

Mapping the impact of educational interventions

A report commissioned by the Education Achievement Service for South East Wales

Final report, July 2017

Mark Priestley¹, Anna Wilson^{1,2}, Andrea Priestley¹, Regina Serpa¹

¹Faculty of Social Sciences, University of Stirling

²School of Social and Health Sciences, Abertay University

Executive Summary

This report addresses impact capture within education systems, with a focus on contexts in which the system, or particular interventions, have been identified as successful in some respect. It was commissioned by the Education Achievement Service for South East Wales, and addressed the following brief:

What do the most effective improvement systems use as a frame for impact capture, and how do they map impact to learner outcomes across short, medium and long term assessments of value and impact?

Literature relevant to the capture of the impact of educational policies and reforms worldwide was reviewed, and a selection of detailed case studies produced, in relation to the following research questions:

- What principles guide impact capture in effective education systems?
- What systems/processes exist for impact capture in effective education systems?
- What is the relationship between impact capture systems and student learning outcomes?
- Is there evidence of short, medium and long term impact on student attainment and outcomes?
- Is there evidence of perverse incentives derived from impact capture systems, and how are these mitigated?

Five case studies were produced, describing policies and impact capture/QA systems in Singapore, Australia, Korea, Ontario and Finland.

The first outcome of this research is the validation of the usefulness of a framework for describing impact capture/QA systems suggested by Scheerens et al. (2007). This framework, based on a context-inputs-process-outputs model of education systems, provides a useful lens for the analysis of impact capture and QA systems. It is also likely to provide a good basis for the design of impact capture systems, since it draws attention to many of the features of an education system that one might need to pay attention to, but allows for flexibility of focus and emphasis and thus tailoring to meet the evaluation needs of specific policy reforms or pedagogical interventions.

The case studies highlight a range of features and potential problems.

Our analysis of Singapore's system suggests that impact capture and QA systems need to be well-aligned with desired educational outcomes. If they are unchanged through a period where intentions are substantially and qualitatively reframed, they may serve to constrain the system so that it cannot effectively shift to the new paradigm. This problem raises the question of how educational outcomes can be compared when they are fundamentally different. The question of whether ideologically neutral metrics can be defined – i.e. whether different educational paradigms can be made commensurate – remains unanswered.

Our analysis of the Australian system highlights the difficulty of untangling multiple factors and ascribing any kind of causal relationship, even with sophisticated, multi-faceted impact capture/QA systems, when several reforms are implemented at once. It also highlights the need to openly recognise the ideological origins of reforms, in order to identify when they are likely to be in tension (as with the market choice and equity-through-standardisation agendas). It also shows how what are intended to be resources and guidelines to support policy implementation can have the effect of removing teachers' capacity to make decisions and teach creatively, even in the context of a strong discourse of autonomy.

The Korean case exhibits similar problems to Singapore, with the huge importance accorded to tests undermining a professed desire to shift towards so-called 21st century skills. In the Korean case, this is made worse by the persistence of a national exam which acts as gate-keeper to the civil service, and the cultural and financial desirability of entering this profession. Arguably, attitudes to education will only change if the social context changes such that either civil service jobs are no longer seen as desirable, or if recruitment to those jobs changes dramatically, to explicitly value creativity and criticality above declarative/content knowledge. This case study also highlights how impact capture/QA measures can fail to capture important elements of school experience (in this case, symbolic and real violence), which are likely to have a profound effect on both attainment in existing tests and students' present and future lives. This suggests that not only are more nuanced measurement processes necessary, but so too are more nuanced approaches to the analysis and reporting of existing measurements.

Our analysis of Ontario's system, like that of Australia, illustrates difficulties in capturing impact when policies and reforms are in tension, especially when standardised tests are (or at least are perceived to be) high stakes. It also highlights the central importance of engaging stakeholders and listening to those who experience the classroom environment on a daily basis. It suggests that however well-intentioned processes may be, if they appear to discount the views and experiences of teachers and rely on external expertise, there will always be both resistance/resentment and a danger of misunderstanding local contexts. However, when

teachers' professionalism is used as an input to process-development, the results can be extremely positive.

The Ontario case also highlights the benefits of in-depth studies that combine quantitative and qualitative data, especially longitudinal/cohort studies and those with an ethnographic element.

Finally, our analysis of the Finnish system presents a case where system reform has been accompanied by reform of the impact capture/QA system. It seems clear that some of the key strengths of the Finnish system are the high levels of teachers' autonomy and their input into both the design and implementation of impact capture/QA processes. Another appears to be the Finnish authorities' willingness to follow its own path, determined by Finnish cultural values, rather than be swept up by the rhetoric of the Global Education Reform Movement.

Overall, it appears that, despite the existence in some cases of apparently complex and rich processes for impact capture, undue weight ends up being placed on single metrics, often the results of national or supra-national standardised tests. These, which are inevitably focused not only on cognitive skills and declarative knowledge, but also, in their generality, reflecting some kind of lowest common denominator, cannot reflect the diversity of learning and teaching that make up a healthy education ecosystem.

Our case studies also suggest that cultural influences also cannot be understated; as the Finland and Korea experiences show, the status of the teaching profession and the respect it is accorded within the community is an important factor. So too are teacher pay and conditions. If teachers are to be asked to take on the extra burden of conducting regular self-evaluation, it needs to be recognised and accounted for as a formal part of their roles.

In addition, the renewed emphasis on so-called 21st century skills, such as creativity and criticality, raises new questions/problems for competence-based or comparative testing. For one thing, there is not widespread agreement in the literature about what critical thinking or creativity are, and how they can/should be taught, let alone how they are manifested in ways that could be subject to the kind of scrutiny required for impact capture.

One important concern, already raised by several authors and emerging again from our analyses above, is about the dangers of policy-borrowing without considering context, both in relation to educational policies and interventions themselves and to the systems put in place to capture their impact.

In relation to this, several authors have raised concerns about the current dominance of the OECD's voice, as promoted through the PISA and PISA for Schools programmes, in shaping

education systems around the world. Neglecting the cultural and historical roots from which an education system develops is problematic, when considering how to transfer certain policies from one 'successful system' to another. In particular, the condensation of complex, contextually-determined policies, reforms and processes into “best practice” by organizations such as the OECD is shown to be problematic.

Rather than relying too much on standardised tests such as those offered by PISA, the experiences in the nations described in this report suggest that detailed, context specific, reflective impact capture is much more informative when there is sufficient emphasis on inputs and processes. However, it is costly, both in terms of direct financing and in terms of time for both operational implementation and time to reflect, interpret and act. It requires adequate training and support, especially when being introduced into contexts where reflective practice and collaborative working are not already norms. It also requires reasonable timescales for evaluation, and therefore patience and the courage to risk failure, which is perhaps one of the main reasons that policy-makers, the media and others tend to fall back onto policy-borrowing and ranking by PISA – if one doesn't, then one has to take responsibility for making policy and practice changes that could have detrimental impacts on learners' current and future lives.

Document contributors

Name	Institution	Contribution
Mark Priestley	University of Stirling	Project leader
Anna Wilson	University of Stirling and Abertay University	Report author
Andrea Priestley	University of Stirling	Draft case studies
Regina Serpa	University of Stirling	Draft case studies

Glossary and key acronyms

Term	Definition
AfL	Assessment for Learning
Education system	A clearly-identified system providing education at any of primary, secondary or tertiary levels, at the national or sub-national level, unified by the use of common curricula, testing regimes, professional qualifications for teaching staff, and/or governance systems.
GERM	Global Education Reform Movement – a “reform” movement based on increasing competition, test-based accountability and performance-related awards.
OECD	Organization for Economic Cooperation and Development
PISA	Program for International Student Assessment – OECD-run bench-marking tests for 15-year olds in reading, mathematical and science literacies.
QA	Quality Assurance

Quality assurance system	A systemic approach to the monitoring and enhancement of the quality of processes and/or outcomes in an education system.
SES	Socio-economic sector
TIMSS	Trends in International Mathematics and Science Study – test-based statistical information managed by the International Association for the Evaluation of Educational Achievement (IEA)

Table of Contents

Executive Summary	i
Document contributors	v
Glossary and key acronyms	v
Report Overview	1
Section 1: Principles.....	3
1.1 Context	3
1.2 What constitutes a good education system?	4
1.3 Impact capture and quality assurance: phases and elements.....	4
1.3.1 The importance of context	6
1.4 Evaluating the elements of education systems	8
Section 2: Detailed construction of impact capture systems	11
Section 3: Case studies	18
3.1 Singapore	19
3.1.1 Context	19
3.1.2 Inputs.....	20
3.1.3 Processes	20
3.1.4 Desired outcomes.....	21
3.1.5 Impact capture.....	21
3.1.6 Critical factors and concerns	22
3.2 Australia.....	24
3.2.1 Context	24
3.2.2 Inputs.....	25
3.2.3 Processes	25
3.2.4 Desired outcomes.....	27
3.2.5 Impact capture.....	27
3.2.6 Critical factors and concerns	29
3.3 Korea.....	33
3.3.1 Context	33
3.3.2 Inputs.....	34
3.3.3 Processes	35
3.3.4 Desired outputs	36
3.3.5 Impact capture.....	36
3.3.6 Critical factors and concerns	37
3.4 Ontario.....	41
3.4.1 Context	41
3.4.2 Inputs.....	42

3.4.3 Processes	43
3.4.4 Desired outputs	43
3.4.5 Impact capture.....	44
3.4.6 Critical factors and concerns	45
3.5 Finland	47
3.4.1 Context	47
3.4.2 Inputs	48
3.4.3 Processes	49
3.4.4 Desired outputs	50
3.4.5 Impact capture.....	51
3.4.6 Critical factors and concerns	51
4. Summary and conclusions	58
References	63

Report Overview

This project has investigated impact capture within education systems, with a focus on contexts in which the system, or particular interventions, have been identified as successful in some respect.

The brief for the project was:

Reframing impact capture and aligning activity to outcome: What do the most effective improvement systems use as a frame for impact capture, and how do they map impact to learner outcomes across short, medium and long term assessments of value and impact?

This project took the form of a literature review, investigating how effective education systems worldwide have approached the issue of mapping the impact of educational interventions. The project was led by Professor Mark Priestley, working with a team of three research assistants. The following research questions were addressed:

1. What principles guide impact capture in effective education systems?
2. What systems/processes exist for impact capture in effective education systems?
3. What is the relationship between impact capture systems and student learning outcomes?
4. Is there evidence of short, medium and long term impact on student attainment and outcomes?
5. Is there evidence of perverse incentives derived from impact capture systems, and how are these mitigated?

This report presents the results of this research. It is organized in four main sections. Section 1 introduces some of the principles that might underpin the design of processes for assessing educational quality, drawing on the work of Scheerens and co-authors (2004; 2011) and prior studies including work in the EU context (EU 2016), to suggest a framework for thinking about impact capture.

Section 2 presents more detailed features of QA/impact capture systems, with an emphasis on the choices that must be made as they are designed.

Section 3 presents evidence concerning the effectiveness of impact capture processes – specifically on whether they are able to show causal relationships between policy changes/educational interventions and changes in student attainment/outcomes. It does this through a series of case studies of impact capture systems in different national or sub-national

contexts. It includes explicit consideration of some potential (and actual) problems with each of the different approaches.

The report concludes with a summary of the main implications for the design of impact capture systems.

Section 1: Principles

1.1 Context

Around the world, education systems at almost every level are increasingly subject to both scrutiny and change.

A growing culture of public accountability and consumer choice has led to the development, in most OECD countries, of systems to monitor and assure the legal compliance and quality of education at compulsory, and in many cases post-compulsory, levels.

At the same time, the perceived importance of education for all is increasing, as is the sense that education systems should prepare children and young people for life in an increasingly complex and technical world. Schools are expected to be more socially inclusive than in the past, providing welcoming and stimulating environments for children from all social and ethnic backgrounds. An academically-focused education is no longer the preserve of an elite; it is no longer enough for only a small fraction of young people to transition to higher education, and all students are expected to have opportunities to learn and develop to their full potential. These changed expectations for education have resulted in efforts to improve educational outcomes, and in particular to reduce previously existing attainment gaps between children and young people from different socio-economic groups. The nature of the changes introduced in schools depends on many factors, such as the local political climate; the availability of resources; the degree of autonomy the school (or school system) has within the larger system; etcetera. They may focus on pedagogical approaches, curriculum design and content, assessment practices, school management and leadership, non-pedagogical issues such as pupil health and well-being, including diet and exercise, and more. The systems which may have initially been introduced to monitor compliance and quality are increasingly being enrolled in the process of measuring the impact of such interventions and policy changes.

However, impact measurement is not a simple process. For one thing, there is the influence of contextual factors, whose complex interactions mean that measurements of school (and indeed student) performance are often difficult to relate to particular pedagogical or other activities or policy changes in a transparent and unambiguous way. Another issue is the question of ethics (and the related question of timescales): if there is no evidence that a changed practice is having a positive effect – or worse, if there is some evidence that it is having a negative effect on at least some pupils – how long should it be allowed to continue, and thus possibly bed in and get over any initial teething problems, before it is discontinued? Where policies and interventions have an ideological flavour, how can we define ideologically neutral measures of success (and indeed are neutral measures desirable in such contexts)? In addition, the introduction of impact

measurement systems is in itself an intervention in an education system, part of what Ng (2008) refers to as the 'delicate balance of quality assurance' (2008, p12). It may itself give rise to changes in student attainment and outcomes, as educational practices and environments are adapted to improve performance in impact capture measurements.

It is thus important to explore the possible principles that might underpin the design of quality assurance and impact monitoring/capture processes.

1.2 What constitutes a good education system?

The first step in exploring these principles is to attempt to define what a good or high quality outcome of an education system might be. Given the breadth of aspirations held out for education – to support social cohesion, equity, employment, health and wellbeing, innovation and growth (EU 2016, Eurydice 2015) in addition to propagating knowledge and accrediting academic attainment – such definitions are inevitably going to vary from context to context. However, a recent study of Quality Assurance (QA) in EU school systems (EU 2016) found that a good education system is widely understood as one in which:

1. key competences are acquired by all pupils;
2. and/or school systems are equitable and inclusive (i.e. low or declining disparities in educational outcomes);
3. and/or the system is driven by excellence (e.g. in teaching or school governance);
4. and/or the system brings pupils to transition successfully from school to the labour market or to further education.

Good education systems are also ones which respond and adapt 'to new challenges (e.g. budgetary cuts, increasing diversity in classrooms, the use of digital resources, etc.) and approaches (e.g. the key competences approach, the learning outcomes approach, the importance of delivering social outcomes of education, etc.)' (EU 2016, p11). Poor outcomes are therefore generally understood as associated with (and potentially measured by) difficulties in transitions between education sectors and education and the labour market, or poor social outcomes.

1.3 Impact capture and quality assurance: phases and elements

The question then arises, what kind of systems and processes can capture the impact of policies and interventions (whether at the leadership, [intra- and extra-]curricular, pedagogical and/or assessment levels) on the achievement of these features of education systems? Such impact

capture and QA systems ideally have three different but inter-related *phases* of activity, as illustrated in Figure 1: standard-setting, accountability and improvement (EU 2016, p13).



Figure 1: Phases in the monitoring of quality and improvement in education

In any given impact capture process, the emphasis is likely to lie more heavily on the accountability phase – that is, the gathering of data with which to demonstrate how a particular policy or intervention is affecting education at the individual, school, or system level. However, it is only possible to develop accountability metrics if objectives have been identified and standards set, and there is little point in measuring if improvement is not sought and, if necessary, adjusted for. The development of objectives and standards inevitably modifies the types of accountability process put in place, and resulting improvements or changes to the system impact on the objectives and standards. Thus in an ideal impact capture system, the three phases are in a relation of continuous feedback and interaction. This report, while focusing on accountability in the form of performance data and measures, thus also includes discussion of the other phases.

The next question that arises is what kinds of data and measures might be usefully employed in impact capture. Following work undertaken for UNESCO in 2005, Scheerens et al. (2011) put forward a comprehensive framework for describing the principles on which indicators of quality (and therefore impact) might be based. Their model is based on the conceptual framework for describing education systems illustrated in Figure 2.

