with an immigration background as a differentiated and emerging multiplicity; indeed, a central aspect of the consecutive PISA Ethnic is distinguishing between different profiles within this population, in terms of correlations between performance and background "factors," and tracking to what extent schools manage to close the "gap" to students with no immigrant background.

Whereas the national test scales, analyzed in the previous sections, materializes student performance in relation to the regular distribution of the norm scale, the probabilistic model refracts future student PISA performance as a series of probable events, contingent upon national test performance. Working through the mechanism of security, this form of governance, according to Foucault (2009), "works on the future, that is to say, [x] will not be conceived or planned according to a static perception" (p. 35). Biopolitics here emerges as an anticipatory form of government that seeks to "reduce or eliminate bad outcomes ex ante" (Guston 2014, 224).

Damvad's predictive model enacts Aisha's future PISA results and invites imagining a different future if her national test results improve. Even if such speculation is outside the scope of the report (indeed, it needs to assume that the effect of teaching is constant in order to develop the probabilistic model), probabilistic analytics also bring about a preemptive form of population governance that spurs new pedagogical interventions (Williamson 2016). This population governance renders "high-risk populations subjects of (...) screening mechanisms (...) designed to optimize their performance according to a specific calculation of risk minimization" (Dean 1996, 62). Interventions aimed at ethnic minorities in Denmark range from training in the Danish language to economic incentives at the level of schools for raising student achievements. In the wake of the most recent PISA Ethnic, for instance, Danish Minister of Education Merete Riisager made fiscal rewards for the poorest performing 120 schools dependent on raising student achievement (Danish Government 2017). Identifying risky populations is thus also used to conduct the work of professionals by introducing economic incentives to make them optimize the potentials believed to exist in all parts of the population.

Making up Europe through PISA relationing of national tests becomes a matter of identifying and tracking those anticipated to achieve poorly in PISA. The relationship between different population objects is refracted through granting student groups considered to have little chance of future employment extra governmental scrutiny, in terms of probabilistic calculations and multivalent interventions.

Conclusion

In this paper, I have explored how large-scale testing materializes a biopolitical logic of government generating multiple and overlapping population objects and student subjects. I have analyzed the percentile, norm, and criteria scales as disciplinary techniques, and I have explored how elements of security emerge with probabilistic estimations of, especially, immigrant population objects. The scales differ in how they enact population objects and student subjects. The norm and percentile scales generate a moving population whose overall and distributed performance can be tracked over time. While a student subject is enacted as a rank, this position does not refer to its own group but rather to the 2010 baseline. The criteria-scale, in turn, enacts a student subject in progression that can be compared to its earlier performance. This enacts a student subject as a dynamic potentiality with a view to improvement, given the right pedagogical intervention. Finally, PISA is considered to anticipate future success in terms of employment and education status; ethnic profiling and an anticipative model further anticipate the student as part of a future (European) workforce, engendering relationality between national test results, population wellbeing in terms of socioeconomic norms, and a viable future economy.

The biopolitical impulse to single out and optimize diverse population objects enacts students of non-Danish ancestry as a matter of governmental concern. Probabilistic and diagnostic models and reports such as the Damvad model and PISA Ethnic stretch into a future where students with immigrant profiles call for extra pedagogical intervention if they are to contribute to the sustainability of future economies. This mode of governing populations is one that aims at inclusion: rather than excluding those "deviant" from the norm, a biopolitical logic takes its population object as dynamic, differentiated, and probabilistic. At the same time, it is a mode of inclusion where certain student profiles are included as risky elements. The intensified calculation and differentiation of students with immigrant background suggests that large-scale testing also enacts (racialized) groups for special treatment and improvement: the non-European "Other," embodied in the fictive example of "Aisha," which is not only a non-Danish but also a non-European name (cf. M'charek, Schramm, and Skinner 2014).

What does the case of large-scale testing of students in Denmark teach us about "The peopling of Europe through data practices?" The Danish student population is not automatically "part" of a European and wider transnational student population. Test data need to be produced, related, and categorized in standardized assessment scales. Indeed, we cannot speak of a delimited European student population; the EU, as well as OECD, is already part of and subject to transnational educational assemblages (Jones in Grek 2009, 32). The assessment scales analyzed in this paper bring about several distinctions: Denmark/other nation states (PISA), Denmark/EU (PISA in EU indicators), population past/present performance (national test percentile- and norm-scale), individual student past/present performance (national test criteria-scale), the Danish/non-Danish student population performance (PISA Ethnic), and national test result present/future PISA result (probabilistic model). These scales organize student populations in terms of both temporality (past/ present, present/future) and ethnicity (Danish/non-Danish). While contributing to the making of Danish, European, and transnational educational populations, standardized testing thus also enacts the European "Other," emerging as a population object in need of extra governmental attention. The people of Europe, in this case of Danish standardized testing, thus rely on multiple and divergent data practices that are at once both the very precondition for thinking and governing multiple and overlapping student populations and imagining their non-European Other.

