To cite this article: Julian Elliott, Lazar Stankov, Jihyun Lee & Jens F. Beckmann (2018): What did PISA and TIMSS ever do for us?: the potential of large scale datasets for understanding and improving educational practice, Comparative Education, DOI: 10.1080/03050068.2018.1545386

What did PISA and TIMSS ever do for us?: The potential of large scale datasets for understanding and improving educational practice.

Julian Elliott, Durham University
Lazar Stankov, University of Sydney, and University of Southern Queensland
Jihyun Lee, University of New South Wales
Jens F. Beckmann, Durham University

****

ABSTRACT

There appears to be something of an intellectual and philosophical gulf between education researchers who seek insights from statistical analyses of complex data-sets such as those provided by the OECD (PISA), and others who seek to develop rich, contextualised socio-historical understandings that can shed light upon why particular classroom practices operate and are sustained within a given milieu. This paper outlines these different perspectives, with particular reference to non-cognitive factors, and concludes that while data from large datasets can assist in gaining greater understanding of cross-cultural differences, to be meaningful, such analyses should be incorporated within complex ecosystemic accounts.

****

The belief that one can use relatively large data sets and sophisticated statistical techniques to guide strategies to improve educational practice and performance – globally and nationally – has grown steadily as a result of increasingly far-reaching large-scale international assessment programmes. Such understandings represent a modern 'hi-tech' version of attempts to learn from overseas practices stretching back to the Victorian age. Then as now, the world's most powerful nations were nervously comparing their industrial, economic and military strength, concerned as to the influence upon these of their respective educational practices. Even then, however, prescient commentators were warning that importing discrete practices from other cultures, without considering the context in which these operated, would not prove efficacious (Cowen, 2014).

Initial small-scale international comparative studies undertaken by the International Association for the Evaluation of Educational Achievement (I.E.A.) and the Education Testing Services' International Assessment of Educational Progress (I.A.E.P.) at the end of the last century alongside small-scale studies comparing American children with those from East Asian countries (Stevenson & Lee, 1990; Stevenson & Stigler, 1992) led to a gradual build-up of unease of comparative performance on the part of Western commentators. Spurring this concern was the rise of the Asian Tiger Economies and the significant attention this attracted in the international media. For many, the quality of schooling appeared to be a key explanatory factor, and there continues to be a strong belief that education is critical for economic growth and survival (although, see Auld & Morris, 2014, for a critique of this).

In the light of the overwhelming evidence of high performance on the tests in several East Asian countries and cities, researchers have sought to identify country- and regional-level characteristics that might offer an explanation. Factors that were suggested included nations' average IQ (Lynn & Meisenberg, 2010), superior memorisation (e.g., Stevenson & Stigler, 1992), parents' expectation and attitudes (Chao & Tseng, 2002; Lam, 2003, 2005; Peng & Wright, 1994; Schneider & Lee, 1990), home learning environment (Schneider & Lee, 1990), tendency to persist and approach to effort (Hau & Salili, 1996), conscientiousness (Vernon, 1982), long-term orientation (Hofstede, 2001), and reward for application (i.e., a belief in return of human investment, Leung, 2010).

Arguably, the most significant event in the evolution of the assessment-based international comparison movement was the publication of the Third International Mathematics and Science Study (TIMSS, Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith, 1996), involving students from several grade levels (for instance, third and fourth grades, seventh and eighth grades, and final year of secondary education in TIMSS 1995) across more than 40 countries. This, together with results from a follow-up study (TIMSS-Repeat, Mullis et al., 2000), highlighted the relative strength of several East Asian countries, together with a number of East and West European countries. While these data largely reflected pre-existing understandings held within the international research community, TIMSS captured the attention of the world's media on a scale not seen before, resulting in much soul-searching in seemingly less successful industrialised nations.

In responding to national concerns over TIMSS, education policymakers tended to focus upon differences in pedagogic practice, perhaps because these are both more salient and also easier elements to reform. In a direct echo of the Victorian traveller inspectors, researchers and policymakers have sought to observe educational practices in high performing countries and subsequently offer judgements as to what factors might be academically influential and open to 'borrowing'. In the final years of the twentieth century, the consensus among those seeking to identify this pedagogic Philosopher's Stone, coalesced around a set of classroom practices that together were termed "whole class interactive teaching" whereby:

> "… the teacher starts with a problem and develops solutions and concepts through a series of graded questions addressed to the whole class" (Reynolds & Farrell, 1996, 56).

This approach was strongly endorsed by the Chief Inspector of Schools in England and Wales who regularly contended throughout the late1990s that 50% of primary school lesson time (60% for mathematics) should take this form. Some English local authority inspectors and schools imported practices from Germany, Switzerland (Luxton & Last, 1997) and Hungary (Burghes, 1996) to great acclaim from politicians and journalists. In the United States, Stigler and Hiebert (1999) offered a series of recommendations drawing upon their observations of Japanese mathematics teaching that was met with similar approbation.

Findings from the various rounds of the Programme for International Student Achievement (PISA) since 2000, and the reformulation of the T in TIMSS from the 'Third' to 'Trends', to enable this powerful brand to continue in subsequent testing phases, have taken international comparison to a whole new level. PISA's (and by extension, the OECD's) success in attracting attention from researchers, policymakers, practitioners and the general public, and its influence upon what is deemed to be important educational factors (e.g. OECD, 2010, 2013a, b) have been remarkable. Indeed, the term 'Pisa Shock' was coined to describe the reaction of some countries that found themselves performing at a level lower than they might have anticipated. PISA has been instrumental in creating a sense of global educational accountability, leading to an increased internal focus upon national accountability mechanisms (Breakspear, 2012; Meyer & Benavot, 2013; Sellar & Lingaard, 2014).

The new rankings that were produced resulted in a well-beaten path to Finland, a consistently high performer often seen by policymakers as more culturally similar to Western contexts than Asian educational systems. Interestingly, it was those aspects of Finnish practice that fitted politicians' existing preconceptions that were most often highlighted: those that did not fit their preselected direction of travel (for example, the investment in a highly qualified and remunerated teaching profession and the absence of high stakes testing) tended to be quietly ignored.

> "Rather than engaging in policy *borrowing*, the government has selectively *referenced* policies in East Asia in an attempt to promote and legitimate its long preferred policy agenda" (You & Morris, 2015, p. 900).

Interestingly, with their eyes on what are perceived as desirable aspects of Western educational practices (Elliott & Nguyen, 2008), supposedly model countries, such as Hong Kong and China, have introduced reforms in a different direction to those advocated by the English policymakers (Adamson, Forestier, Morris, & Han, 2017; You, this volume). Nevertheless, the performance levels of East Asian countries continue to beguile many Western policymakers who appear to be largely undaunted by the many variations among these in terms of educational system-level features including curricula, public expenditures on education instruction time per year, or the time spent on studying out-of-school (Leung, 2002; Schmidt, McKnight, Valverde, Houang, &Wiley, 1997). The only common system-level characteristics were these countries' large class sizes and high population density (Leung, 2002), the long established role of public examinations as a source of social status and mobility (Kim, 2009) and, arguably, rote memorization practices (Leung, 2001).

Such factors can perhaps be more easily disregarded if one focuses solely upon the classroom delivery of academic material in one or two favoured countries. Crossley (2014, 17) quotes the U.K. Secretary of State for Education informing a Parliamentary Committee in 2010 that he had "… been to Singapore and Hong Kong and what is striking is that many of the lessons that apply there are lessons that we can apply here". This was followed by a much publicised visit to England by Chinese

maths teachers and the subsequent introduction of Chinese approaches to the teaching of mathematics and the use of Chinese mathematics textbooks

Notwithstanding headline grabbing initiatives of this kind, the ever-increasing mass of data, made available by TIMSS and PISA, has gradually led to something of a shift in focus from the educational practices of one or two high-scoring nations to the consideration of discrete factors that seem to be common to classrooms across high performing countries and regions (Jensen et al., 2012; Lee, 2014; Schleicher, 2009). In addition, greater consideration has been given to pupil characteristics (e.g. social class), and differing forms of educational administration, structure and expenditure. Recommendations based upon analyses of large datasets include the recruitment of more skilled teachers, increased time devoted to mathematics and science subject matter ('opportunity to learn'), increased investment in the early years of education, decentralization of school management, and the equitable distribution of educational resources (OECD, 2010, 2013b).

