How far is the Global Education Reform Model globally effective?
A comparison of the relationships between student attainment and school accountability, autonomy and institutional diversity within and across national education systems

Colin Gill

A Research & Development Project Submitted for the MSc Educational Assessment 2020

DEPOSIT AND CONSULTATION OF THESIS

One copy of your dissertation will be deposited in the Department of Education Library via the WebLearn site where it is intended to be available for consultation by all Library users. In order to facilitate this, the following form should be completed which will be inserted in the library copy your dissertation.

Note that some graphs/tables may be removed in order to comply with copyright restrictions.

Surname	Gill
First Name	Colin
Faculty Board	Education
Title of Dissertation	How far is the Global Education Reform Model globally effective?

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I understand that the Department requires that I shall deposit a copy of my dissertation in the Department of Education Library via the WebLearn site where it shall be available for consultation, and that reproductions of it may be made for other Libraries so that it can be available to those who to consult it elsewhere. I understand that the Library, before allowing my dissertation to be consulted either in the original or in reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of this thesis belongs to me. I permit limited copying of my dissertation by individuals (no more than 5% or one chapter) for personal research use. No quotation from it and no information derived from it may be published without my prior written consent and I undertake to supply a current address to the Library so this consent can be sought.

3. I agree that my dissertation shall be available for consultation in accordance with paragraph 2 above.

How far is the Global Education Reform Model globally effective ? A comparison of the relationships between student attainment and school accountability, autonomy and institutional diversity within and across national educational systems.

Abstract

The Global Education Reform Movement (GERM) originated in the UK and US four decades ago and its prescriptions are now advanced by international development organisations as the means to raise student attainment across the world. GERM is associated with the promotion of external school accountability through the publication of assessment results, school autonomy, and school diversity to increase parental choice and bolster school independence. Aspects of GERM have been widely adopted across many member countries of the OECD and research in these countries has reported some modest gains in student attainment, but overall findings have been ambivalent. Research investigating the effectiveness of GERM in developing countries has mostly focused on single aspects of the model, such as autonomy, yet has also reported inconclusive and contradictory findings. The present study aimed to address some of the limitations of previous research by making transnational comparisons of the effectiveness of GERM using a balanced sample of non-OECD countries from the 2015 PISA cycle, which were drawn equally from four global regions and from different levels of development, and all of which had implemented GERM related reforms; within this sample the effectiveness of GERM was also investigated at an individual country level in Trinidad & Tobago, the Former Yugoslav Republic of Macedonia and Indonesia. Multilevel modelling was used to explore the associations between external accountability, school autonomy and school diversity and student maths attainment, alongside the variables of accountability to parents, internal accountability (school tracking by regulators), educational material shortages, teacher certification and principals' leadership effectiveness which have also been proposed as factors relevant to student attainment. The main findings suggested that external accountability was modestly associated with higher levels of student attainment in high and middle income countries, but not in the least developed countries. Internal

accountability showed positive associations with student attainment in low income countries. School autonomy was not associated with student attainment in developed countries, but showed a positive association with attainment in middle income countries, and an association with lower levels of attainment in the least developed countries. School diversity did not appear to be associated with higher student attainment in developed countries, but in middle and low income countries students attending independent schools appeared to have markedly lower levels of attainment than peers at public schools. Educational material shortages were associated with lower levels of attainment in middle and low income countries, but not in the most developed countries. Teacher certification showed no associations with student attainment, and principal leadership effectiveness was only associated with student attainment in Indonesia. Overall, the associations between the GERM related variables and student attainment were modest, inconsistent and accounted for a relatively low proportion of the variance in student attainment between and within countries. This, and the fact that other variables appeared to be more salient to attainment in some countries, suggested that GERM might not provide a consistent or comprehensive template for global education reforms to raise student attainment. Therefore, although some of the measures promoted by GERM might be associated with higher student attainment in some countries, it was concluded that GERM related reforms might not be globally effective.

Introduction

Since the early 1980s many nations have adopted a range of educational reforms to promote school accountability, local autonomy and parental choice, a set of measures that Sahlberg (2016) has called the Global Education Reform Movement (GERM). GERM is widely agreed to have begun in the United States and United Kingdom in the early 1980s (Deming & Figlio, 2016) as a reaction to stagnation in educational attainment levels caused, it was argued, by an earlier series of educational reforms (Walford, 1990). The UK led the way with the 1988 Education Reform Act (ERA) which introduced large-scale changes to schools in England and Wales, in the process establishing a template for GERM that the rest of the world followed (Levin & Fullan, 2008). ERA was intended to drive improvements in student attainment by holding schools externally accountable through the publication of standardised test and national examination results, delegating decision making to schools through school based management (SBM), and diversifying school provision to increase parental choice of schools and create an educational market (Bergbauer, Hanushek, & Woessmann, 2018; Epple & Romano, 1998; Figlio & Loeb, 2011; Friedman & Friedman, 1982; Gintis, Cox, Green, & Hickox, 1991; O'Day & Smith, 1993; Walford, 1990; Wößmann, Lüdemann, Schutz, & West, 2007). Elements of GERM have been adopted by many member countries of the Organisation for Economic Co-operation and Development (OECD), and GERM has been actively promoted beyond the OECD by a wide range of multi-lateral educational organisations such as the International Association for the Evaluation of Educational Achievement (IEA), World Bank, Asian Development Bank (ADB) and Inter-American Development Bank (IDB) (Grimmett, 2018; Klees, 2008, Shahjahan, 2016; Verger, Novelli, & Altinyelken, 2018b). The growth of germ has been paralleled by the emergence of international large scale assessments (ILSAs) which have been used to make transnational comparisons of student attainment (Schleicher, 2014, 2007; Winthrop & Simons, 2013), and to inform national policy debates (Baird et al., 2016; Hopfenbeck et al., 2018; Lenkeit, Chan, Hopfenbeck, & Baird, 2015), with the OECD Programme for International Student Assessment (PISA) attracting the most extensive global coverage

(Hopfenbeck, 2016) and taking a preeminent position in the media, academic and policy literature (Lingard & Sellar, 2016; Sellar & Lingard, 2014).

As GERM has become the international framework within which educational policy is decided (Verger, Novelli, & Altinyelken, 2017) it is timely to assess the extent which the primary prescriptions of GERM may actually be globally effective. Therefore, the present study reviews the evidence for the effectiveness of GERM in OECD countries, then examines the model's effectiveness in countries outside the OECD, with more detailed reviews of the evidence available for Trinidad & Tobago, the Former Yugoslavian Republic of Macedonia (FYROM) and Indonesia. These countries have been selected because they represent countries from different global regions, at different levels of economic development, with very different educational systems which have all implemented internationally led GERM-related initiatives, as well as their own domestically led educational reform programmes, and which have been the subject of scholarly interest. The analyses that follow use PISA data for a sample of non-OECD countries which have all implemented GERM based reform programmes and which are balanced across four global regions and levels of economic development, to contrast variables related to GERM inspired reforms and variables from non-GERM related initiatives, and their associations with student attainment.

GERM Within the OECD

School Accountability

In GERM school accountability is "the process of evaluating school performance on the basis of student performance measures" (Figlio & Loeb, 2011, p.384), and these results should be made public because this external accountability provides a simple means to summarise the impact of all other forms of educational accountability, such as political, bureaucratic, professional, moral, and market accountability (Adams et al., 2017). The school is the focal unit because it is assumed that schools determine student attainment and that when they are held externally accountable through published

assessment results they will make whatever changes are necessary to improve student attainment (Geijsel, Krüger, & Sleegers, 2010) so under the aegis of GERM "assessments and accountability are known for 'spreading like wildfire'," (Antoni Verger, Parcerisa, & Fontdevila, 2019, p.260)

The UK published its first ERA inspired school league tables of standardised test and exam results for England and Wales in 1992 (Leckie & Goldstein, 2017), but Scotland never introduced standardised testing and only published exam league tables from 1998, then in 2001 both Scotland and Wales ended league tables (Lemke & Zhu, 2018; Mattei, 2012b). Mattei (2012a) found that external accountability in Wales raised student attainment more steeply than in England and led to positive 'spillover' in subjects not directly contributing to schools' league table positions; in Scotland student attainment increased while external accountability was practised, but after abolition performance remained flat. Goldstein & Leckie (2016) found that the abolition of external accountability led to "a real decline" (p.373-374) in Welsh students' attainment. Burgess, Wilson, & Worth (2013) found that during the period Wales published league tables the performance of students in Welsh and English schools converged, but that once Wales stopped practising external accountability English students' attainment increased more rapidly than that in Wales, leading them to conclude that publishing attainment data is an "extremely cost-effective policy for raising attainment" (p.67).

Carnoy and Loeb (2002) reported similar findings in the US where student scores on the National Association of Educational Progress (NAEP) maths test were significantly higher in those states practising external accountability. In 2001 the US introduced No Child Left Behind (NCLB) which made standardised testing and the publication of school league tables of results mandatory (Linn, Baker, & Betebenner, 2002). Dee & Jacob (2011) contrasted the performance of students on NAEP reading and maths tests before and after the implementation of NCLB, drawing samples from states that had test-based external accountability systems before NCLB, states that did not, and private schools unaffected by any form of external accountability. They found that NCLB did not affect student reading

performance, but that maths attainment increased. In similar studies other US researchers have reported that external accountability significantly raises student attainment (Burgess, 2017; Hanushek & Raymond, 2005; Woessmann, 2009).

Elsewhere the evidence for the effectiveness of external accountability is less clear. In Australia education is the responsibility of individual states but there is also a national testing regime and state exam results are published, yet there seems to have been little consistent impact on student attainment (Marks, 2015; Smith, Parr, & Muhidin, 2019). South Korea started publishing national standardised assessment results in 2008 (Sung & Kang, 2011), but there does not appear to be a relationship with student attainment (Yi, 2015). In Chile school league tables appear to have done little to improve student attainment and it has been suggested they are simply an "approximate ranking based on socioeconomic status" (Mizala & Urquiola, 2013, p.313). Even in the US it appears that external accountability has little effect on high performing schools (Chiang, 2009; Figlio & Rouse, 2006). Nevertheless, of the 65 different educational jurisdictions that participated in the PISA 2012 cycle, 46 had implemented some form of national-level policy for test-based school accountability, though in contrast to the UK and US few countries use external accountability by making the results publicly available or compiling league tables (Yi & Shin, 2018).

Rather than external accountability many countries make accountability data directly available only to parents (Kelly & Kotthoff, 2017; Voldby Beuchert & Brink Nandrup, 2014), often combined with some form of internal accountability whereby school performance data are monitored by a government department or agency against a performance framework and used alongside inspections and other methods of oversight to improve student attainment (Ehren, Altrichter, McNamara, & O'Hara, 2011; Mattei, 2012a). It is noteworthy that students in countries like Estonia, Finland, Ireland and Switzerland which practice forms of internal accountability have consistently performed very well in ILSAs and have generally outperformed jurisdictions like England and the United States which

practise external accountability and this has led some scholars to suggest that external accountability is unnecessary to improve student performance (Holloway, Sørensen, & Verger, 2017; Ladd, 2017), though others have contended that internal accountability systems are actually most effective when the results are made public (Ehren et al., 2015).

School Autonomy

In GERM based reforms school autonomy and accountability are inseparable as "when autonomy and accountability are intelligently combined, they tend to be associated with better student performance," (OECD, 2019, p.306) because it is assumed that when schools are able to exercise sufficient autonomy they will respond to accountability pressure to make whatever changes are needed to improve student attainment (Arar & Abu-Romi, 2016; Bandur, 2012; Cheng & Mok, 2007; Greany, 2018; Klein, 2017; Nir, Ben-David, Bogler, Inbar, & Zohar, 2016). In a study of 1.26 million students attending English government schools granted substantial autonomy as academies, Eyles and Machin (2019) found that within three years of the schools becoming more autonomous student attainment in GCSE exit exams increased significantly even when controlling for socio-economic and other factors, and these improvements were accompanied by a range of positive spillover effects raising attainment in neighbouring schools (Eyles, Hupkau, & Machin, 2016; Machin & Vernoit, 2011). Similar findings have been reported for other OECD countries (Clark, 2009; Falch & Fischer, 2012; Fuchs & Wößmann, 2008; Hahn, Wang, & Yang, 2018; Jeong, Lee, & Cho, 2017; Lastra-Anadón & Mukherjee, 2019; West, Allmendinger, Nikolai, & Barham, 2010) and it is assumed that SBM is effective at raising attainment because it makes schools more responsive to their communities (Gertler, Patrinos, & Rubio-Codina, 2012) and enables them to innovate to improve teaching and learning (Bandur, 2012; Hooge, Burns, & Wilkoszewski, 2012).

However, other researchers suggest that greater independence may lead to a lack of oversight resulting in poorer educational standards (Lubienski & Lubienski, 2014; Malin, Hardy, & Lubienski,

2019; Mbiti, 2016) and point out that the benefits of autonomy are dependent on the effectiveness of principals' leadership (Leithwood, Harris, & Hopkins, 2008) and the broader frameworks within which schools operate (Benton, 2014). Nevertheless, in developed countries a broad political consensus in support of school autonomy has emerged (Christ & Dobbins, 2016) and GERM inspired SBM has become widespread (Allen & Burgess, 2012; Barrera-Osorio, 2006; Caldwell, 2016a, 2016b; Nir, 2002).

School Diversity

The third element of GERM is diversity of school provision to promote parental choice. Some scholars directly link institutional diversity with autonomy on the grounds that there is a continuum from complete government control to full autonomy and the closer a school is to total independence the more flexible, distinctive and responsive it will be (Cheng, Trivitt, & Wolf, 2016; Shakeel & Deangelis, 2017, 2018; Trivitt & Wolf, 2011). Since the 1980s several countries have experimented with various models of school choice and diversity. In England this has taken the form of self-governing schools directly funded by central government and at various times known as grant maintained schools, foundation schools and most recently as academies (West & Bailey, 2013). In Chile and Sweden diversity of school provision has been encouraged through the introduction of school vouchers which have enabled parents to choose from a range of government and independent schools (Böhlmark & Lindahl, 2015; Elacqua, 2012), and in the US vouchers have been accompanied by the introduction of charter schools which are independent of local school board control (Epple, Romano, & Zimmer, 2015).

Alongside the improvements in student attainment reported for English academies above, similar increases in student attainment have been reported in Sweden (Böhlmark & Lindahl, 2015), and in the US where increases in student attainment in independent schools have also been accompanied by positive spillover effects into conventional public schools (Cordes, 2018; Foreman, 2017; Griffith,

2019). However, Hsieh & Urquiola (2006) found that independent schools in Chile performed no better than public schools, though Mizala & Romaguera (2000) suggested that some independent schools did perform slightly better, but only because parents were willing to supplement vouchers with additional donations or discretionary fees. In contrast, Henríquez, Lara, Mizala, and Repetto (2012) showed that when independent schools exercised their autonomy to introduce internal reforms then students drawn from the most deprived communities outperformed students in public and other independent schools.

In a large transnational analysis Deangelis & Shakeel (2017) estimated that a 1% increase in the proportion of students enrolled in schools not governed, controlled or managed by government was associated with a 1.6 point increase in PISA maths scores and a 1.2 point increase in reading scores. It has been suggested that these positive effects are a direct result of competition between schools (Egalite, 2013) and a large metanalysis of 92 studies assessing the impact of school diversity found that "competition resulting from school-choice policies does have a small positive effect on student achievement" (Jabbar et al., 2019, p.22). However, in most western countries such competition is not unconstrained because it takes place within policy frameworks which prioritise student recruitment from particular geographical areas and preclude selection on the basis of prior scholastic attainment (Böhlmark, Holmlund, & Lindahl, 2016; Dee, Jacob, & Schwartz, 2013; Eyles et al., 2016).

Thus, it appears that independent schools may have more autonomy to make internal reforms which may raise attainment, and that these changes may spill over to public schools (Department for Education, 2014; Gleason, 2017). This leads many proponents of school diversity and choice to argue that a mix of government public and independent schools is the most effective way to spur competition leading to improvements in the quality of all schools benefitting all students (Deangelis & Erickson, 2018; Wohlstetter, Smith, & Farrell, 2013).

The evidence so far adduced suggests that external accountability, autonomy, and school diversity are all associated with increased student attainment in some OECD countries, though the evidence is not unambiguous. Nevertheless, the OECD and other international agencies have promoted GERM-related measures to non-OECD countries so the obvious question arises: to what extent has GERM been effective beyond the OECD ?

GERM Beyond the OECD

GERM has perhaps been most widely promoted through the World Bank's Systems Approach for Better Education Results (SABER) initiative which has launched more than 100 projects to promote assessment and school accountability in developing countries (Clarke, 2012; World Bank, 2016).

School Accountability

Alongside the World Bank the OECD, IEA, ADB, IDB and other multilateral organisations have encouraged developing countries to implement ILSAs and domestic testing programmes to promote external accountability, but the available evidence suggests that these projects have had only a small effect on educational policy (Wagner et al., 2012) and there is no evidence suggesting ILSAs or domestic testing have had any direct impact on schools and student attainment (Felipe Barrera-Osorio, Bertrand, Linden, & Perez-Calle, 2011; Liberman & Clarke, 2012; Tobin, Lietz, Nugroho, Vivekanandan, & Nyamkhuu, 2015). Even the World Bank's own SABER project reviews suggest that no accountability project has had any impact on student attainment (Takeda et al., 2014). This may be because external accountability can only be effective as part of a broader accountability system located within a clear standards framework against which improvement can be specified and measured – in effect, an internal accountability system (Altrichter & Kemethofer, 2015). Such frameworks do not simply translate across different societies or cultures (Ehren et al., 2015; Hopmann, 2008) and many developing countries do not have the resources for a full internal accountability system (Mbiti, 2016). However, in developing countries even a very rudimentary system can make a

difference to school performance (Ehren, Eddy-Spicer, Bangpan, & Reid, 2017) and may provide a viable alternative to external accountability (Hopmann, 2008). Some of the practical reasons why external accountability systems may meet with little success beyond the OECD might best be explored by briefly reviewing how GERM related reforms have performed in the three very different non-OECD countries of Trinidad & Tobago, FYROM and Indonesia.

Trinidad & Tobago is an anglophone, high income country with an educational landscape which is essentially that of 1950s Britain with a diverse range of public, independent and private schools (De Lisle, 2012). In the mid-1990s student attainment stagnated (Louisy, 2004) so Trinidad & Tobago launched a GERM based Secondary Education Modernisation Plan (SEMP) from 1999-2009 with the support of the IDB (Younger & Cobbett, 2014). Despite plans for an external accountability system, standardised testing was never introduced and instead a National Certificate of Secondary Education (NCSE) was launched on the grounds that, as it included continuous assessment from K-9 to K-12, with final exams in K-12, it would make schools continuously accountable (Haywood, 2019). However, existing secondary school leaving examinations, the Caribbean Secondary Education Certificate (CSEC), remained in place and most students, teachers and parents considered the NCSE to be a low stakes assessment worthy of little attention or effort (Haywood, 2019). Furthermore, the Ministry of Education (MoE) did not compile NCSE statistics and kept CSEC data in a separate and incompatible database so it was impossible to track student performance or even compare the data easily, effectively undermining the stated purpose of the NCSE (Haywood, 2019). It appears that this happened because the NCSE was developed by MoE to satisfy the IDB (Joseph, 2020) when in fact teachers, parents and politicians were content with the existing examination system, internal and parental accountability, and informal external accountability through the local media (De Lisle, 2015; Joseph, 2010).

FYROM is a middle income European country which emerged from communism in the early 1990s and then suffered the shock of civil war in 2001 leading to polarisation between the majority Macedonian and minority Albanian, Turkish and Serbian populations (Rizova, 2011). Nevertheless, FRYOM joined the World Bank's SABER programme and took part in the IEA 2001 Progress in International Reading Literacy Study (PIRLS) and the 2003 Trends in International Mathematics and Science Study (TIMSS). The very poor results actually helped to begin the process of reconciliation at governmental level as national education policies and practices were critically evaluated and reshaped (Gilmore, 2005; Naceva & Mickovska, 2007). In 2008 attendance at secondary school was made compulsory for the first time (Lyon, 2013) along with a secondary level national exit exam, the *Matura*. Prior to 2008 school leavers would be granted a leaving certificate based on teacher grades which determined access to limited tertiary education places (Murchan, Shiel, & Mickovska, 2012), and this system was open to widescale abuse as students and parents exerted pressure on teachers to secure higher rankings (Grozdanovska, 2007). The government made the Matura a centrally controlled and marked exam with the explicit aim of providing a more objective certification of students' achievement (World Bank, 2012a). However, FYROM did not introduce a national system for external accountability; instead a set of domestically led reforms funded by the European Union introduced a national framework for internal accountability (Murchan et al., 2012). This approach was taken to build school capacity and avoid publicly highlighting performance differences between different language medium schools and possible variations in the standards required for university entrance which could have caused communal tensions (Zendeli, 2015).

Indonesia is a low income developing country occupying the world's largest archipelago with an abundance of different languages, and regional and ethnic identities (Mcnicoll, 2011). Indonesia has participated in several initiatives promoted by the World Bank, ADB and OECD which included reforming the public examinations taken at the end of secondary school, but the results are shared only with parents and are not used in any formal external accountability processes (Harris, Jones,

Adams, Sumintono, & Ismail, 2017). Instead of GERM inspired external accountability, Indonesia introduced its own reforms to develop internal accountability systems and has used PISA to create a national multilogue about student attainment (Stacey, 2011) that has avoided potentially divisive comparisons of attainment across regions or ethnic groups and has thus facilitated discussions about national solutions to raise student attainment (Sulistiani & Zulkardi, 2019).

Thus, in three very different countries it appears that GERM based external accountability systems were rejected because key stakeholders supported existing local approaches to accountability, governments wished to avoid drawing attention to differences between schools which had the potential to be nationally divisive, and because central administrators wanted to prioritise school capacity building and conformity with national standards through internal accountability.

School Autonomy

The evidence that school autonomy may have positive effects on student attainment in some developing countries is at best mixed (Brutti, 2015; Carr-Hill, Rolleston, & Schendel, 2016). Using PISA data over several cycles Hanushek, Link, & Woessmann (2013) suggested that school autonomy may actually reduce student attainment in developing countries, though it should be noted that school autonomy was assessed at country-level based on estimates of national decentralisation rather than on direct measures of autonomy in schools. In a large meta-regression estimating the impact of school autonomy on student mathematics and reading attainment across middle and low income countries in Africa, Asia and Latin America, Carr-Hill, Rolleston, Schendel, & Waddington (2018) found that insignificant estimates outnumbered small positive estimates by more than three to one, and they concluded that school autonomy had little effect because of differences in the ways in which school autonomy was implemented at school level. Gunnarsson, Orazem, Sánchez, & Verdisco (2009) also found no consistent link between autonomy and student attainment in eight South American countries and again highlighted that the effectiveness of school autonomy depends upon the extent

to which schools choose to exercise autonomy irrespective of central policies or regulations. These conclusions have been echoed by researchers in Thailand (Patrinos, Arcia, & Macdonald, 2015) and Kenya (Bold, Kimenyi, Mwabu, Ng'ang'a, & Sandefur, 2018) and the World Bank itself has found that in projects to promote school autonomy it is actually the leadership of individual principals in exercising autonomy that is the most significant factor in raising student attainment (Takeda et al., 2014). This may explain the findings of Hanushek and colleagues (2013) because their estimates of school autonomy were based on published national policies, not on the actual implementation of those policies at school-level, so many of the schools in their sample may simply not have been exercising the levels of school autonomy which had been granted.

In Trinidad & Tobago the SEMP attempt to introduce school autonomy through SBM within a new regulatory framework imposed and monitored by the Ministry of Education (MOE) met with considerable resistance (De Lisle, 2012). There were concerns that SBM might actually result in imposing more central control and regulation leading to the paradoxical situation where a scheme intended to raise school autonomy would actually reduce it and principals would be "left with the task of individually defining their roles with regard to the functioning of their schools and meeting the expectations of the public and the MOE." (Brown & Conrad, 2007, p.185). There were also concerns that SEMP might extend into independent schools the government terms and conditions of employment which meant, for example, that public "school leaders … were powerless to discipline even the most recalcitrant teachers" (Brown & Conrad, 2007, p.188). Ultimately Trinbagonians were content with the *status quo ante* and their independent schools remained independent, rather than becoming 'autonomous' (De Lisle, 2012).

In FYROM the SABER initiative was used to fund a programme of educational decentralisation in 2005 that devolved responsibility for schools to 85 separate municipal governments (Bakiu, Dimitrova, & Brava, 2016). Although this may have appeared to be a policy inspired by GERM, domestically it was

intended to de-escalate tensions by de-politicising the education system so that local communities could provide education in the language of their choice (Bakiu et al., 2016). Furthermore, this apparent increase in autonomy was accompanied by the introduction of a national performance framework intended to make schools comply with national procedures and standards through the internal accountability systems mentioned above (Zendeli, 2011, 2015).

In Indonesia responsibility for public schools and their funding was devolved to local government control in 2000 (Behrman, Deolalikar, & Soon, 2002; Channa, 2016) with SBM and school governing bodies introduced in 2005 (King & Guerra, 2005). "SBM was seen as a panacea" (Sumintono, 2010, p.41), but it was implemented without any clear guidelines (Toi, 2010) leading to considerable confusion within schools, and between schools and local administrators (Vernez, Karam, & Marshall, 2012), so government officials have continued to play a very large part in the management of schools and it is not clear that public schools have really become autonomous (Hutabarat, 2015; Jasmina & Oda, 2018; Nababan, Suyradi, & Hanafi, 2019). More worryingly, it has been suggested that decentralisation has facilitated corruption to the serious detriment of public schools (Suryadarma, 2012), leading some to suggest that school 'autonomy' has had negative effects (Jasmina & Oda, 2018; Sumintono, 2010).

Thus, it appears that the capacity of school autonomy to raise student attainment in non-OECD countries may depend upon the characteristics of individual schools and even the character of individual principals. Where schools already have independence GERM inspired models of autonomy may be seen as restrictive. In educational systems where there is a need to develop national standards and frameworks policy-makers may see establishing conformity as more important than autonomy, and in the wrong conditions implementing SBM may not be a 'panacea' and may even reduce school performance.

