International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Instructional practices and science performance of 10 top-performing regions in PISA 2015

Kwok-chi Lau & Terence Yuk-ping Lam

To cite this article: Kwok-chi Lau & Terence Yuk-ping Lam (2017): Instructional practices and science performance of 10 top-performing regions in PISA 2015, International Journal of Science Education, DOI: 10.1080/09500693.2017.1387947

To link to this article: http://dx.doi.org/10.1080/09500693.2017.1387947

Published online: 12 Oct 2017.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2017
https://doi.org/10.1080/09500693.2017.1387947

Routledge
Taylor & Francis Group

Check for updates

Instructional practices and science performance of 10 top-performing regions in PISA 2015

Kwok-chi Lau and Terence Yuk-ping Lam

Department of Curriculum and Instruction, The Chinese University of Hong Kong, Shatin, Hong Kong

ABSTRACT

This study analysed 10 top-performing regions in PISA 2015 on their science performances and instructional practices. The regions include Singapore, Japan, Estonia, Taipei, Finland, Macao, Canada, Hong Kong, China and Korea. The science performances of the 10 regions and their teaching practices are described and compared. The construct of enquiry-based instruction as developed in PISA 2015 is revised into two new constructs using factor analysis. Then, the relationships of the teaching practices with science performance are analysed using hierarchical linear modelling. Adaptive instruction, teacher-directed instruction and interactive application are found positively associated with performance in all regions, while investigation and perceived feedback are all negative. The regions except Japan and Korea tend to have a high frequency of teacher-directed instruction facilitated by more or less authoritative class discussion in class. A fair amount of practical work is done, but not many of them are investigations. The cultural influences on teaching practices are discussed on how an amalgam of didactic and constructivist pedagogy is created by the Western progressive educational philosophy meeting the Confucian culture. The reasons for investigation's negative association with performance are also explored.

ARTICLE HISTORY
Received 13 April 2017
Accepted 1 October 2017

KEYWORDS
PISA 2015; science teaching practices; science performance; Chinese learners

Introduction

Science was made a major domain of assessment in the last PISA in 2006. Consistent with the past findings, many Asian regions, such as Singapore, Beijing-Shanghai-Jiangsu-Guangdong (hereinafter 'China'), Hong Kong, Chinese Taipei, Macao, Japan, Korea and Vietnam, continue to rank at the top of the 2015 PISA league table. The only Western regions ranked within the top 10 are Finland, Canada and Estonia. It seems that the 'paradox of the Chinese learners and teachers' remains: Chinese learners outperform their Western counterparts despite that they are learning by rote from the authoritarian Chinese teachers using expository and didactic pedagogy (Chan & Rao, 2009; Watkins & Biggs, 2001). 'Chinese' here refers to the students and teachers who are influenced by the Confucian Heritage Culture (CHC) (Watkins & Biggs, 1996). Therefore, Japan and Korea are regarded as CHC regions in some sense. Although the paradox

CONTACT Kwok-chi Lau laukwokchi@cuhk.edu.hk, laukwokchi@gmail.com

has been debunked to some extent (Chan & Rao, 2009; Mok et al., 2001), PISA as a large-scale international study would provide valuable insight into the issue. Lam and Lau (2014) have analysed whether there exists an East Asian Model of science teaching using the PISA 2006 data. The constructs of teaching practices were largely revised for PISA 2015, which provided more information on the science teaching practices of the participating regions, and how the practices are related to science performance.

This study analyses the following 10 top-performing regions in PISA 2015: Singapore, Japan, Estonia, Chinese Taipei, Finland, Macao, Canada, Hong Kong, China and Korea. These regions have ranked within the top 11 regions for science performance. Vietnam, despite ranking eighth, is not included in the study because its assessment is paper based and thus does not include the computer-based items developed for PISA 2015. The analyses aim to reveal how top performance in science is associated with the range of teaching practices shown by the 10 regions. Although an analysis of both the top and bottom regions – an extreme group approach – may give more valid conclusions, the large differences in socioeconomic conditions between the top and bottom performers (OECD, 2016b, p. 333) may make such comparisons problematic. The 10 regions chosen in the study are by contrast more comparable since they are all advanced economies. The choice of these 10 regions would also shed light on the role of culture in education in East Asian (Japan, Korea, Singapore), Chinese (B-S-J-G, Hong Kong, Chinese Taipei and Macao), Western (Canada) and Nordic (Finland and Estonia) regions, despite that the potential conclusions reached by a small number of regions of each culture might limit their generalisability. The research questions the study aims to answer are:

1. How similar are the teaching practices and performances in the PISA 2015 assessment of science amongst the 10 top-performing regions in PISA 2015?
2. How are the science teaching practices of these 10 regions related to science performances in PISA 2015?
3. Are the science teaching practices, science performances, and their relationships associated with the cultures of these regions?

Assessment framework of scientific literacy in PISA 2015

Scientific literacy as defined in PISA 2015 is 'the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen' (OECD, 2016a, p. 20), which requires a 15-year-old student to display three competencies in the personal, local and global contexts. These competencies are affected by a student's attitudes towards science and his/her understanding of the three kinds of scientific knowledge: content, procedural and epistemic. Different kinds of knowledge are related to different competencies: content knowledge mainly supports one to explain phenomena scientifically, while procedural and epistemic knowledge help one to design scientific inquiry and to interpret data. The assessment framework of PISA 2015 is basically aligned with that of PISA 2006, but *procedural and epistemic knowledge* was not differentiated in 2006 and was collectively called *knowledge about science*. Unfortunately, the subscale of epistemic knowledge was not reported separately in PISA 2015 due to inadequate items.

Constructs of instructional practices in PISA 2015

In PISA 2015, both science teachers and students are asked about the teaching practices in science class, but the teachers' reports are deemed less reliable than those of the students, since teachers tend to overstate the positive pedagogy (OECD, 2016c, p. 63). Therefore, the study will use the data from student reports only.

PISA had developed four constructs of science instructional practices according to students' reports in the 2006 assessment. In PISA 2015, four new constructs of instructional practices were developed. Some consist of identical or similar items from 2006, while others are newly constructed (Table 1). Perceived feedback and adaptive instruction are two entirely new instructional constructs, whereas teacher-directed instruction and enquiry-based instruction have remarkable overlap with those in the 2006 assessment.

These constructs of instructional practices, however, are not without their problems. Perceived feedback asks if a teacher provides individual feedback and advice to students *in class*, but it may have missed the feedback a teacher gives *after class*. Chinese teachers tend to concentrate on whole-class teaching during class time and engage students on their academic and personal issues after class. Likewise, adaptive instruction may not happen in class but has been accounted for by the teacher during lesson planning. Even when a teacher adjusts the teaching pace, depth and breadth in class according to student responses, students may not be aware of the adaptations made by the teacher.

But what is particularly problematic is the construct of enquiry-based instruction in PISA 2015. It consists of nine items from the four constructs of teaching practices in PISA 2006, in which six items are identical and three are slightly revised. This makes

Table 1. The four constructs of instructional practices and their assessment items in PISA 2015.