Acknowledgments

I am grateful to my interlocutors in the Danish Ministry of Education for providing access to empirical material and commenting on the analysis, to Jeppe Bundsgaard, the editors, and reviewers of this special issue for providing helpful comments in improving this manuscript. Finally, I want to thank Jakob Wandall for his patience in teaching me about national test assessment scales.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Helene Ratner D https://orcid.org/0000-0002-0842-4049

Notes

- 1. Programme for International Student Assessment (PISA) takes place every third year and compares fifteen-year-old students in a growing number of countries (seventy-one countries participated in the 2012 measurement).
- 2. This, of course, does not mean that standardized testing is *only* configured as a part of Europe. The national objectives and introduction of the criteria-scale and its focus on progression, for instance, were inspired by school reforms in Ontario. PISA, while taken up by the European Union, is also used for Nordic comparisons and Organization for Economic Cooperation and Development's global comparisons (e.g., Greve and Krassel 2017). Thus, while Danish standardized testing contribute to making up "the people of Europe," they also figure as part of national, Nordic, and transnational educational spaces.
- 3. The empirical material has emerged as part of a qualitative study (August 2015 to January 2018) in the Agency for IT and Learning, part of the Ministry of Education.
- 4. When describing the norm scale, the Ministry of Education interchangeably uses the Danish terms "middel" and "gennemsnit," which I here translated as "average," which is not the same as the arithmetic mean value as the category "average" includes 30 percentiles.
- 5. The middle values, in all tests, have improved over the years, but the norm is not updated. For example, in the eighth-grade text comprehension test, the average has changed from the 50 percentile to the 61 percentile (Ministry of Education 2018b).
- 6. Although the national test was tendered with longitudinal measurements, the delays during its development meant that this option was dropped for the 2010 launch.

References

- Benavot, Aaron, and Erin Tanner. 2007. *The Growth of National Learning Assessment in the World*. Paris, France: UNESCO. Accessed January 5, 2017. http://unesdoc.unesco.org/images/0015/001555/155507e.pdf.
- Breakspear, Simon. 2012. "The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System

Performance." *OECD Education Working Papers* 71. Paris: OECD Directorate for Education.

- Bundsgaard, Jeppe, and Morten Rasmus Puck. 2016. "Nationale Test: Danske Lærere Og Skolelederes Brug, Holdninger Og Viden." Copenhagen: DPU, University of Aarhus and University College Lillebælt.
- Damvad. 2014a. PISA-Relatering Af de Kriteriebaserede Nationale Test. Delrapport 1—Formidling Af Resultater. Copenhagen, Denmark: Damvad.
- Damvad. 2014b. PISA-Relatering Af de Kriteriebaserede Nationale Test. Delrapport 2—Teknisk Rapport Og Dokumentation. Copenhagen, Denmark: Damvad.
- Danish Government. 2010. "Danmark 2020: Viden > Vækst > Velstand > Velfærd." The Prime Minister's Office. Accessed January 30, 2019. http://www.stm.dk/mu ltimedia/Danmark_2020_Viden_v_kst_velstand_velf_rd.pdf.
- Danish Government. 2017. "PISA Etnisk." Accessed October 5, 2017. https://www. regeringen.dk/nyheder/pisa-etnisk/.
- Danish Parliament. 2006. "Bekendtgørelse Af Lov Om Folkeskolen Nr. 1195 Af 30. November 2006." Copenhagen: Danish Parliament.
- Dean, Mitchell. 1996. "Putting the Technological into Government." *History of the Human Sciences* 9 (3): 47-68.
- Desrosiéres, Alain. 1998. The Politics of Large Numbers: A History of Statistical Reasoning. Cambridge, MA: Harvard University Press.
- Directorate-General for Education and Culture. 2016. PISA 2015: EU Performance and Initial Conclusions Regarding Education Policies in Europe. Brussels, Belgium: European Commission. Accessed October 9, 2018. https://ec.europa.eu/e ducation/sites/education/files/pisa-2015-eu-policy-note_en.pdf.
- Egelund, Niels. 2008. "The Value of International Comparative Studies of Achievement—A Danish Perspective." *Assessment in Education: Principles, Policy & Practice* 15 (3): 245-51.
- Ekholm, M., P. Mortimore, M. David-Evans, R. Laukkanen, and J. Valijarvi. 2004. OECD-Rapport Om Grundskolen i Danmark. Copenhagen, Denmark: Ministry of Education.
- Ewald, François. 1990. "Norms, Discipline, and the Law." *Representations* 30 (1): 138-61.
- Finn, Matt. 2016. "Atmospheres of Progress in a Data-based School." Cultural Geographies 23 (1): 29-49. doi: 10.1177/1474474015575473.
- Foucault, Michel. 1982. "The Subject and Power." Critical Inquiry 8 (4): 777-95.
- Foucault, Michel. 1991. "Politics and the Study of Discourse." In *The Foucault Effect: Studies in Governmentality*, edited by Graham Burchell, Colin Gordon, and Peter Miller, 53-72. Chicago, IL: The University of Chicago Press.
- Foucault, Michel. 2009. Security, Territory, Population: Lectures at the Collège de France 1977–1978. New York: Palgrave Macmillan.