School Diversity

In some developing countries dissatisfaction with low levels of student attainment in public schools leads many poor families to sacrifice a large proportion of their income to send their children to private or independent schools where levels of attainment appear to be higher (Mizala & Torche, 2017) and the number of such schools is increasing rapidly (Baum, Cooper, & Lusk-Stover, 2018; Kim & Mobrand, 2019). However, despite the fact that it is assumed that independent and private schools can increase levels of student attainment in low income countries (Maulana & Yudhistira, 2019; Morgan, Petrosino, & Fronius, 2015), to date there is a scarcity of robust, large scale primary studies to confirm the point (Null, Cosentino, Sridharan, & Meyer, 2017).

In Trinidad and Tobago there is considerable school diversity with 51% of schools independently operated by religious denominations and receiving state funding, 37% public schools wholly government funded, and the remaining 12% privately funded, usually for profit, receiving no significant state support (George, Mohammed, & Quamina-Aiyejina, 2003; Ministry of Education, 2011). As part of SEMP it was proposed that school admissions should be centrally organised along GERM inspired, western lines which would have brought to an end the existing system whereby parents could nominate three schools from anywhere on their home island with no zoning or residential requirements (De Lisle, Keller, Jules, & Smith, 2009). Given the denominational nature of many schools there was inevitably a degree of self-sorting, but unlike most OECD exemplars some schools also included a selective element based on primary school exit exam results (De Lisle et al., 2009). Trinabogians viewed their system of school choice as transparent, competitive and fair, with places at what were perceived to be the best schools available to all students on an equal basis. As with other SEMP measures, attempts to introduce a centrally regulated school admissions system that Trinabogians saw as restrictive failed (De Lisle et al., 2009).

In FRYOM private schools were first permitted in the post-war constitution, but there was no provision for independent schools funded by the state (Lyon, 2013). This was a conscious act of nation-building intended to ensure equality of opportunity for all students within an agreed national framework that offered the same schooling to all communities irrespective of ethnicity, language or religion (Zendeli, 2011). As yet there are very few private schools, but they are all subject to the same regulations as public schools to ensure national conformity and forestall any educational separatism (Zendeli, 2011).

In Indonesia over 70% of secondary schools are actually privately owned and managed (Holzhacker, Wittek, & Woltjer, 2016) because the government has not been able to meet demand (Stern & Smith, 2016). However, the low quality of many of these schools has led to concerns that private and independent provision may entrench social (Newhouse & Beegle, 2006) and gender inequalities (Asadullah & Maliki, 2018). Rather than encouraging more independent provision the government has introduced the Bantuan Operasional Sekolah (BOS; School Operational Assistance scheme) to bring privately managed schools under government regulatory control (Asadullah & Maliki, 2018). This is actually a form of internal accountability which has made independent schools subject to the same national standards and expectations as public schools (Nababan et al., 2019), with increased government funding for good performance (Sumintono, Subekti, Mislan, Said, & Mohd Tahir, 2014) and funding reductions for lack of compliance (Harris et al., 2017).

Overall, it seems that in some non-OECD countries which have a diverse range of schools and fully open school choice any GERM inspired models may be seen as regressive and restrictive. In countries with diverse and/or polarised populations school conformity with centralised frameworks may be more important than school diversity in supporting the development of national identity and avoiding inter-communal tensions. In poorer countries school diversity may actually dilute educational quality and thus entrench disadvantage.

Thus, it appears that the elements of GERM as promoted by the international development agencies may not always match the priorities and needs of non-OECD countries and may even be counterproductive. Indeed, it is noteworthy that Trinidad & Tobago, FYROM and Indonesia all introduced domestic reform programmes to tackle issues which lay beyond GERM such as initiatives to raise the quality of teachers and principals (Bakiu et al., 2016; De Lisle, Annisette, & Bowrin-Williams, 2019; Pantic, 2012; Pantić, Closs, & Ivošević, 2011; Raihani & Sumintono, 2010) and mitigate the effects of educational material shortages likely to be severer than anything experienced in any OECD country (Blazevski, Petreski, & Ristovska, 2017; Jasmina, 2017). Such measures may be familiar because before the rise of GERM and SABER the World Bank explicitly prioritised 'people and resources' through an emphasis on the development of teachers and principals to raise student attainment (Husén, Noonan, & Saha, 1978; Lockheed & Komenan, 1989). There was also explicit acknowledgement that any improvement programme would only succeed if schools had access to basic educational resources (Fuller, 1987; Lockheed, Vail, & Fuller, 1986) because "Textbooks, writing materials, and teacher quality have been shown in empirical studies to consistently influence student achievement." (Fuller, 1986, p.5). Relatively modest improvements in the availability of educational resources can yield significant impacts on student attainment (Ganimian & Murnane, 2016) as can sufficient expert teachers (Arias Ciro & Torres García, 2018; Glewwe, Hanushek, Humpage, & Ravina, 2011). Yet despite the fact that many developing countries suffer shortages of both teachers (Aizawa, 2019) and basic educational materials (Glewwe & Muralidharan, 2016) these issues are largely absent from the research and debates about GERM. What is more, when the benefits of GERM related reforms in developing countries are discussed, little emphasis is given to parallel domestic initiatives that target other factors and which might actually be the real cause of any modest gains in student attainment.

The Present Research

Overall, the limited amount of evidence for the effectiveness of GERM in non-OECD and developing countries is at best ambivalent. Previous research and the priorities of domestically led educational reforms in the countries discussed above suggest that factors beyond GERM may be associated with improved student attainment. The fact that GERM continues to be promoted by international agencies means the time is ripe to consider to what extent its prescriptions may be globally effective beyond its OECD origins and whether or not there may be alternative factors that are more salient to raising student attainment in non-OECD, especially developing countries. Therefore, the present study examines the extent to which the GERM-related factors of external accountability, school autonomy and diversity of school provision, as well as the alternative factors of internal and parental accountability, teacher quality, principal leadership effectiveness and educational material shortages, may be associated with student attainment. As most previous research has focused on international samples with a heavy bias towards OECD countries, samples drawn from single global regions or samples of countries at the same levels of development, the present study will draw upon a global sample of non-OECD countries, balanced by global region and levels of economic development, which have all implemented GERM related reforms. Furthermore, as Trinidad & Tobago, FYROM and Indonesia all implemented a mix of internationally led GERM-related and domestically led educational reforms, there is an opportunity to directly contrast the associations between different reform measures and school attainment at the individual country level.

Method

PISA Data

The data for the present study were drawn from the PISA 2015 cycle which was the sixth PISA cycle assessing the cognitive skills of 15 year old students across 35 OECD and 37 partner countries. PISA datafiles are publicly available for download from the OECD PISA website (see Appendix I) and are completely anonymised without any information that could identify any school or individual student, so ethical approval was not required for the present research.

PISA focuses on 15 year olds because most students at this age will be approaching the end of their compulsory education in most countries, so sampling this group makes it possible to draw international comparisons (OECD, 2017b). PISA samples are selected in two stages: schools are sampled at the country-level with larger schools having a higher probability of being selected than smaller schools, then an equal probability cluster sample of students is selected from a list of PISA-eligible students within each school (OECD, 2009, p.55).

Every student completes a Student Questionnaire (see Appendix I) which captures information about each student's personal characteristics and also contains the cognitive items. PISA assesses proficiency in reading, mathematics and science using items developed within a variant of the Rasch model which means items are graduated in terms of their difficulty so that student proficiency with respect to the sampling frame can be efficiently established with a relatively small number of items (von Davier & Sinharay, 2014). A rotated survey design is used to present a selection of items to each student and the responses to these items are then statistically contrasted with other students' responses to generate a series of 10 plausible values which are not individual scores, rather they are estimates of the students' proficiency (OECD, 2017b) which "represent the range of abilities that a student might reasonably have, given the student's item responses." (Wu, 2005, p.115).

The principal of every participating school is required to complete a School Questionnaire (see Appendix I), which consists of a series of questions about various features of the school and its activities.

The PISA 2015 Student Questionnaire and School Questionnaire data were downloaded as separate SPSS files from the PISA website (see Appendix I) , loaded into SPSS Version 26 (IBM Corporation, 2019) and merged to create a single datafile. Initial data collation and cleansing was completed in SPSS and, as the focus of the present research was non-OECD countries, data for OECD countries were deleted.

Country Selection

The sample of non-OECD countries for further analysis was selected on three broad criteria; first countries must have introduced a GERM inspired reform programme whether through a domestic reform agenda and/or participation in an international initiative such as the World Bank SABER programme. Second, countries must be equally drawn from the four World Bank defined global regions in which non-OECD PISA partners are located: East Asia and Pacific (EAP), Europe and Central Asia (ECA), Latin America and the Caribbean (LAC), and Middle East and North Africa (MENA). Third countries from each global region should to be equally selected from the World Bank's high, middle and low income categories because income is an effective measure of the level of national economic development (Sanford & Sandhu, 2002). The regional and income information was taken from the Little Data Book 2015 (World Bank, 2015a).

There were only six countries in MENA, with the two highest income countries both being oil producing states with an unusually high proportion of private, mostly international schools unlikely to be representative of the region. Therefore, it was only possible to draw a balanced group of MENA countries with one country at each income level, which meant that a sample balanced by global region would total 12 countries. None of the remaining MENA countries was a high income country, but in 2015 Lebanon had the highest income and was very close to the high income threshold so was selected

for that group; Lebanon was also a SABER partner that began a major reform programme in 2010 (World Bank, 2013b). Jordan participated in SABER and began a GERM educational reform programme in 2007 (World Bank, 2015b); as it was a middle income country it was allocated to that group. Tunisia was a SABER partner which introduced its first educational reforms in 2005 (World Bank, 2015c); as it had the lowest income in MENA and was very close to the low income threshold it was allocated to that group.

In EAP the high income jurisdiction chosen was Hong Kong which, although not a SABER partner, had introduced a range of domestic SBM reforms to public schools in the early 2000s, had a large independent school sector, and had taken measures to increase school accountability and parental choice (Cheng, Ko, & Lee, 2016; Ko, Cheng, & Lee, 2016). Thailand was a SABER partner that began GERM related reforms in 1999 (World Bank, 2012b) and as it was a middle income country it was allocated to that group. Indonesia was selected as the low income country because it was a SABER partner which since 2000 had also implemented GERM inspired reforms supported by the OECD and ADB as discussed above.

In ECA Malta had the highest income and was selected for the high income group; although not a SABER partner Malta enacted its own Education Reform Act in 1988 modelled closely on the UK ERA and began GERM inspired reforms soon after (Bartolo, 2001; Bezzina, 1991). The middle income country selected was FYROM which was a SABER partner and which had implemented a range of GERM related reforms as discussed above (World Bank, 2012a). Moldova was a SABER partner that began its GERM reform programme in 2007 (World Bank, 2014) and as the only low income country in ECA it was allocated to that group.

In LAC Trinidad & Tobago was not a SABER partner, but as discussed above had participated in a large reform programme led by the IDB; as the country with the highest income in the region it was

allocated to that group. Colombia was a SABER partner which began its educational reforms in the mid-1990s World Bank (2013a); as a middle income country it was allocated to that group. Although a middle income country, Peru was the regional country with the lowest income and was very close to the low income threshold, so was allocated to that group; Peru began its GERM inspired educational reform programme in 2006 with support from the IDB (World Bank, 2007).

Student Questionnaire Variables for Analysis

Mathematics

PISA Mathematics scores were selected as the dependent variable for student attainment because maths is the subject that is most learned in school rather than at home (Coleman, 1975) and student learning in maths is less affected by parental, linguistic and cultural background than reading or science (Afridi, Barooah, & Somanathan, 2018; Figlio & Loeb, 2011; Moschkovich, 2012; Wong, Cook, & Steiner, 2009). As already explained, PISA 2015 utilised 10 plausible values based on multiple imputation estimation of the latent construct of interest, in this case Mathematics proficiency, but as one plausible value is sufficient to provide unbiased estimates of population parameters (Jerrim, Lopez-Agudo, Marcenaro-Gutierrez, & Shure, 2017; OECD, 2009) the first plausible value for Mathematics was used as the criterion variable for the analyses following the practice adopted by other researchers (eg. Bol, Witschge, Van de Werfhorst, & Dronkers, 2014; Double, Mcgrane, Stiff, & Hopfenbeck, 2019).

Sex

Sex was included as a control variable because previous research has shown that males tend to outperform females in Mathematics (Byrnes & Miller, 2007) and although this gap may be narrowing (Lindberg, Hyde, Linn, & Petersen, 2010), unfortunately it still persists (Yavuz, Tan, & Atar, 2019).

Immigration Status

Immigration status was included as a control because despite the fact that maths performance is less affected by language and cultural background than reading or science, some performance differences have been found between students who are native to a country and those who are not (Giannelli & Rapallini, 2016; Moschkovich, 2012). PISA captures information about the countries in which students and their parents were born to generate a categorical index of students' immigration status (IMMIG), which codes students as natives (students born in the country with at least one parent born in the country), second-generation immigrants (students born in the country of assessment with parent(s) born in another country), and first-generation immigrants (students born outside the country of assessment with parent(s) also born in another country) (OECD, 2017a, p.281).

Economic, Social and Cultural Status (ESCS)

ESCS was included as a control because there is a large body of evidence that ESCS is associated with student attainment because it effectively captures students' overall background (Harden, Tucker-Drob, & Tackett, 2013; Tucker-Drob & Briley, 2012). The ESCS scale combines information from a range of PISA primary items dealing with parental occupation and education, students' home possessions, number of cultural objects in the home, etc, in a single scale with a mean of zero and a standard deviation of one across OECD countries (OECD, 2017a, p.285).

School Questionnaire Variables for Analysis

School Type

PISA aggregates information from several items into an index, SCHTYPE, with three categories:

Public school – a school managed directly or indirectly by a public education authority, government agency, or governing board appointed by government or elected by public franchise, and which receives its funding from government whether at local, regional, state or national level.

Independent school – a school managed directly or indirectly by a non-government organisation such as a religious body, charitable trust, trade union, business, or other private institution, which receives its funding from government whether at local, regional, state or national level.

Private school – a school owned and managed directly or indirectly by a non-government organisation which receives its funding in the form of student fees, school charges paid by parents, or from other private sources, and which does not receive funding from government.

These categories and definitions were used throughout the present study.

Accountability

PISA uses three dichotomous questions to gain information about the accountability procedures followed by schools. Principals respond Yes (1) or No (2) to indicate if the following forms of accountability are practised in their schools:

External accountability: achievement data are posted publicly (e.g. in the media).

Internal accountability: achievement data are tracked over time by an administrative authority.

Parental accountability: achievement data are provided directly to parents.

All three accountability items were included in the present study.

School Autonomy

Principals respond to 12 questions about where responsibility lies for various aspects of school activity including staffing, budgeting and academic policy, indicating if decisions about these activities are made by the principal, teachers, school governing board, etc or by an external education authority.

Responses are aggregated to create an index of overall school autonomy (SCHAUT) between 0 (no decision-making at school level) and 1 (most decisions made a school level) (OECD, 2016, p.151), which was transformed to a scale with a mean of zero and a standard deviation of one.

Principal Leadership Effectiveness

Principals respond to items about various school leadership activities and their frequency of occurrence during the previous academic year. The individual item responses are aggregated into a single scale (LEAD) with a mean of zero and a standard deviation of one for OECD countries (OECD, 2017b, p.324).

Educational Material Shortages

Principals complete four items relating to the availability of educational material and the quality of school infrastructure, which are then aggregated into a single scale (EDUSHORT) with a mean of zero and a standard deviation of one for OECD countries (OECD, 2016a, p. 286). Higher scores indicate greater levels of educational material shortages.

Teacher Certification

Principals report the number of teachers at their school who are fully certified and the OECD calculates the proportion of teachers certified (PROATCE) from 0 (none) to 1 (all teachers certified) (OECD, 2017b, p.322), which was transformed to a scale with a mean of zero and a standard deviation of one.

Sample Data

As "it is recommended to delete any cases with missing data" (OECD, 2009, p. 209) those cases with missing values for any the variables of interest across the sample were removed from the dataset. The descriptive statistics for the variables used in the analyses are presented in Table 1.

There were slightly more males than females in the sample. In the low income group of non-OECD countries and the individual countries the proportion of immigrants was below 3%. Mean student

Table 1: Descriptive Statistics for the sample of non-OECD countries, high, middle and low income non-OECD countries, and Trinidad & Tobago, FYROM and Indonesia, with standard deviations shown in brackets

Variables	*Non-OECD Countries*	*High Income Non-OECD Countries*	*Middle Income Non-OECD Countries*	*Low Income Non-OECD Countries*	*Trinidad & Tobago*	*FYROM*	*Indonesia*
Income per capita 2015 US dollars	17,475	31,228	12,143	9,055	26,220	11,520	9,270
Mathematics PV1 mean	416.66 (98.15)	481.15 (112.20)	402.05 (85.80)	393.99 (84.80)	423.04 (94.06)	375.26 (93.61)	396.55 (79.57)
Total Students	56974	12533	25012	19429	3014	3189	5599
Male students, %	53	51	53	52	53	49	53
Immigrants 1st generation	1100	773	279	48	49	17	3
Immigrants 2nd generation	1880	1080	698	102	51	43	1
Natives	53994	10680	24035	19279	2914	3129	5595
Student ESCS mean	-0.78 (1.15)	-0.36 (0.99)	-0.68 (1.10)	-1.17 (1.17)	-0.25 (0.94)	-0.24 (0.91)	-1.77 (1.12)
Total Schools	1906	401	812	693	102	67	209
Private Schools	261	79	118	64	3	3	33
Independent schools	286	132	37	117	5	1	55
Public schools	1359	190	657	512	94	63	121
Teacher Certification mean	-0.09 (1.07)	0.17 (0.89)	-0.39 (1.22)	0.13 (0.82)	-0.57 (0.88)	0.11 (0.79)	-0.15 (0.86)
Educational material Shortages mean	0.42 (1.23)	0.06 (1.02)	0.39 (1.23)	0.68 (1.28)	0.76 (1.03)	-0.14 (0.93)	0.71 (1.30)
Principal Leadership Effectiveness mean	0.18 (1.07)	0.06 (0.91)	0.35 (1.06)	0.03 (1.14)	0.51 (0.90)	0.33 (0.91)	0.23 (1.20)
Accountability – external, % schools practising	29	23	39	24	16	33	30
Accountability – internal, % schools practising	82	76	87	83	89	94	96
Accountability – parental, % schools practising	86	77	91	90	77	91	92
School Autonomy mean	0.07 (0.98)	0.26 (0.90)	0.05 (1.03)	-0.02 (0.93)	-0.13 (0.57)	0.20 (0.88)	0.58 (0.70)

ESCS levels appeared to be proportional to country development levels as did school educational material shortages, with less developed countries having lower levels of ESCS and greater educational material shortages. There was a good mix of public, independent and private schools in the sample, income groups and individual countries with the exception of FYROM where 95% of schools were public. The proportion of certified teachers was similar in high and low income countries, but notably lower in middle income countries and Trinidad & Tobago. Principal leadership effectiveness was highest in the middle income and individual countries, but low in the most and least developed countries. There was considerable variation in the accountability practices implemented by schools, with the lowest rates of accountability in the most developed countries. School autonomy was highest in the high income group of countries, yet lowest in high income Trinidad & Tobago. Overall, it appeared that there were variations within and across the sample for all the variables of interest, suggesting that further analyses were justified.

Weights

As PISA is a stratified survey any analyses must be appropriately weighted to reflect differences in the school and student selection probabilities, non-response rates, and variations in sample sizes between schools and countries (Rust, 2014; Leslie Rutkowski, Gonzalez, Joncas, & von Davier, 2010). Once cases with missing data had been removed the student and senate weights for each student case were recalculated as recommended by Jerrim, Lopez-Agudo, Marcenaro-Gutierrez, and Shure (2017) and the OECD (2009): new student weights were calculated by multiplying each student weight by the original number of cases in each country, then dividing the result by the number of cases in the final country sample. The new student weights were then used to calculate new senate weights for each student which summed to a total of 5,000 for each country. In line with the recommendations of the OECD (2009) and Stapleton (2014) the student weights were used when analysing data from single countries. Senate weights were used for transnational analyses because this meant the analyses

"combine[d] data across countries but treat[ed] each country as an equal contributor to an estimate." (Stapleton, 2014, p. 367).

Analytical Strategy

As already discussed, GERM originated in highly developed OECD countries, so the objective of the present study was to establish the extent to which GERM may be globally effective by investigating the associations between maths attainment and the GERM promoted measures of school external accountability, autonomy and diversity in non-OECD countries, and by contrasting these associations with other school-level variables outside the GERM model which have been implicated in student attainment.

As the dataset was drawn from the PISA 2015 cycle and PISA data are hierarchically structured with students nested within schools (OECD, 2017b) the student observations are not independent and the students clustered within schools are more likely to be similar to each other than to students in other schools. In any conventional multivariate analysis it is likely that this clustering will lead to standard errors being underestimated causing a narrowing of confidence intervals and smaller *p* values which make Type 1 errors, that is false positive results, more likely (Snijders & Bosker, 1999). Multilevel modelling (MLM) was chosen for the analyses that follow because it addresses student nesting within schools by examining variables at more than one level, thereby ensuring that the standard errors are not underestimated (Goldstein, 1995). MLM also allows researchers to explore the associations between group-level characteristics with individual outcomes and contrast variability between groups (Kreft & de Leeuw, 1998).

MLMs were developed using the PROC MIXED procedure in SAS© Studio (SAS Institute Inc, 2020) (see Appendix II for details and example output) using random intercepts for schools and countries entered into the models which included several countries, and a random intercept for schools in single country

models. Beginning with the sample of 12 non-OECD countries, then examining the groups of high, middle and low income countries, and finally examining Trinidad & Tobago, FYROM and Indonesia, MLMs were executed to investigate the associations between the student control variables, the school-level variables of interest, and student maths attainment. Each MLM was developed following the iterative process recommended by Hox (2010) and Snijders & Bosker (2011). Model 1 was an unconditional random effect model in which the associations between PISA maths scores for students were able to vary across individual schools and countries, and from which the intraclass correlations (ICC) were calculated to establish the extent to which there was clustering within schools and there were variations between schools and countries. This also established that the ICCs exceeded the generally accepted threshold of 0.05 above which MLM is justified (Heck, Thomas, & Tabata, 2014). In Model 2 the student control variables and school level variables of interest were added as fixed effects. In Model 3 the potential variance between schools and countries was assessed iteratively because "Testing for random slope variation is best done on a variable-by-variable basis. [then after identifying] which of the slopes [had] a significant variance between groups … these [were added] simultaneously to the model." (Hox, 2010, p. 58). In Model 4 potential interactions between variables across and within levels were examined individually, and in those cases where a variable interacted with two or more other variables possible interactions between all the variables were investigated using an "exploratory procedure to arrive at a "good" model" (Hox, 2010, p.58) in which all significant interactions were included. This approach prevented potential "masking effects" (Snijders & Bosker, 2011, p. 107) by revealing any variables which only achieved significant associations when other variables were controlled. However, models containing all possible predictor variables are likely to be over-specified and this increases the risk of Type I errors (Kreft & de Leeuw, 1998; Snijders & Bosker, 2011), so this possibility was reduced by excluding "nonsignificant coefficients from the model. … [in a] backward model fitting process" (Snijders & Bosker, 2011, p. 107) to arrive at a Final Model.

MLM fit is usually judged by making a comparison between the log-likelihood statistics for nested models (Goldstein, 1995), but in a "backward model-fitting process" the earlier models are not nested within the Final Model so it is impossible to compare log-likelihood statistics across models (Hox, 2010). However, the Akaike Information Criterion (AIC) provides a viable alternative to assess model fit because it corrects for model complexity (Liu, Rovine, & Molenaar, 2012; Vrieze, 2012) and is "noted to work well in selecting the true model" (Heck et al., 2014, p. 191) with lower values indicating better fit. Lastly, as an estimate of the amount of variance explained by the Final Model the multilevel version of R^2 recommended by Snijders & Bosker (2011) was calculated as this approach is suitable for models with fixed and random effects.

There are no agreed standards for the presentation of findings from MLMs (Garson, 2019), so the previously cited papers by Double, Mcgrane, Stiff, & Hopfenbeck (2019) and Goldstein & Leckie (2016) were used as guides.

Results

Non-OECD Countries

Bivariate correlations for the whole sample are shown in Table 2; Pearson correlations were used to examine associations between the continuous variables, and Spearman rank correlations when dichotomous or categorical variables were included in the examination of associations between variables.

Table 2: Bivariate correlations between the study variables

Variables	Sex	Immigrant Status	Student ESCS	School Type	Educational Mat. Short.	Teacher Certification	Principal Leadership	Account - external	Account - internal	Account - parental	School Autonomy	Country Income
Mathematics PV1	0	-0.12***	0.33***	-0.24***	-0.22***	0.11***	0.01***	0.08***	0.02***	-0.06***	0.23***	0.27***
Sex		0	0.02***	-0.02***	0.02***	-0.02***	0	0	0	0	0	0
Immigrant status			-0.03***	0.15***	0.06***	-0.11***	0.06***	-0.02***	0	0.08***	-0.04***	-0.22***
Student ESCS				-0.18***	0	0.10***	0.09***	0	0.01***	0	0.02***	0.26***
School Type					0.32***	0.05***	-0.01***	0.04***	0.06***	0.04***	-0.49***	-0.22***
Teacher Certif.						-0.06***	-0.08***	0	-0.02***	0.07***	-0.28***	-0.18***
Educ. Mat. Short							-0.04***	0	0	-0.04***	0	0
Principal Leadership								0.08***	0.16***	0.10***	0.05***	0.02***
Accountability – external									0.19***	0.06***	0.17***	0.07***
Accountability – internal										0.07***	0.04***	0
Accountability – parental											-0.04***	-0.11***
School Autonomy												0.10***

Note: *** p <0.001

As can be seen, there were a variety of moderate to large correlation coefficients suggesting a range of significant associations between the variables of interest and PISA Mathematics scores that were worthy of further investigation.