How often do these things happen in your lessons for this <school science> course?
(*In all lessons/In most lessons/In some lessons/Never or hardly ever*)
Teacher-directed instruction (TD) (ST103)
(a) The teacher explains scientific ideas. (~Application 1)
(b) A whole class discussion takes place with the teacher. (~Interaction 3,4)
(c) The teacher discusses our questions.(~Interaction 2)
(d) The teacher demonstrates an idea. (~Application 1)
Enquiry-based instruction (EI) (ST098)
(a) Students are given opportunities to explain their ideas. (=Interaction 1)
(b) Students spend time in the laboratory doing practical experiments. (=Hands on 1)
(c) Students are required to argue about science questions.(~Interaction 3)
(d) Students are asked to draw conclusions from an experiment they have conducted. (=Hands on 3)
(e) The teacher explains how a <school science> idea can be applied to a number of different phenomena (e.g. the movement of objects, substances with similar properties).(=Application 1)
(f) Students are allowed to design their own experiments. (=Investigation 1)
(g) There is a class debate about investigations.(~Interaction 3, Investigation)
(h) The teacher clearly explains the relevance of <broad science> concepts to our lives. (=Application 3)
(i) Students are asked to do an investigation to test their ideas. (=Investigation 3)
Perceived feedback (PF) (ST104)
(a) The teacher tells me how I am performing in this course.
(b) The teacher gives me feedback on my strengths in this <school science> subject.
(c) The teacher tells me in which areas I can still improve.
(d) The teacher tells me how I can improve my performance.
(e) The teacher advises me on how to reach my learning goals.
Adaptive instruction (AI) (ST107)
(a) The teacher adapts the lesson to my class's needs and knowledge.
(b) The teacher provides individual help when a student has difficulties understanding a topic or task.
(c) The teacher changes the structure of the lesson on a topic that most students find difficult to understand.

Note: Inside the brackets are the items of PISA 2006 that have the same (=) or similar (~) meaning to that item.

enquiry-based instruction an amalgam of four original constructs. Moreover, enquiry-based instruction as a new construct is neither clearly defined nor well justified with theories and literature in PISA 2015. In the PISA report, it is vaguely described as 'about engaging students in experimentation and hands-on activities, and also about challenging students and encouraging them to develop a conceptual understanding of scientific ideas' (OECD, 2016c, p. 69). The first part of this description is clearly about hands-on practical work as emphasised in *Next Generation Science Standards* (NGSS, 2013), but the second part is quite ambiguous. In the assessment framework of PISA 2015 (OECD, 2016a, p. 112), enquiry-based teaching is conceptualised in various ways, including science teaching and learning in contexts that are real and meaningful for learners (Fensham, 1985; King & Stephen, 2012), scientific argumentation (Osborne, 2012) and students' active thinking and drawing conclusions from data (Minner, Levy, & Century, 2010). It seems that the construct of enquiry-based instruction in PISA 2015 comprises a variety of advocated pedagogies in science education without a clear focus.

Given that there is no precise operational definition of inquiry teaching in the literature (Anderson, 2002), PISA has to provide a clear definition of it before it is used in the framework. Besides, it is not desirable for a construct consisting of heterogeneous items. In the PISA 2015 report (OECD, 2016c, p. 71), the nine items of enquiry-based teaching are found to have very different or even opposite, associations with science scores. Given all these empirical and theoretical problems with the construct of enquiry-based instruction, we will conduct factor analysis to revise it before HLM analysis is conducted.

Methods

Sample and data

A nationally representative sample of 15-year-old students in the selected regions participated in PISA 2015. OECD first sampled schools within a participating region, and then sampled students from the selected schools to represent the 15-year-olds in that region. The sampled students completed the two-hour cognitive test, and a student questionnaire of 30–40 minutes. OECD then weighted the participants' test scores and variables to represent the schools and the 15-year-old student populations. (For sampling details, see OECD, 2016d.)

This study examined the data from student tests in science, the student questionnaire and the school questionnaire. The participants were those who completed all questions in the student questionnaire and student tests related to this study. The final sample contained about 4000–20,000 students.

HLM analyses

Variables

The cognitive component of scientific literacy as measured in PISA 2015 is the dependent variable in this study. Using the Item Response Theory and Plausible Value methodology, 10 plausible values were generated for each student to represent their science performance. The reliability coefficient for the plausible value of science knowledge scale is 0.90–0.93 for the 10 regions in this study (OECD, 2016d), indicating a satisfactory reliability of the measure.

The independent variables included in these analyses are drawn from *students* (*background and* teaching practices variables) and *schools.* All items used in the questionnaire to measure these variables were carefully developed through construct validation and item dimensionality by OECD with a calibration sample of 500 students per region (OECD, 2016d).

Student background variables

Student gender (being female) is included in the multilevel models in the study, with negative values indicating lower scores for female students. Parent SES, defined as *family economic, social, and cultural status* (ESCS) in PISA, is calculated from the indices of parents' occupational status, years of education completed and home possessions (for details, refer to the 'Technical report' of PISA 2015). *Immigration status* is based on the place of birth of the student and their parents, which is reflected by three dummy variables: *1st generation*, *2nd generation* and *native*. The grade level of students is also included in the analyses as it was found mediating other variables like immigration status (Pong, 2009).

Teaching practices variables

The construct of enquiry-based teaching as developed in PISA 2015 was first revised using EFA and CFA into two new constructs: interactive investigation and interactive application. In addition to the original three teaching practice constructs, a total of five teaching practices variables in PISA 2015 were included in this study: teacher-directed teaching, adaptive instruction, perceived feedback, interactive investigation and interactive application.

To revise the construct of enquiry-based teaching, the data for Hong Kong were first used to explore different models by exploratory factor analysis (EFA) (Table 2). As shown by EFA (A), the construct of enquiry-based instruction is bi-dimensional with three items (d,b,i) having high cross loadings. When the EFA is run again with only the five items having no cross loadings (g,f,e,h,a) and one item with small cross loading (c), most of the cross loadings disappeared and a higher % of variance is explained by the model. Based on the results of EFA, the six items were further analysed using confirmatory factor analysis (CFA) with the data of Hong Kong. One model that consists of two factors each loading on three items was found satisfactory (Figure 1). The fit indices show that this model fits with the data fairly well: RMSEA < 0.08 (MacCallum, Browne, & Sugawara, 1996), CFI > 0.90 (Hu & Bentler, 1999), NNFI > 0.95 (Hu & Bentler, 1999), and GFI and AGFI > 0.95 (Miles & Shevlin, 1998). Given the large sample size, the significant *Chi*-square value is not a reason to reject the model (Jöreskog & Sörbom, 1993).

To further check if this model is valid for the other nine regions, CFA was done and the fit indices of all the regions are found satisfactory (Table 3).

The original construct of enquiry-based instruction was revised into two new constructs of teaching practices. The first of which is *interactive investigation*, which engages students to design and argue about scientific investigation and science questions; and the second being *interactive application*, in which science is taught in relevance to everyday applications through interaction with student ideas (Table 4). These two constructs of teaching practices are more coherent in meaning and are highly consistent amongst the top 10 regions in their associations with science scores: investigation is negatively associated with scores while application is all positive (Table 8).

Table 2. EFA of the eight items of the construct of enquiry-based Instruction in PISA 2015.