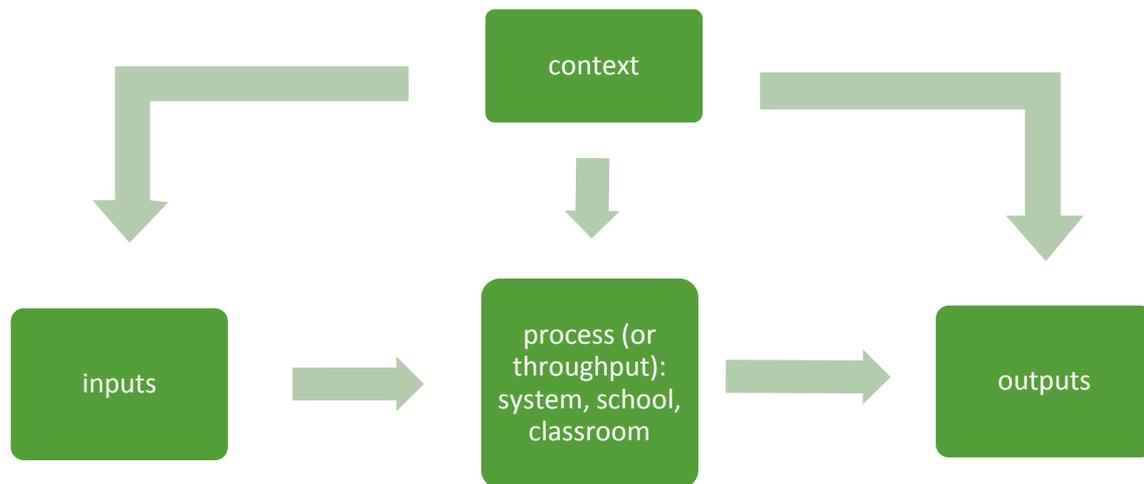


Figure 2: A model for describing education systems (based on Scheerens et al. (2011, p36).

In order to develop an impact capture or QA system, Scheerens et al. (2011) suggest that it one must recognise the following features of the system components:

- The ‘context dimension functions as a source of inputs and constraints but also as a generator of the required outputs that should be produced’ (p36).
- Outcomes should be differentiated ‘into direct outputs, longer-term outcomes and ultimate societal impact’ (ibid.).
- The hierarchical nature of conditions and processes must be recognised, ‘which effectively comes down to considering the functioning of public education as just another example of “multilevel governance”’ (ibid.). This is particularly important when questions of subsidiarity are raised, for example where there are tensions between the apparent equitableness of nationwide standardisation and the adaptive specificity of local governance.
- Choices must be made as to the level at which the central ‘black box’ in the process is to be described. Scheerens et al. suggest that while national, school and classroom levels are obvious and natural choices, other choices are possible, including, for example, explicitly treating students or the local community as separate levels.

1.3.1 The importance of context

As we have already emphasised in relation to the phases of impact capture or QA processes illustrated in Figure 1, Scheerens et al. note that the various elements of education illustrated in Figure 2 are closely inter-related. In particular, they stress the importance of context on all other elements:

the context can be seen as a generator of inputs, as a level that determines or co-determines the definition of desired outcomes that should be generated and as a level that judges quality and provides feedback ... the context dimension [also] gives room for situational adaptation to local conditions (Scheerens et al. 2011, p37).

Indeed, context is of such primary importance and influence on education that Scheerens et al. further suggest sub-dividing it into 'malleable conditions and "given" environmental constraints ... [or] "antecedent" conditions' (p37), where:

Malleable conditions are in the hands of actors on the scene, like national policy planners, local constituencies, school managers and teachers. Antecedent conditions already "exist". Background characteristics of students, such as cognitive aptitude or socio-economic status of their home background, are examples of "given" factors. At higher levels, the school or system level, the distinction becomes more arbitrary. For example, school size could be seen as a given condition, but also, perhaps in a longer-term perspective, as a variable that is subject to change in national policies concerning the desired scale of educational provisions. (Scheerens et al. 2011, p37).

In addition to this differentiation between malleable and antecedent conditions, it may also be important to consider the layered nature of contextual influences. To this end, one might usefully apply Priestley's (2017)¹ conceptual layered model of curricular influences. This model, spanning influences from the supra- to the nano-level, is illustrated in Figure 3.

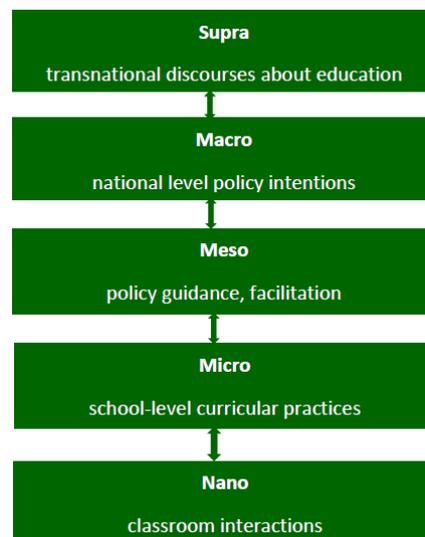


Figure 3: Contextual layers (from Priestley 2017).

The design of any impact capture system must, therefore, take careful account of context.

¹ Adapted from Thijs and van den Akker (2009)

1.4 Evaluating the elements of education systems

The starting point for any kind of impact capture or quality evaluation process relating to education systems has to be the identification and articulation of what the intended outcomes of the system are, or at least which ones the process is intended to measure.

Scheerens et al. (2011) suggest that:

Outcome indicators are central in productivity and effectiveness interpretations of educational quality but also play an indispensable role in assessing the equity, efficiency and responsiveness of schooling (p37).

There are a range of different kinds of academically-related outcome indicator, which might be characterised as more or less tied to specific educational content. This is illustrated in Figure 4.

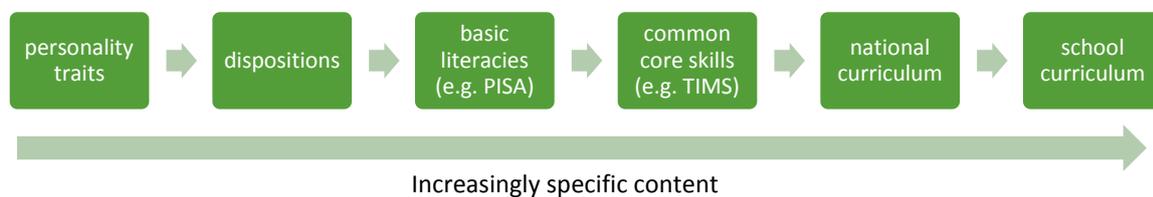


Figure 4: Relation between outcome indicators and specific academic content (adapted from Scheerens et al. 2011)

As well as academic outcome indicators, those interested in assessing impact may also be interested in other kinds of outcome, including attainment indicators such as graduation rates and social impact indicators relating to, for example, social participation rates or social mobility.

Once those designing an impact capture process have decided on the kinds of outcome they are seeking from the education system to be studied (and so started the process of identifying potential outcome indicators to measure), they may need to take into account the inputs into the system as illustrated in Figure 2. According to this basic model for education systems, inputs provide ‘the material and immaterial pre-conditions for the core transformation processes in organisations’ (Scheerens 2011, p41). Scheerens et al. (ibid) suggest that in education systems, the main categories of input are:

- financial and material resources;
- human resources; and
- background conditions of the students.

The difference between inputs and contexts, in Scheerens et al.'s model, is somewhat blurred, and may be most easily understood as depending on how easily the factors can be controlled and how easily they can be related to simple metrics. According to these authors,

... specifying what is meant by the context depends on the level at which the central transformation process is defined. Most of the time transformation processes at school level have been concentrated on. When transformation processes at school are further differentiated to distinguish primary teaching processes at classroom level and secondary, supporting management and organisation processes at school level, a multi-level model results, in which everything "outside" the school is defined as the context. In this kind of conceptualisation "context" could be further subdivided in the direct environment, local community and local/regional administration on the one hand and the national context on the other. ... [Where] education is considered at one level only, the national system level ... the context is defined as the relevant environment of the "education province" as a whole. As such, the general affluence of a country, demographic tendencies, cultural aspects that impinge on values that are important in education and the institutional infrastructure of a nation could be seen as the context of education (ibid, p46).

Thus, as emphasised in the previous section, those designing effective impact capture systems need to take careful account not only of contextual factors, but also of the level at which that context is operating. Depending on that level, some features (for example, teachers' habitus, school culture, or material resources) might be considered as context, or as input resulting from other, identified, contextual factors.

Finally, we need to consider what is meant by process and process indicators. To illustrate what is meant by process and how indicators relating to process might differ from outcome, input and context indicators, it is worth considering the OECD Education Indicators Project (OECD 1998).

The main categories of indicator defined in this project were:

- 1) The demographic, social and economic context of education.
- 2) Financial and human resources invested in education.
- 3) Access to education, participation and progression.
- 4) The transition from school to work.
- 5) The learning environment and the organisation of schools.
- 6) Student achievement and the social and labour-market outcomes of education.

Following Scheerens et al. (2011), we can identify category (1) as relating to context, category (2) as relating to inputs, and category (6) as relating to outcomes. The remaining categories (3-5) relate to process at various levels and phases in the education system.

Thus the description of an education system in terms of the basic process model illustrated in Figure 2 suggests a conceptual framework within which new impact capture systems can be designed, and existing impact capture systems can be analysed.

In the next section, we consider the implications of this model for the specific types of data that might be gathered in relation to each component of the system. Section 3 then presents some selected case studies, which illustrate indicators that are commonly used and also highlight some types of indicator which are commonly neglected or undervalued.

Section 2: Detailed construction of impact capture systems

With the guiding principles set out in the previous section established, it is then possible to consider in more detail the kinds of impact capture systems that could be put in place, and the indicators on which impact measurement might be based.

In designing and implementing an impact capture system, there are various choices to be made, some of which are more likely to be made along ideological lines than others, but all of which have substantial practical impact on what is and what can be known about the system being evaluated.

For example, decisions need to be made about:

- Whether the main aims of the educational policy, intervention or system under study are academic or social.
- The weight placed on non-educational factors such as health and wellbeing, or socio-economic status.

Such decisions have a complex relationship with beliefs about whether the primary function of education is related to economic growth and success or whether education provision is motivated by notions of its intrinsic value and contribution to the holistic improvement of the lives of individuals and society.

Other decisions must be made about:

- Whether the measurements will be criterion or norm referenced (i.e., whether the system looks for absolute standards and competences, or whether it is relative and therefore inherently competitive)
- Whether the data gathered will be kept private or made public, and if so to what extent or at what levels of aggregation.

Such decisions often reflect whether the wider context is one that values “market” accountability or government/authoritative accountability.

One also needs to consider:

- Whether evaluations are externally-conducted or internal to the system under study, to what degree their structure is prescribed through e.g. rules, handbooks and guidelines, and how many different classes of stakeholder are involved.

- Whether evaluation should focus on the identification of risk or good practice.
- Whether the results are linked to financial distributions – that is, whether they are punitive, remedial, incentivizing or intended purely as measurements to inform subsequent reflection.

These decisions may reflect the respect within the wider context for the teaching community, and the degree of professional autonomy accorded to it; the latter is also connected to ideas of accountability, with government accountability more likely to be linked to sanctions and incentives, and market accountability more likely to be divorced from them as the assumption is that providing parents with information and consequent choices will naturally lead to improved practices.

All decisions also involve some costs, for example whether explicitly in terms of funds for external evaluators or implicitly in terms of increased teacher/manager workloads as they carry out self-evaluations.

It is only by considering the various elements described in the previous section and the questions outlined above that coherent impact capture systems can be designed. Unfortunately, it seems to be rare that such complete attention is paid, even in the apparently successful systems described in the next section. A recent EU report (EU 2016) on the QA systems in 41 countries found a general lack of coherence and consolidation. The present report highlights some of the critical factors and concerns that arise in relation to the impact capture and QA systems described in the next section.

Once clarity has been reached in regard to the questions outlined above, the framework outlined in the previous section can be used to identify more detailed areas on the basis of which impact might be captured. Scheerens et al. (2011) suggest several possible areas for observation and/or measurement. These are summarised in Tables 1-4. It is clear from the suggested indicators included in these tables that impact capture/QA systems are likely to have a strong normative character, as the very selection of an indicator and definition of a desired standard for that indicator defines what is counted as “good.”

Table 1: Categories of outcome indicators (adapted from Scheerens et al. (2011))

Achievement measures

- Subject matter based
- Literacies (reading, mathematical, scientific, IT)

- Competencies (e.g. learning to learn)

Attainment measures

- Graduation rates
- Proportion of students graduated without delay
- Drop-out rates
- Class repetition rates

Social participation rates

- % employed at a certain job level
- % enrolled in further education
- Degree of social participation
- Adult literacy rates
- Average income by attainment level; earning differentials
- Skills shortages/surpluses

Table 2: Categories of input indicators (adapted from Scheerens et al. (2011))

System level financial and material resources indicators, e.g.

- Proportion of GDI spent on education
- Expenditure per student
- Public investment in educational R&D
- Total expenditure on disadvantaged students
- Household expenditure and public subsidies to parents
- % spending on salaries/pensions for administrative and educational personal

School level financial and material resource indicators, e.g.

- Proportion of budget acquired in addition of public funding
 - Building facilities, classroom equipment, school supplies
 - Availability of textbooks
 - Basic services (toilets, heating, water, IT)
 - Ancillary services (nutrition, health and wellbeing)
-

Human resources indicators, e.g.

- Teacher background characteristics, such as age, gender/ethnicity distributions
- Full/part-time teacher distributions
- Certification/licence status, formal qualifications, in-service training history
- Years of experience
- Knowledge about pedagogical strategies
- Beliefs and attitudes about teaching and learning
- Flexibility in teaching repertoire
- Teacher salaries and working conditions relative to other professions
- Job mobility
- Working time
- Average class size/ teacher:pupil ratio
- Incentive policies and career structures
- Teacher autonomy, political self-efficacy
- Perceptions of teacher status

Student-related input indicators, e.g.

- General intelligence/scholastic aptitude
- SES, gender, ethnicity, language spoken at home
- Parents' level of educational attainment
- Distance to school
- Amount of time student has to spend in labour out of school
- Home situation, e.g. provision of breakfast, space to study, number of books in the home

Table 3: Categories of context indicators (adapted from Scheerens et al. (2011))

Contextual conditions of education system

- Demographic developments/trends
- The labour market (shortages and surpluses)
- State of the economy
- Cultural aspects
- Institutional infrastructure

- General population health characteristics

Antecedent conditions within the education system, e.g.

- Supply and demand of teachers
- Age distribution of teachers
- % students in school outside the normal age ranges
- Cultural attitude to education
- Cultural expectations of education (e.g. didactic, student-centred, for citizenship)
- Formalisation of rights
- Rules and regulations relating to eg.. teacher absenteeism, teachers providing private tuition

Local community indicators, e.g.

- Existence of a school board
- Existence of local resources centre
- Willingness to engage of e.g. local industry
- Role of local community in financing/in-kind support
- Parents values concerning school participation
- Discrepancies between indigenous and “school” knowledges

Table 4: Categories of process indicators (adapted from Scheerens et al. (2011))

System level process indicators, e.g.

- Teaching time per subject
- Total hours instruction
- Test/curriculum overlap
- Locus of decision-making, degree of school autonomy
- Existence of formal tests at end of each level; targets such as increased completion rates, % of students attaining or exceeding a particular level
- Formal streaming/selection
- Evaluation capacity of the system
- Distribution of public/private schooling, extent to which choice is free

- Provision and magnitude of support structures

School functioning process indicators, e.g.

- Community involvement
- School financial and human resources
- Achievement-oriented policies
- Educational leadership
- Continuity and consensus among teachers
- Efficient use of time
- Absenteeism and delinquency
- Teacher and student ratings of opportunities to learn
- Frequency of (standardised) tests; use made by teachers of test results
- Ratings of teacher quality (by peers and students)

Effective teaching and learning indicators, e.g.

- Opportunities to learn
- Scaffolding and structuring
- Stimulating engagement
- Climate (e.g. mutual respect)
- Monitoring and questioning, feedback and reinforcement
- Modelling (learning and self-regulation)
- Authentic applications and experiences
- Students' learning strategies

As can be seen from these tables, there are many factors which could be taken into account in any system designed to capture impact or assure quality of educational processes. In most cases, the tables do not address the nature of the data that could be gathered, and while several of these indicators may be open to quantitative capture processes, there may also be a need to capture "soft" or qualitative evidence. This is particularly the case with respect to process indicators. However, because any impact capture system is inevitably normative in character, there is a danger that qualitative evidence may be used in what Lewis refers to as 'governing by example' (2017, p286).