Table 3 shows the MLM results for the sample of 12 non-OECD countries; ICCs for Model 1 indicated that approximately a quarter of the variance in maths scores lay between schools ($ICC_{(School)}$=0.24) and a fifth between countries ($ICC_{(Country)}$=0.19) suggesting MLM was justified. The Final Model fitted the data better than earlier models (*AIC*=670510.6) and accounted for a reasonable proportion of the variance in students' maths performance (R^2=0.17).

Males achieved higher maths scores than females (β=0.01, *t*=6.30, *p*<.001). There were no significant differences between immigrants and natives. ESCS was positively associated with maths performance (β=0.15, *t*=7.75, *p*<.001). Attending private schools was associated with higher maths scores (β=0.08, *t*=2.70, *p*<.05) compared to independent or public schools, but there were no significant differences in performance between the latter school types. Shortages of educational materials were linked to lower maths scores (β=-0.07, *t*=-5.66 *p*<.001). All forms of accountability were modestly associated with higher maths scores, external (β=0.03, *t*=2.79, *p*<.01), internal (β=0.03, *t*=2.57, *p*<.01), and parental (β=0.02, *t*=2.13, *p*<.05). Levels of country income were associated with student attainment, with being a student in middle income countries (β=-0.34, *t*=-2.26, *p*<.05) and low income countries (β=-0.26, *t*=-2.07, *p*<.05) both being associated with lower maths scores than studying in highly developed countries. There were no further significant fixed effects.

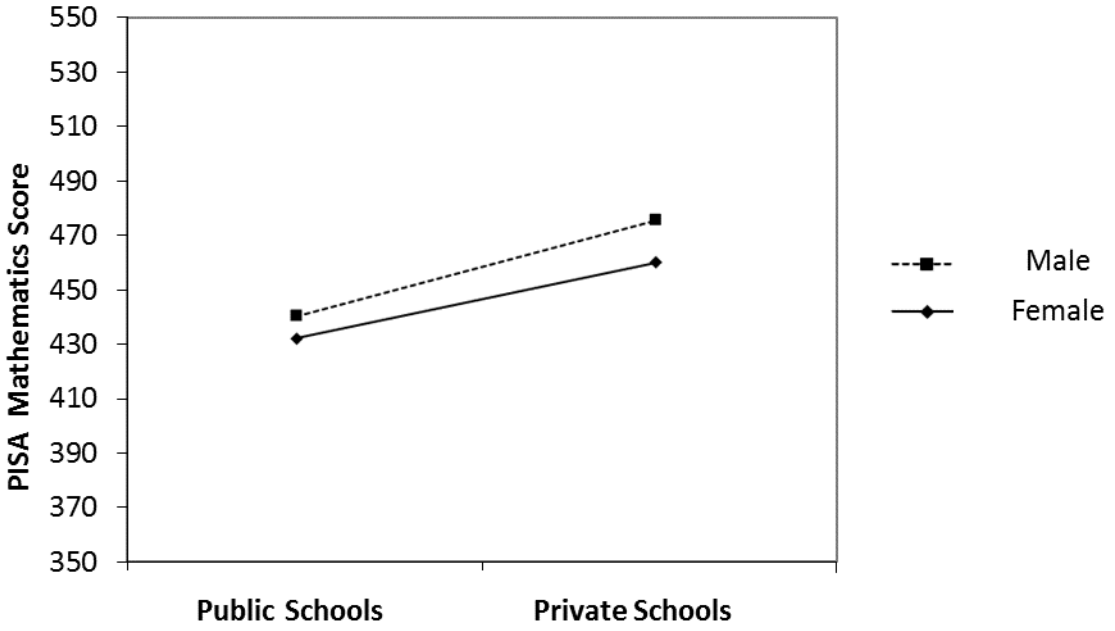
There were considerable variations in the associations between sex and maths attainment across schools (β=0.29, *z*=7.08, *p*<.001). Interactions between sex and school type suggested that attending private schools was associated with higher maths scores for all students (β=0.01, *t*=2.49, *p*<.01) with a slight advantage for males as shown in Figure 1.

Table 3: Multilevel models analysing the relationships between student and school variables and student PISA Mathematics scores in non-OECD countries

	Model 1	Model 2				Model 3				Model 4				Final Model			
	Uncond.	Fixed Effects				Random Slopes				Interactions							
Predictors	*b*	*b*	*β*	CI	p	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*
(Intercept)	413.00	446.50				445.04				437.47				445.73			
Sex (Male)		7.79	0.01	-1.36, 1.38	<0.001	7.62	0.01	-1.59, 1.60	<0.001	5.97	0.01	-1.85, 1.86	<0.001	5.97	0.01	-1.85, 1.86	<0.001
1st Gen Immig[a]		2.69	0.00	-4.46, 4.48	0.24	1.49	0.00	-4.48, 4.49	0.52	1.54	0.00	-4.48, 4.49	0.50				
2nd Gen Immig[a]		-2.20	0.00	-3.63, 3.62	0.23	-2.75	0.00	-3.63, 3.61	0.14	-2.72	0.00	-3.63, 3.61	0.14				
ESCS		12.12	0.14	-0.64, 0.74	<0.001	12.49	0.15	-2.93, 3.40	<0.001	12.51	0.15	-2.93, 3.40	<0.001	12.51	0.15	-2.93, 3.40	<0.001
Private School[b]		22.52	0.08	-7.61, 8.24	<0.001	23.65	0.08	-15.96, 17.34	<0.001	44.97	0.16	-22.37, 26.22	<0.001	22.55	0.08	-15.71, 17.01	<0.05
Indep School[b]		-15.56	-0.05	-9.73, 9.22	<0.001	-7.50	-0.03	-18.56, 18.08	0.48	-1.12	0.00	-23.81, 23.71	0.90	-9.85	-0.03	-18.69, 18.06	0.92
Educ. Mat. Short.		-5.06	-0.06	-1.89, 1.77	<0.001	-5.25	-0.07	-1.94, 1.82	<0.001	-5.33	-0.07	-1.94, 1.81	<0.001	-5.32	-0.07	-1.94, 1.81	<0.001
Teacher Certif.		2.55	0.03	-2.58, 2.65	0.06	2.59	0.03	-2.63, 2.70	0.06	2.47	0.03	-2.63, 2.70	0.07				
Principal Lead		0.47	0.01	-2.07, 2.08	0.66	0.50	0.01	-2.11, 2.12	0.64	0.58	0.01	-2.11, 2.12	0.59				
Account – external		6.45	0.03	-4.76, 4.91	<0.01	6.77	0.03	-4.80, 4.96	<0.01	6.84	0.03	-4.80, 4.96	<0.01	6.91	0.03	-4.78, 4.94	<0.01
Account – internal		8.08	0.03	-5.93, 6.11	<0.01	7.77	0.03	-5.98, 6.16	<0.05	7.70	0.03	-5.99, 6.16	<0.05	7.92	0.03	-5.95, 6.13	<0.01
Account – parental		8.10	0.03	-6.84, 7.02	<0.05	7.46	0.02	-6.90, 7.07	<0.05	7.50	0.02	-6.90, 7.07	<0.05	7.59	0.02	-6.89, 7.06	<0.05
Autonomy		2.91	0.03	-3.50, 3.60	0.11	1.57	0.02	-3.59, 3.65	0.40	1.45	0.01	-3.59, 3.64	0.43				
Countries Low Inc[c]		-61.73	-0.26	-62.90, 48.15	<0.05	-61.44	-0.26	-65.10, 49.89	<0.05	-48.31	-0.20	-64.40, 52.27	0.11	-61.84	-0.26	-66.30, 50.73	<0.05
Countries Mid Inc[c]		-66.30	-0.34	-65.07, 46.07	<0.05	-66.24	-0.33	-67.15, 47.55	<0.05	-56.46	-0.29	-66.85, 49.86	0.06	-67.37	-0.34	-68.50, 48.22	<0.05
Priv Sch*Sex										5.82	0.01	-4.58, 4.65	<0.05	5.88	0.01	-4.54, 4.61	<0.01
Indep Sch*Sex										6.65	0.02	-4.81, 4.89	<0.01	6.65	0.02	-4.81, 4.89	<0.01
Cnty LI*Priv Sch										-39.01	-0.07	-32.12, 29.95	<0.05				
Cnty LI*Ind Sch										-25.18	-0.07	-31.86, 29.59	0.11				
Cnty MI*Priv Sch										-11.53	-0.03	-35.91, 34.91	0.52				
Cnty MI*Ind Sch										-23.81	-0.03	-34.46, 33.35	0.17				
Random Effects																	
σ^2	5423	5321.3				5195.4				5195.3				5195.4			
$\tau_{00\ (School)}$	2460	1821.2				1723				1727.5				1729.68			
$\tau_{00\ (Country)}$	2443.6	1583.2				1563.6				1612.4				1621.99			
ICC (School)	0.24	0.21				0.20				0.20				0.20			
ICC (Country)	0.19	0.18				0.18				0.19				0.19			
Sex (School)						212.35	0.30	-49.03,66.50	<0.001	206.62	0.29	-48.72, 65.54	<0.001	205.60	0.29	-48.64,65.32	<0.001
ESCS (School)						64.34	0.75	-9.51,21.40	<0.001	64.14	0.75	-9.52, 21.35	<0.001	64.12	0.75	-9.52, 21.34	<0.001
ESCS (Country)						29.10	0.34	-20.52, 29.16	<0.01	29.19	0.34	-20.55, 29.24	<0.01	29.63	0.35	-21.02,30.07	<0.01
Sch Type (Country)						252.23	1.21	-99.46, 418.88	<0.01	110.86	0.53	-130.5, 227.5	0.11	113.95	0.55	-126.68,224.42	<0.05
AIC	672503.2	670966.5				670520.6				670515.9				670510.6			

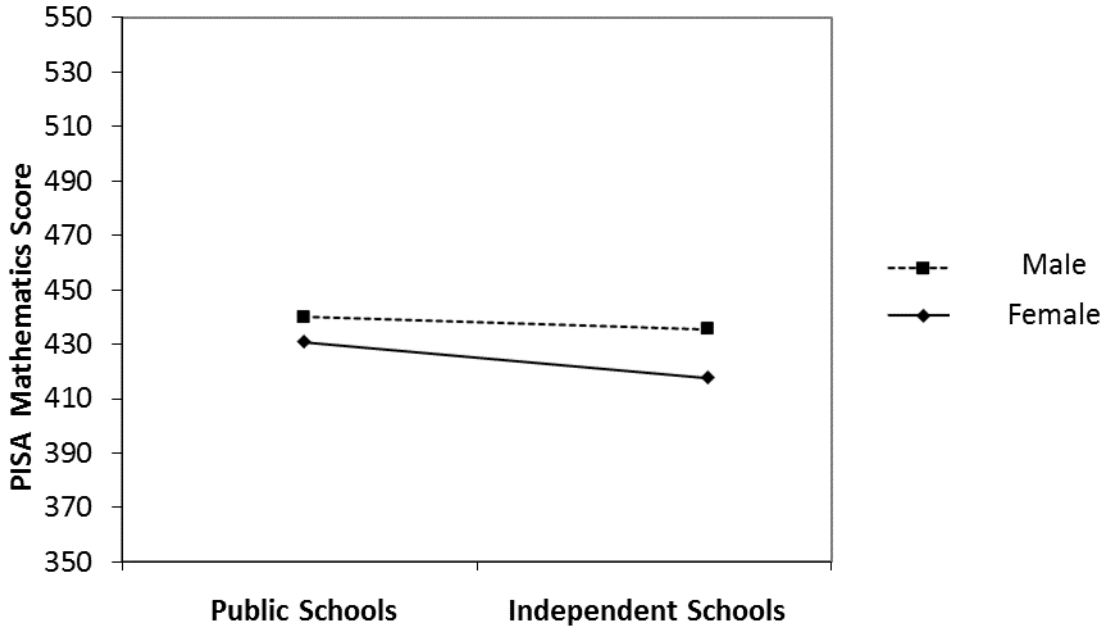
a. reference is native students, b. reference is public schools, c. reference is high income countries

Figure 1: Interaction effect of private schools with student sex on student PISA Mathematics scores in non-OECD countries



Attending independent schools was associated with slightly lower maths scores for both sexes, with females' attainment more adversely affected (β=0.02, t=2.67, p<.01) as shown in Figure 2,

Figure 2: Interaction effect of independent schools with student sex on student PISA Mathematics scores in non-OECD countries



There were no further interactions between variables.

The associations between student ESCS and maths attainment varied between schools (β=0.75, z=7.88, p<.001) and across countries (β=0.35, z=2.30, p<.01), and there were also variations in the associations between school types and attainment across countries (β=0.55, z=1.95, p=0.05).

These findings suggested that in non-OECD countries the association between sex and attainment varied across schools, but males overall achieved slightly higher maths scores that females. Higher ESCS was generally linked with higher attainment, though the nature of this association varied across schools and countries. Attendance at private schools was associated with higher attainment for both sexes, with a slight advantage for males, whereas attendance at independent schools was associated with lower performance for both sexes, with the most adverse association with attainment being among females. Greater educational material shortages were linked to lower student performance. All forms of school accountability were associated with higher attainment. Lower levels of country income, and therefore lower levels of economic development, were linked with lower student attainment.

High Income non-OECD Countries

The MLMs for the four high income non-OECD countries are shown in Table 4. Model 1 indicated that approximately a quarter of the variance in maths scores lay between schools ($ICC_{(School)}$=0.24) and approximately a fifth between countries ($ICC_{(Country)}$=0.21) suggesting MLM was appropriate. There were no significant interactions between variables so Model 4 is omitted. The Final Model fitted the data better than earlier models (AIC=146184.30), but only accounted for a small amount of the variation in students' maths attainment (R^2=0.04).

Table 4: Multilevel models analysing the relationships between student and school variables and student PISA Mathematics scores in high income non-OECD countries

	Model 1 Uncond.	Model 2 Fixed Effects				Model 3 Random Slopes				Final Model			
Predictors	*b*	*b*	*β*	CI	p	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*
(Intercept)	458.98	422.99				421.79				427.32			
Sex (Male)		11.00	0.06	-3.39, 3.59	<0.001	11.08	0.06	-3.39, 3.59	<0.001	11.07	0.06	-3.39, 3.59	<0.001
1st Gen Immig[a]		10.12	0.02	-6.57, 6.74	<0.01	8.80	0.02	-6.59, 6.73	<0.01	8.80	0.02	-6.59, 6.73	<0.01
2nd Gen Immig[a]		2.68	0.01	-5.98, 6.03	0.38	2.26	0.01	-5.98, 6.02	0.46	2.25	0.01	-5.98, 6.02	0.46
ESCS		11.10	0.11	-1.49, 1.67	<0.001	10.14	0.10	-1.84, 2.04	<0.001	10.13	0.10	-1.84, 2.04	<0.001
Private School[b]		38.68	0.14	-22.22, 25.56	<0.001	39.13	0.14	-22.42, 25.84	<0.001	56.82	0.20	-15.64, 19.23	<0.001
Indep School[b]		-3.54	-0.02	-25.24, 24.78	0.78	-4.46	-0.02	-25.49, 24.92	0.73	1.72	0.01	-23.16, 23.37	0.88
Educ. Mat. Short.		-2.07	-0.02	-5.91, 5.78	0.49	-2.81	-0.03	-5.99, 5.82	0.35				
Teacher Certif.		3.86	0.04	-6.48, 6.71	0.25	3.78	0.03	-6.54, 6.77	0.27				
Principal Lead		-1.43	-0.01	-6.89, 6.80	0.68	-1.28	-0.01	-6.94, 6.85	0.72				
Account – external		18.04	0.08	-13.12, 14.26	<0.01	19.02	0.09	-13.19, 14.41	<0.01	20.53	0.09	-12.96, 14.25	<0.01
Account – internal		12.17	0.05	-14.10, 14.84	0.10	12.18	0.05	-14.21, 14.96	0.10				
Account – parental		16.58	0.07	-13.64, 14.67	<0.05	17.32	0.07	-13.72, 14.80	<0.05	17.09	0.07	-13.76, 14.83	<0.05
Autonomy		9.61	0.09	-9.99, 10.93	0.07	9.68	0.09	-10.08, 11.04	0.07				
Random Effects													
σ^2	9165.24	9017.02				8893.19				8893.97			
$\tau_{00\ (School)}$	3936.19	2960.79				2971.35				3035.91			
$\tau_{00\ (Country)}$	3580.74	3397.82				3451.69				4047.84			
ICC (School)	0.24	0.19				0.19				0.19			
ICC (Country)	0.21	0.22				0.23				0.25			
ESCS (School)						96.22	0.97	-21.64,64.09	<0.001	95.49	0.96	-21.68, 63.60	<0.001
AIC	146504.8	146227.8				146185.7				146184.3			

a. reference is native students, b. reference is public schools

Males achieved higher maths scores than females (β=0.06, t=6.22, p<.001). First generation immigrants outperformed natives (β=0.02, t=2.59, p<.01*)* – possible reasons are discussed below. Higher ESCS was moderately associated with higher maths performance (β=0.10, t=10.24, p<.001), though the character of the association varied across individual schools (β=0.96, z=4.39, p<.001). Private schooling was strongly associated with higher maths attainment (β=0.20, t=6.39, p<.001), but there were no attainment differences between independent and public schools. External accountability (β=0.09, t=2.96, p<.01) and parental accountability (β=0.07, t=2.34, p<.05) were associated with higher PISA scores. There were no further fixed or random effects.

The low number of significant associations and the small amount of variance accounted for in the Final Model suggested that the student control and school level variables of interest explained little of the variation in student attainment. Therefore, it appeared that the key factors which might be associated with student attainment in highly developed non-OECD countries were absent from the analyses.

Middle Income non-OECD Countries

The MLMs for the four middle income non-OECD countries are presented in Table 5. Model 1 indicated more variation in maths scores between schools ($ICC_{(School)}$=0.24) than between countries ($ICC_{(Country)}$=0.05) though suggesting MLM was warranted. The Final Model fitted the data better than preceding models (AIC=297921.80) and accounted for a modest amount of the variation in students' maths attainment (R^2=0.10).

Males had higher maths scores than females (β=0.04, t=6.77, p<.001). First generation immigrants achieved slightly lower scores that natives (β=-0.01, t=2.39, p<.05*),* though second generation immigrants were on a par with natives. Higher levels of ESCS were strongly associated with higher maths scores (β=0.16, t=15.58, p<.001*).* In contrast to students in public schools, those in private schools had higher levels of attainment (β=0.06, t=2.99, p<.01*)* and those in independent schools had

Table 5: Multilevel models analysing the relationships between student and school variables and student PISA Mathematics scores in middle income non-OECD countries

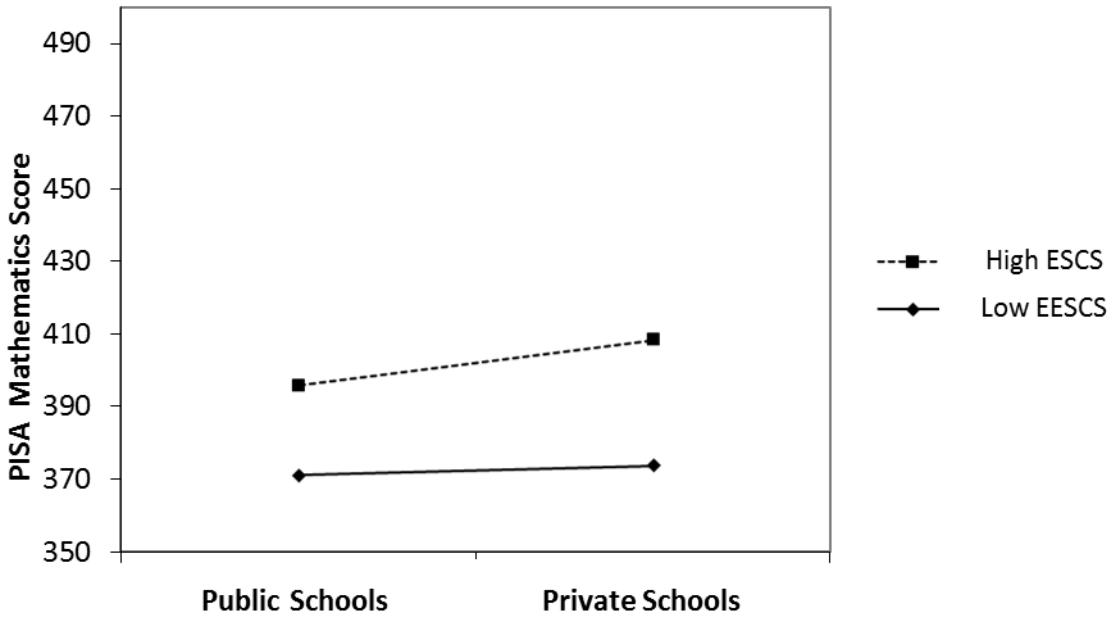
	Model 1	Model 2				Model 3				Model 4				Final Model			
	Uncond.	Fixed Effects				Random Slopes				Interactions							
Predictors	*b*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*
(Intercept)	393.46	397.17				396.25				397.13				396.62			
Sex (Male)		6.64	0.04	-2.05, 2.14	<0.001	7.03	0.04	-2.05, 2.14	<0.001	7.21	0.04	-2.05, 2.14	<0.001	7.22	0.04	-2.05, 2.14	<0.001
1st Gen Immig[a]		-12.03	-0.01	-9.36, 9.22	<0.01	-11.75	-0.01	-9.35, 9.21	<0.01	-11.29	-0.01	-9.34, 9.21	<0.05	-11.32	-0.01	-9.34, 9.21	<0.05
2nd Gen Immig[a]		-3.72	-0.01	-5.65, 5.60	0.19	-3.76	-0.01	-5.63, 5.59	0.19	-4.15	-0.01	-5.63, 5.58	0.15	-4.16	-0.01	-5.63, 5.58	0.15
ESCS		11.77	0.15	-0.94, 1.10	<0.001	12.01	0.15	-1.24, 1.45	<0.001	12.54	0.16	-1.45, 1.71	<0.001	12.55	0.16	-1.45, 1.71	<0.001
Private School[b]		13.87	0.06	-9.72, 10.28	<0.01	13.59	0.05	-9.95, 10.52	<0.01	15.13	0.06	-9.95, 10.58	<0.01	15.58	0.06	-9.90, 10.55	<0.01
Indep School[b]		-32.23	-0.07	-14.64, 13.57	<0.001	-31.48	-0.07	-15.31, 14.22	<0.001	-35.95	-0.08	-16.83, 15.47	<0.001	-36.01	-0.08	-16.72, 15.37	<0.001
Educ. Mat. Short.		-5.37	-0.08	-2.55, 2.36	<0.001	-5.80	-0.08	-2.65, 2.43	<0.001	-8.16	-0.12	-3.04, 2.70	<0.001	-8.15	-0.12	-3.04, 2.69	<0.001
Teacher Certif.		-0.41	-0.01	-3.55, 3.53	0.82	-0.05	0.00	-3.63, 3.63	0.98	-0.22	0.00	-3.62, 3.61	0.91				
Principal Lead		1.58	0.02	-2.78, 2.84	0.27	1.77	0.02	-2.87, 2.94	0.23	1.88	0.02	-2.86, 2.93	0.20				
Account – external		6.56	0.04	-6.13, 6.37	<0.05	6.96	0.04	-6.35, 6.62	<0.05	6.96	0.04	-6.32, 6.59	<0.05	7.17	0.04	-6.21, 6.48	<0.05
Account – internal		-2.08	-0.01	-9.80, 9.72	0.68	-1.72	-0.01	-10.13, 10.07	0.74	-1.73	-0.01	-10.09, 10.02	0.74				
Account – parental		1.25	0.00	-14.82, 14.86	0.87	1.18	0.00	-15.42, 15.47	0.88	0.54	0.00	-15.58, 15.60	0.95				
Autonomy		6.98	0.08	-4.42, 4.81	<0.01	6.37	0.08	-4.54, 4.91	<0.01	5.33	0.06	-4.64, 4.95	<0.05	5.51	0.07	-4.62, 4.95	<0.05
Priv Sch*ESCS										6.59	0.03	-4.38, 4.50	<0.01	6.58	0.03	-4.38, 4.50	<0.01
Indep Sch*ESCS										-1.09	0.00	-7.17, 7.15	0.76	-1.16	0.00	-7.17, 7.14	0.75
Priv Sch*Ed Mat Sh										8.43	0.04	-6.97, 7.29	<0.05	8.13	0.04	-6.96, 7.27	<0.05
Ind Sch*Ed Mat Sh										-5.34	-0.01	-13.85, 13.68	0.45	-5.79	-0.01	-13.64, 13.46	0.40
Ed Mat Sh*ESCS										-2.32	-0.04	-1.13, 1.08	<0.001	-2.32	-0.04	-1.13, 1.08	<0.001
Autonomy*ESCS										-3.47	-0.03	-1.29, 1.25	<0.001	-3.47	-0.03	-1.29, 1.25	<0.001
Random Effects																	
σ^2	4082.0	4004.3				3932.8				3929.7				3929.8			
$\tau_{00\ (School)}$	1821.5	1284.0				1325.1				1313.3				1317.1			
$\tau_{00\ (Country)}$	224.8	315.40				275.4				266.1				259.3			
ICC $_{(School)}$	0.24	0.23				0.24				0.24				0.24			
ICC $_{(Country)}$	0.05	0.06				0.05				0.05				0.05			
ESCS $_{(School)}$						97.66	0.97	-11.23,42.34	<0.001	86.25	0.85	-12.06,37.47	<0.001	86.14	0.85	-12.07,37.43	<0.001
AIC	298752.8	298092.3				297967.3				297928.1				297921.8			

a. reference is native students, b. reference is public schools

lower scores (β=-0.08, t=-4.40, p<.001). Educational material shortages were negatively linked to maths scores (β=-0.12, t=-5.58, p<.001). External accountability was modestly associated with higher maths attainment (β=0.04, t=2.21, p<.05*)*, but other forms of accountability showed no link to performance. Autonomy was positively associated with student attainment (β=0.07, t=2.26, p<.05). There were no further significant fixed effects.