- Gorur, Radhika. 2011. "ANT on the PISA Trail: Following the Statistical Pursuit of Certainty." *Educational Philosophy and Theory* 43 (sup1): 76-93. doi: 10.1111/j. 1469-5812.2009.00612.x.
- Gorur, Radhika. 2016. "Seeing Like PISA: A Cautionary Tale about the Performativity of International Assessments." *European Educational Research Journal* 15 (5): 598-616.
- Gorur, Radhika. 2017. "Statistics and Statecraft: Exploring the Potentials, Politics and Practices of International Educational Assessment." *Critical Studies in Education* 58 (3: The Potentials, Politics and Promises of International Large-scale Assessments): 261-65. doi: 10.1080/17508487.2017.1353271.
- Grek, Sotiria. 2009. "Governing by Numbers: The PISA 'Effect' in Europe." Journal of Education Policy 24 (1): 23-37. doi: 10.1080/02680930802412669.
- Greve, Jane, and Karl Fritjof Krassel. 2017. PISA Etnisk 2015. Copenhagen, Denmark: KORA. Accessed October 5, 2017. https://uvm.dk/internationalt-arbejde/i nternationale-undersoegelser/pisa/pisa-2015/pisa-etnisk-2015.
- Guston, David H. 2014. "Understanding 'Anticipatory Governance." Social Studies of Science 44 (2): 218-42. doi: 10.1177/0306312713508669.
- Hacking, Ian. 1990. The Taming of Chance. Ideas in Context. New York: Cambridge University Press.
- Hamre, Bjørn, Kristine Kousholt, Dorthe Staunæs, and John B. Krejsler. 2015. "Eleverne, Der Skulle Blive Så Dygtige, Som de Kunne: Kategorisering Og Potentialitet i Skolens Lærings- Og Testarrangementer." *Pædagogisk Psykologisk Tidsskrift* 52 (3): 21-43.
- Kousholt, Kristine. 2016. "Testing as Social Practice: Analysing Testing in Classes of Young Children from the Children's Perspective." *Theory & Psychology* 26 (3): 377-92. doi: 10.1177/0959354316641911.
- Krejsler, John Benedicto, Ulf Olsson, and Kenneth Petersson. 2014. "The Transnational Grip on Scandinavian Education Reforms." *Nordic Studies in Education* 34 (3): 172-86.
- Lawn, Martin, and Christina Segerholm. 2011. "Europe through Experts and Technologies." In *Fabricating Quality in Education: Data and Governance in Europe*, edited by Jenny Ozga, Peter Dahler-Larsen, Christina Segerholm, and Hannu Simola, 32-46. Abingdon, UK: Routledge.
- Lingard, Bob, Wayne Martino, and Goli Rezai-Rashti. 2013. "Testing Regimes, Accountabilities and Education Policy: Commensurate Global and National Developments." *Journal of Education Policy* 28 (5): 539-56. doi: 10.1080/ 02680939.2013.820042.
- Lingard, Bob, and Sam Sellar. 2013. "Catalyst Data': Perverse Systemic Effects of Audit and Accountability in Australian Schooling." *Journal of Education Policy* 28 (5): 634-56. doi: 10.1080/02680939.2012.758815.