This, along with the previous sections, establishes a conceptual framework within which to analyse or design impact capture systems. They also offer specific potential measurements that could be made in evaluating the impact of educational systems, policies and interventions. The following section provides a series of case studies, illustrating how some nominally successful educational systems are evaluated. It also identifies critical factors and concerns in relation to each system. In so doing, it highlights both a surprising emphasis placed on only a small subgroup of the indicators outlined above, and a frequent over-reliance on standardised tests as the main indicator of quality.

Section 3: Case studies

In this section, five case studies are presented to highlight the range of approaches to impact capture and quality assurance in apparently successful education systems. The cases have been selected on the following bases:

- Apparently good performance of the system under study, based on international measures such as PISA or on international reputation.
- Variety of social and cultural contexts.
- Variety in stated aims for education.
- Variety in impact capture approaches.

These cases are not intended to represent all possible or indeed enacted approaches to impact capture, but they do serve to raise important questions as to what the impact of impact capture may be.

Each case study is analysed within the framework of context-inputs-process-outputs described in the previous sections. Critical factors and concerns that have been raised by other authors, or that are evident from our analysis, are raised in relation to each.

3.1 Singapore

Singapore is often held up as an example of a country with a highly successful education system. It has done consistently well in international bench-marking processes such as the PISA tests, and Singapore is a relatively stable and economically successful society. It is a favourite of the UK national government, when it comes to identifying countries that the UK could learn from. It therefore provides an interesting starting point for our series of case studies.

3.1.1 Context

If one takes a look at what the Singapore impact capture/QA process currently comprises, one finds that it constitutes a diverse and imposing range of indicators and activities (Isaacs et al. 2015). There are school inspections and a self-evaluation process for schools. Teachers are evaluated and appraised partly on the basis of on student attainment. Output metrics include results in national, standardised academic tests, physical fitness tests, and the percentage of pupils who are overweight.

To understand these components of the impact capture/QA system and their impact, it is worth providing an overview of some of the relevant historical and socio-political context. Deng and Gopinathan (2016) suggest that:

Singapore's present education system is a product of strategic central planning and implementation of educational policies by a strong-willed government in response to unique political, economic and social challenges (Deng and Gopinathan 2016, p456)

Singapore was established as a city-state in the early 19th century in order to serve as a base for the British East India Company. Having gained independence from the UK in 1963, Singapore is a young and effectively one-party state, with a 'powerful elite bureaucracy' (Dimmock and Tan 2013, p321). It is a society which places a high value on civil obedience and citizenship responsibilities.

During the 1960s and 1970s, the country's economy shifted to an export-orientated manufacturing base. To meet the demands of the new economy in terms of workforce, the Single National Education System started to create a workforce that was literate and skilled. Schools were built and teachers were recruited on a large scale (OECD 2016).

At the end of this period, a push towards curriculum standardisation was accompanied by a decision to change the language of instruction to English, and a renewed focus on maths, science and languages, and, to a lesser extent, technical subjects (Deng and Gopinathan 2016).

In the 1980s and 1990s, Singapore's education system could be characterised as meritocratic (Deng and Gopinathan 2016) and exam-driven; a system in which high-stakes testing resulted in concrete rewards that impacted on pupils' lives and futures (Isaacs et al. 2015).

During this period, education in Singapore was seen as key to developing an economically successful society (Deng and Gopinathan 2016). More courses and more differentiation were offered, so that there would be an increased number of students who could progress to secondary schools and beyond (OECD 2016).

Since 1997, the Singapore government has attempted to change the teaching and learning culture in the country (OECD 2016). A changed international socio-economic context meant that Singapore perceived a need to change its own economy to a more creative, high-tech one, in order to compete with neighbouring countries that were able to provide cheaper labour and manufacturing. Two new programmes – Thinking Schools, Learning Nation (TSLN) and Think Less, Learn More (TLLM) – were introduced, with the intention of promoting creativity, innovation and criticality, while retaining a strong focus on citizenship responsibilities 'to family, society and country' (Isaacs et al. 2015, p4).

3.1.2 Inputs

One important input to the system, especially during the late 1970s and early 1980s, has been curriculum standardisation and official support measures for its implementation. To achieve this, the Curriculum Development Institute was set up in 1980 (Deng and Gopinathan 2016).

In terms of human resource inputs, the State itself invests heavily in developing school leaders. There is 'a belief that high quality school leaders can strengthen the quality of classroom teaching' (MoE n.d., p452).

Another extremely important input into the system is money. A substantial amount of investment into teaching and learning comes in the form of the strong parental drive to pay for extra, private tuition (according to Deng and Gopinathan (2016), citing Chow (2012) and Toh (2012), parents spent a billion Singapore dollars on enrichment and private tuition classes). This raises the question, is it the system as provided by the State that leads to Singaporean pupils' good academic performance as measured by standardised tests, or does the system rely on the intensity of private, out-of-system support?

3.1.3 Processes

Teaching processes in Singapore's schools are generally didactic, and have been described as teaching to the test (Hogan et al. 2013). There is substantial use of streaming. Academic ability and performance determine merit. Many processes within the system are highly standardised,

including staff appraisals and the collection of student data. Thus much of the process component of Singapore's education system is characterised by knowledge transfer, testing and standardisation.

3.1.4 Desired outcomes

As mentioned above, the outcomes that appear to be desired by the Singapore government have changed in the last 20 years. Prior to 1997, the emphasis was on academic attainment, the production of responsible and obedient citizens, and ensuring a healthy and physically fit population (Isaacs et al. 2015).

Since 1997, the way that academic success is defined has shifted from mastery of specified content knowledge to creativity, innovation and critical thinking (Isaacs et al. 2015). The 21CC core competencies framework (see <https://www.moe.gov.sg/education/education-system/21st-century-competencies>) promotes the following outcomes:

- Confident person
- Self-directed learner
- Active contributor
- Concerned citizen

According to Isaacs et al. (2015),

Singapore's Ministry of Education introduced a greater focus on creative and critical thinking beginning with its 1997 reforms. The Singapore approach calls for the integration of thinking skills explicitly within core disciplines. Project-based and inquiry approaches were later added to help students make connections across disciplines. In math, the emphasis is on problem-solving and adaptive reasoning; in science and humanities, the focus is on the inquiry process (Isaacs 2015, p12).

TLLM was brought in to help educational practices in classrooms to become less about rote learning, memorisation and didactic teaching and more about exploration and reflection using more constructivist teaching approaches. Policy-makers believed that if they provided more space in the curriculum this would create an environment that might foster these types of approaches. Moreover, this initiative was introduced to help schools move away from teaching to the test and focus more on preparing students for life, lifelong learning and an enjoyment of learning leading to more intrinsic motivation (Liem et al. 2016).

3.1.5 Impact capture

During the 1980s and 1990s, when Singapore's education system was explicitly focussed on producing highly literate graduates who were skilled in languages and more technical subjects,

and on embedding the notion of citizens' responsibilities and loyalty (Deng and Gopinathan 2016), key impact capture measures were secondary school completion, English 'O' level passes, and performance in TIMSS. These suggested that the changed system was indeed achieving its aims, with significantly increased completion rates, with 94% of students completing a least 10 years of education, compared to less than 50% in the 1960s; English 'O' level passes up from 40% in the 1960s to 90% in 1984; and very good performance in TIMSS in 1995 (OECD, 1995).

Despite the change in desired outcomes from the education system since the end of the last century, in many respects Singapore's processes for impact capture and QA have not changed. Significant weight continues to be placed on the results of standardised tests, particularly performance in the international PISA tests, but also the Primary School Leaving Examination (PSLE) which determines which secondary school course each student can take. Teacher performance is still evaluated in part on the basis of student performance, and 'ratings are used to determine bonuses and promotion decisions' (Isaacs et al. 2015, p6).

However, there have been some changes to approaches to impact capture. School appraisals now embrace far more than academic processes, covering management, instructional programmes, extra-curricular activities and pupil welfare programmes. The introduction of a self-evaluation model (School Excellence Model, see Ng (2010)) shifted responsibility from external inspectors to schools themselves. Schools now evaluate themselves yearly, with a five year cycle of validation. The SEM is based on business models, including Singapore's own QA award system for commercial and industrial enterprises.

In the past, schools were ranked based on the various indicators in use (predominantly the results of national tests; indicators of value-added in the form of progress from initial performance; physical fitness tests; and the percentage of overweight pupils). However, the publication in the media of school rankings had unintended consequences, resulting in too much inter-school competition. For this reason, the State decided to stop publishing the results of individual schools.

3.1.6 Critical factors and concerns

Despite claims to the contrary, there is evidence that the Singapore education system does not effectively operate as a meritocracy. Kennedy (quoted in Deng and Gopinathan 2016) notes that there is a long tail in academic attainment for pupils coming from low socio-economic status backgrounds, indicating that Singapore's schools do not level the playing field and allow students from any background to achieve the same levels of academic success.

Liem et al. (2016) suggest that the impact of TSLN and TLLM have been mixed. Schools themselves have suggested positive effects since the start of this initiative claiming that 'students have become more reflective and critical, more participative in class and more intrinsically motivated in pursuing interests beyond their school work' (Ng, 2012, Peh, 2007 cited in Liem et al., 2016, p278).

Hogan et al. (2013) suggest that, while the policies of TSLN and TLLM expect dialogical practice, there is a disconnect between policy and practices, but a 'high degree of pedagogical alignment between system and instruction' (p60). They conclude that:

In effect, the current assessment regime incentivizes and rewards teachers to teach (and student to learn) in ways which maximise assessment performance rather than the kinds of teaching called for in national policy documents and generally associated with teaching for understanding frameworks (ibid, p99).

Other research has suggested that the attempt to import the constructivist view of learning at the heart of the "Knowledge Building" approach met with difficulties because both pupils and teachers had very different beliefs about teaching and learning:

...the activity system of doing school in Singapore ... favours learning well and succeeding in examinations rather than seeking application or meaning from learning to one's everyday life ... Here, an 'imported' cultural form — the [Knowledge Building] pedagogy — found itself in an uneasy relationship in an older and more established activity system that sanctions and rewards education for its exchange-value (Lee 2010, p25).

Hogan et al. (2013) suggest that the high stakes assessment regime in Singapore constrains innovation in teacher practice. Their research suggests that whilst teachers do value different instructional practices aside from traditional instruction and direct teaching, these 'conceptually orientated instructional practices do not have a strong impact on student achievement, given the nature of the current assessment system' (pp98-99). Moreover, these authors raise concerns about the issue of professional development, professional learning communities, instructional beliefs and student outcomes:

[p]rofessional development and professional learning communities reinforce rather than challenge the dominant performative orientation of pedagogical practice in Singapore' (ibid, p100).

3.2 Australia

Australia presents an instructive case study as it mixes both a strong emphasis on constructivist learning and a neoliberal discourse around education. It is one of the key countries in the Global Education Reform Movement (GERM). While it shares some of the characteristics of the Singaporean system, in its emphasis on standardised tests as an important indicator of impact and quality, and Singapore's more recent emphasis on creativity and critical thinking, the political and cultural context is very different. It has a tradition of individualism and a vital, often high-temperature, multi-party political system. It is thus interesting to see how it approaches the problem of measuring impact and quality in education.

3.2.1 Context

Education reforms in Australia should be seen in the context of the Global Education Reform Movement (GERM). The GERM has drawn upon three key trends (Bloxham 2015):

- the shift to cognitive and constructivist approaches in teaching and learning;
- guaranteed effective learning for all as demanded by the public; and
- test-based accountability as the means to raise school performance and the quality of student education.

In order to achieve these aims (particularly the second and third), the Australian government has embarked on a programme of standardisation. The standardisation of educational and pedagogical processes through the introduction of performance standards is the most visible consequence of the GERM in Australia.

At the same time, the neo-liberal inspired perception of an underperforming public education system drives the education reform agenda, thereby legitimising marketisation and the commodification of education and the introduction of competition between schools and students (Bloxham 2015, p369). There has been a general deregulation of education at all levels, justified by a belief in the power of market forces to improve teaching, schooling and student achievement drive these developments, in which Australia is following closely in the footsteps of models developed in the USA and UK (Dinham 2015). Australia already has one of the most privatized education systems in the world, with a much higher level of attendance at non-government schools than in comparable nations. The drift to non-government schools has been facilitated by governments of all political colour over the past few decades (Dinham 2015).

3.2.2 Inputs

Inputs into the Australian education system in terms of financial resources are controlled by the Federal government. Although the constitution assigns responsibility for school education to the states and territories, 'the federal government has played an increasingly powerful role over the last 50 years because it can make grants to states and territories to which conditions are attached ... the states and territories have no power to levy an income tax and are dependent to a large extent on grants from the federal government to operate their schools' (Caldwell 2016, p1172).

In terms of curriculum and curriculum/pedagogical guidance, a major input to the system comes through the Australian Curriculum Assessment and Reporting Authority (ACARA). This body (set up in 2009) has developed a national curriculum, based around standards and competences specified for all subjects at all levels.

In parallel, a national, not-for-profit company owned by all Australian State, Territory and Federal Education Ministers, Education Services Australia (ESA), was given responsibility for supporting national priorities and initiatives in schools (Birenbaum 2015). At the same time, the Australian Institute for Teaching and School Leadership (AITSL) was set up in 2010. This body has responsibility for professional standards and professional development for teachers and school leaders (Birenbaum 2015).

3.2.3 Processes

There is a strong discourse of autonomy and teacher professionalism within the Australian system, consistent with the general national culture of individualism and can-do spirit. In line with this discourse, greater responsibility has ostensibly been handed down to state and local authorities through a process of autonomisation. "Autonomy" is a term that is now widely used in Australia to describe the decentralization of authority and responsibility to public schools within a centrally determined framework of policies, standards, curriculum and accountability' (Caldwell 2016, p1172). This autonomisation process was kick-started through the National Partnership Agreements (NPA) from 2007 until 2013, under a scheme known as 'Empowering Local Schools' that provided substantial funding to all states and territories except Western Australia, and to Catholic and independent schools.

However, at the same time there has been an emphatic attempt at standardisation, through the implementation of the national curriculum as designed by ACARA (see inputs above) and a programme of standardised testing.

There has also been standardisation of teaching and learning processes through the provision of large amounts of supporting material, interpretative resources and tools. For example, at the school level, Queensland introduced a Teaching and Learning Audit Tool (Mills 2016), and at the individual level, the Running Record Converter for literacy results automatically calculates students' error ratios and accuracy rates from data entered by teachers, and provides a classification of how easy or hard the student found the text they were reading (Kerkham 2014).

In addition to these general reform processes, there have also been reforms targeting particular communities such as indigenous and low SES pupils. For example, the Stronger Smarter Learning Communities (SSLC) initiative was implemented following an international comparative analyses of student achievement data, which raised awareness of the under-achievement of Indigenous students and stimulated government spending in this area. The SSLC 'involved a network of schools, based on a 'learning communities' or a 'communities of practice' model, which aimed to spread the key messages about school reform and about education of Aboriginal and Torres Strait Islander students' (Klenowski 2014, p448). Similarly, the Schools in Partnership initiative was set up for schools with high indigenous enrolments. This initiative involves the use of funds from the Smart Schools National Partnerships programme to establish a role of "data keeper," typically undertaken by the school principal, to collate and manage data including Personalised Learning plans (PLPs).

In parallel with processes intended to promote the discourse of autonomy (and hence teacher accountability for student performance) and standardisation, there has also been, as a direct consequence of the constructivist leanings of the GERM movement, a push to develop assessment for, rather than of, learning. This has resulted in attempts to reform pedagogical processes though, e.g., the provision of a Guide to Assessment for Learning developed and distributed by another federal body, the Curriculum Corporation (Birenbaum 2015). According to Fenwick (2017),

Classrooms that are using AfL effectively will include constant dialogue around standards, criteria, tasks and feedback on progress. Clear learning goals, criteria for assessment and feedback, should be used in ways that challenge students to enter into dialogue and express their thinking to both teachers and peers during the learning process. As a result, students will be actively involved in making meaning during lessons, which will lead to deeper understanding. Teachers who use observation on a daily basis will be able to assess how students are using criteria and feedback and to decide if further feedback or modelling of learning is required. The feedback and modelling provided may be from teachers or peers and should be used to generate further dialogue about learning (Fenwick 2017, p44).