Despite this more sophisticated approach to the identification of seemingly effective educational practices, there remain many methodological, contextual, and causal complexities (Auld & Morris, 2016; Hopfenbeck et al., 2018). One problem concerns the use of cross-sectional data to make generalizations about effective elements of educational systems (Carnoy et al., 2016). Snapshots, such as those used by PISA, reduce a student's lifetime of learning to their experience in their present age and school. This can lead to bias in making inferences, in particular, about the effects of the "current" teachers upon the educational performance shaped by multiple years of schooling (and other out-of-school factors). A second criticism, and one voiced frequently by specialists in the field of comparative education, concerns the failure to establish causality and to recognise the importance of the national and local context in which schooling operates. Such a flaw is arguably more obvious when the direct importation of one country's approach to teaching one or more academic subjects is proposed. It becomes rather less salient when the introduction of discrete educational elements that appear to feature in highly successful countries is recommended. Perhaps this helps to explain arguably declined more recently. Now, to the consternation of many education comparativists, "…the genie is firmly back why recognition of the need to take contextual subtleties into consideration has in the bottle: clearly, there are no important differences between Singapore and Hong Kong and Salford" (Cowen, 2014, 287).

While disputing the context-free simplicities served up by the likes of OECD and McKinsey, and readily accepting the observation that PISA is not comparative education (Cowen, 2014), we would like to suggest that it is possible to draw upon findings from large-scale datasets such as PISA to help us understand the operation of education in and across particular contexts. However, we suggest that such data can merely serve to offer insights as part of considerably more comprehensive analyses. To illustrate, Morsy, Khavenson, and Carnoy (2018) provide a detailed analysis of possible reasons why PISA mathematics scores in Australia have declined over the past fifteen years. Their analysis highlights a number of seeming misunderstandings in relation to the reasons put forward by education commentators (e.g., a false belief that a key factor was the quality of the teachers

themselves). However, in line with our position as outlined in this paper, these authors emphasise that the datasets currently available can only offer 'clues' (75) rather than definitive answers.

To gain an adequate understanding of the operation of educational practice in a given context, we need to draw upon data from all aspects of the ecosystem. As Sadler (1900, 50) famously pointed out: "In studying foreign systems of Education we should not forget that the things outside the schools matter even more than the things inside the schools, and govern and interpret the things inside." Unfortunately, analyses drawn from large-scale assessments have been undertaken by those whose disciplinary foundations are not conducive to the forms of contextualised analysis that are required. On the other hand, there is also a significant risk that the potential value of data gained from large-scale studies such as PISA might be overlooked for this very same reason.

We argue that data from a wide variety of levels can provide insights to help us understand learning and development in specific contexts. In advocating an ecosystemic approach, we would wish to highlight Bronfenbrenner's classic theory of human development (1979) which describes the ecological environment as a series of interrelated systems. The microsystem, referring to the various settings in which the individual develops – home, school, recreational locales - operates at the innermost level. The next level, the mesosystem considers how these elements interconnect. Here relations between school and home are crucially important for education. At the third level are exosystems – settings where the individual may not be physically present, but which may still exert significant influence. For example, changes to parental employment or regional conflict may impact significantly upon the child's developing experience at home and at school. Finally, the macrosystem consists of the comprehensive network of relational systems that operates within a given culture. This includes educational and vocational opportunities, socioeconomic and socio-political factors, and the opportunities offered to individuals on the basis of their age, ethnicity and gender. It is important that in analysing these systems at a national level we recognise that there are likely to be significant differences between regions and local communities (Carnoy et al., 2017). Bronfenbrenner's theory does not concern the independent and unidirectional effects of context upon human development. Rather, it emphasises how development as a result of the interactive effects of both context and particular person characteristics. This more complex conception typically requires complex analytical approaches, ideally involving longitudinal designs (Tudge et al., 2016).

If we are to learn from the educational performance of other nations, it would hardly seem necessary to argue that we should avoid the use of simple causal explanations that disregard the interplay of different parts of the ecosystem. The notion that one can lift Chinese or Singaporean mathematics teaching approaches, for example, and drop them relatively unmodified into low attaining schools in the U.K., without recognising that these draw upon systems of human relations that are radically different in the two cultures is clearly problematic. What is necessary is that imported practices are modified in ways that are sensitive to the host culture and its context. Nguyen (this volume) for example, illustrates how cooperative learning, a largely Western development, can be sensitively applied within an Asian (Vietnamese) classroom setting. To work effectively in the

latter context, a number of interpersonal elements, particularly interpersonal confrontation, need to be modified to ensure harmony and the preservation of the public image (face) of all group members.

In the present paper, we argue that many system-level factors – historical and cultural – are typically left out when the implications arising from large-scale assessments are considered, even if the responsible agencies are aware of "bigger" ecosystemic issues. Firstly, it is not possible to include all the encompassing cultural factors which may be unique to one group but not to others, and secondly, cultural subtleties often cannot be captured in respondents' responses to questions that must work for large numbers of languages and countries. This may be particularly true for questions which tap student attitudes and beliefs (so-called non-cognitive factors). In the following sections, as an illustration, we highlight what the literature claims are some of the most important traditional forces driving what are often termed Confucian societies and their members, and present more recent findings from cross-cultural studies that offer insight for analyses of the "unexplained", "uncaptured" stories that figure in current international studies. In this paper, reference to the term Confucian is used to describe East Asian countries or educational systems that are understood to be greatly influenced by Confucianism. It is acknowledged, however, that such use is highly contested.

The promise of non-cognitive factors in large-scale datasets

The broadening of PISA has reflected and fuelled growing interest in the role of non-cognitive factors in student achievement. However, we should be alert to a danger that, in an echo of the errors of the 'what works best' approach, findings will be used to make sweeping assertions about desirable educational practice without arriving at the more sensitive, contextualised understandings that are necessary for informing action. Indeed, without due consideration of ontological and epistemological foundations, even careful and sensitive attempts to introduce classroom approaches from other countries are unlikely to prove successful (Komatsu & Rappleye, 2018).

In signalling the limitations of PISA and similar measures for informing practice, we are not focusing upon the quality of the data derived from large-scale studies of educational performance [although, concerns have been raised in this respect (Hopfenbeck et al., 2018)]. Rather, our contention is that the data that are collected in such investigations only provide description, and this, in itself, does not constitute understanding or evidence of what would work in another national context (including causality). Furthermore, collecting more data will not necessarily result in greater understanding. The underpinning rationale for this statement is that:

(1) Research can be conducted by employing three different lenses, descriptive, explanatory, and through intervention
(2) These lenses are hierarchically ordered. Explanation, or understanding, depends on a thorough description of the phenomenon of interest; effective interventions depend on first achieving a proper understanding. Each set of lenses is reliant upon different methodological approaches

With such caveats in mind, we contend that comparative country-level data about non-cognitive variables, for example, attitudes and beliefs, can inform and enrich contextualised

ecosystemic analyses by laying descriptive foundations for future research studies geared to enable greater understanding.

In examining Confucianism, and then situating this within a wider multi-country perspective, the paper will focus upon two different levels of analysis. Firstly, it will examine discrete factors that large-scale assessments have identified as being influential, or otherwise, in explaining academic performance. It will then shift the scale of analysis to consider broader sociocultural dimensions that offer additional explanatory value. As a result of such analyses, we may gain insights into aspects of supposedly successful nations that also present a range of adverse challenges for their students.

Specific non-cognitive factors and educational achievement

Using several large databases (TIMSS 2003, 2007, and 2011, and PISA 2003 and 2012) Lee and Stankov (2018) sought to identify from 65 non-cognitive variables those which were the best predictors of mathematics performance. While most were found to have a surprisingly low association (including, for example, positive affect towards mathematics, valuing mathematics, engagement with mathematics lesson), those items tapping mathematical confidence (TIMSS) or self-efficacy (PISA) (nb. TIMSS' use of academic 'confidence', is identical to PISA's self-concept) demonstrated a moderately strong effect size (approaching $r = .40$) across both OECD and non-OECD countries. Interestingly, however, when between-country associations were considered, earlier observations concerning a negative relationship between achievement scores and academic self-concept (OECD, 2004; Mullis et al., 2008) were replicated. So why is academic self-concept (or in terms of TIMSS, confidence) seemingly predictive of success at the individual level but not at the country level? Here, contextual analysis sheds light on a seeming paradox.

It appears that the negative between-country relationship is mainly driven by high-performing East Asian students. In spite of their comparatively high performance on tests of mathematics, students from these countries appear not to particularly enjoy this subject and report a low self-concept when this is related to their mathematical ability (e.g., Lee, 2009; Leung, 2002; Lee & Stankov, 2018; Wilkins, 2004) across both PISA and TIMSS surveys. In fact, from more than 40 countries across the world, students from Japan and South Korea were consistently among those cohorts who had the lowest confidence in mathematics, the lowest self-concept in mathematics, the lowest self-efficacy in mathematics, and the highest dislike/anxiety of mathematics (Lee, 2009). The PISA 2012 data that we re-analysed for this paper (presented in Figure 2) show that students from countries strongly influenced by Confucianism, in general, have lower scores on mathematics self-concept, and higher scores on anxiety, than the average of the OECD countries (left-hand panel, N = 32). The same pattern (e.g., lower scores from students in East Asian countries) was present in other, similar, attitudinal measures of PISA, such as attitudes toward school and sense of belonging, (right-hand panel in Fig. 2), among many other non-cognitive PISA 2012 indices.

