What PISA Intends To and Can Possibly Achieve: a critical programme theory analysis

Abstract

This paper advances the enlightened discussion of the nature, logic, and possible effects of the Programme for International Student Assessment (PISA). The purpose is to analyze the assumptions regarding how PISA is to achieve its intended effects, that is, to reconstruct PISA’s programme theory (PT) and to probe the validity of its underlying assumptions. The paper demonstrates that PISA’s PT has low internal validity. However, some PISA assumptions are consistent, for example, the assumption that legitimization activities justify PISA as a transnational benchmarking system measuring education system performance. PISA exemplifies systemic evaluation governance: all actors in the field are expected to use PISA results to react to and reflect on their own practice, compare themselves with others, and then act accordingly to improve education systems and school practice, though no activities or resources are allocated to change school practice. There is no empirical research into how systemic evaluation governance works in practice that can be used to probe PISA’s external validity. PISA’s PT is in line with discourse policy, governance theory, and school effectiveness research, but whether and under what conditions and how PISA helps change education systems and school practice are empirical questions waiting to be answered.

Introduction

In 2000, when the first Programme for International Student Assessment (PISA) test of 15-year-olds’ reading skills was launched by the Organisation for Economic Cooperation and Development (OECD), it marked the start of an OECD-led discourse on the quality of national education systems. Test scores regarding students’ knowledge and their ability to reflect on and apply their knowledge and experience to real-world issues are used to measure the quality of national education systems (OECD, n.d.). PISA is the key pillar in the production of knowledge used to shape policy for steering educational systems (Carvalho, 2012; OECD, 2009a). In 2009, PISA provided the following knowledge:

- A profile of knowledge and skills among 15-year-olds in 2009, consisting of a detailed profile for reading, including digital literacy, and an update for mathematics and science.
- Contextual indicators relating performance results to student and school characteristics.
- An assessment of students’ engagement in reading activities, and their knowledge and use of different learning strategies.
- A knowledge base for policy research and analysis.
- Trend data on changes in student knowledge and skills in reading, mathematics and science, on change in student attitudes and in socio-economic indicators, and also on the impact of some indicators on the performance results. (OECD/PISA, 2009a, p. 25; see also OECD/PISA 2009b)

PISA is conceived and analysed as a programme for the transnational governance of education systems (Grek, 2010, 2012; Lundgren, 2010; Popkewitz, 2011). The ‘PISA movement’, 15 years old by now, has become a global phenomenon permeating education discourse worldwide and has significant impact on policymaking, according to a recent OECD report (Breakspear, 2012).

PISA results are frequently discussed and debated in the policy world and among education researchers. While PISA supporters paint a bright picture of PISA and how it can bolster education in today’s globalized world, its critics draw attention to the negative consequences of PISA. Education has, thanks to PISA, moved away from the enlightenment ideal of promoting personal development and creating reflective and culturally aware citizens, towards an ideal of education in the interest of economic growth, promoting performativity, standardization, and decontextualization – according to some of its critics (cf. Carvalho, 2012; Lawn, 2011; Mangez and Hilgers, 2012). Advocates of PISA do not consider this shift negative. On the contrary, benchmarking education systems and testing the life skills needed in today’s world are claimed to be a great help, informing policy for education system development (Schleicher, 2013).

PISA was not the first effort to promote this shift in education and not the first international measurement system for comparing student knowledge and skills. According to Mangez and Hilgers (2012, p. 195), the first such initiative was taken in 1958 by the educational psychologist Benjamin S. Bloom. His initiative and the TIMSS and PIRLS international measurement systems developed by the International Association for the Evaluation of Educational Achievement (IEA) were largely driven by researchers. In contrast, OECD
developed PISA with support from member states and co-opted researchers into supporting OECD’s mission of developing education in the interest of economic growth. Several actors and bodies constitute PISA: the OECD Secretariat, PISA Governing Board, International Consortium, National Centre, Experts, Open Forum, and OECD Member countries. These actors/bodies have different roles and perform different activities in producing knowledge of education in the interest of policy (Turner, 2007).

When OECD/PISA entered the education discourse, external demands for useful and policy-relevant knowledge of education already existed, reflecting a general trend in the field of knowledge (Gibbons et al, 1994). At present, OECD/PISA possesses high authority and discursive power in the transnational coordination of educational governance and in the standardization of educational systems. The transnational shift in education governance reflects the growing influence of international agencies and ‘a marketization of the field of education in the context of an increasingly multilevel and fragmented arena for educational governance’ (Pereyra et al, 2011, p. 2).