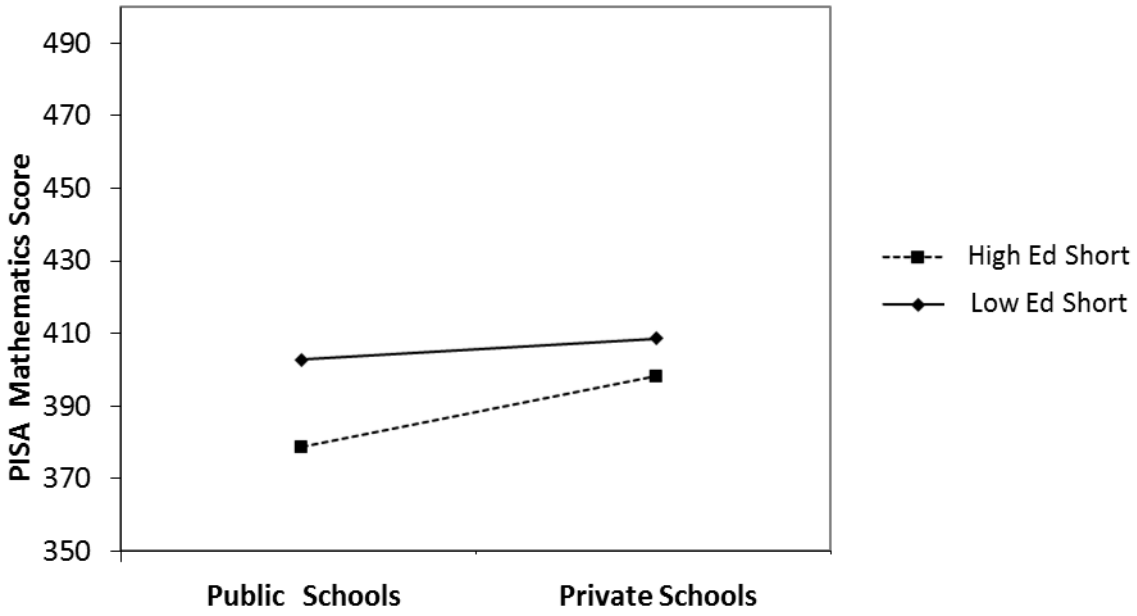
The associations between student ESCS and attainment varied between schools (β=0.85, z=6.82, p<.001). The interaction between private schools and ESCS (β=0.03, t=2.90, p<.01*)* shown in Figure 3 suggested that students attending private schools had higher levels of attainment relative to students with comparable levels of ESCS in public schools.

Figure 3: Interaction effect of private schools with student ESCS on student PISA Mathematics scores in middle income non-OECD countries



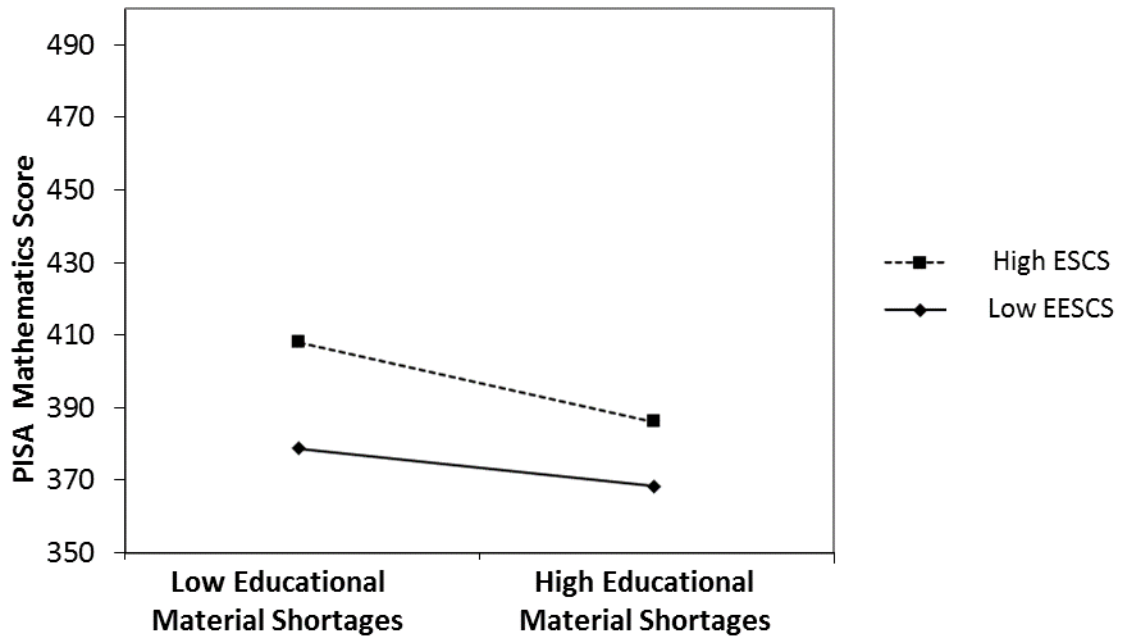
The interaction between private schools and educational material shortages (β=0.04, t=2.24, p<.03) shown in Figure 4 suggested that, in comparison to attendance at public and independent schools, attendance at private schools somewhat attenuated the negative association between material shortages and maths performance.

Figure 4: Interaction effect of private schools with school educational material shortages on student PISA Mathematics scores in middle income non-OECD countries



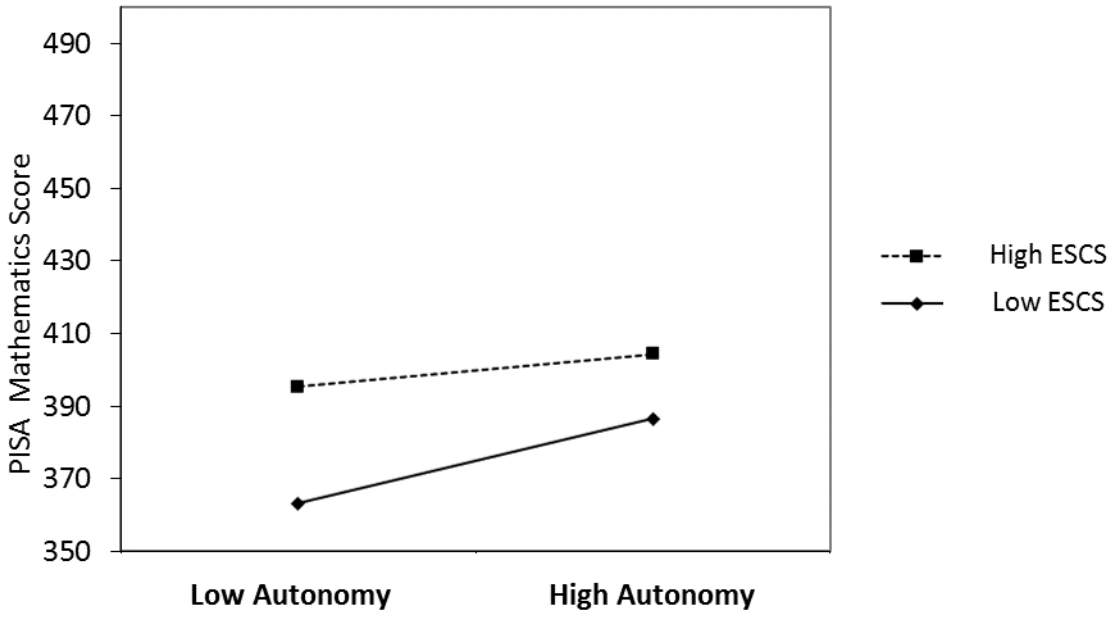
As shown in Figure 5 there was an interaction between school educational material shortages and student ESCS (β=-0.04, t=-4.13, p<.001) which suggested that students with comparable levels of ESCS had lower levels of maths attainment in schools with greater shortages.

Figure 5: Interaction effect of school educational material shortages with student ESCS on student PISA Mathematics scores in middle income non-OECD countries



The interaction between school autonomy and student ESCS (β=-0.03, t=-5.34, p<.001) presented in Figure 6 suggested that students with lower ESCS attending schools with higher levels of autonomy had slightly better levels of maths attainment compared to their peers attending schools with lower autonomy. There were no further significant interactions between the variables.

Figure 6: Interaction effect of school autonomy with student ESCS on student PISA Mathematics scores in middle income non-OECD countries



These findings suggested that student ESCS was more strongly associated with attainment in middle income non-OECD countries than in more developed countries, but varied in its links with performance across schools. Student attainment in private schools was higher than in independent and public schools, but the link was weaker than that found in highly developed countries. In contrast, attendance at an independent school was associated with lower attainment. Educational material shortages were strongly associated with lower attainment irrespective of ESCS, though the association was attenuated in private schools. External accountability and school autonomy were modestly associated with higher student attainment and to some extent autonomy mitigated the adverse link between low ESCS and maths scores. Thus, the GERM related variables of external accountability and

autonomy were positively associated with student attainment, but the links between independent schools and lower student attainment suggested school diversity might not be an effective means of raising performance.

Low Income non-OECD Countries

The MLMs for the four low income non-OECD countries are presented in Table 6. Model 1 indicated nearly a third of the variance in maths scores lay between schools ($ICC_{(School)}$=0.31) with little variance between countries ($ICC_{(Country)}$=0.05), though suggesting MLM was justified. As the countries in this group had fewer than 2% immigrants, this variable was dropped from analysis. The Final Model fitted the data better than the preceding models (*AIC*=223305.90) and accounted for a modest amount of the variation in students' maths attainment (R^2=0.12).

Males scored higher on the PISA maths scales than females (β=0.03, *t*=5.31, *p*<.001). Higher levels of ESCS were very strongly associated with higher maths scores (β=0.24, *t*=20.06, *p*<.001). Compared to public schools, studying in private schools was associated with higher levels of maths attainment (β=0.06, *t*=2.26, *p*<.05), but attending independent schools was associated with lower PISA scores (β=-0.09, *t*=-3.63, *p*<.01). Educational material shortages were linked to lower PISA scores (β=-0.13, *t*=-6.05, *p*<.001). Internal accountability was positively associated with maths performance (β=0.05, *t*=2.83, *p*<.01), but the other forms of accountability showed no similar association and there were no further significant fixed effects.

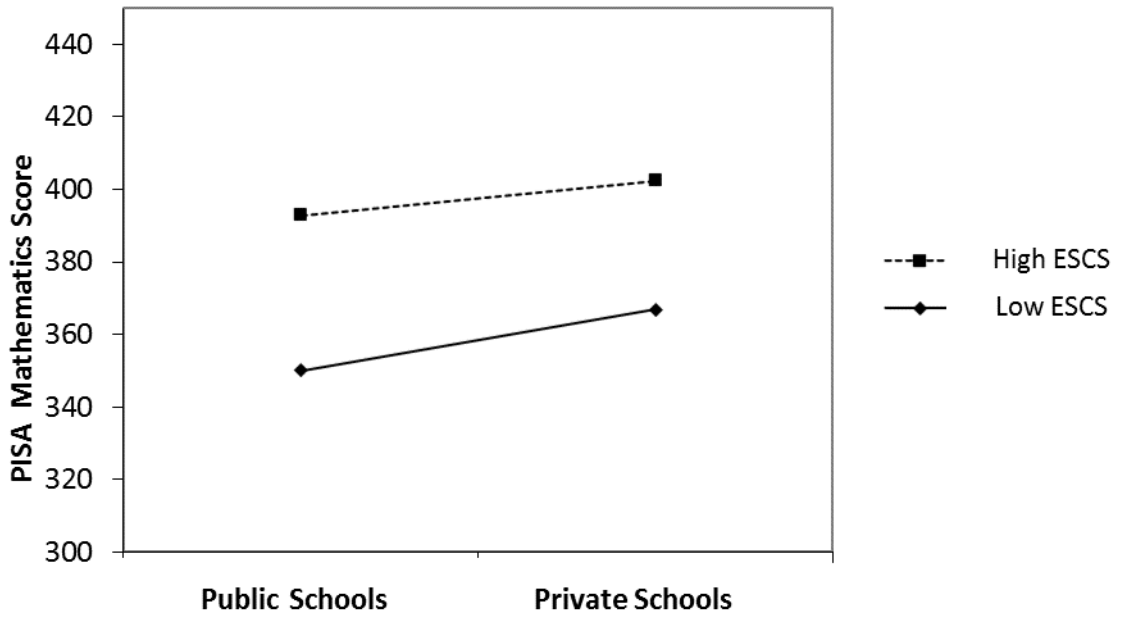
There were variations in the associations between sex and attainment (β=0.80, *z*=3.36, *p*<.001) and ESCS and attainment (β=0.60, *z*=3.93, *p*<.001) across schools. The interactions between school type and ESCS suggested that in comparison to public and independent schools, students with similar levels of ESCS attending private schools had higher levels of attainment (β=-0.02, *t*=-2.04, *p*<.05) as shown in Figure 7.

Table 6: Multilevel models analysing the relationships between student and school variables and student PISA Mathematics scores in low income non-OECD countries.

	Model 1	Model 2				Model 3				Model 4				Final Model			
	Uncond.	Fixed Effects				Random Slopes				Interactions							
Predictors	*b*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*
(Intercept)	387.24	391.33				391.69				395.33				397.13			
Sex (Male)		6.48	0.04	-1.94, 2.01	<0.001	6.26	0.04	-2.16, 2.24	<0.001	5.93	0.03	-2.16, 2.23	<0.001	5.93	0.03	-2.15, 2.23	<0.001
ESCS		13.59	0.19	-0.99, 1.20	<0.001	14.00	0.19	-1.14, 1.39	<0.001	17.25	0.24	-1.48, 1.90	<0.001	17.28	0.24	-1.48, 1.89	<0.001
Private School[a]		22.66	0.09	-11.46, 12.53	<0.01	23.26	0.09	-11.66, 12.78	<0.01	16.74	0.06	-12.29, 13.13	<0.01	14.43	0.06	-12.17, 12.89	<0.05
Indep School[a]		-5.48	-0.02	-13.80, 13.57	0.43	-7.59	-0.02	-14.41, 14.07	0.30	-27.34	-0.08	-17.40, 15.99	<0.001	-30.55	-0.09	-17.28, 15.72	<0.01
Educ. Mat. Short.		-6.38	-0.10	-2.69, 2.44	<0.001	-6.58	-0.10	-2.77, 2.51	<0.001	-8.58	-0.13	-3.06, 2.68	<0.001	-8.86	-0.13	-3.07, 2.67	<0.001
Teacher Certif.		4.00	0.04	-3.98, 4.14	<0.05	4.63	0.04	-4.05, 4.24	<0.05	4.79	0.05	-4.02, 4.21	<0.05				
Principal Lead		1.23	0.02	-2.80, 2.85	0.39	1.54	0.02	-2.87, 2.94	0.30	1.47	0.02	-2.85, 2.91	0.32				
Account – external		-0.18	0.00	-7.36, 7.36	0.96	-0.07	0.00	-7.50, 7.49	0.99	-0.33	0.00	-7.44, 7.43	0.93				
Account – internal		11.65	0.05	-8.16, 8.61	<0.01	11.52	0.05	-8.34, 8.79	<0.01	11.84	0.05	-8.26, 8.72	<0.01	12.13	0.05	-8.18, 8.64	<0.01
Account – parental		0.98	0.00	-9.57, 9.60	0.84	0.75	0.00	-9.77, 9.80	0.88	0.90	0.00	-9.69, 9.72	0.86				
Autonomy		-6.10	-0.07	-5.41, 5.05	<0.05	-6.45	-0.07	-5.52, 5.14	<0.05	-2.20	-0.02	-5.66, 5.52	0.44	-2.08	-0.02	-5.63, 5.50	0.46
Priv Sch*ESCS										-4.82	-0.03	-4.66, 4.54	<0.05	-4.78	-0.02	-4.65, 4.54	<0.05
Ind Sch*ESCS										-12.88	-0.09	-5.75, 5.25	<0.001	-12.55	-0.09	-5.73, 5.25	<0.001
Ed Mat Sh*ESCS										-1.72	-0.05	-0.99, 0.94	<0.01	-1.74	-0.05	-0.99, 0.94	<0.01
Autonomy*Sex										-3.88	-0.02	-2.25, 2.25	<0.01	-3.87	-0.02	-2.25, 2.24	<0.01
Autonomy*ESCS										2.69	0.02	-1.49, 1.52	<0.01	2.70	0.02	-1.49, 1.52	<0.01
Random Effects																	
σ^2	4745.7	4632.18				4541.7				4541.1				4542.2			
$\tau_{00\ (School)}$	2254.8	1523.6				1469.2				1456.6				1472.6			
$\tau_{00\ (Country)}$	380.7	446.1				540.5				496.6				465.7			
ICC $_{(School)}$	0.31	0.23				0.22				0.22				0.23			
ICC $_{(Country)}$	0.05	0.07				0.08				0.08				0.07			
Sex $_{(School)}$						154.34	0.91	-43.48, 118.8	<0.001	138.02	0.81	-46.31,112.0	<0.001	135.07	0.80	-46.73, 110.64	<0.001
ESCS $_{(School)}$						55.06	0.76	-14.32, 32.57	<0.001	43.96	0.61	-14.91, 28.35	<0.001	43.27	0.60	-14.96, 28.15	<0.001
AIC	224103.1	223405.8				223353.2				223307.6				223305.9			

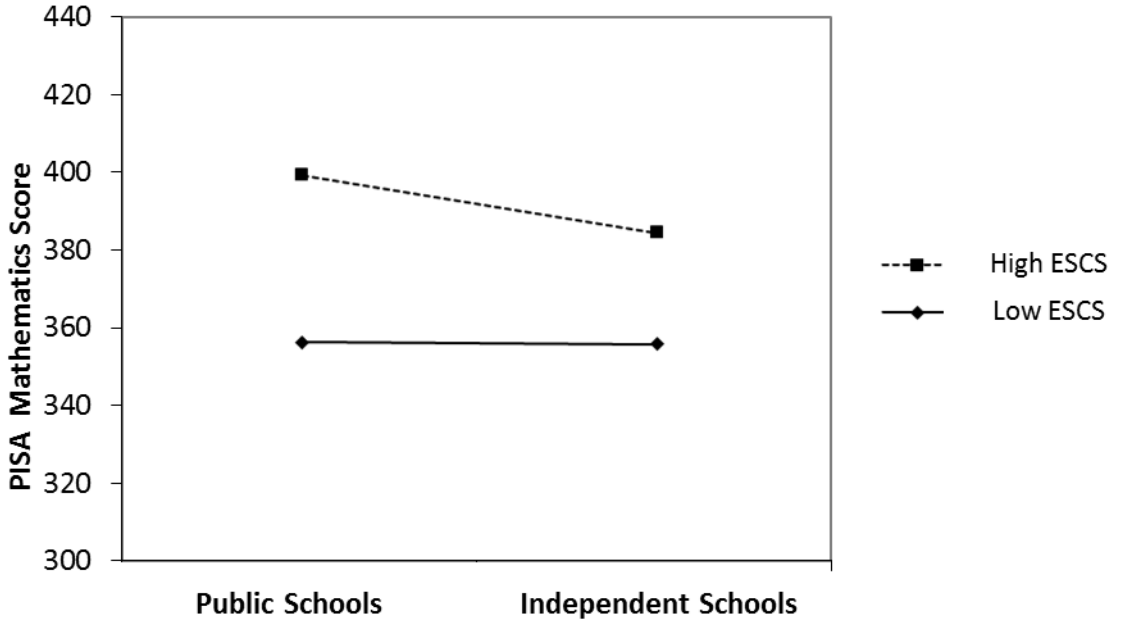
a. reference is public schools

Figure 7: Interaction effect of private schools with student ESCS on student PISA Mathematics scores in low income non-OECD countries



Conversely, as shown in Figure 8, in comparison to private and public schools, students with higher levels of ESCS attending independent schools had lower levels of attainment (β=-0.09, t=-4.48, p<.001) than peers in public schools.

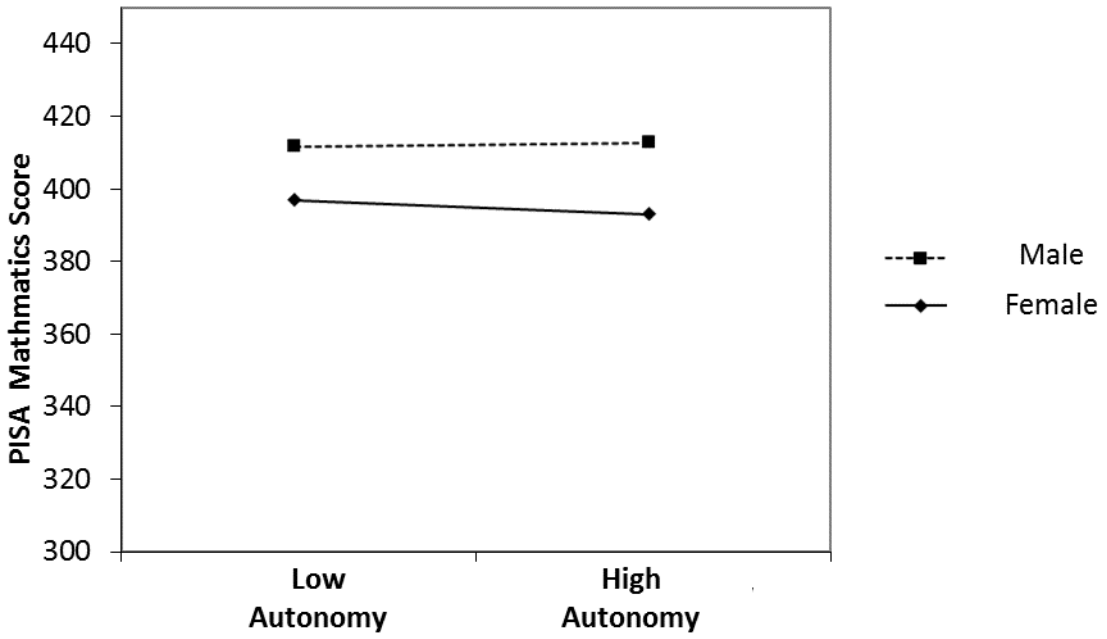
Figure 8: Interaction effect of independent schools with student ESCS student PISA Mathematics scores in low income non-OECD countries



Although school autonomy did not directly show a significant link with student maths performance, it did show associations through two interactions.

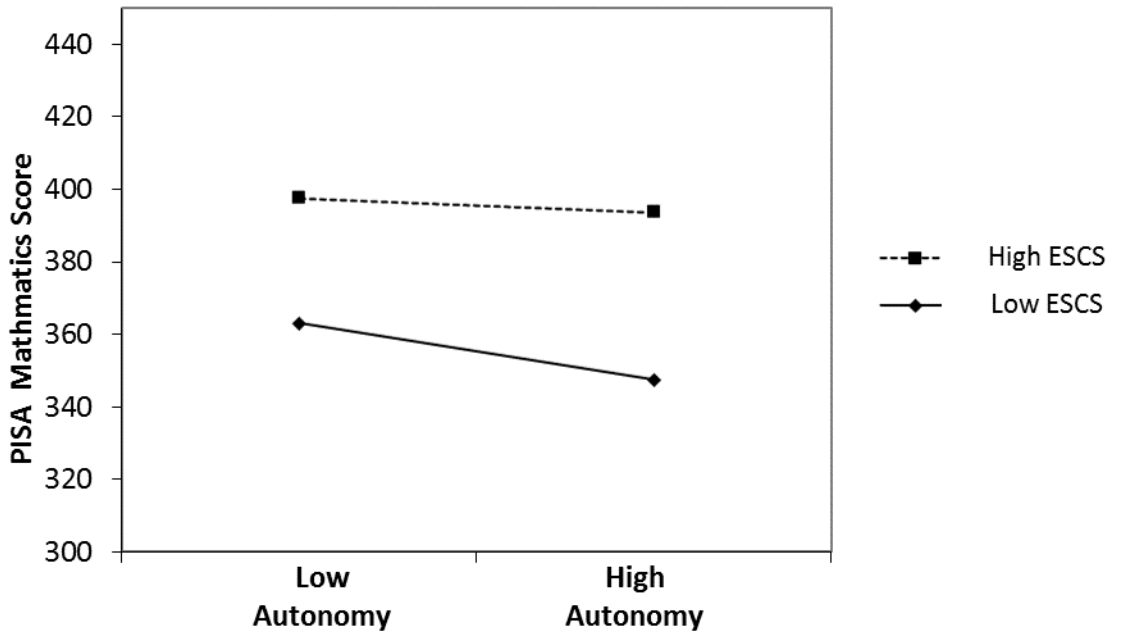
The interaction between school autonomy and sex (β=-0.02, t=-3.38, p<.01*)* presented in Figure 9 suggested that higher levels of school autonomy were associated with a slight widening of the attainment differences between males and females as females' scores were slightly lower in schools with more autonomy.