======= INSERT FIG 1 & 2 ABOUT HERE ========

Such findings have proven puzzling for education policymakers because the main motivational and attitudinal constructs that many educators and some psychologists have long believed to be positive forces in human behaviour did not appear to coincide with the high academic performance of East Asian students. As we argue above, a key strength of adopting an ecosystemic approach to the study of educational climates and practices is its capacity to help contextualisation at many different levels. So how does an ecosystemic analysis help us to understand the paradox of high academic performance and low academic self-concept?

In a series of studies, Stankov (2010, 2013) has emphasized the potentially deleterious nature of various psychological conditions among students exposed to education systems underpinned by Confucianism. Using datasets enabling cross-cultural comparisons in personality, social attitudes, values, social norms, and conservatism (Stankov & Lee, 2009), and a Militant Extremist Mindset (Stankov, Saucier, & Knežević, 2010), Stankov and his colleagues found that East Asian countries held more pronounced levels of negative dispositions (e.g., toughness, maliciousness, and pro-violence) - that he labelled together as unforgiveness – than European countries (i.e., Finland, Netherlands, Denmark, Serbia, Latvia, collectively). Stankov (2010, 2013) then hypothesized that in Confucian societies, unforgiveness may be the source of high levels of anxiety and self-doubt among East Asian students, and tentatively concluded that the roots of the "undesirable" psychological outcomes are to be found in the cultural legacy, interpretation, and modern applications of Confucianism [or neo-Confucianism (Kim, 2009)].

In the following sections, we focus on values that have historically shaped the learning practices, outcomes, and values of East Asian students (while not overlooking the likelihood of significant within-group variability). While these themes have been widely reported in the literature, we contend that not only have these been neglected by policymakers and think tanks (see Auld and Morris, 2014, for detailed discussion of 'New Paradigm' analyses of large scale international comparison of educational performance data) but also point out that potentially negative consequences are typically under-reported. We elaborate on how specific aspects of Confucianism have engendered the fierce desire for high academic achievement, and speculate about how social attitudes summarized as "unforgiveness" might mediate Confucianism and academic achievement and undesirable psychological outcomes among students. Core variables in conceptualizing East Asian educational performance are depicted as below.

======= INSERT FIG 3 ABOUT HERE ========

Historical roots of non-cognitive tendencies among East Asian students

Confucius was a Chinese philosopher, teacher, and political theorist (551 – 478 B.C.). Although Confucianism was promoted as the official "religion" (and there was strong denial of other religions from the government), it was not entirely viewed as a religion but taught as a way of living across ethical, social, political, legal, and philosophical domains (Yao, 2000). Core Confucian values include: humanity, generosity, patience, righteous behaviour, a sense of duty, integrity, wisdom,

ritual, politeness, knowledge, trust, respect, loyalty, filial piety, honesty, kindness, forgiveness, gentleness, harmony, tolerance, modesty, reciprocity, shame, bravery, and frugality. Less positive values might arguably include subservience, elitism, misogyny and patronage. While most of the values in this long list may equally be cited as core values of other world religions, those which militate against personal autonomy and agency seem to be particularly influential in East Asian societies.

At the heart of Confucian teaching is its emphasis on the duty *as* and *toward* government, parents, teachers, and older siblings. One's position and duty in the relationship with others is very important: "A Confucian person is not a free-standing individual but a social being embedded in a social structure" (Nuyen, 2001, p. 61). Two legacies of Confucianism – an emphasis on national academic assessments and the nature of social relationships – have been of particular relevance to education.

National Examination Systems and the Emphasis on Meritocracy

There is a long-standing history of exam-driven educational practice in Confucian Asia (Higgins & Sun, 2002). In China, a bureaucratic examination to select state officials was first implemented in 605 A.D. and thrived during the Ming dynasty (1368-1644 A.D.). Passing the government-initiated examination was viewed as signifying mastery of Confucianism. The local elites of non-aristocratic background were able to obtain a powerful government position by this means (Nuyen, 2001). Thus, the examination was used and, perhaps more importantly, was viewed, as a way to enact social and economic justice and upward social mobility for poor, but intellectually able, people. This has generally been considered to be a fair way to promote equality in a highly structured society (Nuyen, 2001). In contemporary Confucian societies, in part, influenced by Western approaches (Niu, 2007), many types of examinations – both national and commercially available standardized tests intended for use in large-scale settings – are used as the primary method for selection and placement in schools, universities, government positions, and even business sectors. National examinations for university/college entrance and for the selection of government-officials[1] are the best known.

In Confucian societies, it is widely recognized that competitive examinations and meritocracy (i.e., a belief that power should be vested in individuals almost exclusively on the basis of ability and talent) are highly intertwined with each other. Examination systems in such countries tend to be "national", which means that *all* students in the country participate. With dense populations and high parental expectations, these examination systems are *extremely competitive* (Leung, 2002) and often exert a negative psychological impact on students. Highly frequent examinations expose students to *constant judgment* about their academic capabilities. Evaluative judgments ("you are a good/bad student") tend to be frequently given to children from a young age as a form of encouragement by

[1] Singapore is an exception in that it has moved away from using such tests to select government officials.

parents, relatives, parents' friends, teachers, peers, themselves and even neighbours (Schneider & Lee, 1990).

When children are exposed to frequent and competitive exams, their lives are *highly regulated by* the examination schedules. There is a constant and high level of pressure to do well in the next cycle of examinations and, therefore, it is not surprising that TIMSS data have shown that East Asian students do not report sports or having fun are being important to them (Leung, 2002). Perhaps the most troublesome outcome is that "… competitive examinations systems … have left a large number of students classified as failures in their system, and these repeated experiences of a sense of failure may have further reinforced … lack of confidence [and other negative psychological impacts on individuals' self-concept]" (Leung, 2002, 106). It also appears that some students in these countries "over-prepare" for examinations that might possibly be held in the future – thus resulting in life-long learning that goes beyond the "currently required or imposed" examination systems. (Komatsu & Rappleye, 2017).

The feeling of academic failure among students in many Asian countries is a common phenomenon, regardless of their actual academic achievement level. PISA data have consistently shown that students in Confucian societies have the lowest self-beliefs in their academic abilities in spite of their actual high performance (see Figure 2). Perceived academic failure by high-achieving students may result even if they only slip a grade in one academic subject (Schneider & Lee, 1990). It is a prevalent belief that the respectful and "right" way to acquire social approbation is by meeting stringent examination requirements[2]. Performing at a high level in national examinations would appear to be the single most important life goal for most East Asian students and their parents (Lam, 2003, 2005). Consequently, parents and children often prioritise school performance to a degree that is maladaptive for the child. While acting as a strong source of achievement motivation (e.g., Hau & Salili, 1996), and seemingly an important factor in the superior performance of Eastern Asian students (Leung, 2014), a negative consequence is it can result in overwhelming anxiety about the child's scholastic performance (e.g., Lee, 2009). Such pressures lead to undesirable psychological outcomes such as lack of enjoyment in learning, excessive stress, and undue preoccupation for, and fear of examinations [although with the positive changes driven by various countries such as Japan and China, we should be wary of stereotyping the Asian student's experience as being exclusively a joyless 'examination hell' (Komatsu & Rappleye, 2018)].

Social Relationships and Interactions

[2] There are some variations among Confucian Asian countries in relation to second-chance opportunities to re-enter or transfer to a better school/university. In South Korea, for example, it is very hard to transfer to an academic track once a student obtains low scores on the university entrance exams. However, those students in Singapore who are placed into the Normal Academic track (those whose primary school exit examination scores are between the bottom 61th and 85th national-percentile) have the opportunity to take the national examination again, five years later, in order to enter the Advanced Pre-University group.

Three aspects of Confucianism may be particularly relevant for understanding student orientation towards schoolwork - filial piety, teacher-student relationships, and the respect accorded to scholarship. Filial piety (i.e., to respect, be obedient, love, and take care of parents) plays a significant role in the academic achievement of young children and students in East Asian countries. From an early age, children understand that one of their duties is to pay back their parents by achievement of upward social mobility (Lam, 2005; Schneider & Lee, 1990). By doing so, they preserve or upgrade family honor, and avoid shame associated with failure in achievement. People in East Asia traditionally believe that becoming a scholar brings great honour and joy to the family, and parents tend to strongly encourage their children to achieve such a goal (Shek & Chan, 1999). While an increasing number of young people in these societies may deviate from this perspective, being successful in non-academic fields (such as sports, entertainment, or music) is still widely seen as a secondary achievement.

The nature of teacher-student relationships follows that of parent-child relationships. Teachers were to receive great respect from their students for their wisdom, knowledge, and moral conducts (Lau, 1995; Tweed & Lehman, 2002). A teacher is the supreme authority, akin to the idea of rulers in the classroom. While it is unlikely that modern-day East Asian students would hold a similarly high level of respect for their teachers as was the case in the past (Lee & Chung, under review; Biggs, 1996; Pace & Hemmings, 2007), a practice of not questioning teachers' authority continues to dominate. In contrast, US conceptions of ideal teacher-student relationships, marked by an emphasis upon democracy, self-determination and personal agency, are such that contestation is almost inevitable (Alexander, 2000). The implications for the operation of an effective pedagogy in each of these environments are, of course, substantial.