According to Grek:

The PISA charts became the totemic representations of the new governing regime, excluding caveats or any awkward knowledge in order to offer policy makers what they are always after – fast-selling policy suggestions. (Grek, 2012, p. 244).

PISA has recently been assessed in a special issue of the European Educational Research Journal focusing on how PISA is ‘fabricated’, that is, constructed and reconstructed in and across different contexts, involving many actors and kinds of knowledge (Ozga, 2012, p. 166). OECD/PISA is perceived and assessed as a new form of educational governance that uses a ‘mix of policy technologies and constant work by policy actors to maintain connections and coherence in respatialised governing relations’ (Ozga, 2012, p. 167). Carvalho (2012) conceives the making of PISA as a

knowledge-policy instrument as being a part of multidirectional processes that involve reinterpretation, decontextualisation and recontextualisation, and where national, local, regional and international agencies intertwine. (p. 177)
PISA is also characterized as a set of complex, transnational, psychometric, and discursive policy instruments (or technologies) claiming to produce ‘fundamental truths’ about education (Mangez & Hilgers, 2012, p. 196). OECD/PISA links researchers and member countries to legitimize PISA tests as telling truths about education systems and to affirm OECD policy reports as authoritative transnational governance instruments. The assumptions of this governance instrument are those of psychometric research:

The knowledge used and generated by PISA surveys and reports is somewhat disciplined by the so-called ‘literacy theoretical framework’ and by particular assumptions, concepts and methods from the psychometric world. (Carvalho, 2012, p. 180)

Mangez and Hilgers’ (2012) field theory analysis of PISA offers insight into how the ‘field of power’ and the ‘economic pole’ have created demand for the particular policy knowledge that PISA provides. There is a constant struggle over the definition of what constitutes legitimate means of understanding and evaluating a field’s central activity, and PISA has succeeded in becoming ‘a legitimate instrument of measurement for defining the performances of education across different (national) policy fields’ (Mangez & Hilgers, 2012, p. 199). The reverse of this is that national education policy has lost its authority and power to independently define legitimate means of understanding and assessing education. The change also implies that national education and education policy are conceived through their outputs in relation to OECD objectives. The reception and significance of PISA varies between countries and social spaces because PISA is ‘reconstituted or reinterpreted to suit the requirements and the “logic” of that space [read: national education and education policy conditions]’ (Mangez & Hilgers, 2012, p. 199).

When PISA is being ‘consumed’, it is, indeed, also being transformed. And it is being transformed according to the logic of the social space where it is being ‘consumed’.
(Mangez & Hilgers, 2012, p. 199)

This special issue, and other analyses of PISA, has not clarified how PISA’s various activities and assumptions are linked at the transnational programme level, that is, how PISA’s intended effects are assumed to be achieved, and whether its assumptions are consistent. PISA can be described and analysed as a transnational programme implemented in member states, or as a
transnational programme being constructed while being implemented in member states. There is a need to consider PISA from both these perspectives. This article, however, looks into PISA as a transnational programme. Although PISA was partly developed at the OECD level and partly at the national level, the ‘fabrication of PISA’ in different member states (Carvalho, 2012; Ozga, 2012) should be kept apart from OECD policy making, that is, how OECD has set up PISA and what it intends to achieve with the programme. We argue that OECD’s intentions and policy action merit analysis in their own right.

This article aims to reconstruct PISA’s programme theory (PT), probe its underlying assumptions, and discuss how knowledge of PISA’s PT can be used in further research. Programme theory refers to the assumptions as to how a programme can achieve its intended effects and should not be confused with middle-range theories. Special attention is paid to whether, and to what extent, the PT is supported by (middle-range) theories of relevance for PISA’s PT, educational governance, accountability, and school development theories. PISA’s PT can be in accordance with or have more or less support in these theories which the forthcoming analysis of external validity will reveal. A further aim is to advance education policy and governance research, and to promote better-informed discussion of what PISA intends to and possibly can achieve. It is beyond the scope of this article, however, to explore PISA’s effects and consequences in practice. The main reference in this paper is Sweden.

This article demonstrates that some of PISA’s underlying assumptions are consistent: for example, that legitimization activities justify a benchmarking system that reliably measures education system performance, and that PISA activities promote a performance-oriented and decontextualized education discourse. However, some key assumptions are not consistent: for example, publishing PISA results and providing policy recommendations are assumed to induce local actors to take action to improve school practice. Except in the case of participating schools, PISA does not provide accounts of how local schools perform, indicating a gap between PISA activities and what is assumed to happen at the local level.

The paper contributes to the body of knowledge of PISA, which can be used in various ways. Reconstructing and probing PISA’s PT has a value in its own right, and this paper provides a
valid and succinct description of PISA that can advance informed discussion among advocates and critics as to its nature, logic, and possible achievements.