	Factor loading	
	1	2
(A) all items		
e	.856	
h	.798	
a	.715	
d	.662	.494
b	.611	.500
g		.863
f		.838
c	.327	.767
i	.566	.616
Total variance explained after rotation	35.60%	33.80%
Eigenvalue	3.12	2.92
(B) six selected items		
g	.875	
f	.845	
c	.788	.322
e		.866
h		.824
a		.734
Total variance explained after rotation	37.78%	36.09%
Eigenvalue	2.11	2.07

School-level variables

School-level variables used in this study are *school average SES*, boy school and girl school. School average SES is obtained from the aggregation of parent SES in the student level for each school. The two dummy variables, boy school and girl school, are constructed from school questionnaire, and their effects are relative to a co-ed school.

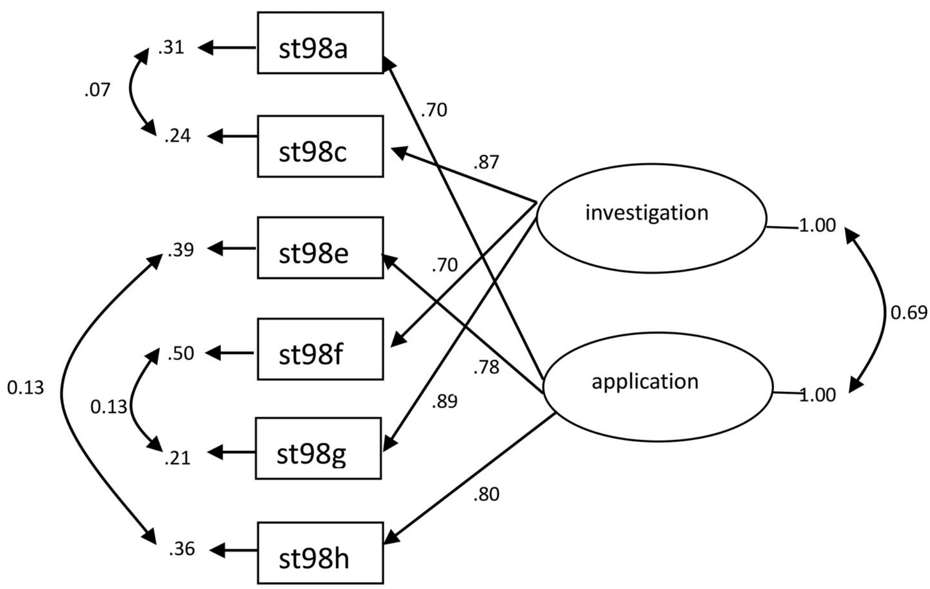

Figure 1. CFA of six items from enquiry-based instruction.

Table 3. CFA model fit indices of the model in Figure 1 for the 10 regions.

	RMSEA	CFI	NNFI	GFI	AGFI
Hong Kong	0.0756	0.994	0.981	0.99	0.958
China (BSJG)	0.075	0.994	0.983	0.99	0.96
Japan	0.0349	0.999	0.996	0.998	0.99
Korea	0.05	0.997	0.992	0.995	0.981
Singapore	0.065	0.993	0.982	0.991	0.969
Taipei	0.0474	0.997	0.993	0.995	0.983
Macao	0.058	0.993	0.982	0.994	0.975
Canada	0.08	0.99	0.975	0.987	0.955
Estonia	0.07	0.989	0.975	0.988	0.964
Finland	0.0593	0.992	0.983	0.991	0.974

Table 4. Two new constructs of teaching practices as revised from enquiry-based instruction of PISA 2015.

Interactive investigation
c. Students are required to argue about science questions.
f. Students are allowed to design their own experiments.
g. There is a class debate about investigations.
Interactive application
a. Students are given opportunities to explain their ideas.
e. The teacher explains <school science> idea can be applied.
h. The teacher clearly explains relevance <broad science> concepts to our lives.

Statistical models and analysis

Hierarchical linear modelling (HLM; see Bryk & Raudenbush, 1992) was used to investigate the associations of teaching practices with students' scientific literacy performance after controlling for background variables. Background variables such as student gender, parent SES and school-mean parent SES are used in this study. In PISA 2015, 10 PV estimates are generated for each student to present students' achievement (OECD, 2016d) and are handled by the software package HLM 6 (Raudenbush, Bryk, Cheong, & Congdon, 2004). The data file is weighted at the student level with 'normalized student final weights'.

The analysis was divided into three parts: (1) HLM analysis of the variation of students' scientific literacy performance amongst schools; (2) HLM analysis of the effects of student background characteristics and school contextual factors on students' scientific literacy performance; and (3) HLM analysis of the effects of teaching practices after controlling for student's background characteristics and school contextual factors.

First, a null model was used to partition the variance of literacy performance into within- and between-school portions. The model is represented by

$$\mathrm{Scie}_{ij} = B_{0j} + R_{ij}, \tag{1}$$

$$B_{0j} = G_{00} + U_{0j}, \tag{2}$$

where Scie_j is the score of science of student i in school j. B_{0j} is the average science score of school j without any adjustment, and G_{00} is the grand mean of science score. The variance (R_{ij}) is the within-school variance of science score, and the variance (U_{0j}) is the between-school variance.

The second model builds on the null model by adding the student background factors and the school contextual factors. This model, Model 1, examines the effect of the student background and the school background on students' science performance. It is represented by

$$\mathrm{Scie}_{ij} = B_{0j} + B_{1j}(\mathrm{girl}) + B_{2j}\,(\mathrm{parent\,SES}) + B_{3j}(\mathrm{1st\,generation}) + B_{4j}(\mathrm{2nd\,generation}) + B_{5j}(\mathrm{grade\,level}) + R_{ij}, \quad (3)$$

$$B_{0j} = G_{00} + G_{01}(\mathrm{school\,mean\,SES}) + G_{02}(\mathrm{girl\,school}) + G_{03}(\mathrm{boy\,school}) + U_{0j}. \quad (4)$$

Finally, the five types of teaching practices were then included to construct Model 2. The equations for Models 2 are similar to that of Equations (3) and (4) as shown above.

Results

Science performances

Table 5 shows the science performance of the 10 regions in PISA 2015. Singapore has a much stronger performance than all other regions by having an overall score of 556, which is 18 points higher than that of the region ranking second, Japan. Singapore not only has the highest proportion of top students but also a small proportion of low performers, indicating that its science education can cater for the needs of both the academically strong and weak students. China and Korea have large proportions of low performers *and* top students, showing a great disparity in performance. Hong Kong and Macao, on the contrary, have small proportions of both the top and low students, indicating fair performances from most of their students. The proportions of top and bottom performers show a region's equity of education as well as how the needs of gifted and students with special education needs are being catered for.

In addition to the total scores, it is important to examine a region's performance on different competencies and knowledge. Although Singapore stands out in all areas amongst the 10 regions, it performs stronger particularly in the areas of *evaluate and design scientific enquiry* (ED) and *procedural and epistemic knowledge* (PEK). The three Western regions, Finland, Canada and Estonia, also perform better in ED than their Asian counterparts. Japan seems to be the only Asian region other than Singapore to perform better relatively in ED. This might reflect a stronger emphasis on scientific inquiry in these regions. On the other hand, the four Chinese regions, Hong Kong, Macao, Chinese Taipei and China, are stronger relatively in the areas connected with science contents: *explaining phenomena scientifically* (EP) and *content knowledge* (CK). Singapore and Japan have done equally well in all areas. The relative performances in different areas might reflect the different emphases of science education of a region amongst science content, scientific inquiry and the nature of science. Singapore and Japan are equally good at all areas.