Aspects of the examination culture detailed above also exist in the other parts of the world. Western societies employ tests for admission to school and many business sectors, and manifest significant levels of in-group pressures and behaviours emanating from social pressure and norms. The differences between the non-Confucian and Confucian Asian cultures are how prevalent and pervasive the problematic phenomena exist in each culture, the extent to which those problems affect individuals' lives, the way the members of the societies perceive and respond to those problems, and how strongly and decisively the early school performance factors determine in later opportunities for success in life, and the extent to which social and educational systems allow for those who wish to go back to school or change career.

Beyond Confucianism: The Potential Role of Non-cognitive Factors across Multiple Countries

We now turn to an examination of a broader range of psychological aspects operating beyond East Asian countries. Here, it is clear that PISA and TIMSS results can be linked and interpreted in the context of some broad cross-cultural dimensions.

Broad Dimensions of Cross-cultural Differences

One of the goals of cross-cultural psychology research is to identify differences between countries in terms of a relatively small number of dimensions. For example, six dimensions of social norms are advocated in Hofstede's (2001) system: power distance, individualism vs. collectivism, masculinity vs. femininity, uncertainty avoidance, long-term orientation, and indulgence vs. restraint. From among the Confucian values that we mentioned above, teacher-student relationship and occupational values can be understood in terms of Hofstede's (2001) power distance dimension, and filial piety is linked to the collectivism/individualism dimension. In fact, Hofstede (2001, p. 217) reports that these two dimensions – power distance and collectivism/individualism – correlate highly (r = .68) and most East Asian countries have high scores on both. On the basis of this information it is tempting to conclude that high achievement and high levels of self-doubt within these societies may, in part, be attributed to their high standing on collectivism and the acceptance of large differences in power exercised by the citizens.

Latitude and Longitude of the Psychological Atlas of the World: Harshness/Softness and Conservatism/Liberalism

Stankov, Lee and van de Vijver (2014) reported that measures of power distance and individualism/collectivism not only have high correlations among themselves but, when placed within a battery of some twenty other measures, they also define the same factor. At the country level, that factor has positive loadings from measures of Toughness, Maliciousness, Self-indulgence and Performance Orientation and negative loadings from all Big Five measures of personality. It was labeled Harshness/Softness and it had a high correlation with Gelfand et al.'s (2014) measures of tightness/looseness. Thus, power distance and individualism/collectivism have commonalities with other psychological constructs (e.g., Toughness, Maliciousness, Self-indulgence), suggesting that East Asian societies are tight and their members have a tendency to treat each other in a harsher way than is common elsewhere in the world today. This indicates that punishment in these countries may be severe if expectations of the group one belongs to are not fulfilled. Harshness as a broader construct than power distance, and individualism/collectivism on their own, may provide a punishment-based motivational source simultaneously for both high achievement and for strong self-doubt.

======= INSERT FIG 4 ABOUT HERE ========

Figure 4 shows the location of 33 countries from Stankov et al. (2014) study in a two-dimensional system. It is apparent that East Asian countries are high on the Harshness pole (vertical line). On the opposite end of this dimension are mostly South American and some South European countries. The horizontal line in Figure 4 represents the Conservatism/Liberalism dimension that proved to be more important both in Stankov et al. (2014) and in subsequent cross-cultural studies, which we elaborate in more detail below.

Cross-cultural Differences in Conservatism/Liberalism

In a recent paper Stankov (2017) summarized the main findings from several cross-cultural studies and argued for the existence of the Conservative Syndrome. This refers to "… a person who attaches particular importance to the respect of tradition, humility, devoutness and moderation (i.e., Traditional values) as well as to obedience, self-discipline and politeness (i.e., Conformist values), social order, family, and national security (Security values) and has a sense of belonging to and a pride in a group with which he or she identifies (In-group Collectivism). A Conservative person also subscribes to conventional religious beliefs (Alphaism) and accepts the mystical, including paranormal, experiences (Deltaism). Such a person is likely to be less open to intellectual challenges (Openness) and will be seen as a responsible 'good citizen' at work and in their society (Conscientiousness) while expressing rather harsh views toward those outside his or her group (Harshness Towards Outsiders)" (see Stankov, 2010; p. 300). In this description, the terms in parentheses are the labels of the well-established scales for the measurement of non-cognitive psychological processes in cross-cultural research.

Conservative Syndrome refers to political and social attitudes whose goal is to preserve traditional philosophy and beliefs – i.e., people are sceptical of change, and believe in maintaining the status quo on social issues. Three main components of Conservative Syndrome identified in Stankov's recent work are: Nastiness[3]/Social Dominance (including Harshness and Unforgiveness), Religiosity, and Social Awareness/Morality. Differences between countries on Social Awareness/Morality tend to be small, so particularly important from the policy point of view are the first two components.

Social conservatism captured by the components of Conservative Syndrome is an important psychological dimension of cross-cultural difference and countries can be classified into three psychological groups. The grouping of these countries was based on close to 50 socio-psychological measures that can be broadly classified into the domains of personality, values, social attitudes, social norms, and social axioms. One can see from Figure 4 that the broad dimension along which the three groups of countries are distributed reflects their levels of conservatism. The following groupings of 33 countries have been reported by Stankov (2017):

- Group 1. Liberal: (mostly Europe): Ireland, Netherlands, United Kingdom, Germany, Spain, Greece, Turkey, Canada and Australia;
- Group 2. In-between: Russia, Poland, Ukraine, the United States, Brazil, Mexico, Egypt, Morocco, Singapore, China, Taiwan, South Korea, and Japan; and
- Group 3. Conservative: (mostly Southern hemisphere): Malaysia, Philippines, Thailand, Nepal, India, Bangladesh, Tanzania, Kenya, Ethiopia, Argentina, and Peru.

As can be seen, East Asian countries belong to the in-between grouping (Group 2), and are clustered together with other countries, such as US, Russia, and Poland. In other words, Confucian

[3] The term 'nastiness' to describe cultural differences may appear overly pejorative to a comparative education readership, but its origin is in cross-cultural psychology, a discipline where the use of such constructs is generally understood in a different fashion.

countries have a lot in common with other parts of the world, when many socio-psychological aspects are simultaneously examined. One strong cultural-heritage component that separates them from the other countries in the same group appears to be their historical roots in Confucianism.

Stankov and Lee's other work (Lee, 2009; 2014; 2016; Lee & Chung, under review; Lee & Shute, 2010; Lee & Stankov, 2013; Stankov & Lee, 2009; 2014; 2016; Stankov et al., 2014; Stankov, 2017) also shows that high achievement in East Asian countries may well be a consequence of much broader, cultural-societal factors (e.g., teacher-student relationships, societal emphasis on academic achievement), rather than a narrow set of variables at the student level. To address the frequent negative non-cognitive student outcomes, policymakers may need to consider major reforms of examination systems, recognise a broader range of forms of educational achievement, and seek to promote greater balance between individual interests and family responsibilities. While most educators in East Asia are fully aware that the push for academic achievement has "gone beyond the optimal level, creating undue pressure upon students, and resulting in all sorts of harmful effects" (Leung, 2001, p.43), social pressure and equity issues based on political, social and economic incentives may need to be evaluated in order to make real change in student expectations and behavioral outcomes shaped by parental and societal attitudes. However, as we repeatedly state in this paper, such recommendations cannot solely rely on international databases from PISA, TIMSS and their like.

Conservative countries (Group 3) tend to be more religious and more intolerant towards outsiders. They also tend to be low in terms of academic performance, at least in maths and science (Stankov, 2018), as do countries that score high on religiosity (Stoet & Geary, 2017). For example, the correlation between the Conservatism/Liberalism dimension and PISA 2012 mathematics scores was r = -.52. Since most countries in the liberal grouping and many countries in the in-between group, have high achievement scores, the main reason for the negative correlation is due to the conservative group (Group 3) that has pronounced low achievement and high conservatism scores. However, it is unlikely that the achievement problems of highly conservative countries can be addressed by tinkering with educational practices because there are likely to be many political and other societal issues that were not captured in the analysis.