Figure 9: Interaction effect of school autonomy with student sex on student PISA Mathematics scores in low income non-OECD countries



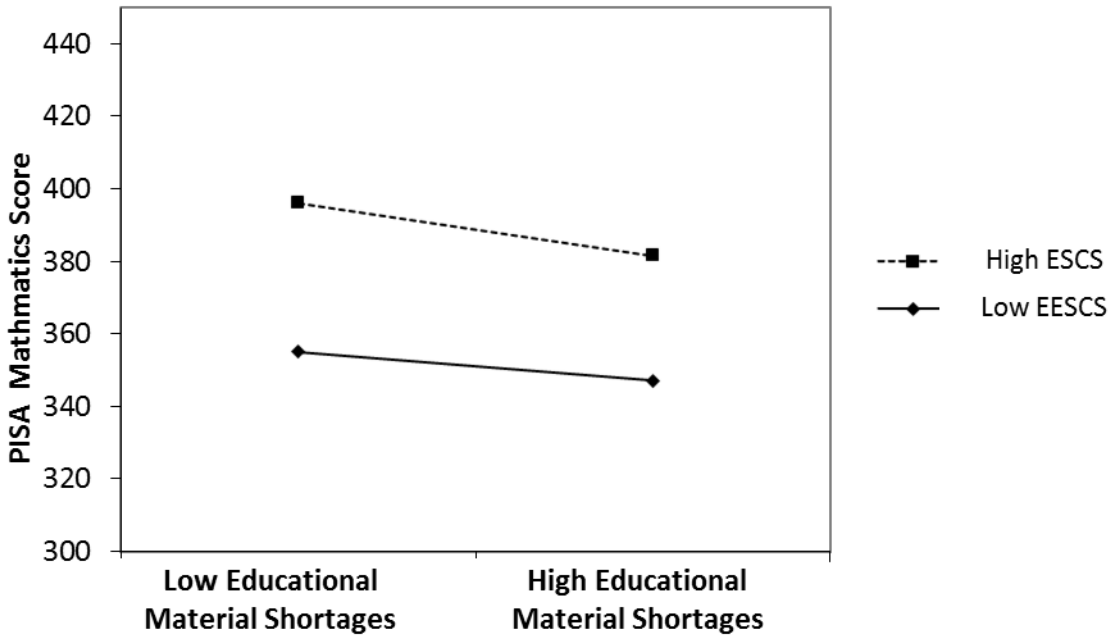
The interaction between school autonomy and student ESCS (β=0.02, t=3.52, p<.01) shown in Figure 10 suggested that higher levels of school autonomy were associated with lower levels of attainment among students with comparable levels of ESCS, with the strongest association among the least advantaged students.

Figure 10: Interaction effect of school autonomy with student ESCS on student PISA Mathematics scores in low income non-OECD countries



As shown in Figure 11 the interaction between school educational material shortages and student ESCS (β=-0.05, t=-3.52, p<.01) suggested that greater shortages were associated with lower levels of attainment among students irrespective of their levels of ESCS.

Figure 11: Interaction effect of school educational material shortages with student ESCS on student PISA Mathematics scores in low income non-OECD countries



There were no further significant interactions between variables.

These findings suggested that student ESCS was more strongly associated with student attainment in low income non-OECD countries than in those with higher levels of development. Attending a private school was associated with higher attainment and to some extent mitigated relative disadvantage. In contrast, independent school attendance was associated with lower levels of attainment and accentuated the negative association between low ESCS and lower student attainment. School educational material shortages had a stronger negative association with student attainment than in richer countries. Internal accountability was moderately associated with higher attainment levels. Autonomy was not directly associated with attainment, but through interactions appeared to be associated with lower maths scores among females and disadvantaged students. Overall, the fact that different school types had positive and negative associations with attainment, school autonomy was implicated in lower levels of performance for some students, and internal rather than external accountability was linked to higher attainment, cast doubt on the effectiveness of GERM in less developed countries.

Trinidad & Tobago

Table 7 shows the MLMs for Trinidad & Tobago, a high income country, and that a considerable amount of the variance in maths scores lay between schools (*ICC*=0.44) suggesting MLM was appropriate. As less than 3% of the Trinibogian students were immigrants this variable was dropped from the analyses. There were no significant random effects or interactions between variables so Models 3 and 4 are omitted. The Final Model fitted the data better than preceding unconditional and fixed effects models (*AIC*=34057.30) and accounted for a modest proportion of the variation in students' maths attainment (R^2=0.10).

There were no differences in maths attainment between the sexes. Higher levels of ESCS were modestly associated with higher maths scores (β=0.07, t=5.06, p<.001). External accountability was strongly linked to higher maths attainment (β=0.18, t=2.52, p<.05) as were higher levels of school autonomy (β=0.23, t=3.40, p<.001). There were no further significant fixed effects.

Table 7: Multilevel models analysing the relationships between student and school variables and student PISA Mathematics scores in Trinidad &Tobago

	Model 1	Model 2				Final Model			
	Uncond.	Fixed Effects							
Predictors	*b*	*b*	β	CI	*p*	*b*	β	CI	*p*
(Intercept)	413.00	375.86				413.45			
Sex (Male)		-2.62	-0.01	-5.80, 5.72	0.37			,	
ESCS		6.91	0.07	-2.58, 2.77	<0.001	3.88	0.07	-2.57, 2.76	<0.001
Private School[a]		-8.24	-0.01	-79.32, 78.23	0.84				
Indep School[a]		-45.58	-0.09	-73.10, 66.43	0.20				
Educ. Mat. Short.		2.24	0.02	-11.83, 12.13	0.71				
Teacher Certif.		6.39	0.06	-13.92, 14.79	0.38				
Principal Lead.		6.53	0.06	-14.74, 15.70	0.40				
Account – external		36.21	0.15	-32.19, 37.35	<0.05	43.56	0.18	-30.89, 36.95	<0.05
Account – internal		23.99	0.07	-41.98, 45.19	0.28				
Account – parental		26.38	0.11	-27.30, 30.68	0.07				
Autonomy		41.32	0.25	-21.17, 27.34	<0.01	37.92	0.23	-19.31, 24.42	<0.001
Random Effects									
σ^2	6382.88	6342.45				6345.84			
τ_{00}	5020.47	3515.58				3881.34			
ICC	0.44	0.36				0.38			
AIC	34093.4	34062.1				34057.3			

Although Trinidad & Tobago has a diversity of schools and there was considerable variation between schools, there were no significant associations between school type and student attainment. However, the very strong links between external accountability, school autonomy and student attainment suggested that aspects of GERM might be more effective in this country than in other high income or less developed non-OECD countries.

FYROM

The MLMs for FYROM, a middle income country, presented in Table 8, showed a fifth of the variance in maths scores lay between schools (*ICC*=0.20) suggesting MLM was appropriate. As less than 2% of the FYROM students were immigrants and 95% of schools were public, these variables were dropped from analysis. There were no significant interactions between the variables so Model 4 is omitted. The Final Model fitted the data better than preceding models (*AIC*=37420.80) and accounted for a modest proportion of the variation in students' maths attainment (R^2=0.10).

Males had higher maths scores than females (β=0.05, t=2.32, p<0.05). Higher levels of ESCS were associated with higher maths scores (β=0.15, t =8.88, p<.001) and shortages of educational material with lower attainment (β=-0.16, t =-3.10, p<.01). External accountability was positively associated with student achievement (β=0.14, t=2.48, p<0.01). There were no further significant fixed effects.

There were substantial differences in the nature of the associations between student ESCS and attainment across schools (β=0.94, z=1.99, p<.05) which remained unexplained by the MLMs.

These findings suggested that the links between attainment and the variables capturing the challenges associated with being a developing country – that is student ESCS, school shortages of educational material – were stronger in comparison to those found in richer countries, and the association with external accountability was weaker. This suggested that aspects of GERM might be effective in FYROM, but that addressing low levels of student ESCS and, school shortages of educational material might be equally effective at raising attainment.

Table 8: Multilevel models analysing the relationships between student and school variables and student PISA Mathematics scores in FYROM

	Model 1		Model 2			Model 3				Final Model			
	Uncond.		Fixed Effects			Random Slopes							
Predictors	*b*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*
Intercept	376.70	368.12				367.01				364.38			
Sex (Male)		9.32	0.05	-5.71, 6.01	<0.01	8.96	0.05	-7.38, 7.75	<0.05	8.92	0.05	-7.35, 7.72	<0.05
ESCS		15.09	0.15	-3.09, 3.59	<.001	15.08	0.15	-3.08, 3.58	<.001	15.10	0.15	-3.08, 3.58	<.001
Educ. Mat. Short.		-15.16	-0.15	-11.72, 10.05	<0.01	-14.88	-0.15	-11.59, 9.97	<0.01	-16.52	-0.16	-11.31, 9.56	<0.01
Teacher Certif.		2.77	0.02	-13.74, 14.08	0.70	4.19	0.04	-13.41, 13.90	0.55				
Principal Lead		1.99	0.02	-11.63, 11.87	0.74	2.43	0.02	-11.41, 11.69	0.68				
Account – external		27.57	0.14	-21.12, 24.31	<0.05	26.67	0.13	-20.77, 23.80	<0.05	28.47	0.14	-20.86, 24.13	<0.01
Account – internal		17.03	0.04	-53.29, 55.55	0.54	16.38	0.04	-52.22, 54.34	0.55				
Account – parental		-24.44	-0.08	-46.84, 43.08	0.29	-22.03	-0.07	-45.89, 42.56	0.33				
Autonomy		5.69	0.05	-11.01, 11.63	0.32	6.20	0.06	-10.82, 11.48	0.28				
Random Effects													
σ^2	10021	9778.3				9680.9				9681.0			
τ_{00}	2509.9	1632.5				1508.2				1592.8			
ICC	0.20	0.14				0.13				0.14			
ESCS						300.13	0.96	-132.53, 709.04	<0.05	293.30	0.94	-150.0, 426.50	<0.0
AIC	37519.0	37434.3				37428.2				37420.8			

Indonesia

The MLMs for Indonesia, a low income country, in Table 9 show approximately half the variance in maths scores lay between schools (*ICC*=0.49) suggesting MLM was justified. As the proportion of immigrants in Indonesia was well below 1% this variable was not included in analyses. The Final Model fitted the data better than preceding models (*AIC*=63590.00), and accounted for a fair proportion of the variation in students' maths attainment (R^2=0.21).

Males had higher maths scores than females (β=0.03, *t*=2.82, *p*<.01). Higher levels of ESCS were associated with higher maths scores (β=0.18, *t*=9.80, *p*<.001). Attending independent schools was very strongly associated with lower maths scores (β=-0.29, *t*=-5.27, *p*<.001) in comparison to attending public or private schools. Educational material shortages were associated with lower student attainment (β=-0.22, *t*=-4.45, *p*<.001). The leadership effectiveness of principals was positively associated with students' maths attainment (β=0.10, *t*=2.60, p<.01). There were no further fixed effects.

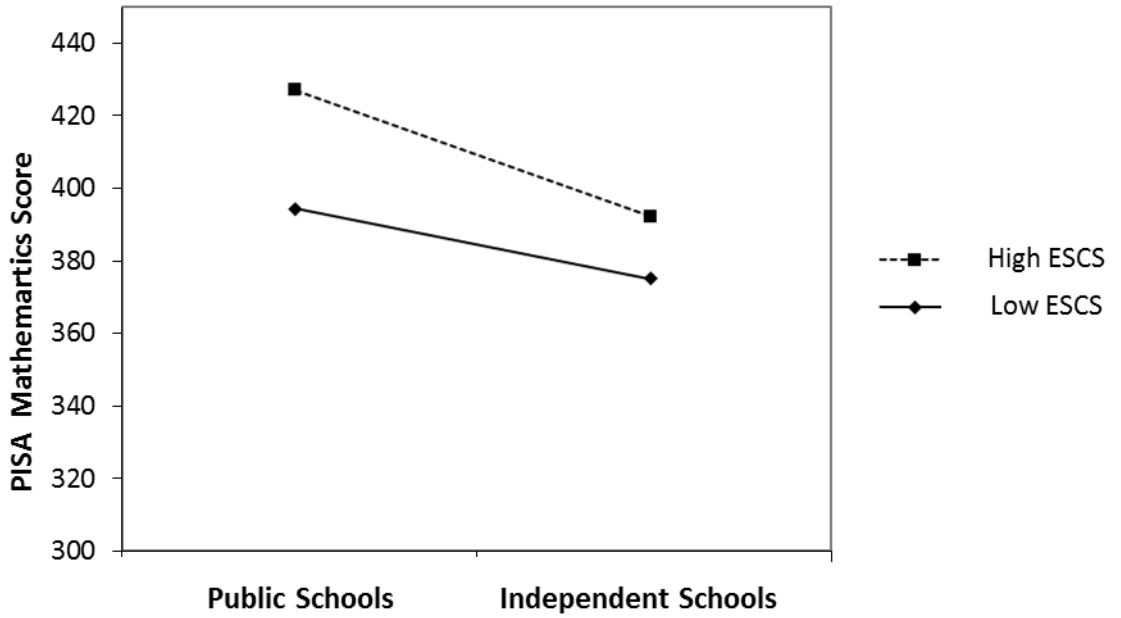
There were significant variations in the associations between ESCS and attainment across schools (β=0.51, *z*=2.35, *p*<.01). As shown in Figure 12, the interaction between independent schools and student ESCS (β=-0.12, *t*=3.95, *p*<.001) suggested that attending an independent school was associated with lower levels of maths attainment in comparison to students with similar levels of ESCS attending public schools.

Table 9: Multilevel models analysing the relationships between student and school variables and student PISA Mathematics scores in Indonesia

	Model 1	Model 2				Model 3				Model 4				Final Model			
	Uncond.	Fixed Effects				Random Slopes				Interactions							
Predictors	*b*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*	*b*	*β*	CI	*p*
(Intercept)	391.25	424.13				424.72				430.45				427.07			
Sex (Male)		4.36	0.03	-3.10, 3.19	<0.01	4.35	0.03	-3.10, 3.18	<0.01	4.50	0.03	-3.09, 3.18	<0.01	4.51	0.03	-3.09, 3.18	<0.01
ESCS		10.26	0.14	-1.66, 1.92	<0.001	10.38	0.15	-1.91, 2.22	<0.001	12.40	0.17	-2.27, 2.72	<0.001	12.48	0.18	-2.27, 2.72	<0.001
Private School[a]		9.22	0.04	-21.19, 22.09	0.40	9.58	0.04	-21.59, 22.54	0.39	9.30	0.04	-22.56, 23.53	0.43	0.33	0.00	-19.57, 19.60	0.97
Indep School[a]		-11.48	-0.06	-20.23, 19.04	0.25	-14.83	-0.08	-21.01, 19.43	0.15	-45.59	-0.24	-27.11, 21.28	<0.01	-55.82	-0.29	-23.84, 17.71	<0.001
Educ. Mat. Short.		-9.02	-0.15	-4.92, 4.23	<0.01	-9.46	-0.15	-5.09, 4.34	<0.001	-13.08	-0.21	-6.63, 5.32	<0.001	-13.64	-0.22	-6.69, 5.32	<0.001
Teacher Certif.		4.60	0.05	-7.40, 7.79	0.24	4.63	0.05	-7.62, 8.02	0.25	5.47	0.06	-7.36, 7.82	0.16				
Principal Lead.		5.87	0.09	-4.89, 5.36	<0.05	6.49	0.10	-5.02, 5.55	<0.05	6.02	0.09	-4.89, 5.37	<0.05	6.81	0.10	-4.87, 5.41	<0.01
Account – external		10.52	0.06	-12.97, 13.80	0.12	10.86	0.06	-13.25, 14.13	0.12	10.73	0.06	-12.87, 13.71	0.11				
Account – internal		-9.20	-0.02	-30.41, 29.72	0.55	-8.81	-0.02	-31.13, 30.46	0.57	-8.76	-0.02	-30.17, 29.52	0.56				
Account – parental		2.39	0.01	-22.16, 22.34	0.83	2.69	0.01	-22.74, 22.96	0.82	2.52	0.01	-22.15, 22.34	0.82				
Autonomy		-5.89	-0.05	-12.03, 11.40	0.32	-6.21	-0.06	-12.34, 11.67	0.31	-6.43	-0.06	-11.97, 11.29	0.28				
Private Sch*ESCS										2.22	0.02	-6.22, 6.33	0.49	2.17	0.02	-6.23, 6.33	0.50
Ind Sch*ESCS										-9.93	-0.12	-5.18, 4.56	<0.001	-9.82	-0.12	-5.18, 4.57	<0.001
Priv Sch*Ed M Short										4.09	0.03	-12.61, 12.94	0.53	3.71	0.02	-12.64, 12.94	0.57
Ind Sch*Ed M Short										12.97	0.12	-10.01, 11.33	<0.05	12.54	0.12	-10.16, 11.45	<0.05
Random Effects																	
σ^2	2893.30	2843.4				2808.65				2807.55				2807.55			
τ_{00}	2789.72	1792.8				1764.02				1653.93				1707.99			
ICC	0.49	0.39				0.39				0.37				0.38			
ESCS						44.20	0.62	-22.00, 42.52	<0.01	35.72	0.50	-22.10, 37.38	0.01	35.84	0.51	-22.19, 37.62	0.01
AIC	63778.9	63619.6				63609.7				63590.0				63590.0			

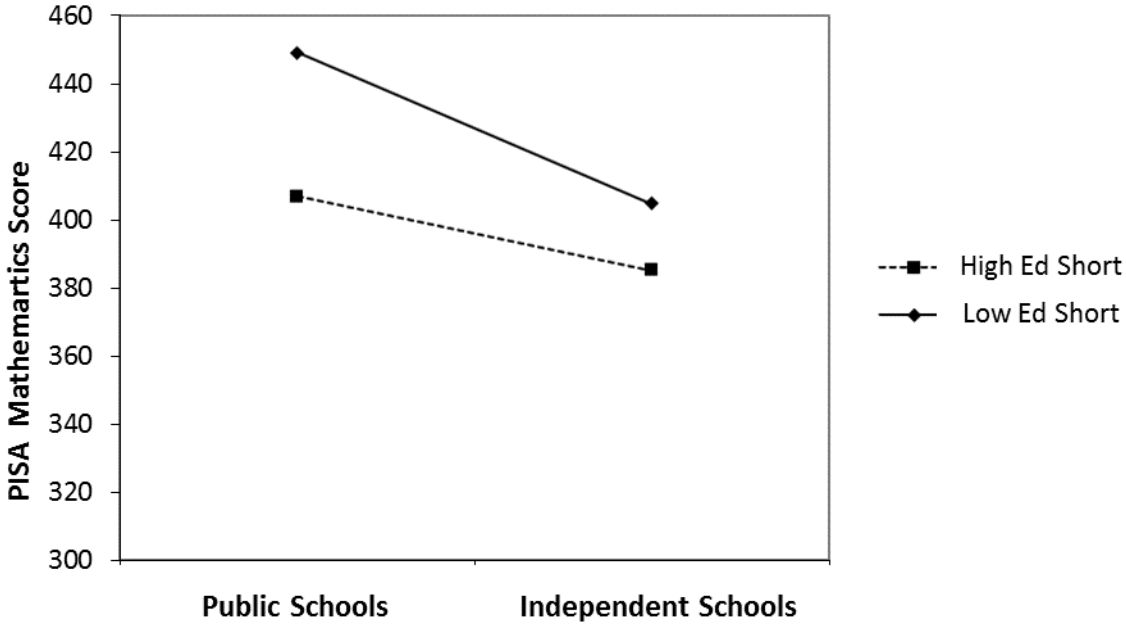
a. reference is public schools

Figure 12: Interaction effect of independent schools with student ESCS on student PISA Mathematics scores in Indonesia



The interaction between independent schools and educational material shortages (β=0.12, t=2.28, p<.05) shown in Figure 13 suggested that attending an independent school compounded the adverse association between material shortages and student maths attainment.

Figure 13: Interaction effect of independent schools with educational material shortages, on student PISA Mathematics scores in Indonesia



There were no further interactions between other variables.

These findings suggested that students attending government supported schools outside direct state control had lower levels of attainment than peers in fee-paying and public schools, and that this negative association might be accentuated for the poorest students in the least resourced schools. There were no links between external accountability or school autonomy and student attainment, so little indication that these aspects of GERM might be effective in Indonesia. Indeed, the fact that student ESCS, the leadership effectiveness of school principals and educational material shortages were all linked to student attainment suggested that traditional 'people and resources' approaches to raising student attainment might be more effective than GERM in Indonesia.

Discussion

Analysing the PISA data for a group of non-OECD countries, all of which had implemented GERM related reforms and which were balanced by global region and level of economic development, highlighted some similarities and potential differences with respect to previous research about the ways in which school accountability, autonomy and diversity may relate to students' maths performance within and between countries. The inclusion of variables relating to teacher certification, school leadership and educational material shortages also suggested that factors which are outside the GERM prescription may also have associations with student attainment in developing countries.

Given the size of the datasets analysed, many of the associations found between the school level variables of interest and students' maths attainment were at best modest. This might have been anticipated in light of the large differences that might be expected between schools drawn from so many varied cultures and contexts, a point underscored by the differences discussed below. However, this was also perhaps an important finding because the fact that most associations reported here between external accountability, school autonomy, school type and student attainment were comparatively modest in a sample of non-OECD countries which had implemented GERM related initiatives might be taken as an indication that these factors central to GERM might not be globally effective.

Student Variables

The factors impacting individual students were not the focus of the present research and the pattern of associations between these student level control variables and maths attainment were broadly consistent with previous research. Males generally scored higher than females in line with previous findings (Lindberg et al., 2010; Yavuz et al., 2019), except in Trinidad & Tobago where equality in scores was likely due to a relative decline in males' performance than a positive increase in females' (Louisy, 2004). First generation immigrants mostly achieved lower maths scores than natives as has been

found elsewhere (Giannelli & Rapallini, 2016; Moschkovich, 2012) with the performance differential disappearing in the second generation. However, among the most developed countries first generation immigrants appeared to have higher levels of maths attainment than natives, a finding that might be explained by the fact that Hong Kong was in this group and had by far the largest proportion of immigrants, of whom more than 80% came from China (World Bank, 2013c). This means that most first generation immigrants to Hong Kong were already highly assimilated linguistically and culturally, and therefore that they were likely to achieve the same high levels of maths attainment found among native Hong Kongers, which in turn would mean that they generally had higher scores than most natives in the other countries sampled. ESCS had a significant impact on student performance, with students from less advantaged backgrounds achieving lower maths scores in common with previous research (Harden et al., 2013; Tucker-Drob & Briley, 2012)

Differences between Schools and Countries

The first step in any MLM exercise is to calculate a model which makes it possible to assess the amount of variance between grouping or clustering units, in the present study that is schools and countries (Heck et al., 2014; Hox, 2010). As this step is generally viewed as a preliminary exercise little attention seems to be given to it, yet it may well have been revealing. Amongst the most developed countries there was variation between schools and countries, among the less developed countries there was variation between schools but little variation between countries. Lower levels of economic development were associated with lower levels of student attainment and as there was also less variation in attainment between these countries this implies that the differences, or lack of differences, between countries may to some extent have been a function of national development. So, for example, increases in national development may not eliminate relative differences in wealth between students within a country, but the absolute standards of ESCS for most students will rise, thus reducing a significant barrier to attainment for all students in that country (Gustafsson, Nilsen, & Hansen, 2018). Higher levels of national development are generally paralleled by improvements in

educational systems and resourcing, leading to better availability of educational materials, enhanced support infrastructure, and the development of standard national practices (Ganimian & Murnane, 2016), so the positive effects of higher national income at the individual level may be multiplied at the school level and lead to some national consistencies. Thus, as a country becomes more developed it may evolve a range of distinct systemic features across schools and even though the relative differences in student attainment between schools within the country may be preserved, the absolute differences in student attainment between schools in that country and schools in less developed countries will increase leading to greater national differentiation. Conversely, poorer developing countries face the challenging combination of low student ESCS, poorly resourced schools and fragmented or non-existent national systems (Ehren et al., 2017; Glewwe, Hanushek, Humpage, & Ravina, 2016; Mbiti et al., 2019) so among countries at lower levels of development the relative differences between schools within countries may simply be mirrored across countries. If this is so, then the constraints associated with national development levels are likely to impact every other aspect of education activity, including reform. This perhaps also implies that in the absence of effective national systems and consistencies among schools it may not be possible to implement complex reforms such as those envisaged in GERM.

Following from the above, it might be expected that the GERM related variables would show their strongest associations with student attainment in the most developed and wealthiest non-OECD countries. However, the high income countries were actually the grouping in which the variables of interest accounted for the least variance in student attainment. This finding might be best explained with reference to the GERM inspired reform programme which took place in Trinidad & Tobago, or more accurately with reference to the fact that Trinidad & Tobago considered it necessary to implement domestically led reforms beyond GERM, suggesting that GERM was not viewed as an all-embracing panacea. GERM arose as a reaction to the perceived educational stagnation in a number of developed countries (Walford, 1990) not as a response to a global problem (Davies & Guppy, 1997).

Thus, in many non-OECD countries GERM may be an historically inappropriate 'solution' to 'problems' these countries never had, and may also ignore a range of other locally specific factors that might already have played some part in raising student attainment (Adamson, Forestier, Morris, & Han, 2017; Layne, Jules, Kutnick, & Layne, 2008). Therefore, the GERM related variables may have accounted for little of the variance in student attainment among the most developed non-OECD countries simply because they were not particularly relevant. Indeed, even in an anglophone, high income country like Trinidad & Tobago which shares many historical and cultural ties with those countries where GERM began, the resistance to the implementation of GERM inspired reforms under the IDB's SEMP initiative was both broad and deep because within the local context the measures were at best seen as unnecessary and at worst as "borrowed, distorted, or non-existent" (De Lisle, 2012, p. 67).

Clearly this suggests two related points: first that GERM may not be suitable for countries outside the OECD because these countries may face different challenges to which alternative, locally relevant reform measures are more suited. Second, that there are factors outside GERM, and also beyond the additional variables selected for the present study, which may be more salient to student attainment in non-OECD countries. In short, beyond the countries of the OECD GERM may be both inapplicable and incomplete.

School Accountability

The most enthusiastic proponents of GERM tend to emphasise the benefits of external accountability (Figlio & Ladd, 2015; Rouse, Hannaway, Goldhaber, & Figlio, 2011), but in the present study measures of parental and internal accountability were also included making some comparisons possible.

The vast majority of schools shared assessment information with parents, yet only in the overall sample and the most developed countries was there an association with student attainment. This

may have been because schools in the high income group of countries had the lowest rates of practising parental accountability, so the association between parental accountability and student attainment may well have emerged because only in such a large sample was there a sufficient number of schools *not* practising parental accountability for the difference in attainment between schools to be revealed. If this is so, then it may be that sharing accountability information with parents is not so much associated with higher student attainment, rather that schools with low levels of student attainment may not practise parental accountability. In any case, in terms of GERM a primary purpose of accountability is to provide information to help parents choose schools and the most potent accountability information is that relating to final exams (Figlio & Ladd, 2015). However, parental accountability only happens *after* a student is enrolled at a school so it is unlikely to affect school choice to any great extent, and information about final exams comes too late for parents to exert pressure for improvement, so how far parental accountability may directly affect student attainment is not clear.