Other processes promoted through the various educational reform policies include outcomes-based learning and revised standards for professional learning (including a shift towards teaching as a Masters level profession on entry).

3.2.4 Desired outcomes

The desired outcomes of all these reforms and implementation processes are, simultaneously,

- an increasingly constructivist and student-centred approach to teaching and learning;
- nominal increases in teacher autonomy with respect to the design of learning activities in context;
- increased use of assessment for formative and feedback/feedforward processes;
- increased description of desired learning outcomes in terms of standards and competences;
- increased standardisation of learning opportunities and practices across the country;
- increased equity and attainment for students from indigenous and other disadvantaged or marginalised communities;
- and increased public accountability delivered through a market/competition environment.

3.2.5 Impact capture

Given these complex and wide-ranging desired outcomes, and the context of multiple simultaneous and sequential reforms processes, it is clear that only a very well-designed impact capture system would be able to successfully correlate cause and effect.

One of the most important impact capture processes in the Australian system is the programme of standardised, national tests known as the National Assessment Program Literacy and Numeracy (NAPLAN). This programme is designed 'to collect, analyse and report nationally comparable data on student achievement in literacy, numeracy, science, ICT and civics and citizenship' (Mills 2016, p123).

National tests in literacy and numeracy have been conducted since 2008 for all students in Years 3, 5, 7 and 9. Design and delivery of NAPLAN have been the responsibility of ACARA since 2009. The results of these tests are highly public, and at a granular level. Individual school results are reported on the MySchool website (www.myschool.edu.au). 'These paper-and-pencil tests are conducted in each school in May, with state-by-state performances reported publicly in August, and school-by-school performances published early in the following year. Parents receive reports on the performance of their children' (Caldwell 2016, p1173).

According to Polesel et al.,

ACARA states that one purpose of NAPLAN is to help schools identify issues within their teaching programmes, and to assist individual students. Parents can also use the results to 'discuss progress with teachers, and teachers can use the results to identify 'outliers', that is, students who may need extra support, or more challenging material' ... systems can use the results to review programmes and target support to schools (Polesel et al. 2014, p649).

NAPLAN scores are thus used in a variety of ways, including as evidence for how well Australian states are performing at the aggregate level (and hence the performance of the nation as a whole). They are also used to distribute funding, within the Smart Schools National Partnerships programme, which made additional funding available to low SES schools contingent on improved NAPLAN scores (Hardy 2015).

Australian states and territories also implement performance evaluations at the school and individual teacher levels, involving both external inspectors and internal self-evaluation. It is through these, as well as ultimately through their students' performance in the NAPLAN tests, that teachers' implementation of strategies such as Assessment for Learning are judged. Despite the discourse of autonomy and decentralisation, the Australian Federal Government has played an increasingly important role in monitoring "teacher quality," 'subsuming the powers of state teaching regulatory bodies through the imposition of national standards for teaching' (Mills 2016, p117), especially through the AITSL. This has resulted in the use and provision of highly detailed expectations of practice. For example, the Australian Professional Standards for School Principals 'articulates what principals need to understand and do in order to achieve excellence in their work' (Bloxham 2015, p356), while the Teacher Development Framework explicitly requires that teachers:

engage in the systematic collection of evidence on their practice as part of the "performance and development cycle" ... In the evaluation of their goals, teachers are required to gather evidence from at least some of the following nominated sources: evidence of the impact of teaching on student outcomes; direct observation of teaching; evidence of the teacher's impact on colleagues and the school as a whole; student feedback; peer/supervisor feedback; parent feedback; teacher self-assessment; evidence of participation in professional learning and teacher reflection on its impact (Groundwater-Smith 2016, p83).

In addition to NAPLAN, the schools inspection system and teacher performance evaluations, the Australian government occasionally commissions one-off reports into specific issues, often those which have been identified as important because of an ideologically-driven reform agenda.

3.2.6 Critical factors and concerns

Critics have suggested that a market-like education service, inspired by notions of efficiency, productivity, and responsiveness, has supplanted professional autonomy, and delivered uniformity and standardisation rather than quality and diversity as promised (Bloxham 2015, p356).

Central to these developments is the belief that public education in its traditional forms has failed and is in crisis. However these developments are not merely a reaction to this 'manufactured crisis' (Dinham 2015, p13), but are said to be actively contributing to the further erosion of confidence in and the dismantling of public education.

Paradoxically, the large amount of supporting material and resources provided by the various new bodies such as ACARA, AITSL, etc., as well as through R&D commissioned from academics, has probably helped to undermine teacher autonomy. For example, to enhance teacher readiness for the various reforms to both system and pedagogical approaches, an in-service package – The Australian Curriculum Preparation Toolkit (Noack 2013) – was provided to school leaders to implement in their schools. The resource's creators describe this toolkit as comprising:

A leaders' resource titled The 5-P Challenge. This resource provides questions for the leadership team related to Pace, Processes, People (existing staff and students), Priorities and Philosophy and facilitates exploration of how these would impact upon preparation for and implementation of an educational reform such as the Australian curriculum; and 2) Additional resources for leaders which allow them to work collaboratively with all staff. Staff are given the opportunity to reflect upon and articulate existing professional values and attributes, as well as support they would require during the change processes. In particular, there are activities to provide teachers with opportunities to explore their personal approaches to change in discussion with peers and leaders (Noack 2013, p459).

This is just one of the many resources and supports that have been provided to teachers and school management to help them understand, interpret and implement the national curriculum, and is an example of the kind of process that, though well-intentioned, arguably decreases teacher autonomy and professionalism by over-stipulating the ways in which teaching and learning, and indeed relationships among teachers and management, should be enacted. As Kerkham notes,

How these government policies are interpreted and mediated in schools is a key question for understanding how educators' work is coordinated, how their words and actions are regulated, and how – in an era of audits and accountability and what

has been called the 'deprofessionalisation' of teachers – educators are able to exercise agency (Kerkham 2014, p344).

Another consequence of the rapid rate and large number of innovations and policy changes is change fatigue. This has been described (in relation to the Western Australian context in particular) as 'tiredness and lack of enthusiasm for change as a result of negative lived work experiences of an overwhelming number of curriculum reforms already mandated and implemented in the last 20 years' (Dilkes 2014, p45).

The rate and volume of reforms and interventions also diminishes the value of standardised tests such as NAPLAN, as genuine indicators of impact. When each year sees a new reform, new resources, or new pedagogical strategies, and teachers are suffering from change fatigue, it is hard to use such tests to determine which factors are important in creating particular outcomes (whether positive or negative).

Even if one assumes they can provide useful diagnostics, concerns have also been expressed about the NAPLAN tests themselves:

The 2009 Masters Report argued there was a need to ensure students were more familiar with the nature of the tests (through exposure to previous papers), and that teachers and students spend time analysing responses to trial tests. In spite of improvements in 2009 results, and subsequent years (registered across all states, not just Queensland), this pressure has continued unabated. As per Masters' (2009) recommendations, this has resulted in a variety of strategies and approaches, including testing students' literacy and numeracy capacities more frequently, and spending more time on test-readiness activities (Hardy 2015, p379).

There are also concerns about the way NAPLAN results are used. The fear is that public school principals who are able to successfully engage with NAPLAN

use it to compete for resources and build the capacity of their school which is not to say that students or community are the beneficiaries. In an education system of finite resources and in an environment of competition and choice, this is most likely to be at the expense of other public schools (Bloxham 2015, p357).

Other authors have provided evidence that schools game the test system through ability grouping, allowing: targeted test preparation; encouragement of students to drop classes they are not performing well in; reducing the range of subjects offered; and tailored pedagogies and even whole school programmes and timetabling. This is all focused on improving test scores (Hardy 2015; Klenwoski 2014).

The NAPLAN tests have also been criticised for failing to measure the outcomes that are ostensibly desired from the Australian system, such as creativity, innovation and critical thinking, or those clearly valued by parents and the community such as artistic or sporting ability, social and community engagement (Klenowski 2014). It has been suggested that 'a preoccupation with national testing (outcomes) may distract attention from the adequate provision of a comprehensive and holistic education for every Australian child in every school, such as highly skilled and qualified teachers, high-quality instructional materials, facilities and a safe and supportive school environment' (Bloxham 2015, p357). Instead, the 'curriculum enacted by teachers in schools is narrowed as they increasingly teach to the test' (Fenwick 2017, p42). This, in turn, increases the likelihood of shallow or surface learning among students (Polesel et al. 2014)

As Klenowski notes,

Constraining pedagogy: greater accountability pressures are resulting in unintended effects of inhibiting and limiting teachers' pedagogic practices. The irony is that the context of increasing diversity of the student population together with equity demands implies the need for alternative, supportive, pedagogic approaches with a variety of assessment types. Quality, alternative, assessment practices are now more significant than ever to gain a more holistic view of the learner and more equitable outcomes for all students (Klenowski 2014, p446).

It seems that the importance accorded to the NAPLAN testing programme undermines the aims of the Assessment for Learning reform programme. Personalised Learning Plans – the individual plans mandated for indigenous and other potentially disadvantaged students in several Australian states and territories – have been suggested as a mean to ensure authentic, relevant and student-centred learning. However, one study, that analysed a sample of these, showed great variation in rigour and implementation (Klenowski 2014), suggesting they are no guarantee of side-stepping the influence of standardised testing. Indeed, the PLP relies on assessment of a student's achievements using benchmark data from basic skills test and NAPLAN results as well as classroom assessment and consultation with parents/carers.

The wide publication of NAPLAN scores and school rankings is also potentially problematic:

The parental perception of schools may be affected by high-stakes testing, with a resulting detrimental effect of parental market choice being exerted on schools that work in disadvantaged communities. Howe, Eisenhart, and Betebenner (2001) also note the detrimental effect of 'white flight' from low socio-economic schools in the USA, as a result of competition arising from the publication of results. Au (2008b,

501) describes the impact of the results of testing in terms of 'sanctions or rewards to students, teachers, administrators, schools, school districts and other official bodies charged with the education of children' (Polesel 2014, p641).

In addition,

Poorer than expected results could impact negatively or very negatively on media reports about a school, the school's reputation, parental perceptions of the school, the school's ability to attract and retain students, and staff morale (Polesel 2014, p651).

Altogether, the evidence presented here suggests that despite Australia's complex and multifaceted impact capture processes, the majority of stakeholders give most weight to the standardised testing system, NAPLAN. This is in many ways counterproductive given the aims and intended outcomes of education outlined above. However, given that there is an inherent tension between the different reform agendas currently being promoted within the system (i.e., accountability, autonomy, competence-based assessment of 21st century skills and standardisation), it is perhaps not surprising that most actors fall back on what is, in the end, the simplest metric to both obtain and interpret.

3.3 Korea

The Korean school system, like that of Singapore, is often held up as an example of a successful system during debates about education in other OECD countries. This is, in part, because of its good performance in PISA tests. Indeed, Korea was ranked first in science and second in maths in the PISA results in 2001. However, preceding this ranking, from 1999, there was much criticism of the public school system and it was suggested that it faced collapse (Kim 2004). It was also suggested that South Korea's success in the PISA rankings was not down to the school system, but to the high number of students that received private tuition and went to private academies (ibid.; c.f. Singapore case study above). Poor behaviour of students and teachers' lack of passion for their work were also raised as problems with the system (ibid.).

This contradictory state of affairs, with an education system that is externally lauded while internal commentary stirs up dissatisfaction and fears of imminent collapse, suggests that the processes used for impact capture in Korea are worth examining.

3.3.1 Context

While Singapore and Korea are often linked in discussions of school systems and performance in the PISA tests, there are major historical and cultural differences that impact on cultural attitudes to and expectations of education. Korea has existed in some form as a recognisable state entity for over two millennia; it has a rich and complex cultural history, including strong influences from China as well as a unique and highly patriotic founding myth/story. In the modern era, South Korea's character is arguably strongly influenced by its violent split from its northern neighbour, North Korea, and subsequent ongoing hostilities and uncertainties, along with its alliance with the USA. It is a large country with a population of over 76 million, as compared to Singapore's population of less than 6 million. Although Korea is often seen as economically successful, its GDP per capita is around \$27,500 per capita, substantially less than the figures of \$53,000 for Singapore and \$50,000 for Australia.

Korea's curriculum has traditionally been standardised and uniform, with the State regulating features such as teaching recruitment, training and pace of instruction (Park et al., 2011). Students are taught in heterogeneous, age-based groups; all receive the same standardised curriculum.

Korea has a historical tradition of testing in education. Its genesis is associated with Confucian philosophy and its 'belief that man is perfectible through education and only the most learned should govern the country and society' (Kwon et al. 2017, p61). The impact of this is seen through various stages in the development of Korea's education system. For example, education is expected to prepare people for the examinations needed to enter the civil service; the

popularity of this career path means that these examinations can come to shape the entire system.

As in Singapore, the purpose ascribed to education by the Korean state underwent a shift in the late 1990s, as global economic conditions changed. Educational reform policies between 1995 and 1999 promoted an ‘open education’ policy (Kim 2004). This included more student-centred learning, diverse teaching methods and after-school non-academic studies. Alongside this a “performance assessment” was introduced to replace existing quarterly, classroom-based pencil and paper tests. The intention of this new testing regime was to introduce an objective, standardised test, and hence to increase consistency within the education system. In 2000 there was a second change of emphasis in policy development, centring not on open education but on “improvement of teaching and learning”. The introduction of the 7th National Curriculum focused on achievement levels.

3.3.2 Inputs

South Korea spends an annual \$7,652 per student, below the OECD average of \$8,868. However, this represents 7.6% of South Korea's GDP, compared to the OECD average of 6.1%. This is the third-highest percentage of GDP spent on education among OECD countries (NCEE 2017) and is almost twice the fraction of GDP spent on education in Singapore. However, as in Singapore, an important input into the Korean education system is the financial contribution of parents and the input of private education provision via the mechanism of private, out-of-hours tuition. This has been recognised in the Korean policy document, Education for the Future (KEDI 2015) as well as several academic studies (Park et al. 2011; Ro 2017; Kwon et al. 2017).

Another factor that arguably provides a financial input into the education system is Korea’s policy on childcare. The state provides childcare to all parents of children in grades 1 and 2 to 5 pm, and if, the families are on low incomes or are a single parent family, until 10pm.

The policy reforms of the late 1990s introduced several factors that might be classified as inputs, for example new targets for class size (with a reduction to a maximum of 35 by 2002) (Kim 2004).

A renewed emphasis on teacher quality led to policies intended to make teaching and educational leadership more attractive professions. For example, school leaders are reported to have a high degree of autonomy in the development of curricula and provision, and in choosing students (OECD 2016) – although they are more constrained around the issues of recruitment of teachers and their pay. In 2016, in an effort to help maintain teacher morale, Korea instigated an opportunity for teachers of at least ten years’ service to take a leave of absence of up to a

year. Because this initiative has been positioned as enabling teachers to undertake training or prepare for retirement, it also has additional impact in terms of teacher quality through enhanced learning or lowered stress.

3.3.3 Processes

In recent years there have been a number of reforms to education processes in middle and high schools. Kwon et al. (2017, p62) list five particular changes:

- The 2009 curriculum revision and educational policy that emphasises education for creativity and character.
- The emphasis on school accountability for the results of the national-level achievement test given to students in the sixth grade of elementary school, third year of middle school and second year of high school.
- The evaluation of teachers and schools.
- Diversification of middle and high schools and expansion of the right to select high schools.
- Changes in policies for entrance to universities from a single test to a multiple assessment portfolio.

One national reform, in 2008, repealed the no-release policy of national test results to a 'restoration of broad-scale national testing and a new test release policy' (Sung and Kang 2012, p53). This was enacted in the climate of the re-emergence of global discourses of accountability in education, including the right to know students' grades and to be able to hold schools and teachers to account for their practices as measured by student success in these national tests.