Instructional practices

The scores of the four indices of the teaching practices reported in PISA 2015 are shown in Table 6. Since the teaching practices are self-reported by students on their prevalence in

Table 5. Mean total science scores and mean scores of different competencies and kinds of scientific knowledge of the 10 regions in PISA 2015 (OECD, 2016b).

Regions	Hong Kong	Macao	China	Taipei	Singapore	Japan	Korea	Canada	Finland	Estonia
Rank	9	6	10	4	1	2	11	7	5	3
Overall science score	523	529	518	532	556	538	516	528	531	534
% low/top performers[a]	9.4/7.4	8.1/9.2	16.2/13.6	12.4/15.4	9.6/24.2	9.6/15.3	14.4/10.6	11.1/12.4	11.5/14.3	8.8/13.5
Scientific competency										
Explaining phenomena scientifically (EP)	524	528	520	536*	553	539	510	530*	534*	533
Evaluate and design scientific enquiry (ED)	524	525	517	525	560*	536	515*	530*	529	535
Interpret data and evidence scientifically (ID)	521	532*	516	533*	556	541*	523*	525	529	537
Scientific knowledge										
content knowledge (CK)	526*	527	520*	538*	553	539	513	528	534*	534
Procedural and epistemic knowledge (PEK)	521	531*	516	528	558*	538	519*	528	528	535

[a]% of students below level 2 (low) or at level 5 or above (top).
*Significantly higher than other competency or knowledge.

Table 6. Mean values of the 4 OECD constructs of instructional practices of the 10 regions in PISA 2015.

	HK Index (SD)	Macao	China	Taipei	Singapore	Japan	Korea	Canada	Finland	Estonia
Teacher directed instruction	0.11 (0.90)	−0.03 (0.84)	0.01 (1.02)	0.17 (1.01)	0.27 (0.95)	−0.21 (0.90)	−0.59 (1.05)	0.37 (1.06)	0.23 (0.94)	−0.05 (0.91)
Adaptive instruction	0.08 (0.93)	−0.12 (0.88)	0.06 (0.92)	0.03 (0.96)	0.41 (0.90)	−0.24 (0.97)	−0.05 (1.01)	0.26 (1.01)	−0.01 (0.92)	−0.17 (0.92)
Enquiry-based instruction	0.10 (0.98)	−0.16 (0.79)	−0.28 (1.10)	−0.45 (1.11)	0.01 (0.85)	−0.64 (1.09)	−0.61 (1.16)	0.27 (0.97)	−0.30 (0.86)	−0.07 (0.83)
Perceived feedback	0.16 (0.92)	−0.12 (0.88)	0.25 (0.93)	0.24 (0.99)	0.31 (0.93)	−0.36 (0.95)	−0.37 (1.05)	0.21 (1.00)	−0.27 (0.92)	−0.08 (0.93)

Source: OECD (2016c). PISA 2015 *Results (Volume II): Policies and practices for successful schools*, Table II.2.17, Table II.2.20, Table II.2.23, Table II.2.27.

Note: The indices are constructed in such a way that the mean of the combined student population from participating OECD countries is set to 0 and the standard deviation is set to 1. Two-thirds of the OECD student population has the values between −1 and 1. A negative value indicates that a group of students responded less positively than all students did, on average, across OECD countries. Likewise, a positive value on an index indicates that a group of students responded more favourably, or more positively, than all students did, on average, across OECD countries.

classroom, comparing these scores across regions needs to account for the variations in student characteristics and culture. Teacher-directed teaching is much more common in the science classroom of Canada, Singapore and Finland, but it is lower than the OECD mean in Japan and Korea. Canada and Singapore also have more adaptive instruction, which means that science teaching in their classrooms, while largely directed by the teacher, is often adapted to the needs of students. Many more students in Singapore, China, Chinese Taipei and Canada reported that teachers give them individual feedback in class compared to the students in Japan, Korea and Finland. For enquiry-based instruction, most regions – Japan, Korea, Chinese Taipei, China and Finland in particular – are below the OECD mean. This teaching practice is most commonly used in Canada, and it is the only region in the study positioned well above the OCED mean.

A closer look at the individual items would reveal greater detail about the teaching practices of the regions. In Table 7, six items are selected to show the mean scores of the

Table 7. Mean scores of the selected items of the constructs of teaching practices of the 10 regions in PISA 2015.

Items	Hong Kong	Macao	China	Taipei	Singapore	Japan	Korea	Canada	Finland	Estonia	OECD mean
1. The teacher explains scientific ideas.	2.90	2.74	2.75	2.77	2.92	2.51	2.12	3.01	2.90	2.53	2.67
2. A whole-class discussion takes place with the teacher.	2.42	2.34	2.55	2.54	2.45	1.69	1.67	2.57	2.46	2.50	2.32
3. Students are given opportunities to explain their ideas.	2.69	2.61	2.97	2.73	2.89	2.47	2.36	3.08	2.97	2.93	2.94
4. The teacher clearly explains the relevance of <broad science> concepts to our lives.	2.54	2.41	2.33	2.32	2.48	2.15	2.38	2.76	2.47	2.65	2.54
5. Students spend time in the laboratory doing practical experiments	2.35	2.06	1.88	1.83	2.15	1.81	1.58	2.19	1.94	1.75	1.94
6. Students are allowed to design their own experiments.	1.83	1.57	1.79	1.55	1.61	1.46	1.54	1.84	1.28	1.56	1.63

Note; Students are asked how often these things happen in their lessons. The responses of *In all lessons/In most lessons/In some lessons/Never or hardly ever* are given scores of 4 to 1, respectively.

teaching practices as reported by the students. In 7 out of the 10 regions, explaining scientific ideas in science classroom by the teacher (Item 1) is more common than the OECD mean, particularly for Hong Kong, Singapore, Canada and Finland. But it does not mean that the teaching is one way and didactic – whole-class discussion is also found to be frequent in most of these regions (Item 2). However, the class discussion involved is more likely to be authoritative than dialogic (Mortimer & Scott, 2003) in most Asian regions in view of the small opportunities given to students to explain their ideas (Item 3). The three Western regions together with Singapore and China tend to allow more student ideas expressed in class. Canada and Estonia see a higher frequency of teachers explaining the relevance of science to daily lives (Item 4). As for the amount of practical work (Item 5), the 10 regions vary widely. Hong Kong, Singapore, Canada and Macao do more experiments than the OECD average. However, practical work may or may not engage students in designing experiments (Item 6), which is one of the main features of scientific investigation. Across the 10 regions, only Hong Kong, China and Canada engage students in experimental design more than the average amongst OECD regions. Japan and Korea are unique in their low scores in almost all items. In Japan and Korea, 14% and 27% of their students, respectively, have reported that their teachers do not explain scientific ideas in class, while more than 50% of the students reported that whole-class discussion never happens. Moreover, 17% and 21% of the students in the Japan and Korea, respectively, reported that they were never allowed to explain their ideas in class, while about 40% and 55%, respectively, reported that they never do experiments.