None of the jurisdictions sampled for the present study had a policy of external accountability through the mandatory publication of assessment results, though a proportion of schools voluntarily practised external accountability. In the high and middle income countries external accountability was associated with higher levels of student attainment but in the low income countries, where the rates of publication were slightly higher than for the most developed countries, there was no similar link. The same pattern was found in the three individual countries examined, with external accountability being positively associated with maths performance in the high and middle income countries of Trinidad & Tobago and FYROM, but not in the least developed, Indonesia. This could be taken as demonstrating that external accountability is an effective means to raise attainment in high and middle income countries as its proponents would argue, yet in educational systems where external accountability is not mandatory it is also possible that a slightly different form of accountability is taking place. When external accountability is not mandatory schools have the choice to publish

assessment data or not, so those with high levels of attainment may choose to publish their assessment results as a means to signal their high quality to prospective parents and to reassure existing stakeholders. However, once a school has embarked upon a policy of publishing its results, to suddenly stop publishing them would be tantamount to an admission that student attainment had declined. So, once a school has started to publish assessment data, what began as voluntary exercise may become a necessity, yet it may also become the guarantor of high levels of student attainment because if results continue to be published then it is necessary to at least maintain the same standards of attainment. This would mean that any form of voluntary external accountability would be likely to be associated with higher levels of student attainment because initially external accountability might be an outcome of higher student attainment, but after the initial publication of results the need to publish results in future would become the reason to maintain higher student attainment.

The majority of schools in the present study were subject to internal accountability and this form of accountability was significantly associated with higher attainment in the overall sample, and specifically among the low income countries where external accountability did not demonstrate similar associations. As internal accountability was practised in most schools the appearance of this association might again have been a function of a sample large enough to make the contrast with the minority of schools not practising effective performance monitoring. Previous research has shown that regulatory oversight is important to ensure standards are maintained throughout schools (Lubienski & Lubienski, 2014; Malin et al., 2019) and when effective internal accountability systems are implemented in less developed countries student attainment rises (Ehren et al., 2017; Mbiti, 2016). In the context of the poorest developing countries the fact that internal accountability takes place at all may be indicative that the broader educational system is also progressing in other ways that support the capacity for improvement (Glewwe & Muralidharan, 2016). This means that internal accountability systems may act both as effective means to raise student attainment and also as a barometer of broader systemic improvements that may lead to more general increases in student

attainment. Therefore, it may be that the association between internal accountability and student attainment emerged amongst the least developed countries precisely because these were the countries in which evolving systems of regulation and accountability still lacked consistency across all schools, thus leading to clear positive impacts on attainment where internal accountability was practised.

The most effective accountability systems may combine both internal and external accountability (Ehren et al., 2015) and internal accountability systems may have elements of external accountability through the publication of reports and other information (Altrichter & Kemethofer, 2015). In developing countries this composite form of accountability may be present through regulatory systems like the Indonesian BOS framework which, it will be recalled, is the national regulatory system used to oversee independent schools (Asadullah & Maliki, 2018). Although GERM style external accountability data are not published in Indonesia, BOS may still make schools externally accountable because of the fact that government financial support is dependent on compliance with the BOS framework (Harris et al., 2017; Sumintono et al., 2014). Any changes in the government grants made to independent schools are apparent to the community (Holzhacker et al., 2016) which means that the financial outcomes associated with the BOS system of internal accountability may also function as a form of local external accountability, but in a way that avoids the potential divisiveness of national league tables.

Thus, it may be that ultimately some form of external accountability, even if it takes a less direct form than the publication of league tables, is the most effective way of raising student attainment. Once student data are in the public domain schools may feel pressure to maintain levels of attainment, but similarly schools whose funding is based on internal accountability may be subject to the pressures of external accountability if the amount of funding given or withheld is publicly known. Although external accountability through the formal compilation of assessment data into published

performance tables may be the most effective approach to accountability in some contexts, perhaps any form of external accountability, even if it is of an informal and local kind, may also be effective and even be preferable in places where league tables could be divisive.

School Autonomy

Whatever the cause which impels improvement in student attainment, GERM prescribes school autonomy as the best way to achieve it (Arar & Nasra, 2020; Barrera-Osorio et al., 2011). Yet, in the most developed countries where it might have been expected that autonomy would be most evident there was no association between autonomy and student attainment, with the exception of an association for Trinidad & Tobago.

Among the middle income countries autonomy was also positively associated with attainment, but neither in these countries nor in Trinidad & Tobago was there any direct association with accountability or school type in the ways that might have been expected if the GERM-related factors were operating together. In the middle income group of countries alone there was an interaction with student ESCS which suggested that students attending schools with higher levels of autonomy had higher levels of attainment than peers with similar levels of ESCS attending less autonomous schools. This may be consistent with previous research which has suggested that some schools in developing countries were able to raise student attainment by using their limited resources more flexibly (Glewwe et al., 2016) and also that autonomy allows innovation in teaching and learning that can to some extent overcome disadvantage (Henríquez et al., 2012).

In contrast, among the least developed countries autonomy did not have a direct association with attainment, rather there was some evidence that school autonomy might be linked to lower attainment among females and students with lower levels of ESCS. Although this might be seen as supporting the conclusions of Hanushek and colleagues (2013) that autonomy is associated with lower

attainment in low income countries, that could almost be taken as suggesting that some schools somehow exercised their autonomy to the disadvantage of their female and poorest students. However, given other evidence it is possible that the association results from the context in which these schools operate. Many schools in the poorer, rural areas of developing countries exercise higher levels of autonomy simply because they are remote from the central authorities and need to make most decisions on their own (OECD/Asian Development Bank, 2015). Such schools are often poorly funded and have a high proportion of students with low ESCS, factors which are likely to reduce attainment even among slightly more advantaged students (Stern & Smith, 2016). Furthermore, in many developing countries female schooling is not prioritised by parents (Jayachandran, 2015), so while sons may be sent to better resourced schools in larger towns, daughters are more likely to attend the less well-resourced village school (Asadullah & Maliki, 2018). Thus, the negative associations between autonomy and attainment may not reflect what schools do, rather it may capture the realities of the challenges faced by schools at the deprived margins.

Overall, it appears that autonomy may not be a contributor to student attainment in the majority of developed non-OECD countries, though autonomy may be beneficial in some middle income countries because it may allow schools the freedom to adapt to local conditions and innovate in teaching and the use of resources. However, in the poorest countries autonomy may be less a benefit that raises student attainment than an indicator and consequence of the challenging context in which some schools are located. Whatever is the case, it appears that school autonomy may only be linked to higher student attainment in certain circumstances and, therefore, that contrary to GERM more school autonomy may not be an effective prescription in every country.

School Diversity

Across higher and middle levels of national development attendance at private fee-paying schools was associated with higher levels of attainment in comparison to public and independent schools, with males appearing to benefit slightly more in terms of higher attainment.

The bulk of students at private schools are drawn from backgrounds with higher levels of ESCS (Parker, Marsh, Jerrim, Guo, & Dicke, 2018) so the higher levels of attainment may simply have been due to this in-built advantage rather than to any intrinsic merits of the schools themselves. As shown in this and other research males generally outperform females in maths and though it is possible that private schools accentuated this trend, it is also possible that a slightly higher proportion of single sex schools amongst private schools in some countries (Asadullah & Maliki, 2018) may have been the cause. Although attending a single sex school does not of itself bestow any educational advantages (Pahlke, Hyde, & Allison, 2014), it may propagate myths about gender roles and differences in the maths aptitude of the sexes (Nuraeni, 2020; Sarouphim & Chartouny, 2017) which may affect students' performance (Kollmayer, Schober, & Spiel, 2018). Thus, rather than private schools conferring a greater advantage on males, it may actually be the case that some private schools slightly supressed the full maths potential of females.

The strength of the association between private schooling and student attainment was weaker in the countries with lower levels of national development, and there were no similar associations between attendance at private schools and attainment at all in Indonesia. The finding that private schools in the poorest societies might not have higher absolute levels of attainment supported the view that any apparent advantage was actually a function of students' higher relative levels of ESCS and also underscored the previously made point that the association between social advantage and student attainment is proportional to the overall ESCS of the societies in which schools operate.

Turning now to independent schools, in the most developed non-OECD countries there was no association between attending an independent school and higher levels of attainment. In middle and low income countries attendance at independent schools appeared to be negatively associated with student attainment. Moreover, there was some evidence which suggested that in the least developed countries independent schools reduced the benefits associated with higher levels of student ESCS and compounded the negative impact of educational material shortages, associations seen most strikingly in Indonesia. This may reflect the fact that independent schools in less developed countries are often the least well-funded (Tooley & Longfield, 2016) and often subject to less regulatory oversight (OECD/Asian Development Bank, 2015) which is why they may actually sustain or even increase social and educational inequalities (Newhouse & Beegle, 2006). This is why the Indonesian government has concluded that increasing the number of independent schools may be counterproductive and it is instead is trying to grow public provision (Nababan et al., 2019).

The findings of the present study suggest that the GERM prescribed increase in school diversity may only be suitable in more developed non-OECD countries which have educational systems sufficiently evolved to support all schools, but in these countries such measures may have little impact. In less developed countries the same GERM related measures which seek to increase school diversity by expanding the number of independent schools might actually be counterproductive. Overall, this would suggest that the GERM prescription for increased school diversity might not offer a globally effective model for reform.

Educational Material Shortages

Educational material shortages were associated with lower levels of student attainment for the whole sample, for the middle and low income countries and for FYROM and Indonesia consistent with previous studies (Ganimian & Murnane, 2016), but there were no similar associations for Trinidad & Tobago or the other highly developed countries.

Across the middle and low income groups there was a close parallel between the size of the association between educational material shortages and attainment and levels of national development such that the shortage of materials had its greatest impact in lower income countries consistent with earlier research (Glewwe et al., 2011). In both middle and lower income countries the negative associations between educational material shortages and attainment were compounded for students with lower ESCS. This pattern of associations was consistent with previous findings from World Bank studies suggesting that educational material shortages are likely to have their greatest impact on the poorest students in developing countries (Fuller, 1987; Lockheed et al., 1986). This may also account for the absence of any link between educational material shortages and attainment in the high income non-OECD countries, because once a country achieves developed status access to educational materials may increase and median levels of student ESCS may be sufficient to blunt the negative impact of any shortages that do occur. If this is the case, then it is no surprise that GERM, with its roots in wealthy OECD countries, pays little regard to educational material shortages because such shortages are comparatively unimportant in those countries. However, this also suggests that in countries where educational material shortages or other basic factors may have an impact on attainment then GERM may not offer a comprehensive or relevant template for reform.

Teacher Certification

Perhaps surprisingly, the proportion of teachers fully certified was not significantly associated with improved student attainment across or within any of the countries. Clearly in any country the presence of better qualified teachers is likely to be beneficial and this is why countries like Indonesia and FYROM have taken steps to upskill their teachers (Bakiu et al., 2016; Fahmi, et al., 2011; Harjanto, et al., 2018; Pantic, 2012; Pantić et al., 2011; Raihani & Sumintono, 2010). It seems reasonable that support for teacher development should form part of any reform package in developing countries (Glewwe et al., 2016), and the lack of an association between teacher certification and student

attainment in the present research may actually indicate support for the view that teacher quality only becomes a significant issue when there are sufficient teachers for performance differentiation (Arias Ciro & Torres García, 2018), though further research would be needed to confirm this point.

Principal Leadership Effectiveness

The effectiveness of principals' leadership did not have an association with student attainment in Trinidad & Tobago or FYROM or any of the transnational comparisons, so only in Indonesia was the leadership of the principal associated with student attainment. This may have been because in Indonesia the poor implementation of SBM in public schools or the large percentage of independent schools has placed a greater onus on principals to manage a portfolio of complex relationships and issues (Koo & Park, 2018). However, in light of the research highlighting the importance of principals' leadership in raising student attainment (Takeda et al., 2014) it seems unlikely that leadership was unimportant elsewhere. Yet, the fact that even in an anglophone, high income country such as Trinidad & Tobago there was no association between principals' leadership effectiveness and student attainment may be instructive. One reason the SEMP programme was unsuccessful was because it attempted to impose international models of school leadership that were not suitable for Trinabogian schools (Joseph, 2010). After SEMP, locally conducted research suggested that successful Trinabogian principals possessed some characteristics not found among principals from western nations (De Lisle et al., 2019). In other non-OECD and developing countries principals' styles of leadership may also differ from what might be considered western norms (Nir & Hameiri, 2014; Takeda et al., 2014; Xia, Gao, & Shen, 2017). This raises the possibility that the lack of association between principals' leadership and student attainment in the present research was not because leadership was unimportant, but because the PISA School Questionnaire was not capturing all the aspects of leadership relevant in some non-OECD countries.

PISA

PISA is a product of the OECD, developed in the context of GERM, which captures information based on western norms and expectations (Elliott, Stankov, Lee, & Beckmann, 2019). If the PISA School Questionnaire did not capture or reflect aspects of leadership that might be relevant to countries outside the OECD, then this may be equally true for other scales. This raises the possibility that the relatively modest associations found between the variables of interest and student attainment may have been due to the fact that the School Questionnaire was not construct valid outside OECD nations. Given that, as Elliott and colleagues (2019) point out, PISA scales are designed to tap constructs that are relevant to GERM, this also implies that GERM may not be valid outside the countries of the OECD. The OECD may already have implicitly acknowledged this possibility through its PISA for Development initiative (OECD, 2016c) which includes a Contextual Questionnaire containing items that "were drawn from PISA and other international and regional studies, or were developed in consultation with the PISA-D participating countries." (OECD, 2019b, 3:5). The inclusion of items from outside the existing PISA could be seen as suggesting that even the most ardent promoters of GERM may be beginning to acknowledge that GERM may not provide a comprehensive template for reform in non-OECD countries.

Limitations

The present study was based upon a single survey at a particular point in time, but using such cross-sectional data means it is impossible to capture any causal relationship and that is why mention of such predictive relationships has been carefully avoided. Although the variables of interest did exhibit a range of associations with student attainment, without replication in other samples, ideally using time series data, any findings reported here are at best tentative.

All the school level variables were drawn from the PISA School Questionnaire and any survey which relies on a single respondent may be unreliable (Blasius & Thiessen, 2015). Principals' faulty memory

for what has taken place over the previous year may lead to inaccurate responses (Walentynowicz, Schneider, & Stone, 2018) and there may even be deliberate distortion of the information provided despite the fact that PISA responses are anonymised (Dalal & Hakel, 2016).

The analyses reported here were constrained by the availability of data sufficient to meet the aims of this research. Across PISA cycles there is a surprisingly large variation in participating non-OECD jurisdictions, as well as frequent changes in the PISA School Questionnaire items and the ways in which aggregate indices are calculated. This made it impossible to construct a consistent sample of non-OECD countries across PISA cycles with which to investigate the associations between all the school level variables of interest and student attainment. Moreover, despite the fact that the deletion of cases with missing data is recommended by the OECD (2009), there is always the possibility that removing cases may impact the quality of the findings (Graham, 2009), though in large samples such as those used here the application of the appropriate sampling weights should correct for missing cases (Cheema, 2014).

Future Research

It would obviously be beneficial if future research were able to address the limitations set out above by constructing a large and consistent time-series dataset. However, previous research which has attempted to do this has had to use widespread imputation and a range of econometric methods which are not suitable for data with a hierarchical structure like the PISA survey (Komatsu & Rappleye, 2017), so it will undoubtedly be a challenge.

As discussed at the outset, most research using PISA data is conducted on large international samples which have a majority of developed OECD countries. The present study suggests that there are variations between countries at different levels of development, and that even what holds true within those broad groupings may not be true at the level of individual countries. This implies that any future

research on PISA should take this diversity into account and rather than simply relying upon senate weights to ensure "each country as an equal contributor to an estimate." (Stapleton, 2014, p. 367) perhaps due regard must also be given to the fact that overall PISA samples may be inherently weighted towards wealthier developed countries that are different in many ways to those which are developing.

The results reported here suggest that there are other variables outside both GERM and the present study which may be salient to student attainment and further research is needed to identify what these may be. Such research might also consider the extent to which the PISA School Questionnaire, and for that matter the PISA for Development Contextual Questionnaire, are capturing information about the key variables associated with student attainment and how far those variables are globally construct valid.

The present study and the literature adduced for it suggest that principals may have an impact on the success of any improvement programme in developing countries and that leadership attributes in non-OECD countries may differ from those in western countries. Future research could seek to clarify exactly which aspects of school leadership may be internationally generic and which specific to levels of national development, global regions or even individual countries. This would not only inform educational reform programmes, it might also help to harness the full potential of the people who may be most important to ensuring that reforms in developing countries succeed (Takeda et al., 2014).

It has been consistently shown that a lack of resources and poor teacher quality have significant impacts on student attainment in developing countries (Fuller, 1987;Glewwe et al., 2011) so it seems only logical that these factors should be given some consideration during the planning of any educational reform package. Therefore, it might be helpful to conduct further research to contrast the costs and potential benefits of implementing GERM inspired initiatives versus programmes such

as those that were domestically led in Trinidad & Tobago, FYROM and Indonesia and which sought to address alternative areas of reform. Such research might identify how international funding could be best directed to those measures likely to have the maximum impact on student attainment in the countries concerned.

Conclusions

The present study set out to assess the global effectiveness of GERM with a particular emphasis on the associations between student attainment and school external accountability, autonomy and institutional diversity in non-OECD countries which had implemented GERM related reforms. The findings discussed above suggest that some features of GERM may have been effective in these non-OECD countries and may, therefore, be potentially globally effective. However, the findings presented here also suggest that some aspects of GERM may at best have been neutral and at worst may have been counterproductive in some of these countries.

All forms of accountability were associated with higher student attainment. External accountability was found to be most positively associated with student attainment in high and middle income non-OECD countries, and in the least developed countries internal accountability was associated with higher levels of student attainment. Although the GERM model of external accountability through league tables is perhaps not globally generalisable to countries where, for example, it could be divisive, the GERM promoted principle of accountability does appear to be globally effective.

In the more developed countries school autonomy was broadly neutral, in middle income countries autonomy appeared to be beneficial, but in the least developed countries autonomy was associated with lower student attainment. The present study does not provide sufficient evidence to explain the underlying causes of these different associations, but it does suggest that school autonomy may not be universally beneficial. Therefore, although the GERM prescription of increasing school autonomy

may be valid in some countries, it does not appear that more school autonomy is likely to be an effective means to raise student attainment in all countries.

Across all levels of national development attendance at fee-paying private schools was associated with high student attainment, but these higher attainment levels appeared to be related to the higher ESCS of students attending private schools and they were to some extent constrained by the economic development levels of the countries concerned. In the high income non-OECD countries school diversity in terms of government-funded independent schools had no clear association with student attainment, and in middle and low income countries attendance at independent schools was associated with lower levels of student attainment. Given that conventional public schools appeared to be associated with higher levels of maths attainment than independent schools across most countries, irrespective of national development levels, it would seem that, contrary to the prescriptions of GERM, public schooling may be the best means of raising educational attainment in some non-OECD countries.

At first sight the principles upon which GERM is based, namely that schools should be accountable, schools should be autonomous, and that there should be a diversity of school provision appear unarguable. However, the present study did not find clear-cut evidence that the measures of external accountability, school autonomy and institutional diversity most associated with GERM were consistently and positively associated with student attainment within and between non-OECD countries at varying levels of national development. Furthermore, despite the fact that the educational systems included in the present research had all undertaken some GERM inspired reforms, the GERM related variables did not explain a substantial amount of the variance in student attainment between or within the countries examined, and it was also clear that other factors such as educational material shortages might be more salient to student attainment in less developed countries. Overall, it appears that aspects of the reforms promoted by GERM may be beneficial in

some countries, but in the absence of any compelling evidence that all the measures associated with GERM are consistently associated with higher levels of student attainment across and within all, or even a majority, of the countries studied in the present research, it must be concluded that GERM may not be globally effective.

References

Adams, C. M., Ford, T. G., Forsyth, P. B., Ware, J. K., Olsen, J. J., Lepine, J. A., … Mwavita, M. (2017). *Next generation accountability: A vision for improvement under ESSA.* Palo Alto, CA: Learning Policy Institute.

Adamson, B., Forestier, K., Morris, P., & Han, C. (2017). PISA, policymaking and political pantomime: education policy referencing between England and Hong Kong. *Comparative Education*, *53*(2), 192–208. https://doi.org/10.1080/03050068.2017.1294666

Afridi, F., Barooah, B., & Somanathan, R. (2018). Improving learning outcomes through information provision: Experimental evidence from Indian villages. *Journal of Development Economics*. https://doi.org/10.1016/j.jdeveco.2018.08.002

Aizawa, T. (2019). Impacts of the community block grant programme on school resources, environment and management in Indonesia. *Education Economics*, *27*(5), 521–545. article. https://doi.org/10.1080/09645292.2019.1641587

Allen, R., & Burgess, S. (2012). *How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England* (article). *IDEAS Working Paper Series*. St. Louis: Federal Reserve Bank of St Louis.

Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, *26*(1), 32–56. article. https://doi.org/10.1080/09243453.2014.927369

Arar, K., & Abu-Romi, A. (2016). School-based management: Arab education system in Israel. *Journal of Educational Administration*, *54*(2), 191–208. article. https://doi.org/10.1108/JEA-09-2014-0118

Arar, K., & Nasra, M. A. (2020). Linking school-based management and school effectiveness: The influence of self-based management, motivation and effectiveness in the Arab education system in Israel. *Educational Management Administration & Leadership*, *48*(1), 186–204. article. https://doi.org/10.1177/1741143218775428

Arias Ciro, J., & Torres García, A. (2018). Economic efficiency of public secondary education expenditure: How different are developed and developing countries? *Revista Desarrollo y Sociedad*, (80), 119–154. article. https://doi.org/10.13043/DYS.80.4

Asadullah, M. N., & Maliki. (2018). Madrasah for girls and private school for boys? The determinants of school type choice in rural and urban Indonesia. *International Journal of Educational Development*, *62*(C), 96–111. article. https://doi.org/10.1016/j.ijedudev.2018.02.006

Baird, J.-A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy. *Educational Research*, *58*(2), 121–138. https://doi.org/10.1080/00131881.2016.1165410

Bakiu, B., Dimitrova, M., & Brava, A. (2016). *How to achieve integrated education in the republic of Macedonia?* Skopje. Retrieved from http://www.epi.org.mk/docs/[D4V]How to Achieve Integrated Education in the Republic of Macedonia.pdf

Bandur, A. (2012). School-based management developments and partnership: Evidence from Indonesia. *International Journal of Educational Development*, *32*(2), 316–328. article. https://doi.org/10.1016/j.ijedudev.2011.05.007

Barrera-Osorio, F. (2006). *The impact of private provision of public education: Empirical evidence from Bogota's concession schools* (Policy Research Working Paper No. 4121). Washington, D.C.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., & Perez-Calle, F. (2011). Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia. *American Economic Journal: Applied Economics*, *3*(2), 167–195. article. https://doi.org/10.1257/app.3.2.167

Bartolo, P. A. (2001). Recent developments in inclusive education in Malta. Mediterranean Journal of Educational Studies. 6(2), 65-91.

Baum, D. R., Cooper, R., & Lusk-Stover, O. (2018). Regulating market entry of low-cost private schools in Sub-Saharan Africa: Towards a theory of private education regulation. *International Journal of Educational Development*, *60*, 100–112. article. https://doi.org/10.1016/j.ijedudev.2017.10.020

Behrman, J. R., Deolalikar, A. B., & Soon, L. Y. (2002). *Promoting effective schooling through education decentralization in Bangladesh, Indonesia, and Philippines*. Manila, Philippines: Asian Development Bank.

Benton, T. (2014). *A Re-Evaluation of the Link between Autonomy, Accountability and Achievement in PISA 2009.* Cambridge.

Bergbauer, A. B., Hanushek, E. A., & Woessmann, L. (2018). *Testing*. Cambridge, Mass.: National Bureau of Economic Research.

Bezzina, C. (1991). Improving the quality of schooling in Malta. International Journal of Educational Management, 5(4), 14-18.

Blasius, J., & Thiessen, V. (2015). Should we trust survey data? Assessing response simplification and data fabrication. *Social Science Research*, *52*(C), 479–493. article. https://doi.org/10.1016/j.ssresearch.2015.03.006

Blazevski, N., Petreski, M., & Ristovska, M. (2017). Breaking Up the Vicious Cycle of Poverty: How Can the School Performance of Children from Low-Income Households in Macedonia Be Improved? *Croatian Economic Survey*, *19*(2), 5–46. article. https://doi.org/10.15179/ces.19.2.1

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does Management Matter in schools? *The Economic Journal: The Journal Of The British Economic Association*, *125*(584), pp647-674. https://doi.org/10.1111/ecoj.12267

Böhlmark, A., Holmlund, H., & Lindahl, M. (2016). Parental choice, neighbourhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform. *Journal of Population Economics*, *29*(4), 1155–1190. article. https://doi.org/10.1007/s00148-016-0595-y

Böhlmark, A., & Lindahl, M. (2015). Independent Schools and Long-run Educational Outcomes: Evidence from Sweden's Large-scale Voucher Reform. *Economica*, *82*(327), 508–551. article. https://doi.org/10.1111/ecca.12130

Bol, T., Witschge, J., Van de Werfhorst, H. G., & Dronkers, J. (2014). Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries. *Social Forces*, *92*(4), 1545–1572. article. https://doi.org/10.1093/sf/sou003

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, *168*, 1–20. article. https://doi.org/10.1016/j.jpubeco.2018.08.007

Brown, L., & Conrad, D. A. (2007). School Leadership in Trinidad and Tobago: The Challenge of Context. *Comparative Education Review*, *51*(2), 181–201. article. https://doi.org/10.1086/512021

Brutti, Z. (2015). *The Political Economy of Public Education: Who Benefits from Decentralization? New evidence from Colombia.* London: London School of Economics.