Next the paper describes the methodology and material used. The following section consists of a reconstruction of PISA’s PT which lays the ground for the analysis of PISA’s PT. The analysis of the PT’s internal validity that follows focuses on the logic and consistence of the assumptions of PISA. The external validity is assessed against theories of governance, accountability and school development. If PISA has high validity, that is, if PISA’s PT is consistent and supported by these theories, PISA is more likely to achieve its intended effects.

The discussion section of the paper contains reflections on problems with PISA, the contribution of the framework to the field and the need for further research. The paper ends with conclusions about the consistency and validity of PISA’s PT.

Framework and methodology

The framework and PT-analysis developed in this article integrates knowledge from three research fields, i.e. policy, governance, and evaluation research. A combination of factors and conditions (including context, institutions, policy instruments, resources, implementation, and legitimacy) affect the conception, effects and consequences of PISA. The framework pays attention to different modes of governance, organizations, actors and functions of evaluation in governance. It conceives PISA as a programme for transnational governance and an evaluation system producing evaluative knowledge for changing national education policy in line with OECD’s objectives. It also pays attention to that evaluation and the evaluation system per se could influence governance. In this article PISA refers to the transnational programme, its components including the evaluation system and its products (e.g. the PISA 2009 test). Knowledge of governance, accountability and school development is also, as mentioned, used in the analysis of the external validity of PISA’s PT.
Figure 1 depicts OECD’s steering of PISA and the interplay between (transnational) governance and PISA (as evaluation system). Using this conceptual framework, OECD has developed PISA as a programme and evaluation system to meet OECD’s knowledge needs for transnational education governance. OECD steers the programme in various ways and filters PISA results and trend data (type I evaluation) through policy reports and synthesis reports (type III evaluation), and provides a knowledge base which integrates knowledge from stand-alone evaluation studies (type II evaluation) to achieve its intentions. The framework, further developed elsewhere (Hanberger, 2012), facilitates the exploration of governance and evaluation systems and also accounts for the unintended effects and functions of evaluation systems.

Insert figure 1

The basic assumptions of this article are that PISA needs ongoing legitimization and justification among policy actors, researchers, and practitioners and that OECD/PISA not only constructs what are conceived as today’s education problems, to which PISA is the solution, but that PISA may be or become part of the problem.

Programme theory analysis

The programme theory analysis (PT analysis) methodology applied here was developed within the research project ‘Consequences of evaluation for school practice’, financed by the Swedish Research Council (grant no. B0615701). Programme theory is a well-established concept used in evaluation research referring to the assumptions as to how a programme achieves its intended effects. There are various approaches to reconstructing and articulating a PT. This paper adopts a policy-scientific approach (Leeuw, 2003) and the PT analysis is inspired by the work of Dahler-Larsen (2012) and Funnel and Rogers (2011).

The PT analysis presented here includes three main steps: reconstructing PISA’s PT; analysing the PT’s internal validity (i.e. the consistency of its assumptions); and analysing the PT’s external validity (i.e. whether it is supported by relevant research and provides feasible knowledge for resolving the problems it is intended to resolve). The PT approach seeks answers to the following questions:

Step 1: Reconstructing PISA’s programme theory
1. What are the assumptions of PISA?
2. What problem(s) was PISA intended to resolve?
3. What are the prerequisites for PISA, and what activities are assumed to produce what intended effects?
4. Are the problems that PISA is to address described and substantiated?
   a. Whose knowledge needs does PISA meet?
   b. Are arguments provided as to why PISA is needed?
5. Are the intended effects clearly described/specification?
   a. Are the intended short- and long-term effects defined?
   b. For whom and where are the effects expected to occur?

Step 2: Assessing the PT’s internal validity
6. Is PISA’s PT consistent?
   a. Is there a logical/coherent description of how programme activities are to achieve the intended effects?
   b. Are there activities for all intended effects?
   c. Are there activities that do not logically match their intended effects?

Step 3: Assessing the PT’s external validity
7. Does PISA have scientific support?
   a. Does PISA reflect/do justice to the phenomena/objects it is intended to measure?
   b. Is PISA’s PT consistent with existing knowledge of conditions/factors that improve teaching/the education system?
   c. Is the PT consistent with governance and accountability theories?
   d. Does the PISA evaluation system provide information about data sources, data quality, non-response, who reports information to the system, etc.?
8. Is the knowledge that PISA produces useful in helping resolve the problems that PISA was set up to address?