Non-cognitive Data and Ecosystemic Analysis

If we are to draw upon non-cognitive information from large international datasets to make recommendations for practice, we must be careful not to make the same type of errors as those who are criticised for offering pedagogic reforms without due consideration of context. Furthermore, we should not confuse correlation with causation, and make recommendations for practice on the basis of the former. For example, Stankov and Lee (2016) and Lee and Stankov (2018) show that PISA and TIMSS data indicate that influential predictors of cognitive performance are self-beliefs, consisting of confidence, self-efficacy, anxiety, and self-concept. Indeed, these can explain more variance in achievement than sociodemographic factors (Pipere & Mieriņa, 2017). Should such findings lead us to conclude that inflating student academic self-beliefs is a sound policy for policymakers to endorse? If

so, how can this be reconciled with the finding that at the country (rather than the individual) level, the relationship operates in the opposite direction for some of these factors? And is there a possibility that boosting academic self-perceptions irrespective of actual effort and performance (a charge often associated with educational practice in the US) will result in a complacency that serves to reduce striving (Elliott, 2002)? It seems that neither high levels of overconfidence nor a significant lack of confidence are likely to be optimal for learning.

Lee and Stankov's (2018) analysis of findings from several PISA and TIMSS studies, incorporated with four other recent meta-analytic reviews of non-cognitive constructs and academic achievement (Hattie, 2009; Lee & Shute, 2010; Richardson et al., 2012; Stankov, 2013), throws up a number of challenging questions. While, as noted above, students' beliefs about their ability to perform academically have predictive capacity, most of the 65 measures of non-cognitive factors examined showed low or zero correlations with academic achievement. Lee and Stankov (2018) note that this outcome is unexpected as the measures employed were those deemed to have most relevance for educational performance. Particularly surprising was the low relationship of measures tapping motivation and engagement. Can this really mean that these are of minor or of no importance for student performance? Or perhaps the sorts of questions that are being asked are subject to differing interpretations by students in different cultures. For example, investing an extra hour a day in self-study might be variously perceived to be normal, below expectations, or excessive.

As is illustrated above in our examination of Confucian societies, in determining educational policy and practice, it is unwise to draw conclusions from large datasets (whether dealing with cognitive or non-cognitive measures) without entering into detailed analysis of all levels of a particular educational ecosystem. To fail to do so is to risk misunderstanding and potentially undermine existing good practice.

An example of such an analysis, albeit in non-Confucian societies, is provided in an account of educational motivation in three 'milieux': St. Petersburg, Russia, Kentucky, USA and Sunderland, England (Elliott et al., 2005). Their ecosystemic accounts, drawing upon a wide range of national and regional historical, sociocultural, and educational influences, led them to question ongoing attempts by policymakers in Kentucky to significantly transform educational achievement. In Kentucky, reforms involved greater emphasis upon learner centred pedagogic practices, backed up by high stakes testing, yet the anti-intellectual, anti-academic attitudes and beliefs of the children, typically reflecting those of their families and their local communities, differed significantly from those encountered by the children who were studied St. Petersburg. In this latter milieu, erudition and high levels of academic striving and achievement tended to be greatly prized at all levels of the ecosystem. In both Kentucky and St. Petersburg, it was possible to demonstrate that these two very different sets of attitudes and beliefs could be traced back through many generations.

Elliott et al.'s study enabled the potential resolution of a number of seeming paradoxes, for example, the Russian children emphasised attributions for ability over effort; the reports of the American children emphasised effort over ability; yet in their analyses of home and school life, the

Russian children, generally worked far harder on educational tasks in school and at home (see Alexander, 2000, for similar findings). By examining the various aspects of each ecosystems the authors were able to offer a number of reasons (most likely interacting with each other) why the American emphasis upon the importance of effort for educational achievement was not reflected by their everyday behaviour. Similarly, they were able to offer an explanation for the puzzling phenomenon that while the Russian children were generally operating at a higher academic level than the Americans, their belief in their abilities and competencies was considerably lower. Of course, this is an issue that has attracted much attention in relation to East Asian settings (Zhao, 2017).

It was immediately apparent that the demands placed upon the Russian children were greater than those of their US peers (see Alexander, 2000, for a similar finding). It is hardly surprising that if academic demand is low in a given educational system, its children will be more likely to believe that they are good at their studies (see Elliott et al., 2005, pp. 81-92 for full discussion of research in this area). Notions of what constituted 'hard work' varied significantly across the three settings, a phenomenon that the authors attributed to the particularly challenging educational environment encountered by Russian children. Thus, in responding to a survey of 3,234 14-15 year olds, 64% (Kentucky), 66% (Sunderland) and 39% (St. Petersburg) of the respondents reported that they usually worked hard in class. This discrepancy bore little resemblance to the researchers' classroom observations, student and teacher interview data, or the wider literature, giving the impression of a significant disjuncture between students' reported perceptions in the survey and all other information obtained from a variety of other sources. An even larger discrepancy was found when student satisfaction with their level of school achievement was sought. Again, in comparison with the Russian students, the students from Kentucky and Sunderland seemed far more satisfied with their current performance, and expressed less confidence that they could improve upon this.

One contributory factor may have been the greater tendency of Russian teachers to communicate negative messages to children about their performance that did not reflect what, in actuality, was their more positive perceptions. This appears to have led the children to underestimate their teachers' true views of them. The reverse was found in Kentucky where children's opinions of their teachers' positive views were more likely to be unrealistically inflated. Such messages, underpinned by skewed grading systems, reflected broader cultural understandings about the importance of praise and broader child socialisation practices (Damon, 1995; Muckle, 1990, Alexander, 2000). In his cross-cultural study of primary education, Alexander (2000) referred to a Russian teacher who, when informed about a poster, '100 ways to praise a child', that was regularly found in US schools, noted that while there were only a handful of praise descriptors in Russian, "… the vocabulary of disapproval is rich and varied" (p. 375). A similar phenomenon was reported in the former East Germany (Marsh et al., 2001) where more negative self-appraisals by students were seemingly influenced by teachers' public and detailed feedback on performance. It was interesting to note that while each of these settings could hardly be considered to be representative of their nation

state, many of the factors that differentiated them from one another were highly reflective of similar findings from other studies undertaken in the same country.

The findings of Elliott et al.'s study demonstrate the value of drawing upon large scale survey data, but also highlight the difficulty of ensuring that cross cultural measures involving student self-report are designed and utilised in an effective and meaningful fashion. Measures such as those used for PISA assume that among many things, firstly, pupils are answering questions that are universally understood in the same way and secondly, that reported self-perceptions are calibrated on a scale that is commonly understood by respondents in all societies. While data from these have potential value, particularly where improved methodological techniques can be deployed to minimise response bias (see also Stankov, Lee, & von Davier, 2018 for criticism of studies claiming to correct response bias), meaningful understanding about educational practices may only be reached when an ecosystemic examination of the respondents in their various milieux is undertaken. Without a fully detailed account, we run the risk of arriving at highly questionable conclusions and, potentially, inappropriate policy formulations. A key problem, however, is that organisations such as the O.E.C.D. seem reluctant to draw upon such data in their analyses and recommendations, leaving the field open for the over-simplified accounts of the media and politicians. Perhaps looking more closely at negative side effects (Zhao, 2017) may be perceived as undermining an enterprise that relies upon positive messages, universalism, the denial of complexity, the ready identification, and promotion of so-called best practices (Grey & Morris, 2018).

Conclusion

The analyses reported in this paper have been reported as means to demonstrate the potential contribution of large education datasets. As we have sought to show with reference to societies heavily influenced by Confucianism, insights from these have the potential to make a valuable contribution to a multi-level ecosystemic analysis; they can serve as one contributory part of a much richer whole.

Humanity's big questions are increasingly being addressed by means of 'Big Data' - vast datasets analysed by advanced analytic methods and powered by high performance computing. Such analyses will surely help us understand many of our most challenging questions, from the origins of the universe to the nature of the genome, and perhaps can be extended to the field of education (Williamson, 2017) yet, at the same time, we need to recognise that it could also distort our focus and understanding of multi-layered human ecosystems (Sobe, 2018). While Sobe (2018) contends that large-scale assessments such as TIMSS, PIRLS and PISA cannot yet be considered as Big Data, we must recognise the dangers of creeping towards an undue reliance upon the computational and algorithmic. Even if we accept that the numbers derived from measures are meaningful (and, as we have sought to show here, this may often not be the case), "… we are stuck with *what is* rather than *what should be* or *what could be"* (Sobe, 2018, 330).

Data from large scale assessments and surveys have the potential to inform and enrich consideration of factors that influence educational and social outcomes. However, researchers working in the field of comparative education, rather more than cross-cultural psychologists, have often sought to challenge the role of such approaches on the grounds that these fail to take into consideration important contextual factors (Elliott, 2014). As this paper has sought to show, there are also important questions as to whether self-report items are understood in a similar fashion in different cultures. Such concerns may mean that the potential of large datasets to contribute to the field of comparative education is overlooked, and a Manichean divide between opposing methodological camps, each dismissive of the other, is the unfortunate result. Surely, this is a situation that we should seek to challenge?

References

Adamson, B., Forestier, K., Morris, P., and C. Han. 2017. PISA, policymaking and political pantomime: education policy referencing between England and Hong Kong, *Comparative Education, 53*: 192-208.

Alexander, R. J. 2000. *Culture and pedagogy: International Comparisons in Primary Education*. Oxford: Blackwell Publishing.