Apart from Japan and Korea, there seems to be a commonality amongst the science classrooms of top-performing regions. Teachers generally spend a substantial amount of class time in explaining scientific concepts, interspersed with some class discussion in which students have fair opportunities to express their ideas. Students do not particularly perceive that the teaching has been adapted to their needs. A fair amount of practical work is done, of which only some are investigations involving the design of experiments. Japan and Korea are unique for having remarkably less direct teaching, class discussion, opportunities for students to express their ideas and feedback given to students. What is shared by the three Western regions is that students have more opportunities to express their ideas in class (Table 4, Item 3), but this also holds true for China and Singapore.

HLM analysis of instructional practices with performance

The full models of HLM of each region are shown in Table 8, where the dependent variable is the scaled mean total science score of a region when the OECD mean score is set to be 500. The numbers for each independent variable represent the changes in mean total science score per unit increase of that variable.

Background variables

The female gender is negatively associated with overall science scores in all regions. In China, the mean score of girls is 22.27 points lower than boys. But Finland sees a much smaller gender inequality in science performance. The relevance of SES with scores reveals partly the educational equity of a region. All regions have seen their student SES positively associated with scores, but in the three Chinese regions, HK, Macao and

Table 8. HLM analysis on the associations of the student background and teaching practice variables with the mean total science scores of the 10 regions in PISA 2015.

	Hong Kong		Macao		China(BSJG)		Taipei		Singapore	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	532.27***	3.00	522.25***	4.45	535.36***	3.00	533.01***	2.18	556.39***	2.49
Student level										
Background										
Girl (relative to Boy)	−18.01***	2.93	−21.47***	2.63	−22.27***	2.16	−14.25***	2.31	−18.99***	2.50
SES	3.71*	1.50	4.90**	1.72	6.97***	1.35	17.40***	1.72	21.95***	1.49
2nd generation of immigrants	3.90	3.40	8.78**	2.73	−49.51	33.65	57.35*	22.23	10.77*	4.41
1st generation of immigrants	19.11***	4.04	17.13***	2.94	−83.07**	27.35	−68.69	39.16	24.17***	4.06
Grade level	33.45***	2.26	37.62***	2.24	30.96***	3.28	12.85**	3.92	47.09***	4.18
Teaching practices										
Perceived feedback	−5.85**	1.88	−8.36***	1.51	−7.53***	1.36	−4.29**	1.31	−14.31***	1.72
Adaptive instruction	4.53*	1.89	3.41*	1.39	7.60***	1.25	−1.69	1.21	14.27***	1.82
Teacher−directed instruction	8.02***	1.82	6.60***	1.95	5.01***	1.26	7.50***	1.17	7.40***	1.56
Interactive Investigation	−20.60***	1.70	−16.08***	1.61	−22.05***	1.46	−30.61***	1.55	−14.19***	1.59
Interactive Application	15.57***	1.57	12.97***	1.51	19.62***	1.47	26.46***	1.45	9.11***	1.74
School level										
School average SES	42.18***	6.54	5.65	7.94	56.82***	3.55	88.99***	6.53	61.12***	9.54
Girl school	3.93	10.53	5.24	13.19	NA		NA		12.75	11.08
Boy school	−8.30	7.66	−77.37***	18.80	NA		NA		20.16	17.42
Between-school variance explained (%)	55.3		60.0		75.3		78.8		71.3	
Within-school variance explained (%)	18.7		24.8		15.1		15.3		14.7	
Total variance explained (%)	30.3		34.9		46.5		38.6		33.7	
	Japan		Korea		Canada		Finland		Estonia	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	540.29***	2.66	521.05***	2.26	535.16***	1.38	542.66***	1.68	541.77***	2.30
Student level										
Background										
Girl (relative to boy)	−19.05***	2.32	−13.22***	3.43	−16.74***	2.02	−5.78*	2.62	−21.94***	2.69
SES	10.39***	1.74	22.22***	2.16	20.64***	1.44	26.33***	1.77	18.50***	1.86
2nd generation of immigrants	−25.25	18.38	na		.32	2.64	−39.10**	10.84	−17.91***	4.38
1st generation of immigrants	−28.23	24.84	−82.01*	37.27	−6.19*	3.15	−37.74**	10.78	−9.27	15.09
Grade level	NA		14.97**	4.94	31.89***	2.36	32.82***	3.31	28.97***	2.59
Teaching practices										
Perceived feedback	−6.81***	1.33	−14.07***	1.45	−12.08***	1.03	−16.08***	1.49	−12.19***	1.83
Adaptive instruction	4.04**	1.29	11.10***	1.52	9.84***	1.29	13.21***	1.74	10.64***	1.57

Teacher-directed instruction	6.88***	1.32	4.23**	1.47	8.39***	1.12	11.13***	1.67	4.90**	1.51
Investigation	−18.29***	1.47	−25.77***	1.73	−26.12***	1.23	−24.72***	1.57	−29.39***	1.46
Interactive Application	6.81***	1.35	12.11***	1.56	13.42***	1.23	14.08***	1.61	14.05***	1.55
School level										
School average SES	126.22***	7.66	80.25***	6.92	34.59***	4.09	26.43***	6.59	26.13***	6.50
Girl school	−38.27***	6.53	6.94	5.18	NA		NA		NA	
Boy school	−19.00	15.29	6.41	6.78	NA		NA		NA	
Between-school variance explained (%)	68.0		72.0		60.5		66.6		55.2	
Within-school variance explained (%)	7.6		12.8		17.1		22.5		19.7	
Total variance explained (%)	34.0		27.1		23.0		25.6		25.2	

*$p < .05$.
**$p < .01$.
***$p < .001$.

China, the relationships are much smaller. This, however, may not indicate that these Chinese regions are more equitable in education. Instead, the Asian regions, particularly Japan, Korea and Chinese Taipei, tend to have their school SES more strongly related to scores than the Western regions. This means that the segregation of students according to their SES happens at school admission. As a result, the effects of individual SES may have been absorbed by the clustering nature of the HLM model that partitions within-school and between-school variances (Pong, 2009). Grade level is also positively related to scores in all regions. One additional year of study is associated with a rise of over 30 points for many regions.

Students who are children of first-generation immigrant parents tend to perform poorer, except for those in Hong Kong, Macao and Singapore. In Hong Kong, when grade level is removed from the HLM model, the positive relation of first-generation students with score is gone, from 33.45, $p < .01$, to 4.87, $p > .05$. Most of the immigrant students in Hong Kong come from the mainland China and they are often admitted to a lower level for their ages because of their poorer English ability. These immigrant students therefore outperform their peers at the same grade because they have had additional years of education in the mainland China (Lam & Lau, 2014).

Teaching practices

The relationships between various forms of teaching practices and performance are very consistent amongst the 10 regions. The three teaching practices, adaptive instruction, teacher-directed instruction and interactive application, are positively associated with performance in all regions except Chinese Taipei. Interactive application is the most positive teaching practice. Investigation and perceived feedback, on the contrary, are negatively related to total science scores in all regions. Investigation is the most negative teaching practice.