After six years of primary education, Korean pupils take three years of high school before taking the College Scholastic Ability Test (CSAT), which determines which university or further education a student can attend. In 2008, the Ministry Of Education in Korea introduced a new policy on English education – the National English Ability Test (NEAT). This test aimed to broaden English education, which was directed towards the elements of the CSAT of listening and reading, to include speaking and writing. Secondly, it sought to reduce the pressure to pay for private English lessons by including two levels; one more practically-based and the other academic to cater for the different aspirations of Korean students (Lee and Lee 2016). The test was also an internet-based test. Lee and Lee (2016) suggest that this test was needed to counteract the influence of the CSAT on high test scores in English rather than good communication skills.

Twenty-one Meister schools, based on the ‘German method of training master craftspeople’ (OECD 2016, p9), were introduced in 2010, with the number growing to 40 in 2015. This was intended to increase the diversification of offer and uptake. The Meister schools are intended to be responsive to the needs of industry and so are (at least in principle) able to adapt their curricula; full scholarships and room and board are provided to students.

To introduce more elements of student well-being into their schools Korea has embarked on a ‘free semester’ in lower secondary education. This is a test-free semester where students can do different activities in different subjects. The apparently positive response in 2015 has meant that it was rolled out to all middle schools in 2016 (OECD 2016).

3.3.4 Desired outputs

As can be seen from the context, inputs and processes described above, it appears that the desired outcomes of the various educational policies and reforms include:

- Increased emphasis on outcomes not related to subject-specific content, such as creativity and character;
- Increased standardisation and accountability; and
- Diversification of educational offerings and a concomitant diversification of school-leaver destinations and career routes.

3.3.5 Impact capture

Korea uses several different strands in its approach to impact capture.

As indicated above, testing has long been an important part of Korean education culture. A National Scholastic Achievement Assessment started in 1998, aimed at capturing the impact of the recent reforms and ensuring national education quality. These tests sampled students on a bi-annual basis and, by 2002, incorporated 1% of students in grades 6, 9, and 10 and 100% of students in grade 3 (700,000 students) (Kim 2004, p130). The testing of all grade 3 students in reading, writing and arithmetic was justified as a way of identifying students who were not at the expected level. However, these results were not reported publicly as there was a concern that this would create a hierarchy of schools (ibid.).

The National Assessment of Educational Achievement test is set by the central government and is given to students in grade 9 and 11. This is used to look at trends and also identify students who are struggling. The CSAT is the most critical test, as discussed elsewhere in this briefing. Korea is currently introducing a new university assessment system which will let universities that perform best have a higher number of enrolments (OECD 2016).

As well as its own standardised tests, Korea makes use of the PISA international bench-marking tests to evaluate the impact of its educational policies and reforms. The OECD produced an Education Policy Outlook paper profiling Korea in 2016, with input and feedback from Korean professionals based in the OECD and the MoE and the Korean Educational Development Institute (KEDI). The highlights of this report state that Korea was a top performer in the OECD's 2012 PISA; the impact of socio-economic background also has less impact on PISA test scores than other countries (OECD 2016).

Another important impact capture measurement used by the Korean system is employment rates. For example, following the introduction of the Meister schools described above, their impact was deemed successful on the basis of over 90% graduate employment in 2014 (OECD 2016).

The Korean government also regularly reviews the system of education and education policies. The Ministry of Education evaluates the municipalities and the municipalities evaluate the local education offices and schools. Schools are assessed every one to three years by the local municipal education offices. Teachers are appraised annually. The Ministry of Education and municipal education offices prescribe the rules and process of the evaluation of teacher performance which is done at school level. Teachers are evaluated on both on their performance (by head teachers and other teaching colleagues) and on their particular speciality (by parents, students and colleagues) (OECD 2016). In 2010 a new teacher evaluation policy was introduced – Teacher Appraisal for Professional Development. This policy introduced of surveys of parents, students and teaching colleagues, including annual, anonymised online surveys conducted at the end of the school year, as well as classroom observations by principal teachers and other teaching colleagues.

3.3.6 Critical factors and concerns

There have been several studies which criticise elements of Korea's impact capture and QA processes for education, as part of wider criticisms of the education system (and attempted reforms of it).

As with other systems relying on high-stakes, standardised testing, the CSAT pre-university test has been criticised by several authors. It is argued that the primary purpose of school for the majority of Koreans is to get a good grade in this test to provide a route into the most prestigious universities and hence obtain a well-paid and respected job (Kwon et al. 2017; Ro 2017). Kwon et al. suggest that the CSAT 'has a strong washback effect on the entire education system, including teaching and learning in the classroom and students' perceived motivation for learning' (2017, p61). The importance of this one test is underlined by some of the practices

that occur on the day of the test, usually in early November. On this one day each year, 'the allocation of buses and subways are expanded to avoid traffic jams, students are escorted by police officers with the siren on, and even aircraft take-offs or landings are forbidden during the listening test of the English section' (ibid., p66). In the classroom, research suggests that there is much teaching to the test and whilst teachers may try to undertake practitioner inquiry, they do it in such a way that it does not affect their capacity to teach to the test also (Ro 2017).

The extreme importance accorded to the CSAT is one of the main drivers of the high uptake of private tuition, which KEDI acknowledges is 'a major social problem in Korea for the last half century' (KEDI 2015, p37). The amount of time students were studying beyond the school day prompted a response from the Korean Government to impose a 10pm curfew on private tuition and to enact the policy of the Education Broadcasting System (Kwon et al. 2017). Park et al. (2011) argue that in Korea's case private tutoring is a form of parental involvement as there is much money and time spent on this activity as there is from the state for official education. Park et al.'s (2011) research suggests that the private tutoring does make a difference to students' attainment in maths and English. They go on to suggest that the particular structural conditions of the Korean schooling system and, arguably the cultural context (Kwon et al. 2017), create the conditions where private tutoring is a highly valued commodity. Statistics Korea reports that 'the higher the household income level, the higher expenditures and participation rate' (2016, n.p.), resulting in a clear advantage for better-off families. The policy of providing a uniform education for all is thus undermined by parents' capacity to pay for private tuition.

Such is the cultural and social value placed on the CSAT test, the NEAT test, intended to mitigate some of the negative impacts of the CSAT test, was never successfully implemented. A number of factors led to its demise. The test was met with suspicion from parents, teachers and students. Teachers' own capacities to introduce the new aspects of speaking and writing to the classroom may be limited, where class sizes are still large and there are reported gaps in the English proficiency of some teachers (Lee and Lee 2016). There were also practical issues regarding material and financial and technical resources to administer the tests. Uncertainties about this new test led parents and students to indicate that they would turn to private institutions to ensure their proficiency, arguably undermining 'NEAT's goal to reduce spending on private institutes and decrease competition' (ibid., p831).

Indeed, although Korea does extremely well on both its own internal standardised tests and the PISA benchmark tests, it has been argued that such measures of educational impact completely fail to capture some important problems that are inherent in Korea's education system. Kwon

et al. (2017a) used an ethnographic approach to understand better Korean students' experiences of school. They suggest that, although the PISA results are excellent, the conditions faced by children and young people in schools are worrying. They identify three themes as particularly important: private education, human rights and school violence. Private education has been discussed above. Human rights violations come through corporal punishment, control of physical appearances and bag inspections. A further aspect, school violence, is particularly concerning as its intensity is said to be growing (ibid.). Kwon et al. describe how:

Overt school violence, by definition, take the form of physical and verbal assaults, bullying, stealing and sexual violence among students, both inside and outside school (KEDI, 2011). Recently about 2.4% of students in south Korea have thought about and tried to commit suicide, and the most critical reason given was depression caused by school violence (Korean Statistical Information Series, 2015 cited by Kwon et al. 2017a, p203).

The heavy emphasis on testing, coupled with increased workloads due to policy reforms such as the 7th National Curriculum, also impacts on teaching practices. Kim (2004) suggests that the diverse range of duties placed on Korean teachers, the assessment of their teaching based on a way of teaching particular texts, and the more specific regulations of the national curriculum constrains how and what they teach. It appears that teachers continue to adopt more traditional didactic teaching, with teachers themselves complaining that their teaching has become more 'technical' (p135), shaped by the national assessments, standardized assessments and pre-prepared texts and worksheets.

While it might be argued that the new teacher appraisal system introduced in 2010 has some merit, as it includes a wide range of stakeholders, it, too, has been subject to criticism. Kim and Young found that teachers are suspicious of the policy and its enactment. It appears that previous iterations of this policy concentrated on those teachers who are interested in promotion, while those with little further ambition have been able to avoid engagement. Indeed, Kim and Youngs note that Korean teachers have secured tenures and suggest, because of this, that 'the evaluation results had little influence on their professionals' lives' (2016, p740).

The surveys used in the appraisal process have also been criticised. The questions asked in the surveys tend to focus on 'teachers' instructional practice rather than their relationship with students, their attitude towards teaching or their performance on school operational tasks' (Kim and Youngs 2016, p736). There seems to be a disconnect between that which Korean teachers consider to be effective teaching, and the questions asked in the evaluation. Principals and teachers considered teacher-student relationships to be important, alongside the teacher's

attitude and management of their classroom (Kim and Youngs 2016), favouring these qualities over pedagogical skills and content knowledge. Moreover, these teachers' views of their duties were directed towards the organisation:

Korean teachers did not perceive themselves as 'individual instructors', but members of an organisation who have responsibilities in terms of the organisation's tasks. Also, tasks for school operation unrelated to 'teaching' were regarded as a major part of teachers' work in Korea (ibid., p735).

Even more significantly, the teachers in this study reported that 'the evaluation results had not affected their practice at all' (ibid., p738). Teachers saw the evaluations as a requirement rather than an opportunity for development. Teachers also tend to rate all their colleagues as 'excellent' (ibid., p726), thus neutralising the risk/weakness identification function of the appraisal process.

Kim and Youngs's analysis illustrates how trying to capture data on teacher performance can result in perverse incentives. In this case, the teachers felt secure enough to participate but in a way that undermined the purpose of the exercise. It seems the policy was introduced in a top-down way that took little regard of what Korean teachers value as effective teaching. The end result is a policy with little or no impact on practice but, arguably, an increase in teacher workload in the administration of the evaluations.

Taken together, the above criticisms and concerns suggest that despite Korea's apparently good performance in bench-marking tests such as PISA, there remain problems with its educational system that are not adequately captured by the metrics most heavily relied on for impact capture and QA.

3.4 Ontario

Canada's school system differs substantially from those described in the previous case studies, and yet has also been perceived as producing good outcomes for the majority of students that pass through it. Like Australia, Canadian culture has a history of individualism and meritocracy, coupled with a pioneering spirit and a special value on the Great Outdoors. Unlike Australia, however, Canadian culture (and Canadian education) is strongly influenced by the presence of its large neighbour to the south, the USA. This influence manifests itself in many ways, but particularly through a widespread resistance to what is seen as an overly-dominant capitalist, materialist and wealth-motivated culture. It has not, however, escaped the reform agenda that has spread through education throughout the OECD. It therefore provides an interesting context in which to examine alternative approaches to impact capture.

3.4.1 Context

Education in Canada is highly decentralised and there exists no federal department or office to oversee the standardisation of education across the country's ten federated provinces. Provincial governments determine educational policy, devise curricula and set funding levels for all levels of the educational system within their respective territories. On a more local level, elected school boards determine an annual budget, appoint staff and have some decision making in the implementation of policy. Provisional oversight resides in the cabinet level ministry of the Department of Education, which communicates policy and curriculum objectives to local districts. Unlike the Minister of Education, who is selected by the ruling parliamentary party, the deputy minister, who administers the bulk of the ministry's work, is a civil servant.

As Canada's most populous province, Ontario has a primary and secondary school system with almost 2 million students and a set of historical and developing challenges in obtaining high educational achievement. Historical issues include: language barriers, addressed by the 1982 Canadian Charter of Freedoms and Rights (Rights 1985); funding for religious schools, which received parity in the 1980s (Brennan 2011); and the entrenched structural inequality faced by Ontario's diverse indigenous communities (Cherubini et al. 2010). More recently, increased immigration has led to an influx of English language learners into the educational system and introduced new challenges and pressures that may arise from a multicultural classroom (Garnett 2010; Levin 2012; Webster and Valeo 2011).

Throughout the 1990s, successive governments implemented reforms aimed at increasing centralisation and standardisation; by the early 2000s, the Canadian vision of education was being driven by economic concerns, with a focus on the need to create a labour force with appropriate skills for work in the 21st century. The National Democratic Party government

created institutional organisations recommended by the Royal Commission on Learning (RCOL), primarily with the aim to increase centralisation and standardisation, such as: the Education Quality and Accountability Office (EQAO), tasked with implementing achievement tests, and a School Board Reduction Task Force to define systemic power structures and centralise education provision and monitoring.

Centralisation of the Ontario education system and neoliberal reform were markedly quickened by the Progressive Conservative (PC) government following their 1995 electoral victory. Vowing a 'common sense revolution', the PC government imposed deep cuts on provisional spending for school boards, with the aim of reducing their number to generate savings. They also passed legislation to centralise education by decreasing control of curriculum at the local level and empowering the EQAO to institute standardised achievement tests.

Friction with the teachers' union, poor policy implementation and a reputation for underfunding schools contributed to the defeat of the PC in the 2003 elections, in which education was a substantial issue (Sattler 2012). The incoming Liberal government, which ran as the education party, was tasked with restructuring funding for education and addressing the policy dissonance that had developed between the provincial ministry, local districts and the teachers' union. While the Liberal government inherited the institutions from the NDP and PC tenures, they promised to listen and work in tandem with education professionals and develop policies and effective intervention strategies to boost educational achievement. However, centralisation continued, and the EQAO was substantially reformed to shift more responsibility to the provincial level, ostensibly to relieve pressure on the grass roots of schools and teachers.

Regardless of the political party in power, over the last 30 years education in Ontario has followed a trend toward centralisation of control, through the standardisation of the curriculum and emphasis on provisional achievement levels. The standardisation of testing has increased in tandem to meet rising expectations, including the introduction of achievement tests which govern student progression and are thus seen as high stakes.

3.4.2 Inputs

In terms of financial inputs to the education system, Ontario has experienced an oscillation between their decrease and increase, according to the ideology of the governing party. In the late 1990s, the PC government introduced tax credits for children attending private schools, in efforts to create market pressure on public schools to improve performance and retain students.

The reforms introduced by the 2003 Liberal government were strongly influenced by a policy paper published by the Ontario Institute for Studies in Education (OISE), *The Schools We Need*:

Recent Education Policy in Ontario and Recommendations for Moving Forward (Leithwood, Fullan and Watson 2003). Suggested reforms included an emphasis on creating clear and coherent policy directives, more robust assessments and standards to cover a wider array of skills, returning a greater level of autonomy to the districts to implement policy and meet ambitious standards, working with the teachers to retain talent and develop skills and supporting evidence-based from data collection of standards achievement, as well as funding independent research to formulate new ideas.

The OISE position paper informed many of the immediate actions of the Liberal government, including a three year influx of \$1.6 million into the education system to promote student funding equity, removing the neoliberal market dynamic encouraged by private education subsidisation and restoring power to elected trustees in the province's largest districts. High schools received provincial money to hire a "Student Success leader," who could assess critical areas for improvement on a school basis and hire the appropriate team to facilitate improvement.

3.4.3 Processes

The Liberal government of 2003 implemented several policies recommended by the OISE through new initiatives, including class size reduction and flexibility to recruit assistance to address underperformance. A Literacy and Numeracy Secretariat was created, primarily with education professionals, to be a conduit between the education ministry and schools to implement initiatives.

Subsequently, the Ontario Focused Intervention Partnership (OFIP) was established to coordinate the approach of the ministry and school districts to target money effectively to lower-performing schools (Gallagher, Malloy and Ryerson 2016). As part of this, the Turnaround programme was introduced to improve literacy; the locus of professional development shifted to local school assessment and planning; and there were attempts to build expertise within boards and schools so that interventions could be developed for specific school challenges.

3.4.4 Desired outputs

The context, inputs and processes described above suggest that the main outputs and outcomes desired from Ontario's education system are:

- Better outcomes for English language learners and indigenous students;
- Reduced high school drop-out rates;
- More autonomy for schools;

- More consultation and influence from teachers; and
- Increased social equity.

The question then arises of whether Ontario's approach to impact capture and QA helps or hinders its progress towards these goals.

3.4.5 Impact capture

In the 1990s, successive Ontario governments created the highly centralised, institutional machinery required for assessing educational achievement across the province on a common curriculum. The Royal Commission on Learning (RCOL) was established to create a strategic vision for education in the province. The RCOL created the first comprehensive review of education in the province and generated *For the Love of Learning*, a report including 167 recommendations for reform (Bégin, 1994).