The PT was reconstructed by the author based on the material described below. The PT’s internal validity was assessed by the author by means of an argumentational analysis (Leeuw, 2003, p. 7) guided by question 6a–c. The PT’s external validity was assessed by the author by assessing the scientific support for PISA (Leeuw, 2003, p. 8), guided by questions 7–8. The
The author declares no potential conflicts of interest with respect to PISA.

**Material**

The PT analysis presented here is based on policy documents, research papers, PISA test results, OECD policy reports, and interviews with two policy actors involved in PISA, one operating at the OECD level and one at the national level. The latter actor is the civil servant at the Swedish National Agency for Education responsible for PISA in Sweden.

**PISA’s programme theory**

Two basic assumptions underpin PISA (Q1): first, policy actors in member countries need recurrent and reliable data on students’ life skills and knowledge that can be used to track change and for comparison between countries and, second, students’ literacy and their knowledge of mathematics and science capture the quality of an education system.

What problem(s) was PISA intended to resolve (Q2)? Before PISA was launched, OECD lacked reliable data with which to compare the quality and performance of various education systems. Today the problem is to maintain the high quality of PISA tests so that their results continue to be perceived as valid and trustworthy data with which to monitor national trends and benchmark quality in education systems. The challenge is to sustain the confidence in PISA tests and in OECD’s policy recommendations so that they continue to be perceived as authoritative and are used to improve the quality of education systems.

What are the prerequisites for PISA and what activities are assumed to produce what intended effects (Q3)? This question is answered by reconstructing PISA’s PT, a simplified version of which is presented below (a more detailed version of the PT appears in the appendix).

**If the following conditions are met:**

- PISA tests are high in quality, have scientific support, and are highly reliable

**And if the following activities are carried out:**

1. Member countries and academics are involved in running and using PISA
2. PISA tests and the policy documents linked to their results are published and disseminated
3. Conferences/meetings are arranged at which to discuss PISA results and the related success factors for education
4. OECD provides policy-relevant knowledge and recommendations for improving education systems

The following short- and long-terms effects should occur:

- (1–4) Legitimize PISA, what is measured and how, and the benchmarking system
- (1–4) Discuss the conditions and factors explaining countries’ PISA results and quality in education systems, and promote learning from high-performing education systems
- (2–4) Education is directed towards the qualities and competences that PISA tests capture and that OECD recommends
- Conditions for learning and the quality of teaching are developed
- Students’ life skills improve, making them better equipped for the future and a globalized world
- (1–4) PISA tests and policy recommendations are used for policy decisions at the national and local levels
- (1–4) Countries develop their education systems so as to minimize social disparities in education
- [Improvements in the education system contribute to] economic growth

Assessment of PISA’s programme theory

Internal validity

Are the problems that PISA is to address described and substantiated (Q4)? The original problem PISA was intended to solve, constructed in the 1980s when OECD initiated the Education Indicators Program (Lundgren, 2011, p. 25), was the lack of reliable data on how
countries’ educational systems performed relative to each other. OECD created a need for this kind of knowledge to support policy, and today policy actors generally perceive a need for PISA, indicating that OECD succeeded in constructing the problem as well as PISA as the solution. The need for this information is taken for granted today. Without recurrent PISA results, national policy makers would lack reliable data with which to improve or maintain their countries’ positions in a competitive world.

Whose knowledge needs does PISA meet (Q4a)? First, PISA meets the needs of OECD and of policy actors in member states, including governments (mainly ministries of education), political parties, national level civil servants, national audit administrations, teachers’ unions, lobby groups, and the media. OECD/PISA created the need for PISA, but whether these actors really perceive a need for PISA or whether they act on PISA results simply because they are paid so much attention in the policy community is not known. So far, there is no indication that local politicians, officials, school leaders, teachers, or parents are asking for this knowledge in Sweden.

OECD provides the following arguments for why PISA is needed (Q4b):

PISA is unique because it develops tests which are not directly linked to the school curriculum and provides context through the background questionnaires which can help analysts interpret the results. The tests are designed to assess to what extent students at the end of compulsory education, can apply their knowledge to real-life situations and be equipped for full participation in society. (OECD, 2013a)

OECD/PISA also stresses that PISA provides governments with a strong policy instrument to support education policy, that many stakeholders need this information, and that it can be useful in developing educational systems:

Parents, students, teachers, governments and the general public – all stakeholders – need to know how well their education systems prepare students for real-life situations. Many countries monitor students’ learning to evaluate this. Comparative international assessments can extend and enrich the national picture by providing a larger context within which to interpret national performance. They can show what is possible in education, in terms of the quality of educational outcomes as well as in terms of equity in the distribution of learning opportunities. They can support setting policy targets by
establishing measurable goals achieved by other systems and help to build trajectories for reform. They can also help countries work out their relative strengths and weaknesses and monitor progress. (OECD/PISA, 2009b, p. 9)