Direct teaching is an efficient way of learning science contents as compared to individualised learning (OECD, 2016c, p. 65). Students in all the regions benefit from this teaching approach to similar extents, while Finland sees the most positive relationship. In Finland, an increase of one unit of this teaching practice is associated with an increase of 11.13 points in the total score. This is seemingly against our common belief that Western students are less accustomed to direct teaching.

Interactive application has the greatest positive relationship with performance, though the relationship varies widely across the regions. Students in Chinese Taipei and China benefit most from this instructional approach, while those in Japan and Singapore the least. This construct shows the extent that science concepts are taught in the contexts of everyday application in a constructivist manner. Utilising this teaching approach more frequently would foster conceptual understanding, as well as the ability to apply science concepts in unfamiliar situations. If teacher-directed instruction is more related to the quantity of content teaching, then interactive application is more about its quality.

Adaptive instruction is crucial in dealing with individual variations by tailoring the class teaching to the needs of students (Hofstein & Lunetta, 2004). This teaching practice is positively associated with science performance in almost all regions, but the association is much smaller or insignificant in the four Chinese regions and Japan. On the other hand, the Western students tend to benefit more from this teaching approach.

Regarding the two teaching practices that have negative associations with performance, investigation is particularly negative in Chinese Taipei, Korea and the three Western regions, while perceived feedback is less negative in Japan and the four Chinese regions. The possible explanations as to why investigation is negatively associated with performance will be discussed in the following section.

The paradox of investigation

Scientific investigation as a pedagogy is advocated in many science education reform documents, such as Next Generation Science Standards (NGSS, 2013). It is generally regarded as an essential part of science education that has modest but positive effects in many cognitive and affective outcomes (Blanchard et al., 2010; Furtak, Seidel, Iverson, & Briggs, 2012; Minner et al., 2010). However, the findings of PISA assessments fly in the face of this notion. The construct of investigation as defined in PISA 2006 was found to have a negative association with science performance in most regions. A related construct in PISA 2015, enquiry-based instruction, is also negatively associated with performance in 56 out of 63 regions, while no positive associations are found (OECD, 2016c, p.72). Even at the item level, the more the students report to be allowed to design experiments, the lower their science scores (OECD, 2016c, p. 73). These findings are generally congruent with other secondary analyses of the PISA data (Lau, Ho, & Lam, 2015; Seidel, Prenzel, Wittwer, & Schwindt 2007; Taylor, Stuhlsatz, & Bybee, 2009).

Would investigation lead to other desired learning outcomes? PISA 2015 also assesses the students' epistemic belief with survey items on the tentativeness of science knowledge and the value of experiment (OECD, 2016b, p. 98). The epistemic belief of students, however, is also negatively or insignificantly associated with the construct of investigation developed in the study (Table 9). Moreover, the construct of investigation is also found negatively or insignificantly associated with enjoyment of science learning in PISA 2006 (Lau et al., 2015).

One explanation for the negative association of the investigation construct with most of the learning outcomes in PISA is that the causality is simply reversed: academically weaker

Table 9. Correlation of the revised construct of interactive investigation with epistemic belief.

	Correlation of interactive investigation with epistemic belief
Hong Kong	.002
Japan	.001
Korea	−.015
Singapore	.051**
Taipei	−.033**
Macao	.038*
Canada	−.040**
Estonia	−.088**
Finland	−.081**
China BSJG	.094**

*$p < .05$.
**$p < .01$.
***$p < .001$.

Table 10. School-level correlation of the investigation construct with mean total science score.

	Correlation of school mean investigation with school mean science score
Hong Kong ($n = 138$)	−.442**
Japan ($n = 198$)	−.211**
Korea ($n = 168$)	−.300**
Singapore ($n = 177$)	.129
Taipei ($n = 213$)	−.413**
Macao ($n = 45$)	−.329*
Canada ($n = 755$)	−.364**
Estonia ($n = 205$)	−.396**
Finland ($n = 162$)	−.164*
China BSJG ($n = 268$)	.035

*$p < .05$.
**$p < .01$.
***$p < .001$.

students are more often engaged in doing investigations, as suggested in the PISA 2015 report (OECD, 2016c, p. 72). The underlying reason may be that teachers in face of a class of under-performing students would choose to do more practical work as a means of engagement while learning science contents for examination is not a major concern. This explanation is suggested in the PISA 2015 report (OECD, 2016c, p. 72). An empirical support for this explanation comes from the finding that the school mean science score is negatively correlated with the school mean index of the investigation construct in eight of the 10 top regions in PISA 2015 (Table 10). That means the poorer the science performance of a school, the more investigations it conducts.

On the other hand, the investigation paradox may stem from the issue of what constitutes an investigation. Some studies that show positive outcomes of scientific investigation are quasi-experimental design with interventions designed by experts (Blanchard et al., 2010; Furtak et al., 2012). PISA, however, is a cross-national study in which students are surveyed on the quantity of 'investigations' in their science class without being given any information on what resembles an 'investigation'. These 'investigations' may vary considerably in their authenticity (Chinn and Malhotra, 2002), such as the level of openness and how data are collected and analysed (Germann, Haskins, & Auls, 1996), which in turn affect the learning outcomes. A review of the effects of practical work by Singer, Hilton, and Schweingruber (2006, p. 88) comes to the conclusion that 'there is almost no direct evidence that typical laboratory experiences that are isolated from the flow of science instruction are particularly valuable for learning specific scientific content' and that 'typical lab experiences have little effect on more complex aspects of scientific reasoning'. Therefore, the sweeping negative association of the investigation construct with science performance in PISA should not be seen as a verdict on its value in science education. Rather, it shows that what is being practised as 'investigation' in schools is probably problematic and ineffective in promoting science learning.

Discussion and conclusions

This study has analysed the science teaching practices of the 10 top-performing regions in PISA 2015 and their association with science performance. Regarding the performances, the Western regions tend to perform better in areas related to scientific investigation, while the Asian regions are better in CK of science. This corroborates with the findings of other studies (Lau et al., 2015; Olsen & Lie, 2009). But Singapore, Japan and Korea seem to perform more similarly to Western regions despite being in Asia, while Finland performs more similarly to Asian regions than Western regions. This roughly forms a continuum with one end of the Chinese regions and the other end of the Western regions.

However, regarding the teaching practices, it is not straightforward to place the regions on such a Chinese-Western continuum. The 10 regions vary widely in their prevalence of teacher-directed instruction, adaptive instruction and perceived feedback, while that of enquiry-based instruction are similar. No cultural clusters could be identified. For example, Canada, Singapore and Finland have the most teacher-directed instruction in science class, while feedback given to students is more common in Singapore, China and Chinese Taipei. It is more common in the Western classroom that students are allowed to express their ideas, but China is similar to the Western regions in that aspect. In general, Canada and Singapore tend to be more student- and enquiry-oriented in science teaching, while Japan and Korea are more traditional and didactic. The Chinese regions and Finland are closer to that of Canada. These observations gain cross-validation from other evidence. Canada's K-12 curricula are found putting great emphasis on enquiry-based, student-directed learning (Eix, 2016). In Japan, except for the Super Science High schools, most schools focus on 'collections of problems designed to prepare students for university entrance' and 'one-way flow of information' (Morimoto, 2015). As for the Chinese teachers, they are found orchestrating a variety of pedagogies to engage students in active learning rather than teaching didactically (Mok et al., 2001).