Auld, E., and P. Morris, P. 2016. PISA, Policy and Persuasion: Translating Complex Conditions Into Education 'Best Practice'. *Comparative Education*, *52*(2): 202-229.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., and T. A. Smith. 1996. *Mathematics Achievement in the Middle School Years. IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Biggs, J. B. 1996. Western misconceptions of the Confucian-heritage learning culture. In D. A. Watkins and J. B. Biggs (Eds.), *The Chinese Learner: Cultural, Psychological and Contextual Influences* (pp. 45–67). Hong Kong: CERC and ACER.

Breakspear, S. 2012. The policy impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance. *OECD Education Working Papers, 71.* 0_1.

Bronfenbrenner, U. 1979. *The Ecology of Human Development*. Cambridge, MA: Harvard University Press.

Burghes, D. 1996. *Education Across the World: The Kassel Project*. Paper presented at conference "Lessons from Abroad: Learning from international experience in education" (London).

Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W., and A. Zakharov. 2016. Revisiting the relationship between international assessment outcomes and educational production: Evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal 53*(4): 1054–1085.

Carnoy, M., Marotta, L., Louzano, P., Khavenson, T., Guimarães, F. R. F., and F. Carnauba. 2017. Intranational Comparative Education: What State Differences in Student Achievement Can Teach Us about Improving Education—the Case of Brazil. *Comparative Education Review 61*(4): 726-759.

Chao, R. K., and V. Tseng. 2002. Parenting of Asians. In M. H. Bornstein (Ed.), *Handbook of parenting: vol. 4: Social Conditions and Applied Parenting* (2nd ed., pp.59-93). Mahwah: Lawrence Erlbaum.

Cheung, K. C., Mak, S. K., and P. S. Sit. 2018. Resolving the Attitude–Achievement Paradox Based on Anchoring Vignettes: Evidences from the PISA 2012 Mathematics Study. *Asia Pacific Education Review*, 1-11.

Christoph, D., and A. K. F. Li. 1985. Cognitive versus social rigidity in old age: Implications for therapy. *Canadian Journal on Aging 4*: 59–65.

Clotfelter, C. T., Ladd, H. F., and J.L. Vigdor. 2006. Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources 41*: 778–820.

Cowen, R. 2014. Ways of Knowing, Outcomes and 'Comparative Education': Be Careful What You Pray For. *Comparative Education 50*(3): 282-301.

Crossley, M. 2014. Global League Tables, Big Data and the International Transfer of Educational Research Modalities. *Comparative Education*, *50*(1): 15-26.

Damon, W. 1995. *Greater Expectations: Overcoming the culture of Indulgence in America's homes and Schools*. New Jersey The Free Press.

Elliott, J., Hufton, N., Willis, W., and L. Illushin. 2005. *Motivation, Engagement and Educational Performance: International Perspectives on the Contexts for Learning*. Dordrecht, The Netherlands: Springer.

Elliott, J.G. 2002. Could do better? The risks of cultivating positive self-esteem, *Human Givens: Radical Psychology Today* 9(1): 38-43

Elliott, J. G. 2014. Lessons from abroad; Whatever happened to pedagogy? *Comparative Education 50*: 27-44.

Elliott, J. and M. Nguyen. 2008. Western influences on the East: Lessons for the East and West. In T. Oon Seng and D. M. McInerney (Eds.) *Research on Multicultural Education and International Perspectives Vol. 7. What the West Can Learn from the East: Asian Perspectives on the Psychology of Learning and Motivation*. Greenwich, CT.: Information Age Publishing

Gelfand, M. J., et al. 2011. Differences between tight and loose cultures: A 33-nations study. *Science 332*: 1100–1104.

Goodstein, L. D. 1953. Intellectual rigidity and social attitudes. *Journal of Abnormal and Social Psychology 48*: 345-53.

Grey, S. and P.Morris. 2018. PISA: Multiple 'truths' and mediatised global governance. *Comparative Education 54*(2): 109-131.

Hattie, J. 2009. Visible learning: A Synthesis of Meta-analyses in Education. London: Routledge.

Hau, K-T., and F.Salili. 1996. Prediction of academic performance among Chinese students: Effort can compensate for lack of ability. *Organizational Behavior and Human Decision Processes 65*: 83–94.

Higgins, L. T., and C. H. Sun. 2002. The Development of Psychological Testing in China. *International Journal of Psychology 37*(4): 246-254.

Hofstede, G. H. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations* (2nd ed.). Thousand Oaks, CA: Sage.

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., and J.A. Baird. 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.

Jensen, B., Hunter, A., Sonneman, J. and T. Burns. 2012. *Catching Up: Learning from the Best School Systems in East Asia*. Melbourne: Grattan Institute.

Kawanaka, T., Stigler, J. W., and J .Hiebert. 1999. Studying mathematics classrooms in Germany, Japan and the United States: Lessons from the TIMSS videotape study. *International comparisons in mathematics education 11*: 86.

Kim, T. 2009. Confucianism, Modernities and Knowledge in *The International Handbook of Comparative Education,* edited by R. Cowen and A. Kazamias, 857-872. Dordrecht, Netherlands: Springer.

Komatsu, H., and J. Rappleye. 2017. A PISA paradox? An alternative theory of learning as a possible solution for variations in PISA scores. *Comparative Education Review 61*(2): 269-297.

Komatsu, H., and J. Rappleye. 2018. Is exam hell the cause of high academic achievement in East Asia? The case of Japan and the case for transcending stereotypes. *British Educational Research Journal*.

Lam, C. M. 2003. Covert parental control: Parent–adolescent interaction and adolescent development in a Chinese context. *International Journal of Adolescent Medicine and Health 15*: 63–77.

Lam, C. M. 2005. Chinese construction of adolescent development outcome: Themes discerned in a qualitative study. *Child and Adolescent Social Work Journal 22*: 111–131.

Lau, D. C. 1995. *Analects*. Mineola, NY: Dover Publications.

Lee, J. 2009. Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Difference 19*: 355–365.

Lee, J. 2014. Universal factors of student achievement in high-performing Eastern and Western countries. *Journal of Educational Psychology 106*: 364-374. doi: 10.1037/a0035609

Lee, J. 2016. Attitude toward school does not predict academic achievement. *Learning and Individual Differences 52*: 1-9.

Lee, J., and K. Chung. (under review). Teacher-student relationships in Confucian countries/economies: Evidence from PISA 2012.

Lee, J., and V. Shute. 2010. Personal and social-contextual factors in K-12 academic performance: An integrative perspective on student learning. *Educational Psychologist 45*: 185-202.

Lee, J., and L. Stankov. 2013. A higher-order factorial structure among fifteen major educational-psychological constructs. *Learning and Individual Differences 26*: 119-130.

Lee, J., and L. Stankov. 2018. Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences 65:* 50-64.

Leung, F. K. S. 2001. In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics 47*: 35-52.

Leung, F. K. S. 2002. Behind the high achievement of East Asian students. *Educational Research and Evaluation 8*: 87-108.

Leung, F. K. S. 2010. Beliefs in Chinese Culture. In M. H. Bond (Ed.), *The Oxford handbook of Chinese psychology* (pp. 221-240). Oxford, United Kingdom: Oxford University Press.

Leung, F. K. S. 2014. What Can and Should We Learn From International Studies of Mathematics Achievement? *Mathematics Education Research Journal 26*(3): 579–605.

Lewis, S., and B. Lingard. 2015. The multiple effects of international large-scale assessment on education policy and research. *Discourse: Studies in the Cultural Politics of Education 36:* 621-637.

Lynn, R., and G.Meisenberg. 2010. National IQs calculated and validated for 108 nations, *Intelligence 38*: 353–360.

Luxton, R. G., and G. Last. 1997. *Under-achievement and pedagogy: Experimental reforms in the teaching of mathematics using continental approaches in schools in the London borough of Barking and Dagenham*. National Institute of Economic and Social Research.

Marsh, H. W., Köller, O., and J. Baumert. 2001. Reunification of East and West German School Systems: Longitudinal Multilevel Modeling Study of the Big-fish-little-pond effect on academic self-concept. *American Educational Research Journal*, *38*(2), 321-350.

Meyer, H. D., and A. Benavot. (Eds.). 2013, May. PISA, power, and policy: The Emergence of Global Educational Governance. Oxford: Symposium Books Ltd.

Morsy, L., Khavenson, T., and M Carnoy. 2018. How international tests fail to inform policy: The unsolved mystery of Australia's steady decline in PISA scores. *International Journal of Educational Development 60*: 60-79.

Muckle, J. 1990. *Portrait of a Soviet School Under Glasnost*. London: Macmillan

Mullis, I. V., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., ... and T.A.Smith.2000. *TIMSS 1999 International Mathematics Report*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberber, E., Arora, A., and J.Galia. 2008. *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

Niu, W. 2007. Western influences on Chinese educational testing. *Comparative Education 43*(1): 71-91.