At the Swedish National Agency for Education’s (NAE) website, these arguments are repeated and the Agency also emphasizes that PISA results can be used in analysing and developing schools:

PISA is also intended to improve our understanding of the causes and consequences of the observed differences in student knowledge and skills. Background information has been compiled using questionnaires completed by students and school leaders. By investigating various correlations in international comparative studies together with background information, countries can identify their own systems’ strengths and weaknesses, which can eventually improve schools. (NAE, 2012; author’s translation)

The intended effects are clearly described (Q5) and specified in the above quotations. A core idea of PISA is that the programme targets a great many actors at various levels of the education system.

PISA’s PT is only partly consistent (Q6). The four main activities are complementary and largely target the same intended effects. However, for some intended effects there are no matching activities, including the effect ‘conditions for learning and the quality of teaching are developed’ and ‘students’ life skills improve, making them better equipped for the future and a globalized world’ (see the appendix for additional assumptions related to sub-objectives). There is no logical connection between, for example, ‘PISA tests and the policy documents linked to their results are published and disseminated’ and the intended effects ‘education is directed towards the qualities and competences that PISA tests capture and that OECD recommends’ and ‘conditions for learning and the quality of teaching are developed’. At best, PISA puts pressure to take action on national actors, who in turn demand action from local actors or help them do so. The assumption that PISA promotes an OECD-oriented education discourse in which PISA results are deliberated on together with knowledge of what characterizes successful schools is consistent at the national level, but PISA is unlikely to promote education discourse at the local level. There is no logical connection between national test results and how students perform in schools that do not participate in PISA. Whether this assumption has support in involved schools in municipalities participating in the
PISA test is not known, but there will always be many students and schools not participating in PISA.

PISA is likely to initiate or trigger a national education discourse on the performance of education systems, but that this will lead to further action is an assumption based on hope rather than fact. No activities and resources are allocated for developing education systems as part of the programme. If the soft steering mechanism of PISA has the intended effects in some countries, this would indicate that other conditions and factors can explain the action taken to improve education. If PISA affects policy making in some countries (Breakspear, 2012), this cannot be explained by PISA alone or by a solid programme theory.

The shaming and blaming mechanism can be expected to encourage policy actors who want change in the education system. They will exploit PISA results while constructing the policy problem and justifying the need to reform the education system in their own way.

PISA uses test methods based on particular measurement theories and psychometrics (Q7) (Carvalho, 2012; OECD, 2004), and the test of life skill competencies is based on OECD’s three identified key competencies needed in today’s societies: ‘using tools interactively’, ‘interacting in heterogeneous groups’, and ‘acting autonomously’ (OECD, 2005; Rychen & Salganik, 2003). However, the assumption that student performance on PISA tests indicates the quality of education systems is not scientifically based, and one needs to share this assumption in order to interpret PISA results in the way OECD/PISA advocates.

Although PISA has achieved legitimacy as a benchmarking tool, there is no consensus as to what ‘quality’ means in education and education systems. It is not claimed that there is no correlation between student performance and education quality, but how students perform on a PISA test depends on more than just the quality of education and education systems.

Does PISA reflect/do justice to the phenomena/objects it is intended to measure (Q7a)? Testing students’ life skills and generic knowledge does not necessarily capture the quality of teaching (e.g. teachers’ knowledge, pedagogical skills, engagement, and ability to support students) or of the education system (e.g. the steering and accountability function, teacher training content, resource allocation, and evaluation at different levels). National education reform may be launched between PISA tests, which implies that the outcomes of an education system in flux will be measured. If so, it will be impossible to tell whether PISA is measuring
the old or new education system. In addition, PISA does not measure the achievement of national education objectives.

**External validity**

Is PISA’s PT consistent with existing knowledge of conditions/factors that improve teaching/the education system (Q7b)? There is little reliable empirical research that can be used to probe whether the PT has scientific support in this respect. For example, whether seminars treating PISA results and success factors for learning, school performance, and education systems, whether education system monitoring, and whether McKinsey reports (Mourshed et al, 2010) help develop teaching or students’ life skills is an empirical question waiting to be answered. Compiling and disseminating this knowledge is not enough. This kind of policy-supporting knowledge, provided together with PISA results, finds support in school effectiveness research (Bottery, 2001; Harris, 2001), that is, high-performing schools and teachers make a difference independently of social class and external factors (Rutter et al, 1979). However, whether teachers and school leaders participating in PISA seminars learn from them, and whether they assimilate this decontextualized knowledge and use it improve their teaching are questions meriting further research, taking into account many factors and conditions. A major change in a country’s PISA results may mistakenly be interpreted as indicating the success or failure of the education system.