Although these top-performing regions do vary in their relative prevalence of the three teaching practices, the associations of the teaching practices with performance are highly consistent within the regions. Adaptive instruction, teacher-directed instruction and interactive application are positively associated with performance in all regions, while investigation and perceived feedback have a negative association. The association of the teaching practices with scores do vary amongst the 10 regions. The scores of students in Finland, Canada and Hong Kong are most positively related to teacher-directed instruction, but the relations are much smaller for Estonia and China. Interactive application has seen the most score gains in Chinese Taipei and China, but the least in Japan and Singapore. Investigation is highly negative to performance in Chinese Taipei, Korea and the three Western regions. Seemingly, there is no clear evidence that students of a particular culture would benefit or suffer more from certain teaching practices. One exception is that the four Chinese regions and Japan are found to have perceived feedback and adaptive instruction much less associated with performance.

So what do we learn from PISA 2015 on science teaching? All the top regions except Japan and Korea tend to have a high frequency of teacher-directed instruction facilitated by more or less authoritative class discussion in science class. Class teaching neither puts a strong emphasis on daily application nor does it heavily adapt to the needs of students. A fair amount of practical work is done, while only some of them are investigations. This mode of instruction is neither didactic nor constructivist, but an amalgam of both that allows for efficient learning of science contents while addressing the ideas of students to a certain extent. It is in between the traditional-reform pedagogy continuum (Anderson, 2002) but seems closer to the traditional end. These findings support that good science performance is a result of the teaching that is highly content-focused (Korsnakova, McCrae, & Bybee, 2009; Thomson, 2009). The TIMSS 1999 Video Study also lends support for it by finding that the classrooms of some low-performing regions such as the United States are filled with activities with no or very limited content learning (Roth et al., 2006).

Practical work, particularly scientific investigations as they are currently practised at school, seems not conducive to PISA performances. But it does not follow that practical work is of no value to science learning – it has positive roles when done properly (Blanchard et al., 2010; Furtak et al., 2012; Singer et al., 2006). But the findings of PISA have alerted us that the desirable outcomes of practical work would not come as granted. Efforts are needed by teachers and curriculum developers to ensure that the practical work is integrated with content learning (Singer et al., 2006) and has a clear focus on the processes and nature of science (Lederman, 2007).

It would be difficult to explain why giving more feedback to students is negative to science learning, given the evidence that constructive feedback can improve student learning and engagement (Lipko-Speed, Dunlosky, & Rawson, 2014). In view of the descriptive nature of the PISA test design, it is probably that the weaker students tend to receive more individual feedback from teachers (OECD, 2016c, p. 68).

As for the roles of cultural and social values in science teaching, caution is needed to attribute the top performance of these regions to particular teaching practices, which are further attributed to particular cultural and social values. There is a host of factors at play other than teaching practices in affecting the educational outcomes of a region, such as demographics, school system, education policy and teacher training. For example, Canada and Finland share many commonalities that account for their top performances: a decentralised education system, high-quality teacher recruitment and training, culture of inclusiveness and strong commitment to public schools (Hargreaves, 2011). The role of cultural and social values in education is obscure. Consider for example the regions influenced by the CHC, i.e. China, Hong Kong, Macao, Chinese Taipei, Japan, Korea and perhaps Singapore (in which ethnic Chinese make up more than 70% of its population). The role of CHC in education is disputed (Wang, 2013). First, CHC is diverse and dynamic, rather than homogeneous and stable (Ryan & Louie, 2007). Therefore, it would be misleading and an overgeneralisation to attribute any educational performance or teaching practices to CHC as a cultural stereotype. Chan and Rao (2009) have debunked the myth of Chinese learners and Chinese teachers. Instead of teaching didactically to a class of passive students who learn merely by rote and memorisation, Chinese teachers are found to use a variety of pedagogies to engage students and to facilitate their understanding, while Chinese learners are actively learning for understanding. Moreover, these CHC regions are not unitary. Hong Kong tends to have more Western influences on its teaching (Chan & Rao, 2009). Even for Canada, given its high proportion of Chinese immigrants, CHC may have played some roles in its performance (Bennett, 2016). All these are supported by some of the findings of the study and can explain why no clear-cut cultural patterns could be found across the 10 regions.

Despite the above, the Western and Asian regions do have some differences in teaching practices and performances that are likely to be rooted in culture. One obvious candidate is the culture of examination. The emphasis on examination by the Chinese has its roots in Confucianism since the time of dynastic China, when national examinations had already been used as a means to promote Confucianism and strengthen meritocratic rule (Wang, 2013). The examination-oriented education in turn constrains the teaching and learning approaches in the CHC regions (Bai, 2010), such as emphasising science contents more than science processes, direct teaching more than experiments and investigations, and whole-class teaching more than small group work. Some of these are supported by the findings of the study.

Another manifestation of the cultural differences is how education is shaped within the liberal and individualist traditions of the Western regions and the conformist and collectivist traditions of the CHC regions. Western teachers tend to enjoy a higher degree of autonomy in a more decentralised education system, be more receptive to learning variability and individual differences and pay greater attention to individual needs. Accordingly, Western teachers have more room to do enquiry-based learning, while adaptive instruction and classroom feedback play greater roles in their teaching. On the contrary, Chinese teachers emphasise the efforts put by students more than individual differences in abilities, which may make adaptive instruction and feedback less related to performance in the Chinese regions.

To conclude, there is no single success formula for science education as shown by the findings of the study. It is an amalgam of didactic and constructivist pedagogy as a result of the Western progressive educational philosophy adapted to the Eastern sociocultural contexts, or the Confucian culture of diligence, high expectation and conformism adapted to the Western sociocultural contexts. But it should be a process of gradual adaptations rather than an implantation. The experiment of the five Chinese teachers teaching in a U.K. school is a case in point (Rumney, 2015).

The above conclusions, however, are limited by the study design that only the 10 top regions are studied. Further study should be done to include some low-performing regions to verify whether the findings hold true. However, extreme caution is needed when comparing the top- and bottom-ranking regions since they may differ hugely in socioeconomic conditions. Further study can also include the attitudinal factors as measured by PISA. Some studies have been done using the PISA 2006 data, where application is found positively associated enjoyment of science, while investigation is negative in the top 10 regions in PISA 2006 (Lam & Lau, 2014; Lau et al., 2015). It is believed that some impacts of teaching practices are through their influence on students' attitudes towards science learning, which calls for path analyses to delineate their relationships.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Kwok-chi Lau http://orcid.org/0000-0002-8729-4309
Terence Yuk-ping Lam http://orcid.org/0000-0002-7429-9098

References

Anderson, R. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, *13*, 1–12.