Burgess, S. (2017). *Education Systems: Assignment, Admissions, Accountability and Autonomy*. *Economics without Borders*. https://doi.org/10.1017/9781316636404.006

Burgess, S., Wilson, D., & Worth, J. (2013). A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, *106*(C), 57–67. article. https://doi.org/10.1016/j.jpubeco.2013.06.005

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity–propensity analysis. *Contemporary Educational Psychology*, *32*(4), 599–629. article. https://doi.org/10.1016/j.cedpsych.2006.09.002

Caldwell, B. J. (2016a). Impact of school autonomy on student achievement: cases from Australia. *The International Journal of Educational Management*, *30*(7), 1171–1187. article. https://doi.org/10.1108/IJEM-10-2015-0144

Caldwell, B. J. (2016b). Professional autonomy, school innovation and student achievement in the 21st century. *Australian Educational Leader*, *38*(4), 9.

Caldwell, B. J. (2016c). Professional autonomy, school innovation and student achievement in the 21st century. *Australian Educational Leader*, *38*(4), 9–13. Retrieved from https://search.informit.com.au/documentSummary;dn=605048286921867;res=IELHSS

Carl, J. (1994). Parental choice as national policy in England and the United States. *Comparative Education Review*, *38*(3), 294–322.

Carnoy, M., & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305–331. article. https://doi.org/10.3102/01623737024004305

Carr-Hill, R., Rolleston, C., Schendel, R., & Waddington, H. (2018). The effectiveness of school-based decision making in improving educational outcomes: a systematic review. *Journal of Development Effectiveness*, *10*(1), 61–94. https://doi.org/10.1080/19439342.2018.1440250

Carr-Hill, R., Rolleston, C., & Schendel, R. (2016). The effects of school-based decision-making on educational outcomes in low-and middle-income contexts: a systematic review. *Campbell Systematic Reviews*, *12*(1), 1–169.

Channa, A. (2016). Popularity of the decentralization reform and its effects on the quality of education. *Prospects*, *46*(1), 131–147. article. https://doi.org/10.1007/s11125-016-9380-7

Cheema, J. R. (2014). A Review of Missing Data Handling Methods in Education Research. *Review of Educational Research*, *84*(4), 487–508. article. https://doi.org/10.3102/0034654314532697

Cheng, A., Trivitt, J. R., & Wolf, P. J. (2016). School Choice and the Branding of Milwaukee Private Schools. *Social Science Quarterly*, *97*(2), 362–375. article. https://doi.org/10.1111/ssqu.12222

Cheng, Y. C., Ko, J., & Lee, T. T. H. (2016). School Autonomy, Leadership and Learning: A Reconceptualisation. *International Journal of Educational Management*, *30*(2), 177. article. https://doi.org/10.1108/IJEM-08-2015-0108

Cheng, Y. C., & Mok, M. M. C. (2007). School-Based Management and Paradigm Shift in Education an Empirical Study. *International Journal of Educational Management*, *21*(6), 517. article. https://doi.org/10.1108/09513540710780046

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, *93*(9), 1045–1057. article. https://doi.org/10.1016/j.jpubeco.2009.06.002

Christ, C., & Dobbins, M. (2016). Increasing school autonomy in Western Europe: a comparative analysis of its causes and forms. *European Societies*, *18*(4), 359–388. https://doi.org/10.1080/14616696.2016.1172716

Clark, D. (2009). The Performance and Competitive Effects of School Autonomy. *Journal of Political Economy*, *117*(4), 745–783. article. https://doi.org/10.1086/605604

Clarke, M. (2012). *What Matters Most for Student Assessment Systems : A Framework Paper*. document, Washington, D.C: World Bank.

Coleman, J. S. (1975). Methods and Results in the IEA Studies of Effects of School on Learning. *Review of Educational Research*, *45*(3), 355–386. article. https://doi.org/10.3102/00346543045003355

Cordes, S. A. (2018). In pursuit of the common good: The spillover effects of charter schools on public school students in New York City. *Education Finance and Policy*, *13*(4), 484–512.

Dalal, D. K., & Hakel, M. D. (2016). Experimental Comparisons of Methods for Reducing Deliberate Distortions to Self-Report Measures of Sensitive Constructs. *Organizational Research Methods*, *19*(3), 475–505. article. https://doi.org/10.1177/1094428116639131

Davies, S., & Guppy, N. (1997). Globalization and educational reforms in Anglo-American democracies. *Comparative Education Review*, *41*(4), 435–459.

De Lisle, J. (2012). Explaining Whole System Reform in Small States: The Case of Trinidad and Tobago Secondary Education Modernization Program. *Current Issues in Comparative Education*, *15*(1), 64–82.

De Lisle, J. (2015). Installing a System of Performance Standards for National Assessments in the Republic of Trinidad and Tobago: Issues and Challenges. *Applied Measurement in Education*, *28*(4), 308–329. article. https://doi.org/10.1080/08957347.2015.1062765

De Lisle, J., Annisette, S., & Bowrin-Williams, C. (2019). Leading high-poverty primary schools in Trinidad and Tobago – what do successful principals do? *Educational Management Administration & Leadership*, 1–19. article. https://doi.org/10.1177/1741143219827304

De Lisle, J., Keller, C., Jules, V., & Smith, P. (2009). When choosing might mean losing: A mixed method study of secondary school choice in the Republic of Trinidad and Tobago. *Caribbean Curriculum*, *16*(1), 131–176.

Deangelis, C. A., & Shakeel, M. D. (2017). Does Private Schooling Improve International Test Scores? An Instrumental Variables Fixed Effects Analysis of the Impact of Private Schooling on PISA Scores. *SSRN Electronic Journal*. article. https://doi.org/10.2139/ssrn.2903523

Deangelis, C., & Erickson, H. (2018). What leads to successful school choice programs: A review of the theories and evidence. *Cato Journal*, *38*(1), 247–263.

Dee, T. S., & Jacob, B. (2011). The impact of no Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418–446. article. https://doi.org/10.1002/pam.20586

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The Effects of NCLB on School Resources and Practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252–279. https://doi.org/10.3102/0162373712467080

Deming, D. J., & Figlio, D. (2016). Accountability in US Education: Applying Lessons from K–12 Experience to Higher Education. *Journal of Economic Perspectives*, *30*(3), 33–56. https://doi.org/10.1257/jep.30.3.33

Department for Education. (2014). *Do academies make use of their autonomy?* London: HMSO.

Double, K. S., Mcgrane, J. A., Stiff, J. C., & Hopfenbeck, T. N. (2019). The importance of early phonics improvements for predicting later reading comprehension. *British Educational Research Journal*, *45*(6), 1220–1234. https://doi.org/10.1002/berj.3559

Egalite, A. J. (2013). Measuring Competitive Effects From School Voucher Programs: A Systematic Review. *Journal of School Choice*, *7*(4), 443–464. https://doi.org/10.1080/15582159.2013.837759

Ehren, M., Altrichter, H., McNamara, G., & O'Hara, J. (2011). Impact of school inspections on improvement of schools—describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, *25*(1), 3–43. https://doi.org/10.1007/s11092-012-9156-4

Ehren, M., Eddy-Spicer, D., Bangpan, M., & Reid, A. (2017). School inspections in low- and middle-income countries: Explaining impact and mechanisms of impact. *Compare: A Journal of Comparative and International Education*, *47*(4), 468–482. https://doi.org/10.1080/03057925.2016.1239188

Ehren, M., Gustafsson, J., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, *51*(3), 375–400. https://doi.org/10.1080/03050068.2015.1045769

Elacqua, G. (2012). The impact of school choice and public policy on segregation: Evidence from Chile. *International Journal of Educational Development*. article, Elsevier Ltd. https://doi.org/10.1016/j.ijedudev.2011.08.003

Elliott, J., Stankov, L., Lee, J., & Beckmann, J. F. (2019). What did PISA and TIMSS ever do for us?: the potential of large scale datasets for understanding and improving educational practice. *Comparative Education*, *55*(1), 133–155.

Epple, D., & Romano, R. E. (1998). Competition between Private and Public Schools, Vouchers, and Peer-Group Effects. *American Economic Review*, *88*(1), 33–62. https://doi.org/10.2307/116817

Epple, D., Romano, R., & Zimmer, R. (2015). *Charter Schools: A Survey of Research on Their Characteristics and Effectiveness* (article No. 212526). *NBER Working Paper Series*. Cambridge: National Bureau of Economic Research, Inc. https://doi.org/10.3386/w21256

Eyles, A., Hupkau, C., & Machin, S. (2016). School reforms and pupil performance. *Labour Economics*, *41*, 9.

Eyles, A., & Machin, S. (2019). The Introduction of Academy Schools to England's Education. *Journal of the European Economic Association*, *17*(4), 1107–1146. https://doi.org/10.1093/jeea/jvy021

Fahmi, M., Maulana, A., & Yusuf, A. A. (2011). *Teacher certification in Indonesia: A confusion of means and ends.* (CEDS No. 2011/3).

Falch, T., & Fischer, J. A. . (2012). Public sector decentralization and school performance: International evidence. *Economics Letters*, *114*(3), 276–279. https://doi.org/10.1016/j.econlet.2011.10.019

Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education Vol. 3* (pp. 383–421). Elsevier Inc.

Figlio, D. N., & Ladd, H. F. (2015). School Accountability and Student Achievement. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (Second ed, pp. 194–210). London: Routledge.

Figlio, D. N., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, *90*(1), 239–255. https://doi.org/10.1016/j.jpubeco.2005.08.005

Foreman, L. M. (2017). Educational attainment effects of public and private school choice. *Journal of School Choice*, *11*(4), 642–654. article. https://doi.org/10.1080/15582159.2017.1395619

Friedman, M., & Friedman, R. D. (1982). *Capitalism and freedom*. Chicago: University of Chicago Press.

Fryer, R. G., & Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210–240. https://doi.org/10.1257/app.2.2.210

Fuchs, T., & Wößmann, L. (2008). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, *32*(2–3), 433–464. https://doi.org/10.1007/s00181-006-0087-0

Fuller, B. (1986). *Raising School Quality in Developing Countries: What Investments Boost Learning?* (World Bank Discussion Papers No. 2). Washington, D.C.: World Bank.

Fuller, Bruce. (1987). What School Factors Raise Achievement in the Third World? *Review of Educational Research*, *57*(3), 255–292. https://doi.org/10.3102/00346543057003255

Ganimian, A. J., & Murnane, R. J. (2016). Improving Education in Developing Countries: Lessons From Rigorous Impact Evaluations. *Review of Educational Research*, *86*(3), 719–755. https://doi.org/10.3102/0034654315627499

Garson, G. D. (2019). *Multilevel Modeling*. New York: SAGE Publications.

Geijsel, F., Krüger, M., & Sleegers, P. (2010). Data feedback for school improvement: The role of researchers and school leaders. *The Australian Educational Researcher*, *37*(2), 59–75. https://doi.org/10.1007/BF03216922

George, J., Mohammed, J., & Quamina-Aiyejina, L. (2003). Teacher Identity in an Era of Educational Reform: The case of Trinidad and Tobago. *Compare: A Journal of Comparative and International Education*, *33*(2), 191–206. https://doi.org/10.1080/0305792032000070048

Gertler, P. J., Patrinos, H. A., & Rubio-Codina, M. (2012). Empowering parents to improve education: Evidence from rural Mexico. *Journal of Development Economics*, *99*(1), 68–79. https://doi.org/10.1016/j.jdeveco.2011.09.004

Giannelli, G. C., & Rapallini, C. (2016). Immigrant student performance in Math: Does it matter where you come from? *Economics of Education Review*, *52*(C), 291–304. https://doi.org/10.1016/j.econedurev.2016.03.006

Gilmore, A. (2005). *The impact of PIRLS (2001) and TIMSS (2003) in low-and middle-income countries.* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Gintis, H., Cox, B., Green, A., & Hickox, M. (1991). Politics , Markets and America ' s Schools. *British Journal of Sociology of Education*, *12*(3), 381–396.

Gleason, P. M. (2017). What's the secret ingredient? Searching for policies and practices that make charter schools successful. *Journal of School Choice: School Choice: Separating Fact from Fiction*, *11*(4), 559–584. https://doi.org/10.1080/15582159.2017.1395620

Glewwe, P., & Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education Vol. 5* (pp. 653–743). Elsevier Inc.

Glewwe, P. W., Hanushek, E., Humpage, S., & Ravina, R. (2016). *School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010* (article). *IDEAS Working Paper Series from RePEc*. St. Louis: Federal Reserve Bank of St Louis.

Glewwe, Paul W, Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). *School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010*. Cambridge, Mass: National Bureau of Economic Research.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed). London: Edward Arnold.

Goldstein, H., & Leckie, G. (2016). Trends in examination performance and exposure to standardised tests in England and Wales. *British Educational Research Journal*, *42*(3), 367–375. https://doi.org/10.1002/berj.3220

Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, *60*(1), 549–576. article. https://doi.org/10.1146/annurev.psych.58.110405.085530

Greany, T. (2018). Innovation is possible, it's just not easy: Improvement, innovation and legitimacy in England's autonomous and accountable school system. *Educational Management Administration & Leadership*, *46*(1), pp65-85. https://doi.org/10.1177/1741143216659297

Griffith, D. (2019). *Rising Tide: Charter School Market Share and Student Achievement.* Washington, D.C. Retrieved from https://fordhaminstitute.org/national/research/rising-tide-charter-market-share.

Grimmett, P. P. (2018). Canada's Trojan Horse: Labor Mobility Legislation: concealing deregulation and an attack on teacher professionalism and the common good. In N. Hobbel & B. L. Bales (Eds.), *Navigating the Common Good in Teacher Education Policy: Critical and International Perspectives.* (pp. 213–224). Abingdon: Routledge.

Grozdanovska, L. (2007). Macedonia: Go to the Head of the Class. *Transitions Online*, (10/30). article.

Gunnarsson, V., Orazem, P. F., Sánchez, M. A., & Verdisco, A. (2009). Does Local School Control Raise Student Outcomes? Evidence on the Roles of School Autonomy and Parental Participation. *Economic Development and Cultural Change*, *58*(1), 25–52. https://doi.org/10.1086/605209

Gustafsson, J.-E., Nilsen, T., & Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*, *57*, 16–30. https://doi.org/10.1016/j.stueduc.2016.09.004

Hahn, Y., Wang, L. C., & Yang, H.-S. S. (2018). Does greater school autonomy make a difference? Evidence from a randomized natural experiment in South Korea. *Journal of Public Economics*, *161*(January 2016), 15–30. https://doi.org/10.1016/j.jpubeco.2018.03.004

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297–327. https://doi.org/10.1002/pam.20091

Hanushek, Eric A, Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, *104*(104), 212–232. https://doi.org/10.1016/j.jdeveco.2012.08.002

Harden, K. P., Tucker-Drob, E. M., & Tackett, J. L. (2013). The Texas Twin Project. *Twin Research and Human Genetics*, *16*(1), 385–390. article. https://doi.org/10.1017/thg.2012.97

Harjanto, I., Lie, A., Wihardini, D., Pryor, L., & Wilson, M. (2018). Community-based teacher professional development in remote areas in Indonesia. *Journal of Education for Teaching*, *44*(2), 212–231.

Harris, A., Jones, M., Adams, D., Sumintono, B., & Ismail, N. (2017). *Leading Turnaround and Improvement in Low Performing Schools in Malaysia and Indonesia.* (THF Working paper, Working Papers Series No. No.2/2017). Singapore: The Head Foundation.

Haywood, M.-M. (2019). Standardized testing in the Caribbean: exploring the relationship between high stakes and low stakes testing in secondary schools in Trinidad and Tobago. *International Journal of Education and Research*, *7*(7).

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and Longitudinal Modeling with IBM SPSS*. Hove: Sussex: Routledge.

Henríquez, F., Lara, B., Mizala, A., & Repetto, A. (2012). Effective schools do exist: low-income children's academic performance in Chile. *Applied Economics Letters*, *19*(5), 445–451. https://doi.org/10.1080/13504851.2011.583208

Holloway. J., Sørensen. T. B., & Verger, A. (2017). Global perspectives on high-stakes teacher accountability policies: An introduction. *Education Policy Analysis Archives*, *25*(84), 1–18. https://doi.org/10.14507/epaa.25.3325

Holzhacker, R., Wittek, R., & Woltjer, J. (2016). *Decentralization and Governance in Indonesia*. Cham.

Hooge, E., Burns, T., & Wilkoszewski, H. (2012). *Looking Beyond the Numbers: Stakeholders and Multiple School Accountability*. Paris. https://doi.org/http://dx.doi.org/10.1787/5k91dl7ct6q6-

Hopfenbeck, T. N. (2016). The power of PISA - limitations and possibilities for educational research. *Assessment in Education: Principles, Policy & Practice*, *23*(4), 423–426. https://doi.org/10.1080/0969594X.2016.1247518

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J.-A. (2018). Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*, *62*(3), 333–353. https://doi.org/10.1080/00313831.2016.1258726

Hopmann, S. T. (2008). No child, no school, no state left behind: schooling in the age of accountability. *Journal of Curriculum Studies*, *40*(4), 417–456. https://doi.org/10.1080/00220270801989818

Hox, J. J. (2010). *Multilevel Analysis Techniques and Applications* (2nd ed.). Hove, Sussex: Routledge.

Hsieh, C.-T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, *90*(8), 1477–1503. https://doi.org/10.1016/j.jpubeco.2005.11.002

Hursh, D. (2018). The rise of authoritarian neoliberalism. In J. Niesche, R. Eacott, S. Wilkinson (Ed.), *Challenges for Public Education: Reconceptualising Educational Leadership, Policy and Social Justice as Resources for Hope* (pp. 145–158). Abingdon: Routledge.

Husén, T., Noonan, R. B., & Saha, L. J. (1978). *Teacher training and student achievement in less developed countries*. Washington, D.C: World Bank.

Hutabarat, W. (2015). Implementation of School Based Management in Indonesia (Affecting Work Motivation, Job-Satisfaction, and Teacher Job-Performance). *International Journal of Sciences: Basic and Applied Research*, *22*(02), 428–440.

IBM Corporation. (2019). IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corporation.

Jabbar, H., Fong, C. J., Germain, E., Li, D., Sanchez, J., Sun, W. L., & Devall, M. (2019). The Competitive Effects of School Choice on Student Achievement: A Systematic Review. *Educational Policy*. https://doi.org/http://10.1177/0895904819874756

Jasmina, T. (2017). Public Spending and Learning Outcomes of Basic Education at the District Level in Indonesia. *Economics and Finance in Indonesia*, *62*(3), 180–190. https://doi.org/10.7454/efi.v62i3.556

Jasmina, T., & Oda, H. (2018). Does Local Government Capacity Hamper Improvement of Basic Education? An Analysis at the District Level in Indonesia. *Journal of Economics and Management*, *12*(S1), 181–196.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, *7*(1), 63–88.

Jeong, D. W., Lee, H. J., & Cho, S. K. (2017). Education decentralization, school resources, and student outcomes in Korea. *International Journal of Educational Development*, *53*, 12–27. https://doi.org/10.1016/j.ijedudev.2016.12.003

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, *61*(C), 51–58. https://doi.org/10.1016/j.econedurev.2017.09.007

Joseph, S. (2010). Lessons learnt from six years of education reform. *The Caribbean Journal of Teacher Education and Pedagogy*, *1*(119–36).

Joseph, S. (2020). The Politics of Education Reform: Practices That Militate Against Change in a Developing Country. *IInternational Journal of Educational Reform*. https://doi.org/https://doi.org/10.1177/1056787920917389

Kelly, P., & Kotthoff, H.-G. (2017). PISA, National and Regional Education Policies and their Effect on Mathematics Teaching in England and Germany. *Research in Comparative and International Education*, *12*(3), 318–332. article. https://doi.org/10.1177/1745499917724286

Kim, H., & Mobrand, E. (2019). Stealth marketisation: how international school policy is quietly challenging education systems in Asia. *Globalisation, Societies and Education*, *17*(3), 310–323. https://doi.org/10.1080/14767724.2019.1571405

King, E., & Guerra, S. (2005). Education reforms in East Asia: Policy, process and impact. In *East Asia decentralizes: Making local government work* (pp. 179–208). Washington, D.C.: World Bank.

Klees, S. J. (2008). A quarter century of neoliberal thinking in education: Misleading analyses and failed policies. *Globalisation, Societies and Education*, *6*(4), 311–348.

Klein, E. D. (2017). Autonomy and Accountability in Schools Serving Disadvantaged Communities. *Journal of Educational Administration*, *55*(5), 589. article. https://doi.org/10.1108/JEA-06-2016-0065

Ko, J., Cheng, Y. C., & Lee, T. T. H. (2016). The Development of School Autonomy and Accountability in Hong Kong: Multiple Changes in Governance, Work, Curriculum, and Learning. *International Journal of Educational Management*, *30*(7), 1207. article. https://doi.org/10.1108/IJEM-10-2015-0145

Kollmayer, M., Schober, B., & Spiel, C. (2018). Gender stereotypes in education: Development, consequences, and interventions. *European Journal of Developmental Psychology*, *15*(4), 361–377. article. https://doi.org/10.1080/17405629.2016.1193483

Komatsu, H., & Rappleye, J. (2017). A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education*, *53*(2), 166–191. article. https://doi.org/10.1080/03050068.2017.1300008

Koo, H., & Park, C. (2018). Foundation of leadership in Asia: Leader characteristics and leadership styles review and research agenda. *Asia Pacific Journal of Management*, *35*(3), 697–718. https://doi.org/10.1007/s10490-017-9548-6

Kreft, I., & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: SAGE Publications.

Ladd, H. F. (2017). No child left behind: A deeply flawed federal policy. *Journal of Policy Analysis and Management*, *36*(2), 461–469. article. https://doi.org/10.1002/pam.21978

Lastra-Anadón, C., & Mukherjee, S. (2019). *Cross-country evidence on the impact of decentralisation and school autonomy on educational performance* (article). *OECD Working Papers on Fiscal Federalism*. Paris: Organisation for Economic Cooperation and Development (OECD).

Layne, A., Jules, V., Kutnick, P., & Layne, C. (2008). Academic achievement, pupil participation and integration of group work skills in secondary school classrooms in Trinidad and Barbados. *International Journal of Educational Development*, *28*(2), 176–194. https://doi.org/10.1016/j.ijedudev.2007.01.001

Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8.' *British Educational Research Journal*, *43*(2), 193–212. article. https://doi.org/10.1002/berj.3264

Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership & Management: The Impact of School Leadership on Student Outcomes*, *28*(1), 27–42. article. https://doi.org/10.1080/13632430701800060

Lemke, M., & Zhu, L. (2018). Successful futures? New economy business logics, child rights, and Welsh educational reform. *Policy Futures in Education*, *16*(3), 251–276. https://doi.org/10.1177/1478210317751269

Lenkeit, J., Chan, J., Hopfenbeck, T. N., & Baird, J.-A. (2015). A review of the representation of PIRLS related research in scientific journals. *Educational Research Review*, *16*, 102–115. https://doi.org/10.1016/j.edurev.2015.10.002

Levin, B., & Fullan, M. (2008). Learning About System Renewal. *Educational Management, Administration and Leadership*, *36*(2), 289–303.

Liberman, J., & Clarke, M. (2012). *Review of World Bank Support for Student Assessment Activities in Client Countries, 1998–2009*. Washington, D.C.

Lindberg, S. M., Hyde, J. S., Linn, M. C., & Petersen, J. L. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin*, *136*(6), 1123. https://doi.org/10.1037/a0021276

Lingard, B., & Sellar, S. (2016). The changing organizational and global significance of the OECD's education work. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The Handbook of Global Education Policy* (pp. 357–373). Chicago: John Wiley & Sons, Ltd. https://doi.org/10.1080/02680939.2019.1588523

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability Systems: Implications of Requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, *31*(6), 3–16. https://doi.org/10.3102/0013189X031006003

Liu, S., Rovine, M. J., & Molenaar, P. C. M. (2012). Selecting a linear mixed model for longitudinal data: repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods*, *17*(1), 15. article. https://doi.org/10.1037/a0026971

Lockheed, M. E., & Komenan, A. (1989). Teaching quality and student achievement in Africa: The case of Nigeria and Swaziland. *Teaching and Teacher Education*, *5*(2), 93–113. https://doi.org/10.1016/0742-051X(89)90009-7

Lockheed, M., Vail, S., & Fuller, B. (1986). How Textbooks Affect Achievement in Developing Countries: Evidence from Thailand. *Educational Evaluation and Policy Analysis*, *8*(4), 379. https://doi.org/10.3102/01623737008004379

Louisy, P. (2004). Whose context for what quality? Informing education strategies for the Caribbean. *Compare: A Journal of Comparative and International Education*, *34*(3), 285–292. https://doi.org/10.1080/0305792042000257121

Lubienski, C., & Lubienski, S. T. (2014). *The public school advantage : why public schools outperform private schools*. Chicago: UCP.

Lyon, A. (2013). Decentralisation and the Provision of Primary and Secondary Education in the Former Yugoslav Republic of Macedonia. *International Journal on Minority and Group Rights*, *20*(4), 491–516. https://doi.org/10.1163/15718115-02004001

Machin, S., & Vernoit, J. (2011). *Changing School Autonomy: Academy Schools and their Introduction to England's Education* (document No. 95/1). London: Centre for the Economics of Education, LSE.