Does the PISA evaluation system provide information about data sources, data quality, non-response, who reports information to the system, etc. (Q7d)? Most of this information is provided in technical reports (e.g. OECD, 2009a) and background documents (2009b). In addition, OECD/PISA also engages in discussions about methods, test design, and how conclusions are drawn.

Is the PT consistent with governance and accountability theories (Q7c)? PISA is an example of what Hanne Foss Hansen refers to as ‘systemic evaluation governance’ (2012, p. 48), that is,

evaluation carried out with steering ambitions and targeted at several actors in a field. Systemic evaluation governance for example may be targeted at several organisations in an organisational field, e.g. a policy sector, as is the case when educational institutions are benchmarked aiming at supporting free user choice, or it may be targeted at several nation
states, as is the case in relation to the Open Coordination Method at the European Union level.

School actors at various levels of the education system should consider PISA results and reflect on their own practices, compare performance, and feel encouraged to develop their education practice within their discretion. Although there is some empirical evidence that PISA has affected national education discourse and policy (Breakspear, 2012; Carvalho, 2012; Mangez and Hilgers, 2012; Ozga, 2012), there is a lack of research into when and why benchmarking and the ‘shame and blame’ mechanism work.

The policy-relevant knowledge that PISA produces can rightly legitimize PISA as a transnational benchmarking system for assessing quality in education systems (Q8), which also can be seen as a system effect (see ‘b’ in Figure 1).

PISA’s value for governance, accountability, and school development at different levels of the education system in different countries merits further research, but, as demonstrated by studies of PISA in the cited special issue, there is evidence that PISA informs policy and governance differently between member countries. This indicates that other factors and conditions – not just those related to PISA – are important when seeking to understand and explain the effects of PISA.

Discussion

PISA’s applied methods are valid for measuring students’ knowledge and life skills as defined by OECD, but whether PISA tests also measure the quality of education systems is open to interpretation and is an assumption one must simply accept uncritically. A correlation between, for example, well-educated teachers, high expectations of students, and high PISA results tells us that there is a correlation between such qualities that OECD/PISA measures, but these factors do not, in themselves, explain PISA results. Furthermore, there is no consensus as to what is meant by ‘quality’ in education systems. If matters such as promoting democracy, human rights, cultural and multicultural awareness, diversity, sustainable development, and critical thinking are conceived as quality in education, PISA tests have low validity as these qualities are not measured. The life skill ‘interacting in heterogeneous groups’ relates to some of these qualities, but does not reflect these qualities specifically. One could also question whether PISA measures change in a fair way, as how individual students
grow and develop is not measured. If progress in individual students’ knowledge and life skills were tested at different points in time, PISA testing would have higher validity; however, PISA is designed to continuously test the skills and competencies of new groups of 15-year-olds.

A related problem concerns the use of PISA for benchmarking of educational systems. If one does not share the assumption that PISA produces valid knowledge of performance of education systems, and neither that PISA helps to improve education in a sensitive and democratic way, the validity and legitimacy of the benchmarking system can be questioned (cf. Grek, 2012). But OECD’s benchmarking of education systems is a political endeavour that member states advocate and that the policy community cannot ignore. One way to respond to PISA is to claim that the validity of PISA and the benchmarking should be interpreted critically and complemented with information about preconditions and other features of education systems (cf. Mangez & Hilgers, 2012). Despite continuing standardisation there are significant differences between education systems which OECD/PISA’s monitoring and ranking of education systems overlooks, that is, objectives and values of national curricula (e.g. critical thinking, democracy, equality, equity, sustainable development). Actors that conceive PISA as an objective account of how education systems perform in comparison with other systems need to share OECD/PISA’s assumptions including that experts have solved existing measurement and comparison problems. A more sensitive, reflective and balanced discussion across nations and education systems will be promoted if knowledge of conditions for learning and teaching from other theoretical perspectives, and evaluations of achievement of objectives of national curricula is considered together with PISA results.