The increased formative and summative assessment envisioned in *For the Love of Learning* was implemented by the EQAO. From its inception, the mandate of the EQAO was to construct a system to administer standardised tests for elementary and secondary schools and collate and report accurate data to the provincial ministry and public (EQAO 2013). This mandate created a system in which the EQAO provided feedback to all levels of the education system, from the teacher up to the ministry, and encourage accountability at all levels via public discussion of results. It also provides an institutional mechanism by which education quality and effectiveness across all levels of hierarchy can be monitored. The EQAO now administers standardised tests at several levels: reading, writing and mathematics assessments to students in grades 3 and 6, mathematics competency in grade 9 and the Ontario Secondary School Literacy Test (OSSLT) in grade 10 (Volante 2007). Unlike most other provinces, the literacy test is critical for students, as it is a basic requirement for graduation and has a direct impact on graduation targets (Klinger, DeLuca and Miller 2008).

The Ontario government has also made use of graduation and retention rates as important indicators of the impact of educational reforms. For example, the fact that high school completion rates increased from 73% to 85% between 2004/2005 and 2011/2012 is seen as evidence for the positive impact of the reforms introduced by the Liberal government in the 2000s (Zegarac and Franz 2007). This was by no means a cheap affair, costing between \$3.5 billion between 2003 and 2008 alone, but this investment funded measurable parameters of student performance.

As mentioned in the Inputs section above, the reforms introduced by the Liberal government in the early 2000s included the creation of the role of Student Success leader, who was responsible

for assessing critical areas for improvement in each school (and who could hire the appropriate team to facilitate improvement). Thus impact capture was undertaken by someone who was effectively an outside expert, brought in to target lower-performing schools.

Like Australia, in addition to standardised testing and school-based evaluations, the Ontario government has also used one-off, commissioned reports to capture impact and monitor the effect and effectiveness of policy reforms and interventions. *A Double Cohort Study* (King 2002), funded by the Ministry in 2002, examined the implementation of various aspects of government reform.

3.4.6 Critical factors and concerns

The *Double Cohort Study* of 2002 found various negative consequences of Ontario's programme of educational reforms (King 2002). It found that the high demands of the new curriculum, and the high stakes of the OSSLT, produced 30% failure rates and placed undue stress and burden on students performing below provincial standards and 'effectively stripping away their motivation' (King 2002). Further challenges have developed for Canada's English language learners, where challenges in 'cultural and contextual aspects of reading and writing' have transformed the OSSLT into a competency test, rather than a literacy test (Cheng and Sun 2015). Although supplementary funding for second language support can help to improve student English language learners enrolment rates until graduation, which has worked for several Canadian provinces, including Ontario (Garnett 2010), it can be argued that the students would be benefited by maintaining standardised testing as a barometer for achievement, rather than discriminate against the advancement of English language learners (Ryan and Joong 2005).

Ontario teachers have also described unintended consequences of achievement assessment. The common curriculum details the content and expectant competencies that teachers should prepare their students for in a standardised test. Despite reforms from the Liberal government and the aim of positive pressure rather than punitive costs, the Canadian media has used the scores to draw sweeping conclusions about schools and districts that affect property values and increase parental pressure (Després 2013). EQAO administrators have also admitted candidly that the assessment can be used to coerce teachers into implementing certain aspects of the curriculum (Koch 2013). In extreme cases, some schools have moved inappropriately to improve scores and engaged in cheating (Pinto 2016).

Some corrective action has been taken to make teachers and schools comply with testing instructions and motivate students to take testing seriously, such as allowing testing scores to be incorporated as a percentage of final grades for grade 9 maths (Auditor General of Ontario 2009). There have also been performance audits when more nefarious manipulation is possible

(Pinto 2016). However, a sense of lingering problems lie in teachers' attitudes towards "management by spreadsheet", focused on supporting government reports rather than student and teacher interests (Pinto 2015).

Just as the use of high-stakes testing drives teacher and learner behaviour towards action that will produce good test results rather than student-centred education, other government reforms also had the unintended effect of undermining teachers' autonomy and professional expertise. With the Turnaround programme and the introduction of Student Success leaders, a large amount of impact capture and QA was in the hands of an external expert who worked with schools to analyse data and devise plans for improvement. This raised questions about the sustainability of both gains and monitoring/reform processes, when reliant on outside expertise (Gallagher, Malloy and Ryerson 2016). It was also felt to diminish the role of educators and question their decision making, despite their close attachment to school and challenges. Thus despite their professed intentions to work closely with educational professionals, the Liberal government responsible for this impact-capture-based reform effectively undermined teachers' sense of professionalism and autonomy (Gallagher, Malloy and Ryerson 2016). In response to teachers' concerns, the processes were adapted with a shift from external expertise to initial closer working with the schools to train in-house expertise to maintain a sustained skill set for assessment and corrective action.

The effort to combat school disengagement in the 2000s demonstrates the critical importance of working with educators to resolve the problem through policy, rather than relying on data and external expertise alone. Following on from King's wide ranging cohort studies (King et al. 2004), the province commissioned *Early School Leavers: Understanding the Lived Reality of Student Disengagement from Secondary School* investigate the risk factors for secondary school student disengagement prior to graduation (Ferguson et al. 2005). Despite the multiple risk factors the report identified, *Early School Leavers* suggested a simple remedy; every at-risk student needed a mentor, one that would be directly involved in their well-being, as a protective factor. This was the result of cooperative work including the authorities, teachers, and academics, and demonstrates the importance of such rich, qualitative research, involving the full range of stakeholders, in not only capturing impact but also diagnosing problems and developing solutions.

3.5 Finland

Of all European countries, Finland is most often held up as an example of a highly successful education system. In the UK, it has frequently generated headlines such as 'Finland's schools flourish in freedom and flexibility' (Guardian 2010). In 2006, it was ranked 1st among OECD countries in mathematical and scientific literacy and 2nd in reading literacy, and it has been the only country in which girls perform better in science than boys. However in 2015, while it had retained its ranking as 2nd in reading literacy, its absolute score had dropped by 21 points from 547 to 526, and it had dropped to 3rd in mathematical and 7th in scientific literacy. Despite what are still excellent scores, this has resulted in a change of tone in media reporting, with headlines such as 'Finland's schools were once the envy of the world. Now, they're slipping' (Washington Post 2016). A variety of reasons have been put forward for the apparent downward trend in performance, including poorer performance among boys, reduced reading for pleasure, and austerity policies following the economic crisis of 2008, resulting in cuts to education spending (Washington Post 2016). However, there have also been substantial changes to Finland's education impact capture and QA processes over the past twenty years. It thus provides another interesting context in which to explore the relationship between education policies, impact capture and attainment.

3.4.1 Context

Isaacs et al. (2015) have produced a detailed description of Finland's educational system. Below are some brief details to set the context.

Sahlberg (2012) suggests that Finland's education system has developed in three particular stages: 1945-1970 – the transition from agricultural to an industrialised nation and the enhancement of education for all; 1965-1990 – the Nordic welfare system and so the development of public comprehensive schooling; 1985-present – the improvement of the education system.

Teaching in the early years after WW2 was teacher-directed (ibid.) and there were a number of different types of schools (for example, grammar schools, civic schools, primary schools). However, in 1972 *peruskoulu* – the new comprehensive school system (ibid., p22) was starting to be rolled out across the country. This policy saw the end of the primary school to grammar or civic school streams.

Finland's education system is free at all levels. Compulsory education in Finland has a number of features that it promotes as 'special' compared to other countries. These include (FNBE 2016):

- Teaching is a very popular profession.
- Teachers feel valued by society.
- Work conditions for teachers are good: school time is relatively short (190 working days; 10 week summer breaks; 9 years of basic education = 6,300 hours per student), and class sizes are low (average=19).
- Since 1998, there has been no external school inspection regime.
- There are no formal, external teacher evaluations.
- There are no national exams (teachers assess pupils' learning: assessment outcomes at lower grades are written, while at upper grades they are represented numerically).
- The amount of homework is low.
- The number of lessons in art, music and physical education has been increased.

There has also been a strong focus on processes. It is notable, that while many education systems copy Finland's policies, in an attempt to emulate Finland's success, fewer adopt Finland's well developed structures to support sense-making and curriculum development (e.g. see Soini et al. 2017).

Children receive free pre-primary education at age 6 and start their basic education at age 7. At the age of 15, compulsory schooling comes to an end. Children are generally taught by the same class teacher for the first 6 years (up to 13) and then by separate teachers. Students with additional needs are taught in mainstream education but receive appropriate high-quality support to make this happen. Finland has two official languages, Finnish and Swedish, and some 5% of students attend schools in which Swedish is the main spoken language.

Funding for schools is part of the basic services monies that are transferred by the government to local municipalities. The municipality decides how it allocates this funding. It is based on the number of school-aged children (6-15) living in the area.

In terms of the time spent in school on certain subjects, schools in Finland have autonomy in how they use the allocated time (Creese et al., 2015).

3.4.2 Inputs

In terms of human resource inputs, the high status of teaching as a profession in Finland has meant that the education profession has been able to attract skilled, competent and motivated people. In fact, it is highly competitive to attain a teacher education place – more so than law or medicine (Chung 2016). This is often put forward as one of the reasons for Finland's success in PISA.

The Finnish education system also has a large range of financial and material resource inputs. State spending on education is above 6% of GDP – higher than the OECD average. Private schools are rare, but those that do exist (mostly religious schools) are granted the same amount of government funding as public schools, and in return are required to use the same admissions policies. Children in primary school receive free transport to school, a free school meal and textbooks. In secondary school, students continue to get a free meal but need to purchase their textbooks. Textbooks are chosen by the teachers (Creese et al. 2015).

In terms of curricular inputs, although education in Finland is fairly free and flexible, a new national core curriculum (produced by the Finnish National Education Board, FNEB, in 2014) was implemented locally in autumn 2016.

3.4.3 Processes

The Ministry of Education and Culture is responsible for education policy and the FNBE holds responsibility for the implementation of these policy aims. The latter develops, with the Ministry, the ‘educational objectives, content and methods for early childhood, pre-primary, basic, upper secondary and adult education’ (Ministry of Education and Culture 2017, p12). The intention is that these core curricula leave ‘room for local interpretation’ (Isaacs et al. 2015b, p9). Assessment is school-based until students take a national college entrance exam at the end of high school.

Finnish education processes focus on much more than the transmission of declarative knowledge or the acquisition of technical skills. Creese et al. (2016) suggest that Finland’s approach to skills for the 21st century is

... unique in that it includes such themes as cultural identity and internationalisation and responsibility for the environment. The Finnish National Board of Education expects teachers to incorporate broad cross-curricular themes such as active learning, technology and society, active citizenship and media skills into their instruction, without prescribing exactly how they are taught (Creese et al. 2016, p14).

The comprehensive and equal opportunity principles are particularly important in Finnish education. Sahlberg (2012) states that the priority given to these has led to three distinct features in the educational system. Firstly, they underline the belief that all students should be able to access and be successful in their learning. Because of this, there has been no strategy of ability groupings since 1985 (Creese et al. 2016). Secondly, counselling and guidance about career progression is a fundamental part of the school curriculum. Lastly, teacher education has been reformed with more emphasis on research and on teacher professional development; it is

now a master's level course (c.f. the moves towards Masters level professional status in Australia).

School choice in Finland has been part of the education system since 1998. Parents can choose a school based on the curriculum and any special character. There are limited private schools (Varjo and Kalalahti 2015) so parents choose within the publicly funded system. As a result, municipal schools offer specialised classes in certain subjects or themes and provide more study time in these areas. Students from across the municipality are selected for such classes on the basis of aptitude tests, rather than their academic achievements (Varjo and Kalalahti 2015).

The core principle of equity in education is also seen through Finland's inclusion policy. Children who have learning difficulties are provided with extra help, through 'part-time special education in small groups led by specialist teachers' (Creese et al. 2016, p17).

3.4.4 Desired outputs

Education is seen as an extremely important process in Finnish society, and not merely a means to an important end. Isaacs et al. suggest that, through education, 'Finland's goal is to create a democratic society, empowering individual students to create an egalitarian society' (2015b, p8).

Finnish Education in a Nutshell (FNAE 2017) details a number of goals of Finland's education system. These are (ibid., pp4-5):

- Equity in education – education is free at all levels and every student has the right to educational support
- Education system based on trust and responsibility – most education is publicly funded, educational autonomy is high and quality assurance is not based on control but steering.
- Early childhood and basic education as part of life-long learning – basic education provided in a single structure, timetables are local but school year the same everywhere, assessment part of daily school work.
- General and vocational pathways at upper secondary level – flexible system, most students continue
- Higher education with a dual structure – most university students graduate with a Master's degree, polytechnic students gain practical professional skills
- Highly educated teaching professionals – all teachers hold a Master's degree and continuing teacher education is encouraged

These desired features and outcomes are also prominent in the best practices highlighted in an OECD report produced for its *PISA for Schools* programme (cited in Lewis 2017).

3.4.5 Impact capture

In terms of using test data for impact capture, Finland has one national exam at the end of the basic education element (i.e. at the end of high school). This is preceded by some school-based testing and some sample testing.

Prior to 1998, Finland had a centralised school inspectorate that was responsible for impact capture and QA monitoring in Finnish schools. In 1998, the external system of school inspections was abolished and legislation was adopted which places a duty on those providing basic education and general upper secondary education to use self-evaluation as a way to reflect on their practices. There is no prescribed method for this; rather providers can choose their own focus and methods for evaluation and reporting of outcomes. The objective of this exercise is development and improvement of education. External control and monitoring has also been reduced to sample-based student assessment and municipal level inspection of schools as well' (Creese et al. 2016, p11).

In 2015, an evaluation was carried out by the Finnish Education Evaluation Centre (FINEEC) and the Ministry of Education and Culture of self-evaluation processes (FINEEC 2017). The objective of this was to produce a development plan to support providers in the implementation of self-evaluation and quality management methods and processes. The framework used to evaluate practices is based on management practices intended to help companies increase their quality and productivity (ASQ 2017).

3.4.6 Critical factors and concerns

According to Sahlgren (2015),

The standard policy explanations for [Finland's] rise include a focus on equity, with the comprehensive school reform of the 1970s as the bedrock, and the absence of standardised tests, accountability, and market reforms. Other explanations highlight comparatively little school and homework, and the country's current teacher education system (Sahlgren 2015, pii).

However, Sahlgren disputes these explanations and instead points to Finland's historical context, attributing the 2001 success to the centralised education system that continued in the 1990s and the dips in performance in PISA results in 2009 and 2012 to the economic and policy reforms of the late 20th century, including 'the traditional and teacher-centred educational culture ... being replaced by more pupil-led ways of working' (ibid., piv).

In relation to impact/QA processes that rely on testing, a study by Kasanen et al. (2009) suggests that children in Finland, who begin getting tested in class from grade 3 (9/10 years old), become less optimistic about their academic abilities as they progress through school. The study interviewed 58 pupils from 9 -12 years old (28 from grade 3 and 30 from grade 6). The grade 6 pupils were more uncertain about their academic ability than the grade 3 students. The authors related this to previous studies which suggest that the changing educational environment with its increase in formal testing, competition and normative feedback has an effect on how these pupils perceive their abilities (ibid., p36).

Finland's attitude to the public availability of impact capture and QA data could perhaps be described as ambivalent. Varjo and Kalalahti (2015) report on interviews with education officials on the '*policy of equalising*' (ibid., p324). This is about allocation of resources (including financial incentives) based on data about school populations and it is a way for local education authorities through their policies to ensure heterogeneous classes and control admission policies. Their study suggests this needs to remain out of the public sphere:

This is crucial in Finland, where in compulsory education national examinations for the whole age cohort are not held, and neither governmental organisations nor the mass media publish league tables. All the interviewees share the sentiment that in order to prevent league tables, test results and other school-based performance indicators must remain both confidential and for administrative purposes only (see also Kauko and Varjo, 2008) (Varjo and Kalalahti 2015, p324).

Moreover, politicians are also of the opinion that information about poor school performance should stay with the local education authorities:

As a member of the Board of Education, I don't expect the local education authorities to deliver a map of the weakest schools in our city to the Board meeting. That would just not be clever. (Interviewee #9 cited by Varjo and Kalalahti 2015, p324).