- M'charek, Amade, Katharina Schramm, and David Skinner. 2014. "Topologies of Race: Doing Territory, Population and Identity in Europe." *Science, Technology, & Human Values* 39 (4): 468-87. doi: 10.1177/0162243913509493.
- Miller, Peter, and Nikolas Rose. 1990. "Governing Economic Life." *Economy and Society* 19 (1): 1-31. doi: 10.1080/03085149000000001.
- Ministry of Education. 2018a. Om Fælles Mål. Copenhagen, Denmark: Ministry of Education. Accessed August 13, 2018. https://uvm.dk/folkeskolen/fag-timetalog-overgange/faelles-maal/om-faelles-maal.
- Ministry of Education. 2018b. Den Nationale Præstationsprofil. Copenhagen, Denmark: Ministry of Education. Accessed July 13, 2018. https://uvm.dk/statistik/ grundskolen/karakterer-og-test/nationale-test/national-praestationsprofil.
- Ministry of Education. 2018c. Om Nationale Test. Copenhagen, Denmark: Ministry of Education. Accessed August 17, 2018. https://www.uvm.dk/folkeskolen/ele vplaner-nationale-test-og-trivselsmaaling/nationale-test/om-de-nationale-test.
- National Agency for Education and Quality. 2017. Hovedresultater Fra PISA Etnisk 2015. Copenhagen, Denmark: Danish Ministry of Education, Ministry of Education. Accessed October 5, 2017. https://uvm.dk/internationalt-arbejde/interna tionale-undersoegelser/pisa/2015/pisa-etnisk-2015.
- NordicMetrics. 2017. *National Reference for de Nationale Test. Notat*. Copenhagen, Denmark: NordicMetrics. Accessed September 24, 2018. http://www.nordicme trics.com/wp-content/uploads/2014/05/National-reference-for-de-nationale-te st_12122017.pdf.
- Pedersen, Peder-Holm, and Camilla Mehlsen. 2010. "Drømmen Om Den Perfekte Målestok." Asterisk 53 (juni): 6-9.
- Piattoeva, Nelli. 2015. "Elastic Numbers: National Examinations Data as a Technology of Government." *Journal of Education Policy* 30 (3): 316-34. doi: 10. 1080/02680939.2014.937830.
- Ratner, Helene, Bjarke Lindsø Andersen, and Simon Ryberg Madsen. 2019. "Configuring the Teacher as Data User: Public-private Sector Mediations of National Test Data." *Learning, Media and Technology* 44 (1): 22-35. doi: 10. 1080/17439884.2018.1556218.
- Ratner, Helene, and Christopher Gad. 2018. "Data Warehousing Organization: Infrastructural Experimentation with Educational Governance." *Organization*, October. doi: 10.1177/1350508418808233.
- Regeringen, Venstre, and Dansk Folkeparti. 2013. "Aftale Mellem Regeringen (Socialdemokraterne, Radikale Venstre Og Socialistisk Folkeparti), Venstre Og Dansk Folkeparti Om et Fagligt Løft Af Folkeskolen." Copenhagen: Danish Parliament.
- REVIEW-panelet. 2007. *De Nationale It-Baserede Test i Folkeskolen*. Copenhagen, Denmark: REVIEW-panelet.

- Rosdahl, Anders. 2014. "fra 15 år til 27 år: Pisa 2000-eleverne i 2011/12." *SFI—Det nationale forskningscenter for velfærd*. Accessed October 5, 2017. https://uvm. dk/-/media/filer/uvm/udd/folke/pdf14/jun/140617-fra-15-27-aar.pdf.
- Ruppert, Evelyn. 2011. "Population Objects: Interpassive Subjects." Sociology 45 (2): 218-33. doi: 10.1177/0038038510394027.
- Ruppert, Evelyn. 2012. "The Governmental Topologies of Database Devices." *Theory, Culture & Society* 29 (4–5): 116-36. doi: 10.1177/0263276412439428.
- Sellar, Sam, and Bob Lingard. 2013. "The OECD and Global Governance in Education." *Journal of Education Policy* 28 (5): 710-25. doi: 10.1080/02680939. 2013.779791.
- Sellar, Sam, and Greg Thompson. 2016. "The Becoming-statistic: Information Ontologies and Computerized Adaptive Testing in Education." *Cultural Studies* ↔ *Critical Methodologies* 16 (5): 491-501. doi: 10.1177/1532708616655770.
- Villadsen, Kaspar, and Ayo Wahlberg. 2015. "The Government of Life: Managing Populations, Health and Scarcity." *Economy and Society* 44 (1): 1-17. doi: 10. 1080/03085147.2014.983831.
- Wandall, Jakob, and Martin Wåhlin. 2014. "De Nationale Test—Percentiler Og Progression." Folkeskolen (iss. januar). January. Accessed September 20, 2018. https://backend.folkeskolen.dk/~/4/1/kronik-nationale-test-2.pdf.
- Williamson, Ben. 2016. "Digital Education Governance: Data Visualization, Predictive Analytics, and 'Real-time' Policy Instruments." *Journal of Education Policy* 31 (2): 123-41. doi: 10.1080/02680939.2015.1035758.

Author Biography

Helene Ratner is an associate professor at Danish School of education. In her research, she explores intersections between educational governance and data. She is currently researching datafication of educational governance, drawing on ethnography and science and technology studies.