Although some educational research, some of which is cited above, questions what good PISA has done and can do for education and education systems, there is a lack of empirical research into the consequences of PISA that can be used to probe the net effect of PISA. According to the cited special issue, a clear PISA effect prevails in some countries. PISA has undoubtedly had an impact on the education policy discourse and construction of the policy problem in many countries. National differences, however, indicate that many factors and conditions explain whether and when PISA contributes to change in education policy and governance.
Though PISA opponents have highlighted the negative consequences of PISA for national education, surprisingly, many researchers use PISA results largely without questioning the validity of PISA tests and league tables. PISA results are taken at face value, conceived and used as valid measures of student performance and education system quality in light of which conditions and factors explaining national differences can be explored. Today we have a situation in which many actors advocate or benefit from PISA; these actors include policymakers, the media, and researchers who share PISA’s basic assumptions and use PISA test results in their own knowledge production. On the other hand, PISA’s critics among policymakers, practitioners, and researchers (some of whom benefit from PISA by publishing critical articles) are worried about how PISA has damaged education and disempowered national education policy and education systems. What the present PT analysis has demonstrated is that PISA’s effects should not be exaggerated, mainly because PISA contains no built-in activities for changing practice. Still, PISA can have a significant impact on the construction of the policy problem, and as political ammunition in the policy discourse.

PISA has become an ‘institutional order’ in which OECD and co-opted policy actors/researchers construct education (system) problems to match the knowledge OECD/PISA can provide, that is, PISA results constitute performance data with which to benchmark education systems (Grek, 2012). Furthermore, this order legitimizes an OECD-led discourse on what constitutes good education and effective educational governance. PISA produces decontextualized knowledge of education systems, employs ‘governance by comparison’, and promotes the belief that effective policy making should be based on statistics, evidence, and learning from other countries in order to improve education (Grek, 2012). As the PT analysis reveals, however, PISA’s PT is largely inconsistent: PISA includes measures and significantly affects OECD’s framing of the policy discourse, conception of the policy problem, and efforts to maintain demand for PISA’s knowledge production; however, activities to change education practice are not part of the programme.

What, then, can we learn about PISA from this PT analysis? First, the analysis provides a description of PISA that can be inter-subjectively shared and that has value in its own right. It helps advocates and critics to deliberate about the nature of PISA and the consistency of its PT theory. Second, the PT analysis explores the significant intended effects when exploring the real effects of PISA on governance, accountability, and school development. Third, analysing the internal and external validity of PISA’s PT builds knowledge of PISA as a transnational governance instrument. Fourth, the PT analysis provides a summary of the PT’s
underlying assumptions for use in further study of whether, when, and to what extent PISA has achieved its intended effects. The PT can be used for further research and evaluation exploring PISA’s effects on educational governance, accountability, and school development in different countries and at different levels of government.

The assessment of PISA’s impact on education policy and education systems reflects the position of OECD/PISA’s board members in transnational and national education systems (Breakspear, 2012; OECD, 2013b). Whether and how PISA has helped develop education (systems) and its consequences for school practice merit further research. PISA is not the only evaluation system and policy instrument used in educational governance. Hence, empirical research into the real effects and consequences of PISA needs to take into account the impact of other evaluation systems and governance instruments.

The present PT analysis started from the assumptions of PISA’s creators; however, whether this evaluation system meets the needs of the broader accountability environment (Behn, 2001) or serves the knowledge needs of the main stakeholders of national education governance (Foss Hansen, 2012; Ozga, 2012) deserves discussion. PISA was constructed to create a need for PISA knowledge and, as indicated, OECD/PISA has succeeded in creating this need among national policy actors, though not among local school actors.

Peter Dahler-Larsen (2012) suggests that 19 questions should be asked before any evaluation system is developed. These questions do not start from programme makers’ assumptions, but consider the need for an evaluation system from a societal perspective. Some of these questions could also be applied to existing systems: Is the evaluation system infused with an overarching ideology that tends to make the evaluation system self-justifying (question 12)? Have the consequences of the evaluation system been investigated (question 15)? Does the evaluation system incorporate learning mechanisms and ways to ensure a responsiveness to critique that are meaningful and appropriate compared to the institutional power invested in the evaluation system (question 19)? Given that the answers for PISA are ‘yes’ on question 12 and ‘no’ on questions 15 and 19, there is a need to further discuss OECD’s legitimacy in education governance and PISA’s authoritative role in the policy discourse.

Conclusions
That PISA can measure the quality of education systems is an assumption one must share if PISA results are to be conceived as valid for this purpose. There is no consensus as to what quality in education systems means, and many relevant aspects of education systems are not measured by PISA tests.

Overall, PISA’s PT has low internal validity. However, some of its assumptions are consistent, for example, the assumption that legitimization activities help support and legitimize PISA as a transnational benchmarking system that measures education system performance. This assumption can be interpreted as an effect of OECD’s high authority. The assumption that PISA activities promote a performance-oriented and decontextualized national education discourse is also consistent (see the appendix). However, there is no logical connection between, for example, the activities of publishing PISA results (activity 2) and providing policy recommendations (activity 4) and the assumption that these activities will help actors, particularly at the local level, learn about their own practices and take action to improve education. National policy actors need knowledge about the education system as a whole, but for most municipalities and schools, aggregated results do not validly describe the quality of education in their jurisdiction. The internal validity of PISA’s PT as a whole is low at the national level and very low at the local and school levels. As OECD/PISA lacks the needed mandate, resources, and activities to support action to achieve its intended effects, PISA can at best function as an alarm system and a facilitator of policy change at the national level.