Nuyen, A. T. 2001. Confucianism and the idea of equality. *Asian Philosophy 11*: 61–71.

OECD 2004. *Learning for Tomorrow's World: First results from PISA 2003*, Paris: OECD Publishing.

OECD 2010. *Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II)*, Paris: OECD Publishing.

OECD 2013a. *Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (volume II)*, Paris: OECD Publishing.

OECD 2013b. PISA 2012 Results: What Students Know and Can Do (Vol. 1). Paris: OECD Publishing

Pace, J. L., and A. Hemmings. 2007. Understanding authority in classrooms: A review of theory, ideology, and research. *Review of Educational Research 77*: 4-27.

Peng, S. S., and D. Wright.1994. Explanation of academic achievement of Asian American students. *Journal of Educational Research 87*: 346–352.

Pipere, A., and I. Mieriņa. 2017. Exploring non-cognitive predictors of mathematics achievement among 9th grade students. *Learning and Individual Differences 59*: 65-77.

Rappa, A. L., and S.H. Tan. 2003. Political Implications of Confucian Familism. *Asian Philosophy 13*: 87–102.

Reynolds, D., and S. Farrell, S. 1996. Worlds apart? A review of international studies of educational achievement involving England. London: HMSO for OFSTED.

Richardson, M., Abraham, C., and R. Bond. 2012. Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2): 353-387.

Sadler, M. E. 1900. *How Far Can We Learn Anything of Practical Value From the Study of Foreign Systems of Education?*. Printed at the Surrey Advertiser Office.

Schleicher, A. 2009. Securing quality and equity in education: Lessons from PISA. *Prospects 39*: 251-263.

Schmidt, W. H., McKnight, C., Valverde, G.A., Houang, R.T., and D.E.Wiley. 1997. *Many Visions, Many Aims, Volume I: A Cross-national Investigation of Curricular Intentions in School Mathematics*. Dordrecht/Boston/London: Kluwer.

Schneider, B., and Y. Lee. 1990. A model for academic success: The school and home environment of East Asian students. *Anthropology and Education Quarterly 21*: 358 – 377.

Sellar, S., and B. Lingard 2014. The OECD and the Expansion of PISA: New Global Modes of Governance in Education. *British Educational Research Journal*, *40*(6): 917-936.

Shek, D. T. L., and L.K.Chan. 1999. Hong Kong Chinese parents' perceptions of the ideal child. *Journal of Psychology 133*: 291-302.

Sobe, N. 2018. Problematizing comparison in a post-exploration age: Big data, educational knowledge, and the art of criss-crossing. *Comparative Education Review,*

Stankov, L. 2010. Unforgiving Confucian culture: A breeding ground for high achievement, anxiety and self-doubt? *Learning and Individual Differences 20*: 555–563.

Stankov, L. 2013. Depression and life satisfaction among European and Confucian adolescents. *Psychological Assessment 25*: 1220-34.

Stankov, L. 2017. Conservative Syndrome: Individual and cross-cultural differences. *Journal of Cross-cultural Psychology 48*:1–11. doi: 10.1177/0022022117709984

Stankov, L. 2018. Psychological processes common to social conservatism and terrorism. *Personality and Individual Differences 120*: 75–80.

Stankov, L., and J. Lee. 2018. Conservative syndrome and the understanding of negative correlations between religiosity and cognitive abilities. *Personality and Individual Differences 131*: 21–25.

Stankov, L., and J. Lee. 2009. Dimensions of cultural differences: Pancultural, etic/emic, and ecological approaches. *Learning and Individual Differences 19*: 339-354.

Stankov, L., and J. Lee. 2014. Quest for the best noncognitive predictor of academic achievement. *Educational Psychology: An international journal of experimental educational psychology 34*: 1-8.

Stankov, L., and J. Lee. 2016. Nastiness, morality and religiosity in 33 nations. *Personality and Individual Differences 99*: 56-66.

Stankov, L., and J. Lee. 2017. Self-beliefs: Strong correlates of mathematics achievement and intelligence. *Intelligence 61*: 11-16.

Stankov, L., Saucier, G., and G. Knežević. 2010 Militant extremist mindset: Pro- violence, vile world, and divine power. *Psychological Assessment 22*: 70-86.

Stankov, L., Lee, J., and F. J. R. van de Vijver. 2014. Two dimensions of psychological country-level differences: Conservatism/Liberalism and Harshness/Softness. *Learning and Individual Differences 30*: 22–33.

Stankov, L., Lee, J., and M. von Davier. 2018. A note on construct validity of the anchoring method in PISA 2012. *Journal of Psychoeducational Assessment*, *36*: 709-724.

Stevenson, H. W., and J.W.Stigler. 1992. *The Learning Gap: Why our Schools are Failing and what we can Learn from Japanese and Chinese Education*. New York: Simon and Schuster.

Stevenson, H. W., and S.Y. Lee.1990. Contexts of achievement. *Monographs of the Society for Research in Child Development 55* (Serial No. 221).

Stigler, J. W., and J. Hiebert. 1999. *The Teaching Gap*. New York: Free Press.

Stoet, G., and D. C. Geary. 2017. Students in countries with higher levels of religiosity perform lower in science and mathematics. *Intelligence 62*: 71–78.

Tudge, J. R., Payir, A., Merçon-Vargas, E., Cao, H., Liang, Y., Li, J., and L. O'Brien, L. 2016. Still Misused After All These Years? A Reevaluation of the Uses of Bronfenbrenner's Bioecological Theory of Human Development. *Journal of Family Theory & Review*, *8*(4), 427-445.

Tweed, R. G., and D. R. Lehman. 2002. Learning considered within a cultural context. *American Psychologist 57*:89-99.

Vernon, P. 1982. *The Abilities and Achievement of Orientals in North America*. New York: Academic Press.

Wilkins, J. M. 2004. Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education 72*: 331–346.

Williamson, B. 2017. *Big Data in Education: The Digital Future of Learning, Policy and Practice.* London: Sage.

Yao, X. 2000. *An Introduction to Confucianism*. Cambridge: Cambridge University Press.

You, Y., and P. Morris. 2015. Imagining school autonomy in high-performing education systems: East Asia as a source of policy referencing in England. *Compare: A Journal of Comparative and International Education 46*: 882–905.

Yu, A. 1996. Ultimate Life Concern, Self, and Chinese Achievement Motivation. In M. H. Bond (Eds.), *The Handbook of Chinese Psychology* (pp. 227–246). Hong Kong: Oxford University Press.

Zhao, Y. 2017. What works may hurt: Side effects in Education. *Journal of Educational Change*, *18*(1): 1-19.

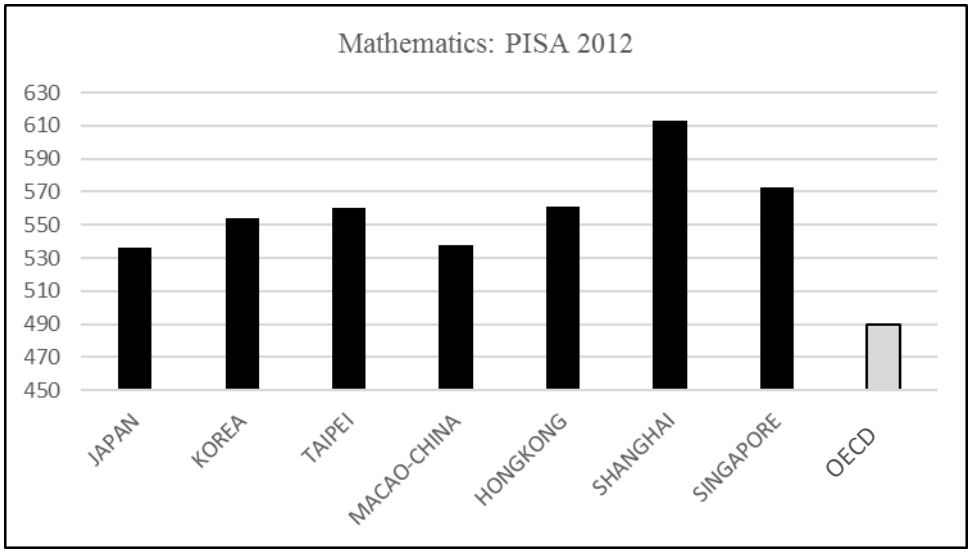

Figure 1. Country-level performance on measures of mathematics achievement in PISA 2012

Note. The "OECD" is based on 32 OECD countries (excluding Japan and South Korea).