Reports of failing schools in the Finnish system in 2011(ibid.) have shaken the principles of the comprehensive system but these authors argue that the response has been to place more control at the local level to enable them to manage these concerns through their local policies.

In relation to the self-evaluation processes introduced at the turn of the century, the outcomes of the FINEEC evaluation of self-evaluation processes in the school sector (FINEEC 2017) highlighted that self-evaluation systems and systematic assessment cultures were found lacking in many providers. A number of key development needs were identified by this evaluation to support improvements in self-evaluation and quality management practices. This suggests that

self-evaluation, even in an environment where teaching is a highly-respected profession and teachers/managers are used to high degrees of autonomy, requires particular professional skills and knowledge that teachers may not initially have. Effective self-evaluation processes thus need to be introduced with appropriate guidance and support.

Lappalainen and Lahelma (2016), in their analysis of Finnish curricula documents regarding upper secondary education from 1970-2010s, suggest that the 'commitment to work on educational equality has been vague and unstable' (ibid., p666). They argue for more 'critical analyses on how equality and social justice are understood and argued in statements that oblige education providers' (ibid., p666).

Berisha and Seppanen (2017) consider whether ability grouping has entered Finland's system by the route of the selective school classes described in the Processes section above. They suggest that 'schools with a special emphasis form one type of ability grouping because of their selectiveness' (ibid., p249). They go on to suggest 'that selective school classes are attended more often by children from upper and upper-middle classes and the child's family background is an even stronger predictor than school performance' (ibid., p249). Varjo and Kalalahti argue this choice of school is controlled and arguably 'restricted by the local education authorities' (2015, p325).

These are the kinds of effect that Finland's impact capture and QA processes, with their strong emphasis on self-evaluation at the school level, are unlikely to identify.

A study by Modin et al. (2015) considered whether Finland's school success impacts on student health in terms of stress-related conditions. It used data from both Sweden and Finland and applied a two-level random intercepts model. In their conclusion the authors comment on the difference between stress-related complaints between 'differently' performing schools.

Students who attend (relatively) low performing schools appear to fare better from a health perspective than those who attend the more high performing school in Finland. These findings give further support to the notion that the undivided attention paid to learning in Finnish school may result in stressful comparisons between students with different levels of performance (Modin et al. 2015, p483).

As with the findings relating to equity and universality, these findings relating to student stress and optimism have not been reflected in the published results of official impact capture and QA processes. It may be necessary to adjust these processes in order to better understand the experiences of students and, indeed, parents.

The FINEEC (2017) report on self-evaluation highlighted a range of strengths and areas for improvement. These are reproduced in Tables 5 and 6.

Table 5: Key strengths of the self-evaluation process (adapted from FINEEC (2017))

- **Quantitative monitoring data and feedback are collected from a range of areas** in the operations, particularly from the management and the employees. The more developed the providers assessed their level of quality management and self-evaluation to be, the more comprehensive and systematic they reported (in 2010–2015) their collection of quantitative monitoring data and feedback, collected from different groups on several aspects, to be.
- **Persistent quality assurance efforts generate systematic and robust results.** The duration of quality management and self-evaluation shows a clear link to the level at which the provider sees their quality management and self-evaluation to be. Development is simply based on persistent practice: the provider must be committed to developing their quality management and self-evaluation over many years with clear targets in mind.
- **Evaluation increases community spirit and cooperation.** The outcomes indicate that networking, such as local or regional cooperation in planning and developing quality management, brings synergies to any quality assurance efforts, whether carried out by small or large providers. Such benefits can be seen, for example, in a systematic approach to operations; stability; established practice; inclusivity; sharing of responsibility; and consistency.
- **Development projects are worth the effort.** Some two thirds of the providers had implemented internal quality management or self-evaluation development projects in 2010–2015 and approximately half of the providers reported to have participated in regional or national development projects. The providers who had participated in such projects exhibited a higher level of quality management than others. The effectiveness of project activities was also reflected in the fact that, in many cases, the practices initiated in the projects had become an established part of the provider’s operations.
- **External evaluations bring benefits.** Most of the providers who had carried out the self-evaluation (85.7%) considered it useful. For example, the evaluation helped them to form

an overall picture of their situation, identify development needs, and draw attention to self-evaluation and quality management practices. The self-evaluation in itself developed the activities and gave the participants time that they could spend together.

Table 6: Development needs for the self-evaluation process (adapted from FINEEC (2017))

- **Differences in leadership** – quality management cannot succeed without visible commitment. The outcomes pinpointed major differences in how well the providers ensured management commitment to quality management. In the group where the development of quality management remained at a rhetorical level, weak leadership in quality management was highlighted, as well as a lack of methods to ensure that the management is committed to quality management and continuous development and improvement.
- **Prerequisites for self-evaluation and quality management were found to be insufficient.** The weakest aspect of quality management among the providers was related to the prerequisites, which were widely considered to be insufficient (at the emerging level). Attention was particularly drawn to the deficiencies in evaluation skills and the allocation of time by the management and employees to quality assurance and evaluation efforts. None of the providers assessed their prerequisites to be advanced.
- **Self-evaluation does not necessarily result in improvement and development.** The evaluation indicated that the effectiveness of evaluation efforts can appear somewhat weak: collection and analysis of evaluation data may fail to influence decision making, development measures, and monitoring of the effectiveness of development measures.
- **Room for improvement was found in quality management documentation.** The outcomes demonstrated a clear link between the number of quality management documents and the level of quality management: respondents who did not document their activities were found at the lowest level, while those who could provide a variety of documentation were at the highest level.
- **Problems in Swedish-speaking teaching and education were highlighted in units with small pupil and student numbers.** Swedish-speaking providers with less than 500 learners had evaluated their level to be lower than that of the providers who had more

learners. The self-evaluations of small Swedish-speaking providers were also found to be less coherent and uniform than those of larger Swedish-speaking providers.

- **Inclusivity could be increased** and responsibilities shared not only among the management and the employees. According to the evaluation outcomes, school employees participate and are included in the development of quality management and self-evaluation to a moderate extent (developing), while learners, guardians and stakeholders only have emerging-level inclusion in these activities. Responsibility for the activities is largely assumed by the management; on average, the sharing of responsibility for quality management and self-evaluation was assessed to be at the emerging level.

Arguably, a strength of the Finnish approach is that issues such as those highlighted above are at least recognised, if not fully addressed. Indeed, the Finnish system is not static, and in response to a rapidly changing global environment the Finnish government is currently proposing six key reforms. These reforms are:

1. **New learning environments and digital materials to comprehensive schools**
This reform promotes digital learning and learning outside the classroom and suggests that new pedagogical approaches are needed to support this. These will be developed through continuing professional development. Moreover, Finland has identified inactivity as an issue within its child population and has set a goal of an hour of activity a day for its secondary school students. It has a *Schools on the Move* project running which promotes being active in lessons rather than sitting and at break times but not increasing formal P.E. time.
2. **Reform of vocational education**
Acknowledgement here of the financial squeeze on educational resources and so a change on the market model used for vocational education. A streamlining of VET is proposed including a move from a supply-orientated model to a demand-driven model, examination of the education provider network and decreasing the number of qualifications available, but increasing the qualification content.
3. **Development of tertiary education**
There is an aim to increase the length of working careers in Finland. Hence, HE will reform their selection processes so that students take less gap years between upper secondary and HE. Also HE will offer education all year.

4. Access to art and culture

Increased opportunities for all children to art and crafts to help foster creativity and recognise the benefits to well-being.

5. Cooperation between higher education and business life will be strengthened to bring innovations to the market

The aim is to promote the export of Finnish educational resources and research.

Moreover, from August, 2017 Finland will introduce tuition fees for HE to students outside the EU and EEA.

6. Youth guarantee to support young people

Finland is guaranteeing a place to study for all students finishing comprehensive schools. For young people who do not go on to study support is offered through coaching workshops, outreach youth work and an advisory service.

Clearly, the impact of such reforms will need careful evaluation. It will be interesting to see how the Finnish state goes about this process.

4. Summary and conclusions

There are several lessons to draw from the findings described above.

The framework adopted in this report, drawn from the context-input-process-output model put forward by Scheerens et al. (2007), provides a useful lens for the analysis of impact capture and QA systems. It is also likely to provide a good basis for the design of impact capture systems, since it draws attention to many of the features of an education system that one might need to pay attention to, but allows for flexibility of focus and emphasis and thus tailoring to meet the evaluation needs of specific policy reforms or pedagogical interventions.

The case studies presented in Section 3 highlighted a range of features and potential problems.

Our analysis of Singapore's system suggests that impact capture and QA systems need to be well-aligned with desired educational outcomes. If they are unchanged through a period where intentions are substantially and qualitatively reframed, they may serve to constrain the system so that it cannot effectively shift to the new paradigm. In Singapore's case, we see how the retention of standardised tests and the weight accorded to performance in PISA drive system behaviour away from the desired change towards prioritising creativity and criticality. As Ng suggests,

There is now a dilemma between public accountability based on a standardised framework and creating a platform to allow institutions to break through the existing educational paradigm. There may also be dilemmas of how far decentralisation should take place and whether the institution's self-evaluation and validation reports be made public to increase "customer" information and choice. (Ng 2008, p117).

This problem, however, raises the question of how educational outcomes can be compared when they are fundamentally different. If the currently-used standardised tests were to be abandoned, critics would likely discount apparently good performance against any new approach to measuring outcomes as the system awarding itself a gold star. The question of whether ideologically neutral metrics can be defined – that is whether different educational paradigms can be made commensurate – remains unanswered.

Our analysis of the Australian system highlights the difficulty of untangling multiple factors and ascribing any kind of causal relationship, even with sophisticated, multi-faceted impact capture/QA systems, when several reforms are implemented at once. It also highlights the need to openly recognise the ideological origins of reforms in order to identify when they are likely to be in tension (as with the market choice and equity-through-standardisation agendas). It also

shows how what are intended to be resources and guidelines to support policy implementation can have the effect of removing teachers' capacity to make decisions and teach creatively, even in the context of a strong discourse of autonomisation.

The Korean case exhibits similar problems to Singapore, with the huge importance accorded to tests undermining a professed desire to shift towards so-called 21st century skills. In the Korean case, this is made worse by the persistence of a national exam which acts as gate-keeper to the civil service, and the cultural and financial desirability of entering this profession. Arguably, attitudes to education will only change if the social context changes such that either civil service jobs are no longer seen as desirable, or if recruitment to those jobs changes dramatically, to explicitly value creativity and criticality above declarative/content knowledge.

This case study also highlights how impact capture/QA measures can fail to capture important elements of school experience (in this case, symbolic and real violence), which are likely to have a profound effect on both attainment in existing tests and students' present and future lives. In fact, understanding the quality of student life is one of the aims of the PISA tests. Yoon and Jarvinen (2016) analysed PISA (2012) data directly relating to student's quality of and compared the responses of students in Korea and Finland. Their findings showed that whilst Finnish students' outlooks tended to be more positive than Korean students, their views on the relationships with their teachers was less positive. This suggests that not only are more nuanced measurement processes necessary, but so too are more nuanced approaches to the analysis and reporting of existing measurements.

Our analysis of Ontario's system, like that of Australia, illustrates difficulties in capturing impact when policies and reforms are in tension, especially when standardised tests are (or at least are perceived to be) high stakes. It also highlights the central importance of engaging stakeholders and listening to those who experience the classroom environment on a daily basis. It suggests that however well-intentioned processes may be, if they appear to discount the views and experiences of teachers and rely on external expertise, there will always be both resistance/resentment and a danger of misunderstanding local contexts. However, when teachers' professionalism is used as an input to process-development, the results can be extremely positive.

The Ontario case also highlights the benefits of in-depth studies that combine quantitative and qualitative data, especially longitudinal/cohort studies and those with an ethnographic element. The value of longer-term studies raises the question of timescales in an era where reforms are introduced with every new election and every new minister of education.

Finally, our analysis of the Finnish system presents a case where system reform has been accompanied by reform of the impact capture/QA system. It may be that the slight slippage in PISA results reflects a decreased relevance of those tests, since they focus on academic literacies rather than the creation of democratically- and socially-empowered citizens. It seems clear that some of the key strengths of the Finnish system are the high levels of teachers' autonomy and their input into both the design and implementation of impact capture/QA processes. Sahlberg (2011) compares the developments in education of many other countries to Finland and suggests that Finland does not have any elements of the GERM that has been so influential in, for example, Australia. This does not imply that the GERM is without merit, but it does perhaps imply that a good educational system can be created using alternative policies.

Overall, it appears that, despite the existence in some cases of apparently complex and rich processes for impact capture, undue weight ends up being placed on single metrics, often the results of national or supra-national standardised tests. These, which are inevitably focused not only on cognitive skills and declarative knowledge but also, in their generality, reflecting some kind of lowest common denominator, cannot reflect the diversity of learning and teaching that make up a healthy education ecosystem.

Our case studies also suggest that cultural influences also cannot be understated; as the Finland and Korea experiences show, the status of the teaching profession and the respect it is accorded within the community is an important factor. So too are teacher pay and conditions. If teachers are to be asked to take on the extra burden of conducting regular self-evaluation, it needs to be recognised and accounted for as a formal part of their roles.

In addition, the renewed emphasis on so-called 21st century skills such as creativity and criticality raises new questions/problems for competence-based or comparative testing. For one thing, there is not widespread agreement in the literature about what critical thinking, or creativity, are, and how they can/should be taught, let alone how they are manifested in ways that could be subject to the kind of scrutiny required for impact capture. As Mills (2016) puts it,

Neoliberal educational paradigms at odds with high performance/high-stakes testing for schools; narrow accountability regimes; a lack of clear communication between the 'centre' and schools; a de-professionalisation of teachers; auditing mechanisms and partisan political interference into educational processes. These factors all work against one of the common features common to the curriculum in 'high-performing' countries, a concern with the competencies required for the twenty-first century (Mills 2016, 129).

One important concern, already raised by several authors and emerging again from our analyses above, is about the dangers of policy-borrowing without considering context, both in relation to educational policies and interventions themselves and to the systems put in place to capture their impact.

A major reason to be cautious is the difficulty in understanding the timescale for impact of particular policy reforms. Given the kind of inertia (and even resistance) described in some of our case studies, it is unlikely that current measurement outcomes reflect current policies, or at least current policies alone. Le Grand (in Sahlgren 2015) points to the precarious nature of international policy comparisons, suggesting that:

... the main problem with the traditional explanations of the Finnish education “miracle” was that they originated from the idea of “best practice”, an approach that highlights current arrangements in high-performing countries as the key determinant without adequate consideration of whether these are causally linked to performance. Consequently, the policy lessons drawn from this approach are not particularly reliable and might in fact do more harm than good (Le Grand in Sahlgren 2015, pvi).

In relation to policy-borrowing, several authors have raised concerns about the current dominance of the OECD’s voice, as promoted through the PISA and PISA for Schools programmes, in shaping education systems around the world. Neglecting the cultural and historical roots from which an education system develops is problematic when considering how to transfer certain policies from one 'successful system' to another. As Lewis suggests,

... such forms of policy borrowing enable the governing of local schooling through the OECD's definition of what works, rather than necessarily promoting meaningful and sustainable policy learning in the light of philosophical considerations around the purposes of schooling. Even while the dominance of the OECD's policy 'voice' does not necessarily prevent alternative discourses around schooling accountability and performance from being promoted, its status as the global authority does arguably restrict the ways in which educators and policy-makers might otherwise understand what is meant by 'effective' education (Lewis 2017, p298).

By sounding this note of caution about unquestioningly accepting the validity of the OECD’s definition of good quality education, Lewis prompts other questions. For example, we might ask why we should accord more authority to a supra-national organisation whose primary mission is to promote economic growth and financial stability over the voices of teachers, students, parents and local communities, whose primary objective may be to promote learning, well-being and empowered forms of social cohesion in their local context.