Malin, J. R., Hardy, I., & Lubienski, C. (2019). Educational neoliberalization: the mediatization of ethical assertions in the voucher debate. *Discourse: Studies in the Cultural Politics of Education*, *40*(2), 217–233. article. https://doi.org/10.1080/01596306.2019.1569880

Marks, G. N. (2015). School Sector Differences in Student Achievement in Australian Primary and Secondary Schools: A Longitudinal Analysis. *Journal of School Choice*, *9*(2), 219–238. https://doi.org/10.1080/15582159.2015.1028827

Mattei, P. (2012a). Market accountability in schools: policy reforms in England, Germany, France and Italy. *Oxford Review of Education*, *38*(3), 247–266. https://doi.org/10.1080/03054985.2012.689694

Mattei, P. (2012b). Raising educational standards: national testing of pupils in the United Kingdom, 1988-2009. *Policy Studies*, *33*(3), 231–247. https://doi.org/10.1080/01442872.2012.658260

Maulana, R. A., & Yudhistira, M. H. (2019). Do schools compete with each other? Empirical evidence from Jakarta, Indonesia. *Applied Economics Letters*. 26(12), 983-986. https://doi.org/10.1080/13504851.2018.1527437

Mbiti, I. M. (2016). The Need for Accountability in Education in Developing Countries. *Journal of Economic Perspectives*, *30*(3), 109–132. https://doi.org/10.1257/jep.30.3.109

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. *The Quarterly Journal of Economics*, *134*(3), 1627–1673. https://doi.org/10.1093/qje/qjz010

Mcnicoll, G. (2011). Achievers and laggards in demographic transition: a comparison of Indonesia and Nigeria. *Population and Development Review*, *37*(Suppl 1), 191. https://doi.org/10.1111/j.1728-4457.2011.00384.x

Meyer, H.-D. (2014). Imagining PISA's Policy Futures: A Postscript and Some Extensions to the Open Letter to Andreas Schleicher. *Policy Futures in Education*, *12*(7), 883–892. https://doi.org/10.2304/pfie.2014.12.7.883

Meyer, H. D., & Zahedi, K. (2014). An Open Letter: To Andreas Schleicher, OECD, Paris. *Mitteilungen Der Gesellschaft Für Didaktik Der Mathematik*, *40*, 31–36.

Ministry of Education. (2011). *Administrative Report*. Port of Spain:Government of Trinidad & Tobago.

Mizala, A., & Romaguera, P. (2000). School Performance and Choice: The Chilean Experience. *The Journal of Human Resources*, *35*(2), 392. https://doi.org/10.2307/146331

Mizala, A., & Torche, F. (2017). Means-Tested School Vouchers and Educational Achievement: Evidence from Chile's Universal Voucher System. *The ANNALS of the American Academy of Political and Social Science*, *674*(1), 163–183. https://doi.org/10.1177/0002716217732033

Mizala, A., & Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *The Journal of Development Economics*, *103*, 313–335.

Morgan, C., Petrosino, A., & Fronius, T. (2015). The impact of school vouchers in developing countries: A systematic review. *International Journal of Educational Research*, *72*, 70–79. https://doi.org/10.1016/j.ijer.2015.04.010

Moschkovich, J. N. (2012). *Mathematics, the Common Core, and language: Recommendations for mathematics instruction for ELs aligned with the Common Core.* Stanford: Understanding Language.

Moutsios, S. (2010). Power, politics and transnational policy-making in education. *Globalisation, Societies and Education*, *8*(1), 121–141.

Murchan, D., Shiel, G., & Mickovska, G. (2012). An education system in transition: assessment and examinations in the Republic of Macedonia. *Assessment in Education: Principles, Policy & Practice*, *19*(4), 487–502. https://doi.org/10.1080/0969594X.2012.710195

Nababan, H. S., Suyradi, & Hanafi, I. (2019). Evaluation of Policy Implementation of Basic Education Minimum Service Standards in Bekasi Municipality, Indonesia. *Advances in Social Science, Education and Humanities Research*, *337*, 73–82.

Naceva, B., & Mickovska, G. (2007). The impact of PIRLS in the Republic of Macedonia. In K. Schwippert (Ed.), *Progress in reading literacy. The impact of PIRLS 2001 in 13 countries* (pp. 165–180). Munster: Waxman Verlag GmbH.

Newhouse, D., & Beegle, K. (2006). The effect of school type on academic achievement. *Journal of Human Resources*, *41*(3), 529–29. article.

Nir, A., Ben-David, A., Bogler, R., Inbar, D., & Zohar, A. (2016). School Autonomy and 21st Century Skills in the Israeli Educational System: Discrepancies between the Declarative and Operational Levels. *International Journal of Educational Management*, *30*(7), 1231. https://doi.org/10.1108/IJEM-11-2015-0149

Nir, A. E. (2002). School-based management and its effect on teacher commitment. *International Journal of Leadership in Education*, *5*(4), 323–341. https://doi.org/10.1080/13603120210134616

Nir, E., & Hameiri, L. (2014). School principals' leadership style and school outcomes: The mediating effect of powerbase utilization. *Journal of Educational Administration*, *52*(2), 210–227. https://doi.org/10.1108/JEA-01-2013-0007

Null, C., Cosentino, C., Sridharan, S., & Meyer, L. (2017). *Policies and Programs to Improve Secondary Education in Developing Countries.* Princeton, NJ.

Nuraeni, S. F. C. (2020). Single-sex education: does it make a difference to students' academic achievement? *International Journal of Indonesian Education and Teaching (IJIET)*, *4*(2), 296–305.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. Furnham (Ed.), *Designing coherent education policy: Improving the system* (pp. 250–312). San Franscisco: Jossey-Bass.

OECD/Asian Development Bank. (2015). *Education in Indonesia: Rising to the Challenge*. Paris: OECD Publishing. https://doi.org/Http://dx.doi.org/10.1787/9789264230750-en

OECD. (2009). *PISA Data Analysis Manual SAS* (2nd ed.). Paris: Organisation for Economic Co-operation and Development.

OECD. (2016a). *PISA 2015 Results: Policies and Practices for Successful Schools (Volume II)*. Paris: Organisation for Economic Cooperation and Development (OECD). https://doi.org/https://doi.org/10.1787/9789264267510-en

OECD. (2016b). *PISA 2015 Results (Volume 2): Policies and Practices for Successful Schools*. Paris: Organisation for Economic Cooperation and Development (OECD). Retrieved from https://www.oecd.org/pisa/publications/

OECD. (2016c). *PISA for Development*. Paris: Organisation for Economic Cooperation and Development (OECD).

OECD. (2017a). *PISA 2015 Results: Excellence and Equity in Education (Volume I)*. Paris: Organisation for Economic Cooperation and Development (OECD). https://doi.org/https://doi.org/10.1787/9789264266490-en.

OECD. (2017b). *PISA 2015 Technical Report*. Paris: Organisation for Economic Cooperation and Development (OECD). Retrieved from https://www.oecd.org/pisa/data/2015-technical-report/

OECD. (2019a). *Education at a Glance 2019: OECD Indicators*. Paris: Organisation for Economic Cooperation and Development (OECD). https://doi.org/https://doi.org/10.1787/f8d7880d-en.

OECD. (2019b). *PISA for Development Technical Report*. Washington, D.C: Organisation for Economic Cooperation and Development (OECD).

Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: a meta-analysis. *Psychological Bulletin*, *140*(4), 1042–1072. article. https://doi.org/10.1037/a0035740

Pantic, N. (2012). Teacher Education Reforms between Higher Education and General Education Transformations in South-Eastern Europe: Reviewing the Evidence and Scoping the Issues. *CEPS Journal*, *2*(4), 71.

Pantić, N., Closs, A., & Ivošević, V. (2011). *Teachers for the future teacher development for inclusive education in the Western Balkans.* book, Luxembourg: Publications Office. https://doi.org/10.2816/16766

Parker, P. D., Marsh, H. W., Jerrim, J. P., Guo, J., & Dicke, T. (2018). Inequity and Excellence in Academic Performance: Evidence From 27 Countries. *American Educational Research Journal*, *55*(4), 836–858. https://doi.org/10.3102/0002831218760213

Patrinos, H., Arcia, G., & Macdonald, K. (2015). School autonomy and accountability in Thailand: Does the gap between policy intent and implementation matter? *Prospects*, *45*(4), 429–445. https://doi.org/10.1007/s11125-015-9368-8

Raihani, R., & Sumintono, B. (2010). Teacher Education in Indonesia: development and challenges. In K. G. Karras & C. C. Wohlter (Eds.), *International Handbook on Teacher Education Worldwide Vol 2* (pp. 181–197). Athens: Atropos.

Rizova, T. P. (2011). A case of contested sovereignty: explaining ethnic conflict in the former Yugoslav Republic of Macedonia, 1991-2002. *Review of European Studies*, *3*(2), 73. https://doi.org/10.5539/res.v3n2p2

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2011). Feeling the Florida Heat How Low-Performing Schools Respond to Voucher and Accountability Pressure. *American Economic Journal: Economic Policy*, *5*(2), 251–281. https://doi.org/10.1257/pol.5.2.251

Rust, K. (2014). Sampling, weighting, and variance estimation in international large-scale assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-scale Assessment: Background, technical issues, and methods of data analysis* (pp. 117–153). Boca Raton: CRC Press.

Rutkowski, Leslie, Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

Sadler, M. (1900). Address given at the Guildford Educational Conference on Saturday 20 October 1900. In J. H. Higginson (Ed.), *Selections from Michael Sadler: Studies in world citizenship* (pp. 48–51). Liverpool: Dejall & Meyorre.

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The Handbook of Global Education Policy* (pp. 128–144). Chichester, West Sussex: Wiley Blackwell.

Sanford, J. E., & Sandhu, A. (2002). *Developing Countries: Definitions, Concepts and Comparisons* (document). Washington, D.C: Library of Congress. Congressional Research Service. Retrieved from https://digital.library.unt.edu/ark:/67531/metadc822497/

Sarouphim, K., & Chartouny, M. (2017). Mathematics education in Lebanon: gender differences in attitudes and achievement. *Educational Studies in Mathematics*, *94*(1), 55–68. https://doi.org/10.1007/s10649-016-9712-9

SAS Institute Inc. (2020). SAS Studio. Cary, NC.

Schleicher, A. (2014, May). Response to Dr Heinz-Dieter Meyer and others. *The Guardian*.

Schleicher, Andreas. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change*, *8*(4), 349–357. https://doi.org/10.1007/s10833-007-9042-x

Sellar, S., & Lingard, B. (2014). The OECD and the expansion of PISA: New global modes of governance in education. *British Educational Research Journal*, *40*(6), 917–936. https://doi.org/10.1002/berj.3120

Shahjahan, R. A. (2016). International organizations (IOs), epistemic tools of influence, and the colonial geopolitics of knowledge production in higher education policy. *Journal of Education Policy*, *31*(6), 694–710.

Shakeel, M. D., & Deangelis, C. A. (2017). Who is More Free? A Comparison of the Decision-Making of Private and Public School Principals. *Journal of School Choice*, *11*(3), 442–457. https://doi.org/10.1080/15582159.2017.1345235

Shakeel, M. D., & Deangelis, C. A. (2018). Can private schools improve school climate? Evidence from a nationally representative sample. *Journal of School Choice: The Best of the 2018 International School Choice and Reform Conference*, *12*(3), 426–445. https://doi.org/10.1080/15582159.2018.1490383

Smith, C., Parr, N., & Muhidin, S. (2019). Mapping schools' NAPLAN results: a spatial inequality of school outcomes in Australia. *Geographical Research*, *57*(2), 133–150. https://doi.org/10.1111/1745-5871.12317

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis : an introduction to basic and advanced multilevel modeling*. London: SAGE.

Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis : an introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Stacey, K. (2011). The PISA View of Mathematical Literacy in Indonesia. *IndoMS-Journal on Mathematics Education*, *2*(2), 95–126. article. https://doi.org/10.22342/jme.2.2.746.95-126

Stapleton, L. M. (2014). Incorporating sampling weights into single-and multilevel analyses. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-scale Assessment: Background, technical issues, and methods of data analysis* (pp. 363–388). Boca Raton: CRC Press.

Stern, J. M. ., & Smith, T. M. (2016). Private secondary schools in Indonesia: What is driving the demand? *International Journal of Educational Development*, *46*, 1–11. https://doi.org/10.1016/j.ijedudev.2015.11.002

Sulistiani, L., & Zulkardi. (2019). Developing PISA-Like Mathematics Problems Using Indonesian Food Context. *Journal of Physics: Conference Series*, *1315*, 012078. https://doi.org/10.1088/1742-6596/1315/1/012078

Sumintono, B. (2010). School-Based Management Policy and Its Practices at District Level in the Post New Order Indonesia. *Journal of Indonesian Social Sciences and Humanities*, *2*(2014), 41–67.

Sumintono, B., Subekti, N., Mislan, N., Said, H., & Mohd Tahir, L. (2014). In Search of the Excellent School: A Case Study of an Indonesian International Standard School in Improving Its Capacity Building. *The Asia-Pacific Education Researcher*, *23*(3), 565–576. https://doi.org/10.1007/s40299-013-0130-9

Sung, Y.-K., & Kang, M. O. (2011). The cultural politics of national testing and test result release policy in South Korea: a critical discourse analysis. *Asia Pacific Journal of Education*, *32*(1), 53–73. https://doi.org/10.1080/02188791.2011.621691

Suryadarma, D. (2012). How corruption diminishes the effectiveness of public spending on education in Indonesia. *Bulletin of Indonesian Economic Studies*, *48*(1), 85–100. https://doi.org/10.1080/00074918.2012.654485

Takeda, N., Demas, A., & Shibuya, K. (2014). *World Bank Portfolio Review of School Autonomy and Accountability: Operations and Analytical Work, FY2003-2013.* Washington, DC: World Bank.

Tell, S. (2015). Charter Schools Appearance versus Essence. In A. T. Costigan & L. Grey (Eds.), *Demythologizing Educational Reforms: responses to the political and corporate takeover of education* (pp. 133–152). Oxford: Routledge.

Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R., & Nyamkhuu, T. (2015). *Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region*. Melbourne: Centre for Global Education.

Toi, A. (2010). An empirical study of the effects of decentralization in Indonesian junior secondary education. *Educational Research for Policy and Practice*, *9*(2), 107–125. https://doi.org/10.1007/s10671-010-9081-6

Tooley, J., & Longfield, D. (2016). Affordability of private schools: exploration of a conundrum and towards a definition of "low-cost." *Oxford Review of Education*, *42*(4), 444–459. https://doi.org/10.1080/03054985.2016.1197830

Trivitt, J. R., & Wolf, P. J. (2011). School Choice and the Branding of Catholic Schools. *Education Finance and Policy*, *6*(2), 202–245. article. https://doi.org/10.1162/EDFP_a_00032

Tucker-Drob, E. M., & Briley, D. A. (2012). Socioeconomic status modifies interest-knowledge associations among adolescents. *Personality and Individual Differences*, *53*(1), 9–15. https://doi.org/10.1016/j.paid.2012.02.004

Verger, A., Novelli, M., & Altinyelken, H. K. (2017). Global education policy and international development: A revisited introduction. In A. Verger, M. Novelli, & H. K. Altinyelken (Eds.), *Global Education Policy and International Development: New Agendas, Issues and Policies* (2nd ed., pp. 1–34). Lomdon: Bloomsbury Academic.

Verger, Antoni, Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms. *Educational Review*, *71*(1), 5–30. article. https://doi.org/10.1080/00131911.2019.1522045

Vernez, G., Karam, R., & Marshall, J. H. (2012). *Implementation of School Based Management in Indonesia*. Santa Monica, CA.

Voldby Beuchert, L., & Brink Nandrup, A. (2014). *The Danish National Tests – A Practical Guide*. Aarhus.

von Davier, M., & Sinharay, S. (2014). Analytics in international large-scale assessments: Item response theory and population models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-scale Assessment: Background, technical issues, and methods of data analysis* (pp. 155–174). Boca Raton: CRC Press.

Vrieze, S. I. (2012). Model Selection and Psychological Theory: A Discussion of the Differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).(Report)(Author abstract). *Psychological Methods*, *17*(2), 228. https://doi.org/10.1037/a0027127

Wagner, D. A., Lockheed, M., Mullis, I., Martin, M. O., Kanjee, A., Gove, A., & Dowd, A. J. (2012). The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, *42*(3), 509–545. https://doi.org/10.1080/03057925.2012.670480

Walentynowicz, M., Schneider, S., & Stone, A. A. (2018). The effects of time frames on self-report. *PloS One*, *13*(8), e0201655. article. https://doi.org/10.1371/journal.pone.0201655

Walford, G. (1990). The 1988 Education Reform Act for England and Wales: Paths to privatization. *Educational Policy*, *42*(2), 127–144.

West, A., Allmendinger, J., Nikolai, R., & Barham, E. (2010). Decentralisation and Educational Achievement in Germany and the UK. *Environment and Planning C: Government and Policy*, *28*(3), 450–468. https://doi.org/10.1068/c0992

West, Anne, & Bailey, E. (2013). The Development of the Academies Programme: "Privatising" School-Based Education in England 1986-2013. *British Journal of Educational Studies*, *61*(2), 137–159.

Winthrop, R., & Simons, K. A. (2013). Can International Large-Scale Assessments Inform a Global Learning Goal? Insights from the Learning Metrics Task Force. *Research in Comparative and International Education*, *8*(3), 279–295.

Woessmann, L. (2009). *School accountability, autonomy and choice around the world*. Cheltenham: Edward Elgar.

Wohlstetter, P., Smith, J., & Farrell, C. C. (2013). *Choices and Challenges: charter school performance in perspective*. Cambridge, Mass: Harvard Education Press.

Wong, M., Cook, T. D., & Steiner, P. M. (2009). *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series*. Institute for Policy Research.

World Bank (2007). Toward high-quality education in Peru: Standards, accountability, and capacity building. Washington DC: World Bank.

World Bank. (2012a). *FYR Macedonia SABER Report*. Washington, D.C.: World Bank. Retrieved from http://saber.worldbank.org/index.cfm?indx=2&ctrn=MK

World Bank (2012b) Thailand SABER Country Report. Washington, D. C: World Bank. Retrieved from http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/SAA/SABER_SAA_Thailand_CR_Final_2012.pdf

World Bank (2013a) Columbia SABER Country Report. Washington, D. C: World Bank. Retrieved from wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/ECD/SABER_ECD_Colombia_CR_Final_2013.pdf

World Bank (2013b) Lebanon SABER Country Report. Washington, D. C: World Bank. Retrieved from http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/SAS/SABER_SA_Lebanon_CR_Final_2013.pdf

World Bank. (2013c). *Bi-lateral immigration statistics*. Washington, D.C.: World Bank. Retrieved from pubdocs.worldbank.org/pubdocs/publicdoc/2015/10/38881445543162029/bilateral-migration-matrix-2013-0.xlsx

World Bank (2014) Republic of Moldova SABER Country Report. Washington, D. C: World Bank. Retrieved from http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/TCH/SABER_Teachers_Moldova_CR_Final_2014.pdf

World Bank. (2015a). *The Little Data Book*. Washington, D.C: World Bank.

World Bank (2015b) Hashemite Kingdon of Jordan SABER Country Report. Washington, D. C: World Bank. Retrieved from http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/SAA/SABER_SAA_Jordan.pdf

World Bank (2015c) Tunisie Rapport Pays SABER. Washington, D. C: World Bank. Retrieved from http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/ECD/SABER_ECD_Tunisia.pdf

World Bank. (2016). *School Autonomy and Accountability*. Washington, D.C.: World Bank. http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/brief/SABER_SAA_Brief.pdf

Wößmann, L., Lüdemann, E., Schutz, G., & West, M. R. (2007). *School accountability, autonomy, choice, and the equity of student achievement: International evidence from PISA 2003*. *OECD Education Working Papers*. Washington, DC & Paris. https://doi.org/10.1787/246374511832

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*(2–3), 114–128. https://doi.org/10.1016/j.stueduc.2005.05.005

Xia, J., Gao, X., & Shen, J. (2017). School Autonomy: A Comparison between China and the United States. *Chinese Education & Society*, *50*(3), 284–305. https://doi.org/10.1080/10611932.2017.1331010

Yavuz, E., Tan, Ş., & Atar, H. K. (2019). Effects of Students and School Variables on SBS Achievements and Growth in Mathematic. *Journal of Measurement and Evaluation in Education and Psychology*, *10*(1), 96–116. article. https://doi.org/10.21031/epod.493297

Yi, P. (2015). Do school accountability and autonomy affect PISA achievement? Evidence from South Korea. *KEDI Journal of Educational Policy, 12(2),*197–223.

Yi, P., & Shin, I. S. (2018). Multilevel relations between external accountability, internal accountability, and math achievement: A cross-country analysis. *Problems of Education in the 21st Century*, *76*(3), 318–332.

Younger, M., & Cobbett, M. (2014). Gendered perceptions of schooling: classroom dynamics and inequalities within four Caribbean secondary schools. *Educational Review*, *66*(1), 1–21. https://doi.org/10.1080/00131911.2012.749218

Zendeli, F. (2011). Educational reforms and administration of the education in Macedonia. *Procedia - Social and Behavioral Sciences*, *15*, 4071–4075. https://doi.org/10.1016/j.sbspro.2011.04.417

Zendeli, F. (2015). Decentralizaton of Power in the Republic of Macedonia as Problem-Solving in the Inter-Ethnic Relations. *Revista de Științe Politice. Revue Des Sciences Politiques*, (48), 42–51.

Appendix I – PISA 2015 Materials

PISA 2015 database : https://www.oecd.org/pisa/data/2015database/

PISA 2015 Student Questionnaire:
https://www.oecd.org/pisa/data/CY6_QST_MS_STQ_CBA_Final.pdf

PISA 2015 School Questionnaire: https://www.oecd.org/pisa/data/CY6_QST_MS_SCQ_CBA_Final.pdf

PISA 2015 Questionnaire Codebook:
https://www.oecd.org/pisa/data/2015database/Codebook_CMB.xlsx

PISA 2015 Student Questionnaire SPSS datafile:
https://webfs.oecd.org/pisa/PUF_SPSS_COMBINED_CMB_STU_QQQ.zip

PISA 2015 School Questionnaire SPSS datafile:
https://webfs.oecd.org/pisa/PUF_SPSS_COMBINED_CMB_SCH_QQQ.zip

Appendix II – SAS PROC MIXED

SAS© Studio is made freely available to researchers:
https://www.sas.com/en_us/software/on-demand-for-academics.html

The SAS PROC MIXED procedure fits a variety of mixed linear models to data and allows the estimation of fixed-effects parameters and covariance parameters for nested data:
https://support.sas.com/documentation/onlinedoc/stat/141/mixed.pdf

Example code and output:

```
/*COMBINED FINAL MODEL*/
PROC MIXED DATA=MEA.COMBINED NOCLPRINT METHOD=ML COVTEST;
WEIGHT NSENWT;
CLASS cntschid schltype cntryid  incgp;
MODEL PV1MATH =  gender escs  schltype edushort aext aint apar incgp gender*schltype /solution;
RANDOM INTERCEPT gender escs /subject=cntschid (cntryid);
RANDOM INTERCEPT escs schltype /subject=cntryid;
RUN;
```

The Mixed Procedure

Model Information

Data Set	MEA.COMBINED
Dependent Variable	PV1MATH
Weight Variable	NSENWT
Covariance Structure	Variance Components
Subject Effects	CNTSCHID(CNTRYID), CNTRYID
Estimation Method	ML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions	
Covariance Parameters	7
Columns in X	16
Columns in Z per Subject	815
Subjects	12
Max Obs per Subject	8575

Number of Observations	
Number of Observations Read	57074
Number of Observations Used	57074
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Log Like	Criterion
0	1	690091.75229319	
1	4	670715.40645009	.
2	1	670583.39666491	0.00015739
3	1	670526.57017037	0.00007433
4	1	670499.08827639	0.00004126
5	1	670483.85391161	0.00002155
6	1	670475.98198341	0.00001000
7	1	670472.37742934	0.00000395
8	1	670470.98370087	0.00000112
9	1	670470.60999230	0.00000014
10	1	670470.56606916	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	**CNTSCHID(CNTRYID)**	1729.68	69.0389	25.05	<.0001
GENDER	**CNTSCHID(CNTRYID)**	205.60	29.0861	7.08	<.0001
ESCS	**CNTSCHID(CNTRYID)**	64.1232	7.8827	8.15	<.0001
Intercept	**CNTRYID**	1621.99	734.69	2.21	0.0136
ESCS	**CNTRYID**	29.6314	12.6842	2.30	0.0109
SCHLTYPE	**CNTRYID**	113.95	89.5662	1.27	0.1025
Residual		5195.40	31.8942	162.89	<.0001

Fit Statistics

-2 Log Likelihood	670470.6
AIC (Smaller is Better)	670510.6
AICC (Smaller is Better)	670510.6
BIC (Smaller is Better)	670520.3

Solution for Fixed Effects

Effect	SCHLTYPE	INCGP	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			445.73	21.7323	0	20.51	.
GENDER			5.9682	0.9467	1583	6.30	<.0001
ESCS			12.5131	1.6144	0	7.75	.
SCHLTYPE	**1**		22.5529	8.3463	0	2.70	.
SCHLTYPE	**2**		-9.8529	9.3740	0	-1.05	.
SCHLTYPE	**3**		0	.	.	.	.
EDUSHORT			-5.3240	0.9580	52E3	-5.56	<.0001
AEXT			6.9090	2.4803	52E3	2.79	0.0053
AINT			7.9182	3.0830	52E3	2.57	0.0102

Solution for Fixed Effects

Effect	SCHLTYPE	INCGP	Estimate	Standard Error	DF	t Value	Pr > |t|
APAR			7.5939	3.5626	52E3	2.13	0.0331
INCGP		**1**	-61.8354	29.8547	52E3	-2.07	0.0383
INCGP		**2**	-67.3667	29.7736	52E3	-2.26	0.0237
INCGP		**3**	0	.	.	.	.
GENDER*SCHLTYPE	**1**		5.8794	2.3551	52E3	2.49	0.0102
GENDER *SCHLTYPE	**2**		6.6413	2.4731	52E3	2.67	0.0075
GENDER *SCHLTYPE	**3**		0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
GENDER	1	1583	85.72	<.0001
ESCS	1	0	60.07	.
SCHLTYPE	2	0	5.74	.
EDUSHORT	1	52E3	30.91	<.0001
AEXT	1	52E3	7.76	0.0053
AINT	1	52E3	6.60	0.0102
APAR	1	52E3	4.54	0.0331
INCGP	2	52E3	3.15	0.0430
GENDER *SCHLTYPE	2	52E3	5.79	0.0031