Bai, L. M. (2010). Human capital or humane talent? Rethinking the nature of education in China from a comparative historical perspective. *Frontiers of Education in China*, *5*(1), 104–129. doi:10.1007/s11516-010-0008-z

Bennett, P. W. (2016, December 8). PISA 2015 results: Why is Canada riding high on its stable, lopsided student performance? *Educhatter*. Retrieved from https://educhatter.wordpress.com/2016/12/08/pisa-2015-results-why-is-canada-riding-high-on-its-stablelopsided-student-performance/

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, *94*(4), 577–616.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis*. Newbury Park: Sage.

Chan, C. K. K., & Rao, N. (2009). Moving beyond paradoxes: Understanding Chinese learners and their teachers. In K. K. C. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing contexts, changing education* (pp. 3–34). Dordrecht: Springer.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*, 175–218.

Eix, S. (2016, December 9). B.C.'s redesigned inquiry-based curriculum: A wonder-inducing teaching method. *Huffpost*. Retrieved from http://www.huffingtonpost.ca/sandy-eix/bc-inquiry-based-curriculum_b_11935894.html

Fensham, P. (1985). Science for all: A reflective essay. *Journal of Curriculum Studies*, *17*(4), 415–435. doi:10.1080/0022027850170407

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, *82*(3), 300–329.

Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. *Journal of Research in Science Teaching*, *33*, 475–499.

Hargreaves, A. (2011, January 26). Canada's culture of excellence in education. *The Star*. Retrieved from https://www.thestar.com/opinion/editorialopinion/2011/01/26/canadas_culture_of_excellence_in_education.html

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, *88*(1), 28–54. doi:10.1002/sce.10106

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International Inc.

King, D., & Stephen, M. R. (2012). Learning science through real-world contexts. In B. J. Fraser, K. Tobin, & C. McRobbie (Eds.), *Springer international handbooks of education – second international handbook of science education* (pp. 69–79). Dordrecht: Springer.

Korsnakova, P., McCrae, B. J., & Bybee, R. W. (2009). Improving science teaching and learning. In R. W. Bybee, & B. J. McCrae (Eds.), *PISA science 2006: Implications for science teachers and teaching* (pp. 59–68). Arlington, VA: NSTA Press.

Lam, T. Y. P., & Lau, K. C. (2014). Examining factors affecting science performance of Hong Kong in PISA 2006 using hierarchical linear modelling. *International Journal of Science Education*, *36*(15), 2463–2480. doi:10.1080/09500693.2013.8.

Lau, K. C., Ho, E. S. C., & Lam, T. Y. P. (2015). Effective classroom pedagogy and beyond for promoting scientific literacy: Is there an East Asian model. In M. S. Khine (Ed.), *Science education in East Asia* (pp. 13–40). Dordrecht: Springer.

Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research in science education* (pp. 831–879). Mahwah, NJ: Lawrence Erlbaum.

Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2014). Does testing with feedback help grade-school children learn key concepts in science? *Journal of Applied Research in Memory and Cognition*, *3*(3), 171–176. doi:10.1016/j.jarmac.2014.04.002.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modelling. *Psychological Methods*, *1*, 130–149.

Miles, J., & Shevlin, M. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, *25*, 85–90.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction: What is it and does it matter? Results from a research synthesis 1984 to 2002. *Journal of Research in Science Teaching*, *47*(4), 474–496.

Mok, I., Chik, P. M., Ko, P. Y., Kwan, T., Lo, M. L., & Marton, F. (2001). Solving the paradox of the Chinese teacher? In D. A. Watkins & J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp. 161–179). Hong Kong/Melbourne, VIC: Comparative Education Research Centre, The University of Hong Kong/Australian Council for Educational Research.

Morimoto, S. (2015). *The major problem is the polarization of science education*. My vision, 10. Nippon Institute for Research Advancement.

Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead: Open University Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Retrieved from http://www.nextgenscience.org/

OECD. (2016a). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy, PISA*. Paris: OECD Publishing. doi:10.1787/9789264255425-en

OECD. (2016b). *PISA 2015 results (volume I): Excellence and equity in education, PISA*. Paris: OECD Publishing. doi:10.1787/9789264266490-en

OECD. (2016c). *PISA 2015 results (volume II): Policies and practices for successful schools, PISA*. Paris: OECD Publishing. doi:10.1787/9789264267510-en

OECD. (2016d). *PISA 2015 technical report, PISA*. Paris: OECD Publishing.

Olsen, R. V., & Lie, S. (2009). Knowledge of and about science. In R. W. Bybee & B. J. McCrae (Eds.), *PISA science 2006: Implications for science teachers and teaching* (pp. 59–68). Arlington, VA: NSTA Press.

Osborne, J. (2012). The role of argument: Learning how to learn in school science. In B. J. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education, Springer international handbooks of education* (pp. 933–949). Dordrecht: Springer.

Pong, S. L. (2009). Grade level and achievement of immigrants' children: Academic redshirting in Hong Kong. *Educational Research and Evaluation*, *15*(4), 405–425.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., … Gallimore, R. (2006). *Highlights from the TIMSS 1999 video study of eighth-grade science teaching (NCES 2006-17)*. U.S. Department of Education, National Center for Education Statistics, Washington, DC: U.S. Government Printing Office.

Rumney, B. (Producer). (2015, August). *Are our kids tough* enough? [Television broadcast]. UK: BBC.

Ryan, J., & Louie, K. (2007). False dichotomy? 'Western' and 'Confucian' concepts of scholarship and learning. *Educational Philosophy and Theory*, *39*(4), 404–417. doi:10.1111/j.1469-5812.2007.00347.x

Seidel, T., Prenzel, M., Wittwer, J., & Schwindt, K. (2007). Unterricht in den aturwissenschaften [Instruction in the natural sciences]. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, K. Klieme, & R. Pekrun (Eds.), *PISA 2006. Die ergebniss der dritten internationalen vergleichsstudie* [PISA 2006: The results of the third international comparative study] (pp. 147–179). Munster: Waxmann.

Singer, S. R., Hilton, M. L., & Schweingruber, H. A. (Eds.) (2006). *America's lab report investigations in high school science*. Washington, DC: National Academies Press.

Taylor, J. A., Stuhlsatz, M. A. M., & Bybee, R. W. (2009). Windows into high-achieving science classrooms. In R. W. Bybee & B. J. McCrae (Eds.), *PISA science 2006: Implications for science teachers and teaching* (pp. 59–68). Arlington, VA: NSTA Press.

Thomson, S. (2009). Teaching and learning science: PISA and the TIMSS video study. In R. W. Bybee & B. J. McCrae (Eds.), *PISA science 2006: Implications for science teachers and teaching* (pp. 59–68). Arlington, VA: NSTA Press.
Wang, J. (2013). Confucian heritage cultural background (CHCB) as a descriptor for Chinese learners: The legitimacy. *Asian Social Science*, *9*(10), 105–113.
Watkins, D. A., & Biggs, J. B. (Eds.). (1996). *The Chinese learner: Cultural, psychological and contextual influences*. Hong Kong/Melbourne, VIC: Comparative Education Research Centre, The University of Hong Kong/Australian Council for Educational Research.
Watkins, D. A., & Biggs, J. B. (2001). The paradox of the Chinese learner and beyond. In D. A. Watkins, & J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp. 3–23). Hong Kong, China/Melbourne, VIC: Comparative Education Research Centre, The University of Hong Kong/Australian Council for Educational Research.