Equally, we might return to the question raised in relation to Singapore: can (and should) the success of education systems operating in different contexts and with differing goals and intended outputs be measured on the same criteria? Lewis also addresses this point, describing how 'enumerations of education performance help to produce commensurable spaces of measurement, in which different schools and/or schooling systems can be known, compared and evaluated by reference to a common metric, such as PISA' (Lewis 2017, p286), even if such metrics do not reflect the "true" value of the system as perceived by students, teachers and the wider community. Lewis expresses concern that,

... the ensuing 'evidence' from such comparative assessments have now occupied a central importance in educational policy-making, at both the school and schooling system level, and especially when these forms of evidence are used to legitimate looking abroad to, and borrowing from, (supposedly) more successful systems (Rappleye 2012). These processes, for the ostensive purpose of becoming world class, have become so normalised as to become what Auld and Morris (2014) have termed the 'New Paradigm', which is underscored by three assumptions. Namely, these are that (1) the aims and outcomes of different schooling systems are directly commensurable; (2) system performance on such comparative testing is directly correlative to future economic success; and (3) causal factors are universal and absolute (134–135). Importantly, these assumptions all overlook the numerous problems inherent in promoting both best practice and policy transfer (Lewis 2017, p286).

The condensation of complex, contextually-determined policies, reforms and processes into "best practice" by organizations such as the OECD is thus 'extremely problematic, with PISA for Schools inducing a more normative policy borrowing than an educative and contextually informed policy learning' (Lewis 2017, p297).

Rather than relying on PISA alone, the experiences in the nations described above suggest that detailed, context specific, reflective impact capture is much more informative, when there is sufficient emphasis on inputs and processes. However, it is costly, both in terms of direct financing and in terms of time for both operational implementation and time to reflect, interpret and act. It requires adequate training and support, especially when being introduced into contexts where reflective practice and collaborative working are not already norms. It also requires reasonable timescales for evaluation, and therefore patience and the courage to risk failure, which is perhaps one of the main reasons that policy-makers, the media and others tend to fall back onto policy-borrowing and ranking by PISA – if one doesn't, then one has to take responsibility for making policy and practice changes that could have detrimental impacts on learners' current and future lives.

References

- ASQ. (2017). *W. Edwards Demings 14 Points for Quality Management*. Available online at <http://asq.org/learn-about-quality/total-quality-management/overview/deming-points.html>
- Auditor General of Ontario. (2009). *Annual report of the office of the Auditor General of Ontario*. Available online at <http://www.auditor.on.ca/en/content/annualreports/arbyyear/ar2009.html>.
- Bégin, M. (1994). *For the love of learning: report of the Royal Commission on Learning: a short version*. Ontario: The Royal Commission on Learning.
- Berisha, A. and Seppänen, P. (2017). Pupil selection segments urban comprehensive schooling in Finland: composition of school classes in pupils' school performance, gender, and ethnicity. *Scandinavian Journal of Educational Research*, 61(2), 240-254.
- Brennan, T.-L. K. (2011). Roman Catholic schooling in Ontario: Past struggles, present challenges, future direction? *Canadian Journal of Education*, 34(4), 20-33.
- Caldwell, B.J. (2016). Impact of School Autonomy on Student Achievement: Cases from Australia. *International Journal of Educational Management*, 30(7), 1171-1187.
- Cheng, L., and Sun, Y. (2015). Interpreting the impact of the Ontario Secondary School Literacy Test on second language students within an argument-based validation framework. *Language Assessment Quarterly*, 12(1), 50-66.
- Cherubini, L., Hodson, J., Manley-Casimir, M., and Muir, C. (2010). Closing the gap at the peril of widening the void: Implications of the Ontario ministry of education's policy for Aboriginal education. *Canadian Journal of Education*, 33(2), 329-355.
- Chung, J. (2016) The (mis)use of the Finnish teacher education model: 'policy-based evidence-making'? *Educational Research*, 58 2), 207-219.
- Creese, B., Gonzalez, A., and Isaacs, T. (2016) Comparing international curriculum systems: the international instructional systems study. *The Curriculum Journal*, 27 1), 5-23.
- Deng, Z. and Gopinathan, S. (2016). PISA and high-performing education systems: explaining Singapore's education success. *Comparative Education*, 52 (4), 449-472.
- Després, S. A. (2013). *Real Accountability Or an Illusion of Success?: A Call to Review Standardized Testing in Ontario*. Available online at: http://testingillusion.ca/wp-content/uploads/2013/01/illusion_of_success_EN.pdf

- Dilkes, J., Cunningham, C. and Gray, J. (2014). The New Australian Curriculum, Teachers and Change Fatigue. *Australian Journal of Teacher Education*, 39(11), Article 4.
- Dimmock, C. and Yong Tan, C., (2013). Educational leadership in Singapore: Tight coupling, sustainability, scalability, and succession. *Journal of Educational Administration*, 51(3), 320-340.
- EQAO. (2013). *EQAO: Ontario's Provincial Assessment Program Its History and Influence, 1996-2012*. Toronto: Queen's Printer for Ontario.
- EU (2016). *Comparative study on quality assurance in EU school education systems Policies, procedures and practices*. Directorate-General for Education, Youth, Sport and Culture, Education, Youth, Sport and Culture (European Commission). DOI: 10.2766/422920.
- Eurydice. (2015). *Assuring Quality in Education: Policies and Approaches to School Evaluation in Europe*. European Commission/EACEA/Eurydic Report. Luxembourg: Publications Office of the European Union.
- Fenwick, L. (2017). Promoting assessment for learning through curriculum-based performance standards: teacher responses in the Northern Territory of Australia. *The Curriculum Journal*, 28(1).
- Ferguson, B., Tilleczek, K., Boydell, K., Rummens, J. A., Cote, D., and Roth-Edney, D. (2005). Early school leavers: Understanding the lived reality of student disengagement from secondary school. *Final Report submitted to the Ontario Ministry of Education*. Available online at:
<https://www.tcdsb.org/FORSTAFF/NewTeacherInduction/Documents/schoolLeavers.pdf>
- FINEEC (2017). *The evaluation of the quality management and self-evaluation practices of basic education and general upper secondary education providers*. Helsinki: Finnish Education Evaluation Centre. Available online at
https://karvi.fi/app/uploads/2017/01/KARVI_0417.pdf (note: includes a summary in English).
- FNAE (2017). *Finnish Education in a Nutshell*. Helsinki: Finnish National Agency for Education. Available online at
http://www.oph.fi/download/146428_Finnish_Education_in_a_Nutshell.pdf.
- FNBE (2016). *Compulsory education in Finland*. Helsinki: Finland National Board of Education. Available online at
http://www.oph.fi/download/180148_Compulsory_education_in_Finland.pdf

- Gallagher, M. J., Malloy, J., and Ryerson, R. (2016). Achieving excellence: Bringing effective literacy pedagogy to scale in Ontario's publicly-funded education system. *Journal of Educational Change*, 17(4), 477-504.
- Garnett, B. (2010). Toward understanding the academic trajectories of ESL youth. *Canadian modern language review*, 66(5), 677-710.
- Groundwater-Smith, S., Mitchell, J. and Mockler, N. (2016). Praxis and the Language of Improvement: Inquiry-Based Approaches to Authentic Improvement in Australasian Schools. *School Effectiveness and School Improvement*, 27(1), 80-90.
- Guardian (2010). Finland's Schools Flourish in Freedom and Flexibility. The Guardian, 05/12/2010. Available online at <http://www.guardian.co.uk/world/2010/dec/05/finland-schools-curriculum-teaching>.
- Hardy, I. (2015). Education as a "Risky Business": Theorising Student and Teacher Learning in Complex Times. *British Journal of Sociology of Education*, 36(3), 375-394.
- Hogan, D., Chan, M., Rahim, R., Kwek, D., Maung Aye, K., Loo, S.C., Sheng, Y.Z. and Luo, W., (2013). Assessment and the logic of instructional practice in Secondary 3 English and mathematics classrooms in Singapore. *Review of Education*, 1(1), 57-106.
- Isaacs, T., Creese, B., and Gonzales, A. (2015) *Aligned Instructional Systems: Cross-Jurisdiction Benchmarking Report*. Available online at: <http://ncee.org/wp-content/uploads/2015/09/AIS-Cross-Jurisdiction-Report.pdf>
- Isaacs, T., Creese, B., and Gonzales, A. (2015b) *Aligned Instructional Systems: Finland*. Available online at <http://ncee.org/cieb/instructional-systems-study/>
- KEDI. (2015). *Education for the Future*. Korea: Korean Education Development Institute.
- Kerkham, L. and Nixon, H. (2014). Literacy Assessment That Counts: Mediating, Interpreting and Contesting Translocal Policy in a Primary School. *Ethnography and Education*, 9(3), 343-358.
- Kim, J.W. (2004). Education reform policies and classroom teaching in South Korea. *International Studies in Sociology of Education*, 14(2), 125-146.
- Kim, J. and Young, P. (2016). Promoting instructional improvement or resistance? A comparative study of teachers' perceptions of teacher evaluation policy in Korea and the USA. *Compare: A Journal of Comparative and International Education*, 46(5), 723-744.
- King, A. (2002). *Double Cohort Study–Year 2 Report*. Toronto: Ministry of Education and Training.

- King, A. J., Warren, W. K., Boyer, J., and Chin, P. (2004). *Double cohort study: Phase 4 report*. Ontario Ministry of Education. Toronto: MoE.
- Klenowski, V. (2014). Towards fairer assessment. *The Australian Educational Researcher*, 41(4), 445-470.
- Klinger, D. A., DeLuca, C., and Miller, T. (2008). The Evolving Culture of Large-Scale Assessments in Canadian Education. *Canadian Journal of Educational Administration and Policy*, 76, 1-34.
- Koch, M. J. (2013). The Multiple-Use of Accountability Assessments: Implications for the Process of Validation. *Educational Measurement: Issues and Practice*, 32(4), 2-15.
- Kwon, S.K., Lee, M. and Shin, D. (2017). Educational assessment in the Republic of Korea: lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice*, 24 (1), 60-77.
- Kwon, S., Kristjánsson, K. and Walker, D.I. (2017). Misery in dark shadows behind the high achievement scores in South Korean schooling: an ethnographic study. *Educational Review*, 69(2), 201-217.
- Lappalainen, S. and Lahelma, E. (2016). Subtle discourses on equality in the Finnish curricula of upper secondary education: reflections of the imagined society. *Journal of Curriculum Studies*, 48 (5), pp. 650-670.
- Lee, H. and Lee, K. (2016). An Analysis of the Failure(s) of South Korea's National English Ability Test. *Asia-Pacific Educational Research*, 25(5-6), 827-834.
- Leithwood, K., Fullan, M., and Watson, N. (2003). *The Schools We Need: Recent Education Policy in Ontario and Recommendations for Moving Forward*. Ontario: Ontario Institute for Studies in Education.
- Levin, B. (2012). System-wide improvement in education. *Education Policy Series*, 13.
- Lewis, S. (2017). Governing schooling through 'what works': the OECD's PISA for Schools. *Journal of Education Policy*, 32(3), 281-302.
- Liem, G.A.D., Lau, W.K. and Cai, E.Y.L. (2016). Promoting Mastery-Approach Goals to Support the Success of the "Teach Less, Learn More" Educational Initiative. In *Building Autonomous Learners* (pp. 277-302). Singapore: Springer.
- Mills, M. and McGregor, G. (2016). Learning not borrowing from the Queensland education system: lessons on curricular, pedagogical and assessment reform. *The Curriculum Journal*, 27(1).

- Modin, B., Karvonen, S., Rahkonen, O. and Östberg, V. (2015). School performance, school segregation, and stress-related symptoms: comparing Helsinki and Stockholm. *School Effectiveness and School Improvement*, 26(3), 467-486.
- NCEE. (2017). *South Korea: System and School Organization*. National Center on Education and the Economy. Available online at <http://ncee.org/what-we-do/center-on-international-education-benchmarking/top-performing-countries/south-korea-overview/south-korea-system-and-school-organization/>
- Ng, P. T. (2008). The phases and paradoxes of educational quality assurance: The case of the Singapore education system. *Quality assurance in Education*, 16(2), 112-125.
- Ng, (2010). The evolution and nature of school accountability in the Singapore education system. *Educational Assessment Evaluation and Accountability*, 22(4), 275.
- OECD. (1998). *Education at a glance*. Paris: Organisation for Economic Cooperation and Development.
- OECD (2016). *Education Policy Outlook: Korea*. Accessed on 30th June. Available online at: www.oecd.org/edu/policyoutlook.htm
- Park, H., Byun, S. and Kim, K. (2011). Parental Involvement and Students' Cognitive Outcomes in Korea: Focusing on Private Tutoring. *Sociology of Education*, 84(1), 3-22.
- Pinto, L. E. (2015). Fear and loathing in neoliberalism: School leader responses to policy layers. *Journal of Educational Administration and History*, 47(2), 140-154.
- Pinto, L. E. (2016). Tensions and fissures: the politics of standardised testing and accountability in Ontario, 1995–2015. *The Curriculum Journal*, 27(1), 95-112.
- Polesel, J., Rice, S. and Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: a teacher perspective from Australia. *Journal of Education Policy*, 29(5).
- Priestley, M. R. (2017). *Approaches to Specifying Curriculum Areas of Learning*. Paper prepared for the groups developing the Welsh Successful Futures AoLEs, May 2017.
- Rights. (1985). Canadian charter of rights and freedoms, 183-229. Toronto: Carswell
- Ro, J. (2017). A journey from the classroom to the world of educational reform: a study of three Korean teachers' practitioner inquiry. *Asia Pacific Journal of Education*, 37(1), 28-41.
- Ryan, T. G., and Joong, P. (2005). Teachers' and Students' Perceptions of the Nature and Impact of Large-Scale Reforms. *Canadian Journal of Educational Administration and Policy*.
- Sahlberg, P. (2011). The fourth way of Finland. *Journal of Educational Change*, 12(2), 173-185.

- Sahlberg, P. (2012). A Model Lesson: Finland Shows Us What Equal Opportunity Looks Like. *American Educator*, 36(1),
- Sahlberg, P. (2014). *Finnish lessons 2.0: What can the world learn from educational change in Finland?* Cambridge, Ma.: Teachers College Press.
- Sahlgren, G.H., (2015). Real Finnish Lessons. *The True Story of an Education Superpower*. London: Centre for Policy Studies.
- Sattler, P. (2012). Education Governance Reform in Ontario: Neoliberalism in Context. *Canadian Journal of Educational Administration and Policy*.
- Scheerens, J., and Hendriks, M. (2004). *Benchmarking the quality of education*. European educational research journal, 3(1), 101-114.
- Scheerens, J., Luyten, H., and van Ravens, J. (2011). Measuring educational quality by means of indicators. In *Perspectives on educational quality* (pp. 35-50). Springer Netherlands.
- Soini, T., Pietarinen, J. & Pyhältö, K (2017). Shared sense-making in curriculum reform – transforming the core curriculum to the local school development initiatives. Paper presented at the *European Conference on Curriculum Studies*, Stirling, 17 June, 2017.
- Statistics Korea. (2016). *Private Education Expenditures Survey in 2016*. Available online at <http://kostat.go.kr/portal/eng/pressReleases/1/index.board?bmode=read&aSeq=359882>.
- Thijs, A. and van den Akker, J. (2009). *Curriculum in Development*. Enschede, the Netherlands, SLO.
- Varjo, J. and Kalalahti, M. (2015). The conceivable benefits of being comprehensive – Finnish Local Education Authorities on recognising and controlling the social costs of school choice. *European Educational Research Journal*, 14(3-4), 312-330.
- Volante, L. (2007). Educational Quality and Accountability in Ontario: Past, Present, and Future. *Canadian Journal of Educational Administration and Policy*, 58, 1-21.
- Washington Post. (2016). *Finland's schools were once the envy of the world. Now, they're slipping*. Washington Post 8/005/2016. Available online at https://www.washingtonpost.com/local/education/finlands-schools-were-once-the-envy-of-the-world-now-theyre-slipping/2016/12/08/dcf0f56-bd60-11e6-91ee-1adddf36cbe_story.html?utm_term=.df02a6780f97
- Webster, N. L., and Valeo, A. (2011). Teacher preparedness for a changing demographic of language learners. *TESL Canada journal*, 28(2), 105-128.

Yoon, J. and Järvinen, T. (2016). Are model PISA pupils happy at school? Quality of school life of adolescents in Finland and Korea. *Comparative Education*, 52(4), 427-448.

Zegarac, G., and Franz, R. (2007). *Secondary school reform in Ontario and the role of research, evaluation and indicator data*. Available online at:

<http://www.edu.gov.on.ca/eng/research/SSreform.pdf>