The external validity of PISA’s PT has been assessed in relation to governance, accountability, and school development theories. Knowledge of success factors for teaching and learning is provided and integrated into OECD’s policy reports and recommendations, and supported by school effectiveness research (Bottery, 2001; Harris, 2001). What the PT analysis has demonstrated is that PISA activities, such as seminars, dissemination of policy reports, and recommendations, are intended to promote integration of this knowledge, but that there are no activities and resources allocated to change education systems and school practice. Whether and, if so, how PISA has contributed to advancing school development is an empirical question waiting to be answered.

PISA exemplifies what Foss Hansen (2012) refers to as systemic evaluation governance and soft governance. In such governance, all actors in a field are expected to react to and reflect
on their own practice, compare themselves with others, and take action to develop their own practice and the field within their discretion; however, no empirical research examines how such governance works in practice. There is empirical support for the assumption that PISA can promote a national and transnational policy discourse on quality in education (systems) and can construct the policy problem in education (Breakspear, 2012; Lawn, 2011). PISA’s PT is in line with discourse policy and governance theories, which helps explain its validity for legitimizing a national and transnational education policy discourse (Carvalho, 2012; Ozga, 2012). Whether and, if so, why and under what conditions PISA contributes to change in education systems and school practice is an empirical question waiting to be answered.
References


Key: Monitoring and Evaluation (M&E) system refers to indicator-based information systems, producing continuous evaluative information on inputs, processes and outputs/outcomes of policies and programs (Hanberger, 2012).

* The specific prerequisites for fulfilling particular functions differ, though a few prerequisites are the same for almost all functions. Most functions require that the evaluation should maintain acceptable quality and be accurate, relevant, and credible (see Hanberger, 2011 for a discussion of the common and specific prerequisites for various functions).

---

1 “Individuals need to be able to use a wide range of tools for interacting effectively with the environment: both physical ones such as information technology and socio-cultural ones such as the use of language. They need to understand such tools well enough to adapt them for their own purposes – to use tools interactively. Second, in an increasingly interdependent world, individuals need to be able to engage with others, and since they will encounter people from a range of backgrounds, it is important that they are able to interact in heterogeneous groups. Third, individuals need to be able to take responsibility for managing their own lives, situate their lives in the broader social context and act autonomously.” (OECD, 2005, p. 5)
Appendix. Assumptions about how OECD/PISA can achieve its objectives: PISA’s programme theory

Basic assumptions: OECD’s PISA meets member countries’ and policy makers’ needs and demands for regular and reliable data on students’ knowledge and life skills that can be used to assess the performance of a country’s education system in comparison with others.

1. Legitimization and capacity building
   - Get member countries and academics involved in running PISA
   - Develop analytical competence in the PISA team and at the national level
   - OECD monitors education systems and demands responses to questions about education systems

2. Dissemination of PISA results
   - Announce previous and upcoming PISA tests
   - Publicize & disseminate PISA results and policy notes
   - Demonstrate what students need to know in a competitive knowledge society

3. Conferences and meetings
   - Provide knowledge of PISA results and trends in different countries
   - Provide interpretations of policy-relevant knowledge, e.g. success factors for student and education performance

4. Provide knowledge base and recommendations
   - Provide knowledge of successful education systems (e.g. McKinsey reports)
   - Provide policy advice to improve education systems (e.g. in policy briefs and speeches)

CONDITIONS FOR LEARNING AND THE QUALITY OF TEACHING ARE DEVELOPED

- Develop national education system with high equity in distribution of learning outcomes
- Improve all students’ life skills and knowledge for a globalized and competitive world

ECONOMIC GROWTH

SHORT TERM

- Legitimize PISA, what is measured and how, and the ranking system
- Pay attention to what is being measured/tested
- Create the notion of a self-motivated lifelong learner
- Learn about one’s own practice
- Conditions for learning and the quality of teaching are developed

LONG TERM

- Promote a benchmarking culture and results-based governance
- National discourse on factors creating successful schools and students, learning from high-performing countries
- Local actors (e.g. municipal school boards, school owners, school leaders, and teachers) are held to account for the performance of the education system
- Implement evidence-based policy promoting successful schools and students
- Develop transnational education governance for a globalized market and world
- Economic growth