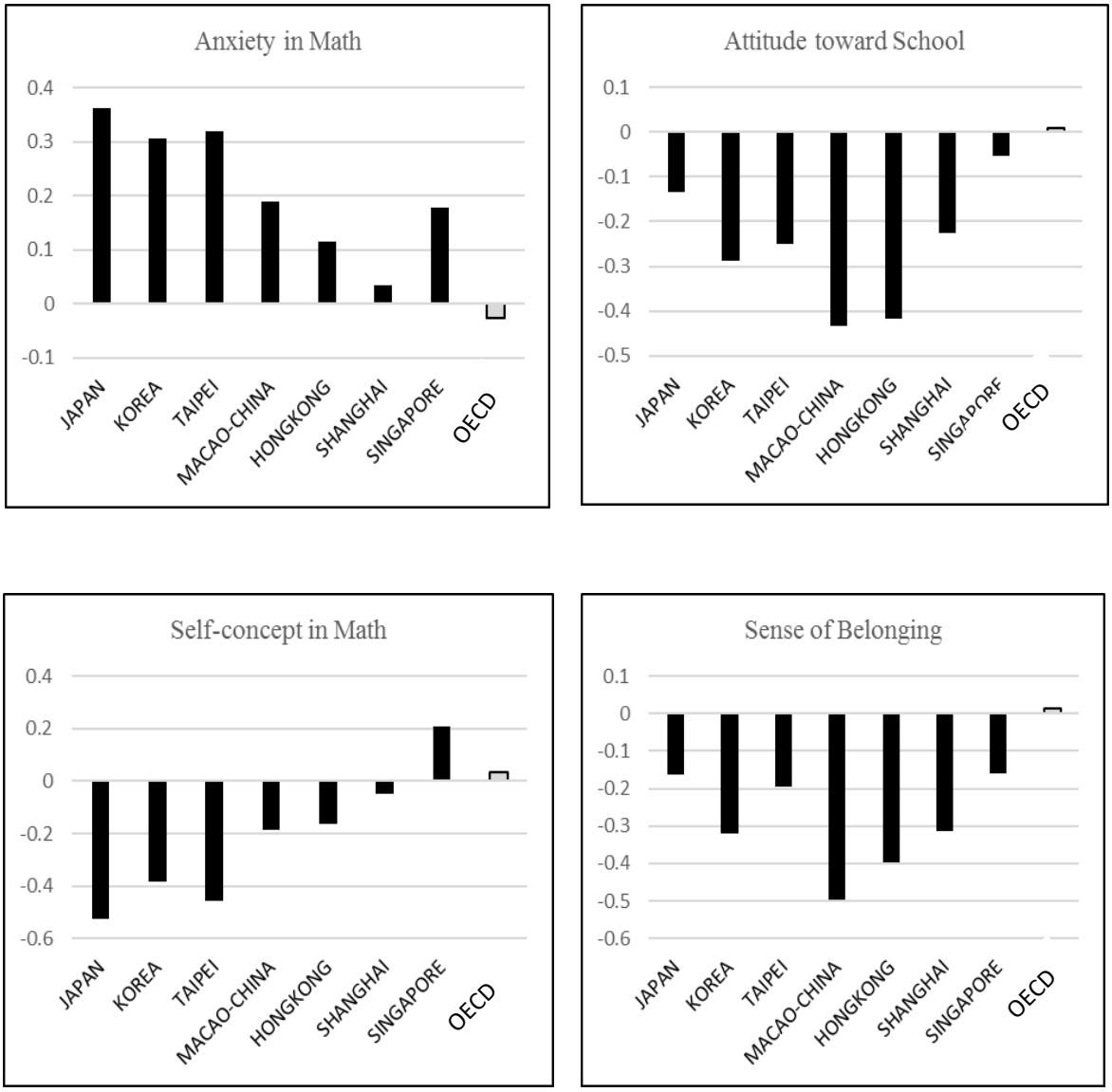

Figure 2. Country-level scores on anxiety and self-concept (left panel) and attitude toward school and sense of belonging (right panel) in PISA 2012

Note. The "OECD" is based on 32 OECD countries (excluding Japan and South Korea).

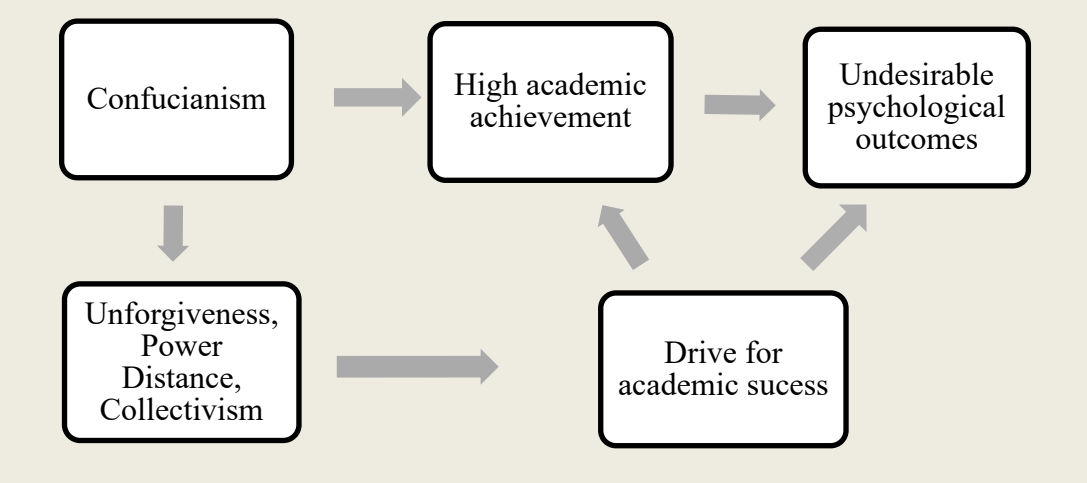

Figure 3. Factors underpinning academic achievement in Confucian countries

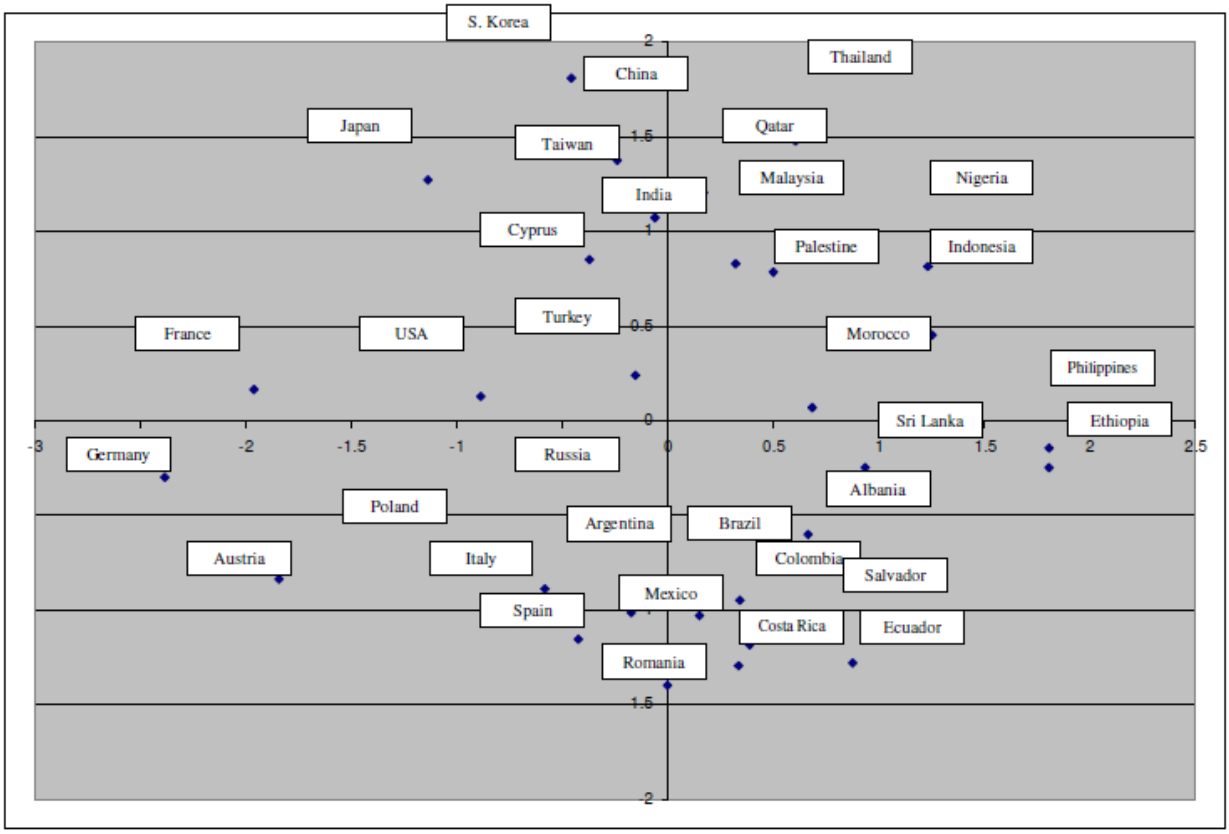

Figure 4. Countries' locations on two dimensions – broad Conservatism/Liberalism (X-axis) and Harshness/Softness (Y-axis). (From Stankov, Lee and van de Vijver, 2014)